DETERMINATION OF SENSITIVITY OF HYSPLIT BACK TRAJECTORIES FOR INPUT DATA COMMONLY USED IN TURKEY

TÜRKİYE'DE YAYGIN OLARAK KULLANILAN HYSPLIT MODEL GİRDİLERİNİN GERİ YÖRÜNGELER ÜZERİNDEKİ HASSASİYETİNİN BELİRLENMESİ

FİRDEVS DOĞRUSEVER

DR. DERYA DENİZ GENÇ TOKGÖZ

Supervisor

Submitted to

Graduate School of Science and Engineering of Hacettepe University

as a Partial Fulfillment to the Requirements

for the Award of the Degree of Master of Science

in Environmental Engineering.

September 2023

ABSTRACT

DETERMINATION OF SENSITIVITY OF HYSPLIT BACK TRAJECTORIES FOR INPUT DATA COMMONLY USED IN TURKEY

FİRDEVS DOĞRUSEVER

Master of Science, Department of Environmental Engineering

Supervisor: Dr. Derya Deniz Genç Tokgöz

September 2023, 105 pages

The Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model (web-version) was used to simulate hourly 96-hr back trajectories arriving at a rural site in the Eastern Mediterranean, at altitude of 1500 m from the surface, for each day between 2010 and 2013 years. Two meteorological data archives (NCEP/NCAR Reanalysis and GDAS1) and vertical velocity methods (isentropic and model vertical velocity) were used as model inputs as these are the most widely used input variables. The sensitivity of trajectories to model inputs was measured by the absolute horizontal transport deviation (AHTD), the absolute vertical transport deviation (AVTD) and the relative horizontal transport deviation (RHTD) statistics. Both the meteorological archive and vertical transport method significantly influenced the trajectories. Trajectories simulated by NCEP/NCAR Reanalysis archive were less sensitive to the vertical transport method than trajectories simulated by the GDAS1 archive. Cluster Analysis by SPSS (k-means technique) was applied for each back trajectory data set to classify them into similar groups (clusters). Based on their speed, back trajectories in each data set were classified into five clusters. To examine the dependence of potential source directions of pollutants (i.e. cluster centroids) to the HYSPLIT

model inputs, differences of sulfate concentrations were discussed. The results indicated that there were discrepancies in the interpretation of the source-receptor relationship when different inputs were used to run the HYSPLIT model.

Keywords: Back trajectory, Clustering, HYSPLIT, GDAS1, NCEP/NCAR Reanalysis, Air quality modelling

ÖZET

TÜRKİYE'DE YAYGIN OLARAK KULLANILAN HYSPLIT MODEL GİRDİLERİNİN GERİ YÖRÜNGELER ÜZERİNDEKİ HASSASİYETİNİN BELİRLENMESİ

FİRDEVS DOĞRUSEVER

Yüksek Lisans, Çevre Mühendisliği Bölümü

Tez Danışmanı: Dr. Derya Deniz Genç Tokgöz

Eylül 2023, 105 sayfa

HYSPLIT modeli web sürümü, 2010 ile 2013 yılları arasında her gün için Doğu Akdeniz'de yüzeyden 1500 m yükseklikteki kırsal bir bölgeye varan 96 saatlik geri yörüngeleri saatlik olarak simüle etmek için kullanıldı. En yaygın kullanılan girdi değişkenleri olmaları nedeniyle model girdileri olarak iki meteorolojik veri arşivi (NCEP/NCAR Reanaliz ve GDAS1) ve dikey hız yöntemleri (izentropik ve model dikey hız) kullanılmıştır. Yörüngelerin model girdilerine duyarlılığı, mutlak yatay aktarım sapması (AHTD), mutlak dikey aktarım sapması (AVTD) ve bağıl yatay aktarım sapması (RHTD) istatistikleriyle ölçüldü. Hem meteorolojik arşiv hem de dikey taşıma yöntemi yörüngeleri önemli ölçüde etkiledi. NCEP/NCAR Reanaliz arşivi tarafından simüle edilen yörüngeler, dikey taşıma yöntemine karşı GDAS1 arşivi tarafından simüle edilen yörüngelere göre daha az duyarlı bulunmuştur. Her bir geri yörünge veri seti için bunları benzer gruplara (kümelere) sınıflandırmak üzere SPSS ile Küme Analizi (k-ortalamalar tekniği) uygulandı. Hızlarına bağlı olarak, her veri setindeki geri yörüngeler 5 kümeye ayrıldı. Kirleticilerin potansiyel kaynak yönlerinin (yani küme merkezlerinin) HYSPLIT modeli girdilerine bağımlılığını incelemek için sülfat konsantrasyonlarındaki farklılıklar tartışıldı. Sonuçlar, HYSPLIT modelini çalıştırmak için farklı girdiler kullanıldığında kaynak-alıcı ilişkisinin yorumlanmasında farklılıklar olduğunu göstermiştir.

Anahtar Kelimeler: Geri yörünge, Kümeleme, HYSPLIT, GDAS1, NCEP/NCAR Reanaliz, Hava Kalitesi Modellemesi

ACKNOWLEDGEMENT

I would like to thank and express my deepest gratitude to my supervisor Assistant Professor Deniz Genç Tokgöz for her support, guidance, and encouragement provided during the entire research process. I would like to thank Prof. Dr. Gülen Güllü, Prof. Dr. Merih Aydınalp Köksal, Assoc. Prof. Dr. Fatma Öztürk, and Assoc. Prof. Dr. Güray Doğan for their support and contributions.

I would like to express my sincere appreciation to my precious parents Cemile Nur Doğrusever, Mehmet Adil Doğrusever; and my lovely siblings Eymen Doğrusever and Salih Doğrusever, and my dear cousin Yusuf Taha Toprak for their everlasting support, understanding, and patience not only during this study but also throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

AGL	Above Ground Level
AHTD	Absolute Horizontal Transport Deviation
AMSL	Above Mean Sea Level
ARL	Air Resources Laboratory
AVTD	Absolute Vertical Transport Deviation
BTH	Beijing-Tianjin-Hebei
CLRTAP	Convention on Long-Range Transboundary Air Pollution
CWT	Concentration Weighted Trajectory
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA-40	ECMWF Reanalysis-40 years
FLEXPART	Flexible Particle Dispersion Model
FNL	Final Analysis
GDAS	Global Data Assimilation System
GMM	Gaussian Mixture Model
HYSPLIT	Hybrids Single Particle Lagrangian Integrated Trajectory Model
K-W	Kruskal-Wallis test
LRT	Long Range Transport
MURA	Multi-Receptor
MSL	Mean Sea Level
NCEP	National Centers for Environmental Prediction
NCAR	National Center for Atmospheric Research
NE	Northeast
NH_3	Ammonia
NMC	US National Meteorological Center
NOAA	National Oceanic and Atmospheric Administration, United States
NO	Nitric oxide
NO_x	Nitrogen oxides

NO_2	Nitrogen dioxide
NW	Northwest
O_3	Ozone
$PM_{2.5}$	Particulate Matter < 2.5 μm in diameter
PM_{10}	Particulate Matter < 10 μm in diameter
POPs	Persistent Organic Pollutants
PSCF	Potential Source Contribution Function
RHTD	Relative Horizontal Transport Deviation
SA	Similarity Algorithm
SO_2	Sulfur dioxide
SO_4^{-2}	Sulfate
SOM	Self-organizing maps
SPSS	Statistical Package for the Social Sciences
STILT	Stochastic Time-Inverted Lagrangian Transport Model
SW	Southwest
TRMSD	Total Root Mean Square Deviation
TSV	Total Spatial Variance
U.S.S.R.	The United Socialist Soviet Republic
UTC	Coordinated Universal Time
VOCs	Volatile Organic Compounds
3-D	Three dimensional

1. INTRODUCTION

The negative effects of air pollution on human health (Chen et al., 2013) and ecosystems (Compton et al., 2011; Cooter et al., 2013) have led to tighter standards for ambient air quality (*REVIHAAP Project*, 2013). Identifying pollutants and their sources is essential to develop effective mitigation strategies. For a particular pollutant, if the pollutant concentration is above the limit values of air quality standards, some control measures must be taken by the authorities to improve the air quality. These control measures can be determined if the sources of this particular pollutant are known. The relation between the pollutant emission and resulting concentrations are complex (Fenger, 2002). An air pollutant is emitted into the atmosphere and then it is dispersed and diluted (i.e. mostly due to the meteorological conditions, especially wind speed, wind direction, turbulence, and atmospheric stability) (Lyons & Scott, 1990) and undergo some physical and chemical reactions (Mayer, 1999). Furthermore, as the atmosphere is the primary pathway for air pollutants, both local and distant sources may contribute to the observed pollutant concentrations. Depending on the synoptic scale meteorological conditions, an air pollutant that is emitted into the troposphere can travel long distances (i.e. 100 km to 1000 km from the source) and during this travel it can undergo gradual mixing with the background and eventually arrives the receptor site and contributes the pollutant concentration in the receptor site. This is called "Long Range Transport of Pollutants (LRT)". In the literature, depending on the time (i.e. a few days to a few weeks) or horizontal scale (i.e. 100 km to 1000 km) of pollutant transport, different names such as continental, transboundary or regional transport can be used to refer long range transport of pollutants. European Environment Agency defines the LRT as the atmospheric transport of air pollutants within a moving air mass for a distance greater than 100 kilometers (European Environment Agency, 2017). Actually, long range transport of air pollutants is first recognized in 1960s with the acidification of lakes in UK and Scandinavia (Mylona, 1996). There were no emission sources that may cause acidification problem in the remote areas of the UK and Scandinavia. However, sulfur dioxide emissions transported from continental Europe caused acidification problem in these countries. As a result, the first multilateral agreement namely "Convention on Long-range Transboundary Air Pollution (CLRTAP)" was established in 1979 to address transboundary air pollution problem. At present, 51 Parties (i.e. countries) ratified the CLRTAP convention and a regional framework has been created to reduce transboundary air

pollution and better understand the air pollution science (U.S.A Department of State-Office of Environmental Quality, n.d.) by adopting emission reduction targets for specific pollutants (i.e. sulfur dioxide (SO2), nitrogen oxides (NOx), ammonia (NH3), volatile organic compounds (VOCs), ozone (O3), persistent organic pollutants (POPs), and heavy metals). All Parties have to achieve these targets to combat the resulting transboundary air pollution.

Previous LRT (long-range transport) studies in Turkey were mostly conducted in the Eastern Mediterranean region of Turkey (Al-Momani et al., 1997; G. Dogan et al., 2010; G. Güllü et al., 2004, 2005; G. H. Güllü et al., 1998, 2000; Günaydin & Tuncel, 2003; Im et al., 2012; Koçak et al., 2004, 2009a, 2012; Koçak, Kubilay, et al., 2007; Koçak, Mihalopoulos, et al., 2007b, 2007a; Kubilay et al., 2000, 2005; Öztürk et al., 2012; Sciare et al., 2003, 2005; Theodosi et al., 2010; Türküm et al., 2008). In these studies, eastern Turkey and the eastern and central regions of northern Africa are identified as the main source regions for crustal elements, and the countries surrounding the Mediterranean, the Balkan countries and the former USSR countries are identified as the main source regions for anthropogenic pollutants.

To establish the influence of long range transported pollutants at a receptor site, air mass back trajectories must be examined. Air masses have the potential to carry both natural and human-made pollutants as they travel. Establishing the history of air masses across time is necessary to determine the influence of local sources of pollution. Trajectory is the path that air masses travel as they move through the atmosphere. Trajectories are often used to understand how pollutants disperse over time and distance, and to identify potential sources of pollution.

In order to simulate air masses, back trajectory calculations are commonly used. By tracing the paths of air masses leading up to their arrival at the receptor site, back trajectory analysis provides a simulation of their movement. Several models (HYSPLIT, FLEXPART, LAGRANTO, NAME, STILT, TRAJ3D, METEX) can simulate back trajectories (Bowman et al., 2013). The Hybrid Single Particle Lagrangian Integrated Trajectory Model (HYSPLIT) (Draxler et al., 1998; Stein et al., 2015) has been widely utilized by air quality researchers to identify the main emission sources influencing the particular site (i.e. receptor site) through statistical analyses (i.e. trajectory

statistics) and to interpret episode events such as dust transport from desert or emissions from forest fires (Vitali et al., 2017).

The HYSPLIT model requires meteorological data (archive) as input data and various parameters, including a vertical transport velocity method, arrival time, arrival height, and time length. Back trajectories may differ depending on the model inputs (Gebhart et al., 2005; Harris et al., 2005; Su et al., 2015).Therefore, all the input variables used for trajectory simulation shall be reported. Those studies indicated that trajectories also vary depending on the receptor sites' geographical properties. It is well known that for a particular region (or receptor site), different model inputs shall be used to simulate back trajectories to determine the sensitivity of trajectories to input variables. However, sensitivity analyses are rarely conducted. Errors up to 100% can also occur in critical flow scenarios, according to Stohl (1998).

In Figure 1 and Figure 2, 96 h (i.e. 4 days back) back trajectories generated for 8 August 2010 using different input data (i.e. different meteorological archieve and vertical transport method) are presented as an example to illustrate how trajectories can vary when different model inputs are used. In Figure 1, the back trajectory is simulated using the NCEP/NCAR Reanalysis meteorological archieve and isentropic vertical method. According to Figure 1, air masses originated from Libya and Egypt and then travelled through the Mediterranean Sea before intercepting the receptor site (arrival site). In Figure 2, GDAS1 meteorological archieve and model vertical velocity (three dimensional (3-D)) is used as input data to simulate the back trajectory. According to Figure 2, air masses originated from the Agean Sea and then transported to the receptor site. Although these two back trajectories are simulated for the same day, they indicated totally different source regions for the air masses.

Quantifying the differences in the HYSPLIT trajectories produced by different model inputs is therefore crucial; however, such an examination has not been conducted in Turkey. Actually, different model inputs have been used in the back trajectory studies conducted over Turkey. NCEP/NCAR Reanalysis archive and GDAS1 archieve are the most commonly used meteorological archives, while isentropic method and model vertical velocity are the most commonly used vertical transport methods in the studies conducted over Turkey. Hence, this study

is designed to assess the sensitivities of HYSPLIT back trajectories simulated with the most widely used input variables for a particular receptor site in Turkey.

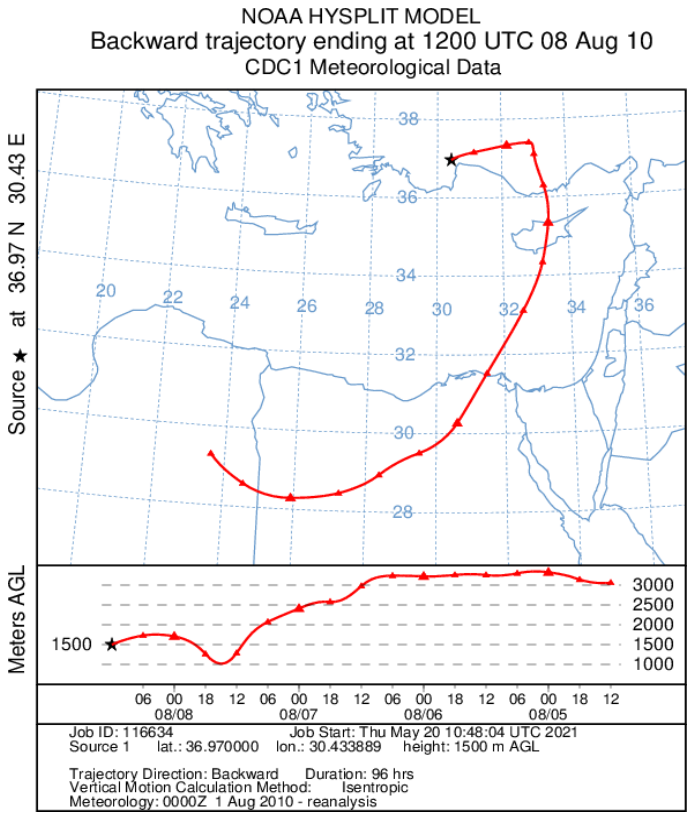


Figure 1. 96-hr back trajectory generated by HYSPLIT Model using NCEP/NCAR Reanalysis archive and isentropic vertical method

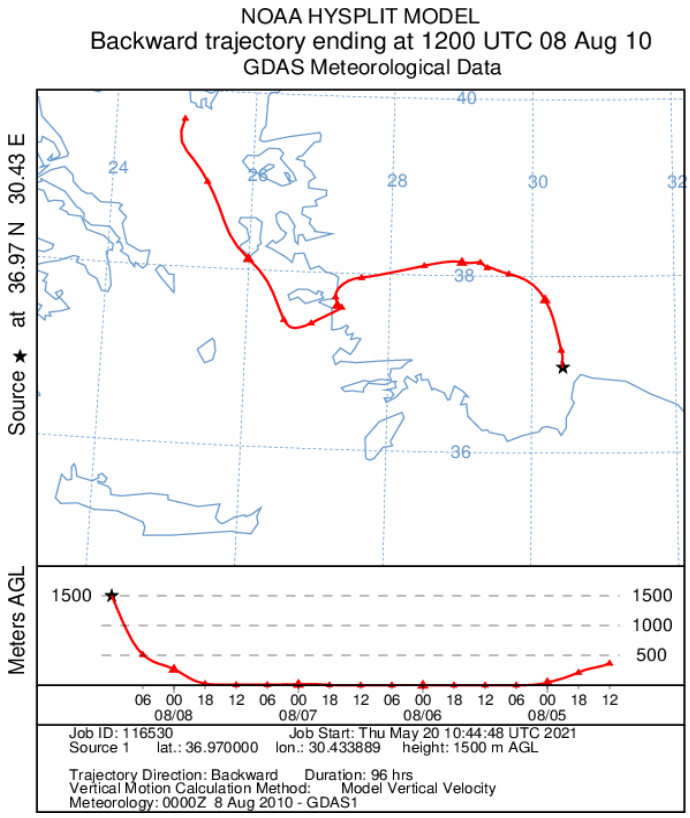


Figure 2. 96-hr back trajectory generated by HYSPLIT Model using GDAS1 archive and model vertical velocity

1.1. Purpose and Scope of the Study

The primary goals of this study are to compute the differences between HYSPLIT back trajectories simulated with the most common meteorological archives and vertical velocity methods by conducting sensitivity analysis; to compare the most dominant patterns of back trajectories generated from these common input variables by using cluster analysis; and last but not least, to examine whether the differences in the predominant patterns impact the interpretation of the source-receptor relations. It's crucial to note that this study refrains from declaring one meteorological archive or vertical velocity method as superior to the other. Actually, this study is designed to inform the HYSPLIT users in Turkey to consider that some differences could arise when back trajectories are simulated with different inputs and to highlight the importance of reporting input variables that were used for running the HYSPLIT model.

1.2. Structure of the Thesis

In Chapter 1, the problem is defined. The primary goals and scope of the study are listed.

In Chapter 2, the literature review provides background information on back trajectories, model inputs, cluster analysis, and sensitivity analysis.

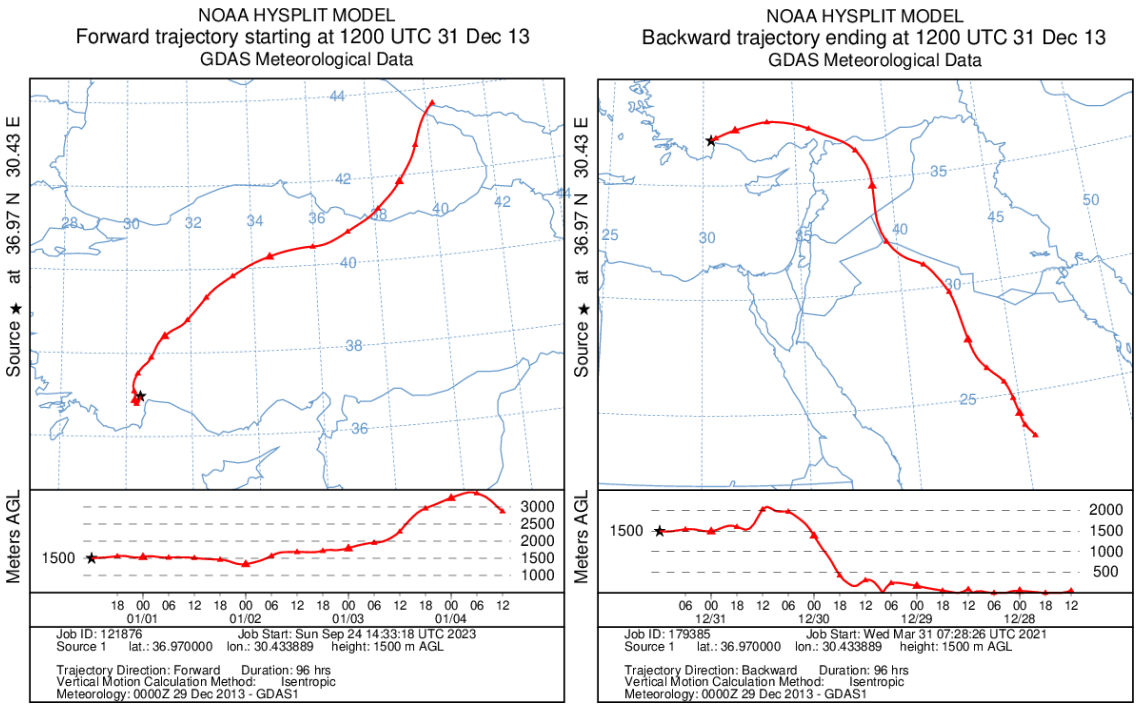
In Chapter 3, the materials and methods of this thesis were described. The study region and the input data utilized were explained.

In Chapter 4, four different back trajectory data sets were compared with sensitivity analysis. The variation of sensitivity parameters with respect to time and season were explained. One of the trajectory statistics methods, namely Cluster Analysis were applied to the back trajectory data sets. The variation of transport pattern and the variation in the interpretation of the source-receptor relationship with respect to the different trajectory data sets were discussed.

Chapter 5 contains the conclusion of this thesis and recommendations for future studies.

2. LITERATURE REVIEW

Air pollution has been a significant concern for human health and the environment for many decades, prompting a considerable amount of research on the topic. Identifying sources of pollution has become crucial to determine policies and strategies to prevent the impact of pollution. The trajectory is a set of vectors that shows the possible routes of air parcels forward or backward in the atmosphere in latitude, longitude, and altitude. The trajectory followed by an air parcel in the backward direction is called the back trajectory, and the trajectory followed in the forward direction is called the forward trajectory. Both trajectory types have particular usage in air pollution studies. The forward trajectory actually simulates the potential path of air masses using the forecast meteorological data i.e. where the air pollutant which is emitted from the source can go at a given time and place (Figure 3a). Forward trajectories can be used to forecast the transport of forest fires, nuclear or chemical pollutants (Stein et al., 2015). Back trajectories also known as backward trajectories uses archived meteorological data and simulates where the air masses come from before arriving the receptor site (Figure 3b). In general, back trajectories are combined with the air pollution measurement at a particular site and the air mass history is tracked by means of trajectory statistics so that source-receptor relation can be established. Therefore, back trajectories are an important tool in air pollution research, providing valuable information about the sources and transport of pollutants.



(a) (b)

Figure 3. (a) Forward trajectory and (b) Back trajectory

Forward and back trajectories can be calculated by using HYSPLIT trajectory model. There are two versions of the HYSPLIT model, a web version, and a PC version.

Hybrid-Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Draxler et al., 1998) developed by the National Oceanic and Atmospheric Administration's (NOAA) Air Resources Laboratory (ARL). The HYSPLIT model is utilized to examine the atmospheric transport, dispersion, and accumulation of pollutants and hazardous substances on the surface of the Earth. HYSPLIT uses include monitoring and predicting the release of radioactive material, volcanic ash, forest fire smoke, and pollutants from emission sources.

The computation method of the model combines elements of the Eulerian and Lagrangian approaches. In the Lagrangian approach, when the air parcels move from their starting positions, a moving reference frame is utilized while the Eulerian approach utilizes a stationary three-dimensional grid. The model utilized the Lagrangian approach to compute the trajectories,

advection, and diffusion while using the Eulerian approach to calculate pollutant concentrations (NOAA, 2011).

2.1. Input Variables

Executing the HYSPLIT trajectory model requires access to a substantial meteorological database, typically at a medium to global scale. Meteorological data is obtained by combining meteorological models and meteorological measurements. The HYSPLIT trajectory model provides users with various options for meteorological archives, enabling them to select the most suitable data set for their specific needs. Table 1 shows the meteorological archives that are available within the HYSPLIT trajectory model.

Table 1. HYSPLIT Meteorological Archives (NOAA, 2023)

Data set		**Horizontal Resolution (km-approx.)**	**Full-grid dimensions**	**Temporal resolution (hrs)**	**Vertical Levels**	**Period of each file**	**Availability**
Global	GFS - 0.25º	27	1440 x 721	3	56	1 day	Jun 2019 -> present
	GDAS - 0.5º	55	720 x 361	3	56	1 day	Sep 2007 -> Jun 2019
	GDAS - 1º	111	360 x 181	3	24	1 week	Dec 2004 -> present
	Global Reanalysis - 2.5º	278	144 x 73	6	18	1 month	1948 -> present

As seen in Table 1, there are four meteorology archives available globally in HYSPLIT. GFS 0.25 is an archive of combined short-term NCEP Global Forecast System (GFS) model output having a quarter-degree latitude-longitude grid on native model levels (NOAA, 2019a). The GFS archive is available from 2019 to the present. GDAS 0.5 daily archive files include global 3-D gridded meteorological model output. The output file includes 3-hourly data, at a half-degree latitude by

half-degree longitude resolution, on hybrid sigma-pressure surfaces (NOAA, 2012). GDAS 0.5 archive is available between 2007-2019. The Reanalysis archive has a horizontal resolution of about 2.5 degrees latitude. The Reanalysis archive is available for the globe, from 1948 to the present (NOAA, 2019b). GDAS1 has 1 degrees latitude by 1 degrees longitude spatial resolution (NOAA, 2019). The GDAS1 archive is available from 2004 to the present.

Using different meteorological archives affects the resulting back trajectory. Su et al. (2015) calculated 72-hr back trajectories four times a day (02 UTC, 08 UTC, 14 UTC and 20 UTC) for the year of 2011 at 500 m arrival height by using HYSPLIT in the Northern Hong Kong. In this study, two meteorological data sets were compared (GDAS1 and GDAS0.5). There are some differences between GDAS1 and GDAS0.5 in terms of horizontal resolution and vertical velocity field, despite the fact that they assimilate the same observations. In the study, vertical velocities with the observations and the performance in obtaining PM contributions from various directions were compared. As a result, GDAS1 archive concluded more proper in back trajectory analysis for that region. They also evaluated the seasonal variations of trajectories and found that differences in trajectories are greater during winter and less during summer.

Harris et al. (2005) compared two different meteorological data (ERA-40 and NCEP/NCAR Reanalysis). Trajectories was compared in this study. Study have been conducted to three geographically diverse sites. Deviation statistics have been calculated. These are the absolute horizontal transport deviation (AHTD), the absolute vertical transport deviation (AVTD), and the relative horizontal transport deviation (RHTD). Deviation statistics results revealed that the sensitivity to the meteorological archive was 30–40%.

Kahl et al. (1989) examined Arctic trajectory types. They concluded that at times trajectories were more sensitive to different meteorological archives than they were to variations in vertical transport method.

In HYSPLIT model, not only meteorological archives but also the vertical velocity methods, arrival hour, arrival height, total run time are required as input variables. In the following

paragraphs, studies conducting using different vertical velocity methods, and other input variables are presented.

In HYSPLIT trajectory analysis, users can select the vertical velocity method for calculating back trajectories. The speed of air moving either upward or downward is known as vertical velocity. Understanding the large-scale dynamics of the atmosphere, such as regions of upward and downward motion, can be aided by knowing the vertical velocity (ECMWF, 2023). As vertical velocity method, HYSPLIT desktop version offers nine options. These are model vertical velocity, isobaric, isentropic, constant density, isosigma, from divergence, remap MSL (Mean Sea Level) to AGL (above ground level), average data, and damped magnitude option (NOAA, 2007). HYSPLIT web version offers only three of them, model vertical velocity, isobaric and isentropic.

Generally, the meteorological archive includes the vertical motion field in most situations. These data fields can be used directly in trajectory calculations as the vertical velocity method. However, the HYSPLIT model offers other velocity methods that may be required for special situations by assuming the pollutant is transported on another surface. This is achieved by calculating the velocity (Wη) needed to keep a parcel on the chosen η surface, taking into account the surface's local rate of change and slope (NOAA, 2007).

$$W\eta = (-\partial\eta/\partial t - u\,\partial\eta/\partial x - v\,\partial\eta/\partial y) / (\partial\eta/\partial z) \qquad \text{(Equation 1)}$$

Where $W\eta$ = velocity (Wη) required to maintain a parcel on the selected η surface

η = surface can be either isobaric (p), isosigma (σ), isopycnic (ρ), or isentropic (θ).

x,y,z = locations of the trajectories

Using different vertical velocity methods as well as different meteorological archives also affects the resulting back trajectory. Harris et al. (2005) compared the isentropic and model vertical velocityMetin girmek için buraya tıklayın veya dokunun., two different meteorological data were used (ERA-40 and NCEP/NCAR Reanalysis) to compare isentropic and model vertical velocity

tropospheric trajectories. Deviation statistics have been calculated. These are AHTD, AVTD, and RHTD. At all three locations, model vertical velocity trajectories reached greater altitudes and wind speeds compared to isentropic trajectories. Additionally, 3-D trajectory deviations exceeded isentropic trajectory deviations due to uncertainties in the vertical wind fields and greater wind speeds in three-dimensional trajectories. Deviation statistics results revealed that the sensitivity to the vertical transport method was 18–34%, the sensitivity to the meteorological archive was 30–40%, and the sensitivity to combined two-way discrepancies in vertical transport method and the meteorological archive was 39–47%.

Martin et al. (1990) compared isentropic trajectories and three-dimensional (3-D) trajectories. It was concluded that although isentropic trajectories better symbolize vertical motion, the model is suited to intensive applications in pollution investigations.

Three-dimensional and isentropic trajectories are computed by Stohl & Seibert (1998). Isentropic trajectories are found to be more affected by dynamical discrepancies between meteorological fields, while 3-D trajectories are determined to be the most accurate ones.

According to Draxler (1996), isentropic and kinematic trajectories are often similar to one another (90%), but they can diverge significantly when they enter baroclinic parts of the troposphere.

The distances between the sources and destination regions and the routes that the air mass takes affect how long back trajectories are. Unless the trajectories are examined as part of a tracer study, the level of accuracy for any single trajectory is typically unknown (Markou & Kassomenos, 2010).

In addition, total trajectory run time can also be determined by the user. In the previous, back trajectory cluster analysis has been widely employed to interpretation of pollution (Abdalmogith & Harrison, 2005; Borge et al., 2007a; Cabello et al., 2008) and precipitation composition (Dorling et al., 1992a, 1992b). Most of these studies employ back trajectories with timescales between 3 and 5 days. This represents a compromise between having enough time to characterize the long-range transport and the individual back trajectories' deteriorating accuracy the further back they

are in time (Stohl, 1998). As total run time increases, uncertainty increases. Cabello et al. (2008) observed that differences between trajectories have grown linearly for a minimum up to 48-hr and have shown faster growth after 72-hr. Furthermore, when selecting the total run time, sufficient time must be given if secondary pollutants and their sources are to be investigated (Baker, 2010).

Another input variable influencing the back trajectories is arrival height. Saunders et al. (2013) examined the sensitivity of back trajectories using four different arrival heights (i.e. 500 m, 1000 m, 1500 m, and 2000 m AGL) and found that back trajectory calculations are affected by the choice of the arrival height.

Gebhart et al. (2005) discovered directional biases in trajectories began at different elevations from the studied region. In the study, it was seen that trajectories began at higher elevations was in tendency to move faster and back farther than those that began at lower elevations. Cabello et al. (2008) found that the number of clusters of each clustering scenario varies for the trajectories reaching 1500 and 3000 m.

2.2 Application of Back Trajectories

Back trajectories are the most commonly calculated type of air trajectories. Trajectories are used in cyclones (Gozzo et al., 2013), and synoptic meteorology studies (Hondula et al., 2010), atmospheric moisture (Knippertz & Wernli, 2010), clouds (Feingold et al., 2003), and precipitation studies (Tošić & Unkašević, 2013); in air chemistry applications such as common air pollutants (Asaf et al., 2008), ozone (MacDonald et al., 2011), trace gases studies (Ferrarese et al., 2002), to transport of hazardous substances such as radionuclide transport (Srinivas et al., 2012), insecticides/pesticides/persistent organic pollutants (POPs) (Qin et al., 2012), toxic metal studies (Cheng et al., 2013); to determine sources and transport of aerosols such as particulate matter studies (Mahapatra et al., 2013), forest fire (Chan et al., 2006), and biomass burning studies (Ortiz-Amezcua et al., 2014,Pérez, Artuso et al. 2015).

Zhao et al. (2020) investigated the features and distribution of extensive air pollution episodes during the COVID-19 pandemic in the studied region using back trajectory analysis and cluster

analysis. The 72-hr back trajectories were calculated by using HYSPLIT, at 50, 100, and 500 m AGL using GDAS1 meteorological archive. For each of the three cities, a total of 432 trajectories were obtained. The trajectories were then classified into four groups. During the first pollution episode in the studied region, the back trajectories have resulted from the local emission sources. Mitigation strategies were determined for the studied region. HYSPLIT, PSCF and CWT models was used to study the source distribution and transport paths. The study aimed to gain insights into the transport paths and source distribution of air pollutants during this period marked by significant changes in human activities and emission patterns.

Ma et al. (2019) examined the transport paths and sources of atmospheric pollution in the city of Shenyang by using back trajectories. The study utilizes the HYSPLIT model to analyze the movement of pollutants and gain insights into their origins and dispersion patterns within the region.

Sturman & Zawar-Reza (2002) were employed back trajectories in atmospheric modeling techniques to identify and define fresh air zones within urban areas. By analyzing the backward movement of air masses, the study aimed to delineate regions within cities that experience relatively cleaner and less polluted air. This approach provided valuable insights into the spatial distribution of air quality and informed urban planning and policy decisions aimed at preserving and enhancing air quality in densely populated areas. The integration of back trajectories and atmospheric modeling demonstrated its potential in promoting healthier and more sustainable urban environments.

Generally, back trajectories are used in air pollution studies to identify the source regions of pollutants and to understand the transport of air masses. Back trajectories are calculated by tracking the movement of air masses in reverse from a receptor site (where the air pollution is being measured) to the potential source regions. This information can be used to inform policies and regulations aimed at reducing emissions in those areas.

2.2.1. Literature Review in Turkey

The HYSPLIT model has been extensively utilized in studies conducted in Turkey. Forty-three studies using HYSPLIT in Turkey were examined, and information on model inputs is given in Table 2. Among these studies, the NCEP/NCAR Reanalysis archive was employed in five of them; GDAS (Global Data Assimilation System) was used in nineteen; WRF-ARW was used in one; and the FNL (Final) archive was utilized in two studies. The isentropic method was chosen as the preferred approach in five studies, while the model vertical velocity method was selected in eleven studies. However, the meteorology archive used in sixteen studies was not explicitly specified; and in twenty-seven studies, the vertical velocity method was not specified.

When considering the data presented in Table 2, which examines the studies conducted in Turkey, it becomes evident that the most commonly used meteorology archives are GDAS and NCEP/NCAR Reanalysis. Additionally, the preferred vertical velocity methods are the model vertical velocity and the isentropic method.

As observed in Table 2, it is noteworthy that sixteen of the forty-three studies examined did not provide any information regarding the meteorology archive utilized. Similarly, in more than half of the studies, there was no mention of the vertical velocity method employed. Since differences due to model inputs affect the back trajectories, information on which inputs were used in the studies should be provided. Otherwise, it is not possible to compare the results with other studies or to continue the study for a specific region.

The absence of information about both the meteorology archive and the vertical velocity method raises questions regarding the potential effects these factors may have on the outputs and conclusions of the studies.

Table 2. Studies conducted in Turkey using HYSPLIT back trajectories

Study	Subject	Period	Trj. Days	Location	Met. Archive	Vertical Velocity Method	Arrival Height
(Dinçer et al., 2003)	SO_2	2000	24-hr	Izmir	–	–	–
(İm et al., 2006)	O_3, NOx, VOC	2001-2003	2-days	Istanbul	FNL	Model Vertical Velocity	500 m
(Im et al., 2008)	O_3, NOx, VOC	2001-2005	24-hr	Istanbul	FNL and GDAS	Model Vertical velocity	–
(Alp & Hanedar, 2009)	O_3	2001-2004	–	İstanbul	–	–	100 m, 500 m, 1000 m, 1500 m
(Koçak et al., 2009b)	PM_{10}	2001-2002	3-days	Erdemli, Mersin	–	–	1000 m, 2000 m, 3000 m, 4000 m
(Öztürk, 2009)	PM_{10}	1993-2001	5-days	Antalya	Reanalysis	Isentropic	100 m, 500 m, 1500 m
(Karaca et al., 2009)	PM_{10}	2009	3-days	Istanbul	–	–	1500 m
(Genç Tokgöz & Tuncel, 2008, 2011)	Particulate matter	2008	5-days	Istanbul	–	–	500 m, 1000 m, 1500 m
(Uygur et al., 2010)	Rain-water samples	2007-2008	5-days	Istanbul	GDAS	Model vertical velocity	1500 m
(Genç Tokgöz & Tuncel, 2008, 2011)	Particulate matter	2006-2007	5-days	Northwest of Turkey	Reanalysis	Isentropic	100 m, 500 m, 1500 m

Table 2 cont. Studies conducted in Turkey using HYSPLIT back trajectories

Study	Subject	Period	Trj. Days	Location	Met. Archive	Vertical Velocity Method	Arrival Height
(Zemmer et al., 2012)	Ragweed polen	2007	5-days	Istanbul	GDAS	Model Vertical velocity	100 m
(Papayannis et al., 2012)	Volcanic ash	2010	7 to 13-days	Istanbul, Athens	GDAS	–	–
(Ozdemir et al., 2012)	Particulate matter	2009	96-hr	Istanbul	GDAS0.5	Model Vertical velocity	–
(Im et al., 2013)	O_3	2007-2009	3-days	Istanbul	–	–	1000 m
(Uygur & Saral, 2013)	Particulate matter	2007-2008	2-days	Istanbul	–	–	500 m,
(Kuzu et al., 2013)	Particulate matter	2009-2010	–	Istanbul	GDAS	Model Vertical velocity	500 m
(Oğuz & Dündar, 2014)	Dust transport	2013	2-days	South of Turkey	GDAS1	Model vertical velocity	10 m, 50 m, 1500 m
(Toros et al., 2014)	PM_{10}	2019	72-hr	Istanbul	GDAS	Model Vertical velocity	–
(Ağaç, 2016)	PM_{10}	2012-2014	3-days	Istanbul	–	–	1500 m
(Sari et al., 2016)	Ozone	2013-2014	72-hr	Biga Peninsula	GDAS0.5	–	500 m
(Ünal, 2016)	Dust transport	2015	3-days	Diyarbakir	–	–	100 m, 500 m, 1500 m

Table 2 cont. Studies conducted in Turkey using HYSPLIT back trajectories

Study	Subject	Period	Trj. Days	Location	Met. Archive	Vertical Velocity Method	Arrival Height
(Topuz & Karabulut, 2017)	Dust transport	2015	3-days	Antakya	Reanalysis	Model vertical velocity	500 m, 1500 m, 3000 m
(Anil et al., 2017)	Rain-water samples	2009	5-days	Istanbul	GDAS	Model vertical velocity	1500 m
(Balcılar, 2018)	Particulate matter	2011-2012	5-days	Eastern Black Sea	Reanalysis	Isentropic	100 m, 500 m, 1500 m
(Kasparoglu et al., 2018)	O_3, NO, NO_2	2013-2016	3-days	Istanbul	GDAS0.5	–	500 m
(Alan et al., 2019)	Ragweed polen	2015-2016	2-days	Zonguldak	GDAS	–	50 m
(Özdemir, 2019)	PM_{10}	2012	48-hr	Central Mediterranean	GDAS1	–	10 m, 1500 m, 3000 m
(Celenk, 2019)	Ragweed polen	2014	2-days	Bursa	GDAS1	–	500 m
(Baltaci et al., 2020)	PM_{10}	2007-2017	72-hr	Istanbul	GDAS1	–	–
(Dörter et al., 2020)	Rainwater sample	2013	–	Bolu, Black Sea Region	–	–	–
(Mutlu, 2020)	PM_{10}	2016	72-hr	Balikesir	GDAS1	Model Vertical velocity	500 m
(Rastgeldi Dogan & Yalcin, 2020)	PM_{10}	–	–	Sanliurfa	–	–	–

Table 2 cont. Studies conducted in Turkey using HYSPLIT back trajectories

Study	Subject	Period	Trj. Days	Location	Met. Archive	Vertical Velocity Method	Arrival Height
(Sari et al., 2020)	Ozone	2013-2015	72-hr	Biga Peninsula	WRF-ARW	–	500 m
(T. R. Dogan et al., 2021)	Particulate matter	2019	–	Southeastern Anatolia	–	–	500 m, 1000 m, 1500 m, 2000 m, 2500 m
(Çapraz & Deniz, 2021)	Particulate matter	2015	–	Istanbul	–	–	–
(Baltaci et al., 2022)	PM_{10}	2007-2017	72-hr	Canakkale	GDAS1	–	1000 m
(Baltaci & Ezber, 2022)	PM_{10}	2014-2019	72-hr	Southeastern Anatolia	GDAS1	–	–
(Oruc, 2022)	PM_{10}	2019-2020	72-hr	Kırklareli	GDAS1	–	1500 m
(Dogan Rastgeldi & Atbinici, 2022)	PM_{10}, SO_2	2010-2020	–	Southeastern Anatolia	–	–	500 m, 1000 m, 1500 m
(Yavuz et al., 2022)	Particulate matter	2019	96-hr	Istanbul	–	–	250 m, 500 m, 1000 m
(Eşsiz & Acar, 2023)	Snowstorm	2004	2-days	Canakkale	Reanalysis	Isentropic	–
(Kilic & Kilic, 2023)	Rainfall samples	2021	72-hr	Antalya	–	–	100 m, 1000 m, 1500 m
(Kilic & Pamukoglu, 2023)	Rain samples	2020	72-hr	Antalya	–	Isentropic	100 m, 1000 m, 1500 m

2.3. Trajectory Statistics

Back trajectory statistics are employed to display the direction and sources of air pollution at a receptor location. In this field, it is common to apply trajectory statistical methods like clustering (Harris & Kahl, 1990), potential source contribution function (PSCF) (Ashbaugh et al., 1985), and concentration weighted trajectory (CWT) (Seibert et al., 1994). The conditional probability that an air mass with a pollutant concentration above a threshold level will reach at a receptor area following its passage across a particular geographic area is called PSCF. The PSCF model combines chemical data with meteorological data by analyzing chemical concentrations and air trajectories to provide an indication of the regions where pollutants are emitted (Ashbaugh et al., 1985). Previous studies have been widely employed PSCF to locate potential sources (G. Dogan et al., 2010; Genç Tokgöz, 2013; G. Güllü et al., 2005; Karaca et al., 2008; Uygur et al., 2010). In CWT, the concentration values at the receptor location conferred to the corresponding back trajectories. Each grid cell's residency time is weighted according to its mean/logarithmic mean concentration that is computed (Zhou et al., 2004). In this study, among the trajectory statistics, Cluster analysis will be used to relate the atmospheric transport pathway and pollution sources with the receptor. Therefore, cluster analysis is discussed in detail.

2.3.1. Cluster Analysis

Cluster analysis is a common technique used in back trajectory analysis. This method is used to analyze the sources of air pollution and the transport of pollutants. It is based on the assumption that similar back trajectories are likely to originate from the same source and travel through similar pathways. With clustering analysis, back trajectories are categorized according to the similarities in transport velocity and directions. Each category created is called a cluster. While each cluster is similar in itself in terms of speed, direction, curvature and length, there are differences with other clusters. The method of analysis involves combining trajectories that are close to one another and using the mean trajectory to represent those groups called clusters. Trajectory differences within a cluster are decreased, whereas differences between clusters are increased. Cluster analysis is especially used to investigate the relationship between pollutant concentrations and back trajectories (Harris & Kahl, 1990).

Piñero-García et al. (2015) performed cluster analysis of back trajectories to investigate the transportation of radioactive aerosols in the southeastern region of Spain. The results revealed distinct clusters representing different pathways and dispersion patterns of the aerosols, providing valuable insights into their transport and potential impact on the study area.

Markou & Kassomenos (2010) used 4-days back trajectories clustered at 750m, 1500m, and 300m AMSL (above mean sea level) for 5-year period. Three clusters were chosen with respect to their lengths as SSM (Short-Slow-Moving), MM (Medium-Moving), LFM (Long-Fast Moving). Cluster analysis was applied again. The findings indicate that the number of back trajectories in the SSM category reduced with height raises, while LFM category increased.

In order to study the effects of air back-trajectories on aerosol optical characteristics in Hornsund, Poland, Rozwadowska et al. (2010) categorized the back trajectories into clusters based on the resembling velocity and direction of advection.

Baker (2010) performed cluster analysis on four-day back trajectories for the years 1998- 2001 to understand pollution meteorology affecting Birmingham, UK. K-means cluster analysis technique was used in this study.

Cabello et al. (2008) studied back trajectory differences and sensitivity to the meteorological archive. Seven years of back trajectories arriving in Southeast Spain at 300, 1500 and 500m are studied. Two different meteorological archives were used (NCEP/NCAR Reanalysis and FNL). Trajectory differences have grown linearly for a minimum up to 48-hr and have shown faster growth after 72-hr. K-means cluster analysis was applied. It was found that the number of clusters of every clustering scenario varies for the trajectories reaching 1500 and 3000 m. Mostly, the input meteorological data has a greater impact on trajectory membership to the detected flows than does the initial cluster centroid choice.

It can be said that cluster analysis is a subjective method, that is, it varies according to the choices made by the user. These are choice of trajectory technique, choice of distance, and choice of optimum cluster number.

2.3.2. Clustering Techniques

Cluster analysis consists of two types that are hierarchical and non-hierarchical (Borge et al., 2007a; Kassomenos et al., 2010a; Sirois & Bottenheim, 1995). Hierarchical technique builds a hierarchical nested clustering tree by determining how similar various types of data points are to one another. The lowest layer of a cluster tree is made up of the initial data points for various clusters, while the highest layer is the root node of a cluster. A cluster tree can be made using either top-down splitting or bottom-up merging. The merging technique gathers the two most similar points into a single cluster by measuring similarity. The cluster performs another calculation using new points. Up until all the data is gathered into a single cluster, this process is repeated. The merging algorithm calculates the distance between clusters to identify similarity. As the distance reduces, the degree of similarity raises. To form a cluster tree, the two closest points or clusters are combined (Cui, Song, & Zhong, 2021). Ward's method is a frequently used method among hierarchical clustering methods. By combining clusters with similar variances, Ward's method creates homogeneous clusters. It is assumed that the variables have a multivariate normal distribution and that clusters have a tendency to be distinct and of comparable size (Shannon, 2007).

A non-hierarchical clustering technique is one where data points are directly assigned to distinct clusters without being grouped according to hierarchical categories. Using non-hierarchical clustering methods, data points are grouped according to their similarities or distance measures in order to optimize a clustering criterion. Among the nonhierarchical clustering algorithms are k-means clustering, Fuzzy C-means clustering, Gaussian Mixture Model (GMM), and Self-Organizing Maps (SOM).

The k-means clustering algorithm has been widely used among clustering algorithms because of its simplicity and efficiency (Hartigan & Wong, 1979). In a comprehensive review of literature

covering the period from 1980 to 2019, the clustering methods employed in air pollution studies were analyzed (Govender & Sivakumar, 2020). The findings of this study revealed that the k-means technique was utilized in 70% of the conducted studies aimed at identifying pollutant sources and analyzing the routes of air trajectories. Based on the studies reviewed, it is evident that the k-means clustering technique has a wide range of uses for back trajectories. The K-means approach was first developed as an iterative categorization process utilizing a predefined number of sample trajectories, or "seeds" (Dorling et al., 1992b). In order to be used, this approach necessitates seeding with initial values for the clusters' centers, which requires foreknowledge of the number of clusters, k. It has been demonstrated that these initial seed values play a significant role in determining how data are ultimately assigned to clusters. In other words, the initial seed selection for the value of cluster centers has a significant impact on k-means clustering (Peña et al., 1999).

In fuzzy C-means clustering technique, a cluster is given a point based on the distance between the data point and the cluster center (Menze et al., 2015). The further a data point is from a cluster center, the lower its membership value becomes. The membership summation for every data point should equal one (Baid et al., 2016).

In GMM, every cluster is regarded as a mean- and variance-containing generative model. To find the variance and mean of probability distributions, mixture models are employed (Baid et al., 2016).

SOM is a type of self-organizing neural network recommended by Kohonen (1990). The network involves an input and output layer. The input layer's number of neurons is determined by the input network's input layer's vector's dimensions (Tawadrou & Katsabani, 2005). Each output neuron in the clustering analysis corresponds to a single cluster. The same cluster is formed when various signals activate the same excitatory neuron (Cui, Song, & Zhong, 2021).

Generally, different cluster analysis techniques have been utilized to categorize back trajectories based on meteorological conditions and identify common atmospheric circulation patterns.

Cui et al. (2021) compared clustering techniques, namely, k-means, hierarchical, and SOM. Authors concluded that, K-means and SOM are superior to the hierarchical and HYSPLIT models. Similar to other hierarchical clustering methods, the HYSPLIT clustering method utilizes a bottom-up combining approach. However, diversely specific data pretreatment techniques are utilized.

The long-range transport and regional sources of $PM_{2.5}$ in Beijing between 2005 to 2010 were investigated by L. Wang et al. (2015). TrajStat program which employs Ward's method based on a hierarchical clustering algorithm was used to perform cluster analysis in this study. It is concluded that long-distance transport comprised one-third of the annual $PM_{2.5}$ contribution.

Kassomenos et al. (2010) used 5-days model vertical velocity back trajectories used to compare different clustering algorithms (k-means, hierarchical, and SOM) for 4-years time period. Model vertical velocity back trajectories of air masses arrived at receptor site at 12.00 UTC, in three various elevations. HYSPLIT was used to simulate back trajectories. The findings suggest that although all three clustering techniques demonstrate a dependence on arrival height, the extent of this correlation varies significantly among them. The results indicated that fast-moving trajectories had the strongest correlation with arrival height in hierarchical clustering, with SOM showing a lower level of dependence in comparison. The analysis revealed that k-means exhibited the lowest level of dependence on arrival height among the considered clustering techniques.

Karaca & Camci (2010) examined 5-day back trajectories in Istanbul for episodic events in 2008. Self-Organizing-Maps (SOM) technique was used for clustering analysis. Eight cluster groups were obtained. These data were analyzed and interpreted together with the measured PM_{10} data. Using an artificial neural network structure, SOM is used to represent input data in a lower-dimensional manner.

Borge et al. (2007) used 4-days model vertical velocity back trajectories for 3 years time period. A two-stage clustering algorithm was used to examine the back trajectories that reached three different locations. The study's objective was to examine the transport patterns of PM_{10} in an urban environment. Two-stage atmospheric clustering included non-hierarchical k-means. As a result of

the second stage of the analysis, specific trajectory clusters originated from North Africa for Madrid and Athens.

3. MATERIAL AND METHODS

In this chapter, the methodology of the study was described. The study region and the input data utilized were explained.

3.1. Study Area and Data Description

In this study, the NCEP/NCAR Reanalysis and GDAS1 meteorology archives were utilized. These archives were chosen due to their widespread usage and relevance within the context of studies conducted in Turkey. Additionally, model vertical velocity and isentropic methods were used since these are the most commonly used vertical velocity methods in Turkey. Therefore, it is aimed to calculate the back trajectories using these two different archives and 2 different vertical velocity methods and to numerically show the differences between the back trajectories. For this purpose, it was necessary to select a receptor site to simulate back trajectories. In the location selection, the Mediterranean region, which is the region where LRT studies are most studied, was chosen. Figure 4 presents a visual representation of the study area under examination.

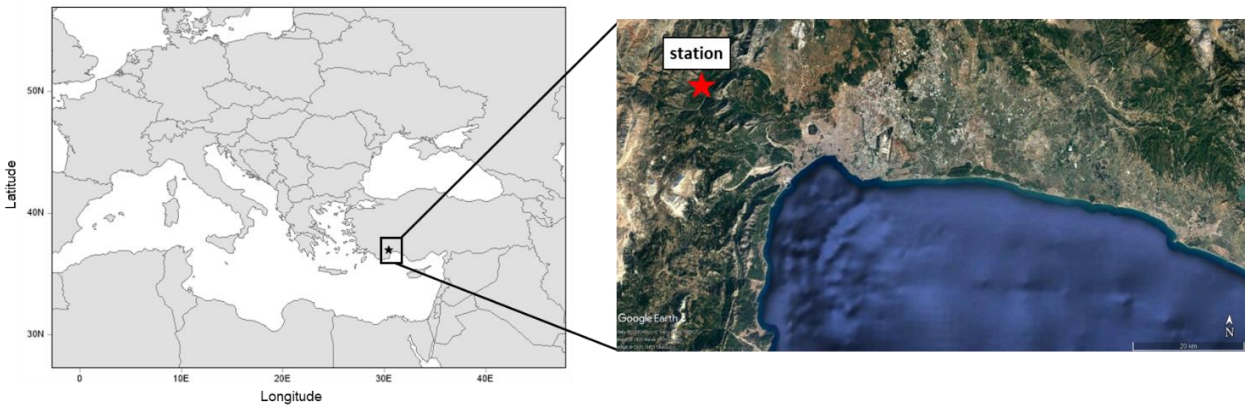


Figure 4. The study area *(*Google Earth, *2023)*

96-hr back trajectories were calculated one time a day (12 UTC) from January 1 2010 to December 31 2013, using HYSPLIT-web version. The starting location for the trajectories is in the Eastern Mediterranean (36^0 58’ 12” N and 30^0 26’ 2” E) at an elevation of 1500 m AGL.

The reason for choosing the period between 2010 and 2013 is that we have the chemical composition of PM measured in that range and at that point. 1500 m AGL was chosen due to the decrease in the potential effects of topographic changes on lower arrival heights, and commonly used in long-distance transport. 12 UTC selection was due to the particulate matter sampling is often conducted daily in many long-distance transport studies. By selecting 96-hr back trajectories, it was aimed to compare them with studies in the literature.

As stated above, the most used archives are Reanalysis and GDAS meteorology archives. Therefore, in this study, the effects of different inputs on trajectories were determined by using these meteorological archives. The details of meteorological archives used in this study are explained in the following paragraphs.

The production purpose of reanalysis to create new atmospheric analyses based on previously collected data (*ARL - Global Reanalysis Data Archive,* 2023). Reanalysis meteorology archive has a resolution of T62 (209 km). Six-hour interval results are provided. More than 80 variables are involved, such as the geopotential height, temperature, relative humidity, u-wind, v-wind and w-wind components, etc. in multiple different coordinate systems, such as 17 pressure level stack on 2.5 by 2.5 degree grids, 28 sigma level stack on 192 by 94 Gaussian grids, and 11 isentropic level stack on 2.5 by 2.5 degree grid. Radiative heating, convective heating, and precipitation rate variables are also provided (*NCEP/NCAR Global Reanalysis Products, 1948-Continuing - Dataset - DASH Search - Production*, 2023). Most of the Reanalysis output is in GRIB-1 which is a WMO standard. (*Reanalysis Data Sources*, 2007). The detailed information of the NCEP/NCAR Global Reanalysis Data Archive is presented in **APPENDIX A.**

Key strengths of Reanalysis meteorology archive are that it is a globally available data set, the long-termed reanalysis utilized rawinsonde data, and used in numerous studies therefore it serves as a guide for many calculations. The Reanalysis meteorology archive has several drawbacks, including being a model that is outdated, having low temporal and spatial moisture variability across the oceans, and performing relatively poorly in the Southern Hemisphere (Kalnay et al., 1996; Kistler et al., 2001).

The NCEP model output is utilized for air quality transport and dispersion modeling at NOAA's Air Resources Laboratory. The GDAS is one of NCEP's operational systems. Global Data Assimilation System, (ftp://arlftp.arlhq.noaa.gov/pub/archives/gdas1/). The GDAS archive is a latitude/longitude global grid with a resolution of 1 degree (*NOAA - Air Resources Laboratory* -, 2019) GDAS archive include main fields such as the u-wind, v-wind, and w-wind components, temperature, and humidity. W-wind is the vertical velocity wind field. Four times a day, at 00, 06, 12, and 18 UTC, the GDAS is executed. The analysis time and the 3, 6, and 9-hour forecasts are included in the model output. Through NCEP post-processing of the GDAS, the data is transformed from sigma levels to required pressure levels and from the spectral coefficient form to 1-degree latitude-longitude (360 by 181) grids. GRIB format is used for model output (*ARL - Global Reanalysis Data Archive*, 2023). The detailed information of the GDAS1 data set is presented in **APPENDIX B.**

There are differences between meteorology archives. These differences arise from spatial and temporal resolution and data coverage. The NCEP/NCAR Reanalysis has a coarser spatial resolution, 2.5 degrees latitude by 2.5 degrees longitude (≈ 278 km), GDAS has a higher spatial resolution, 1 degrees latitude by 1 degrees longitude (≈ 111 km). In addition, while a reanalysis archive data includes one month, GDAS1 data includes one week. Reanalysis archive is available from 1948 to present, when GDAS1 archive is available from December 2004 to present. Besides, the atmosphere is a continuous, three-dimensional, time-varying field. The atmosphere is not accurately represented by gridded meteorological fields. A single trajectory is impacted by variations in the horizontal grid, vertical levels, time intervals, and vertical velocity.

As mentioned earlier, the most used vertical velocity methods are isentropic and model vertical velocity. The trajectory model has several options that control the computation. According to HYSPLIT, for the majority of applications, it is necessary to leave some of them at their default values. To illustrate, calculations should utilize the vertical motion field (model vertical velocity) that is included in the data file except modeling the flow along constant pressure (isobaric) or constant theta (isentropic) surfaces is required (*HYSPLIT Trajectory Model Configuration*, 2007).

Model vertical velocity option utilized the vertical velocity field in the meteorological data file. In this method, the trajectory moves with the vertical velocity wind fields (Draxler, 1996).

Potential temperature is utilized in the isentropic assumption to limit the vertical motion of trajectories (Stohl, 1998). Conservation of potential temperature, as assumed for isentropic trajectories, is fulfilled only for adiabatic and inviscid motions (Stohl & Wotawa, 1995). An idealized thermodynamic process that is adiabatic and reversible is called an isentropic process in thermodynamics. Heat or matter are not transferred (Horlock, 1967). The presumption of isentropic flow is no heat is added to or removed from the flow. Since it is assumed that the flow is thermodynamically reversible, all viscous and other irreversible effects are disregarded (*One-Dimensional Isentropic Flow*, 2006). The US National Meteorological Center (NMC) gridded analyses' three-dimensional wind field is "collapsed" into two dimensions by vertically interpolating the winds to isentropes, or surfaces with constant potential temperature.

The slope of the isentropes thus implies dry adiabatic vertical motion. The main advantage of this model is that it can take vertical shears into account as the isentrope samples winds at various heights. However, its main disadvantage is that in the presence of significant gradients in static stability, such as fronts, isentropic surfaces become undefinable. Because of this issue, isentropic trajectories frequently finish before they are supposed to, providing an inadequate representation of atmospheric transport (Kahl et al., 1989).

The trajectory is kept on a constant pressure surface when using the isobaric option. In constant density option, trajectories remain on surfaces with constant density. In isosigma option, there is no vertical motion, and the trajectory maintains its internal sigma. In “from divergence” option, by vertically integrating the velocity divergence, the vertical motion is calculated (NOAA Air Resources Laboratory, 2007).

In total, 1461 back trajectories were simulated for each trajectory data set, representing the air mass transport patterns over the specified study period. The description of the trajectory data set generated in this study can be found in Table 3, providing information about the simulated trajectories. Note that numbers (i.e. 1 and 2) refers to meteorological archive and the letters (i.e. a

and b) refers to vertical motion method. The NCEP/NCAR Reanalysis archive represented by the number 1, GDAS1 archive represented by 2; the model vertical velocity method presented by the letter a, and isentropic method represented by the letter b.

Table 3. Trajectory data set simulated in this study

Symbol of data	**Data Description**		**Total number of back trajectories**
	Meteorology Archive	**Vertical Velocity Method**	
1a	NCEP/NCAR Reanalysis	Isentropic	1461
1b	NCEP/NCAR Reanalysis	Model vertical velocity	1461
2a	GDAS1	Isentropic	1461
2b	GDAS1	Model vertical velocity	1461

The 1a-1b comparison was used to determine the Reanalysis archive's sensitivity to the velocity method. To identify the sensitivity of the GDAS1 archive to the velocity method, 2a-2b comparison was used. Using 1a-2a and 1b2b comparisons, sensitivity to the meteorological archives was determined. Lastly, 1a-2b and 1b-2a comparisons were utilized to identify the sensitivity to both meteorological archive and velocity methods. Descriptions of the comparisons are included in Table *4*.

Table 4. Descriptions of the comparisons

Comparison	**Description**
1a-1b	Sensitivity of the Reanalysis archive to the velocity method
2a-2b	Sensitivity of the GDAS1 archive to the velocity method
1a-2a	Sensitivity to meteorology archive
1b-2b	
1a-2b	Sensitivity to both meteorology archive and velocity method
1b-2a	

3.2. Methodology

The methods employed in this investigation are presented based on the background data provided. Initially, various model inputs were used to calculate back trajectories. The back trajectories were subjected to a sensitivity analysis. Then, cluster analysis was applied to back trajectories. Sensitivity analysis was performed to cluster centers with different datasets. Seasonal variation was also examined. In addition, pollution data and the results of cluster analysis were combined. Statistical tests were applied to the combined pollution data. The methodology of this study is given in Figure 5.

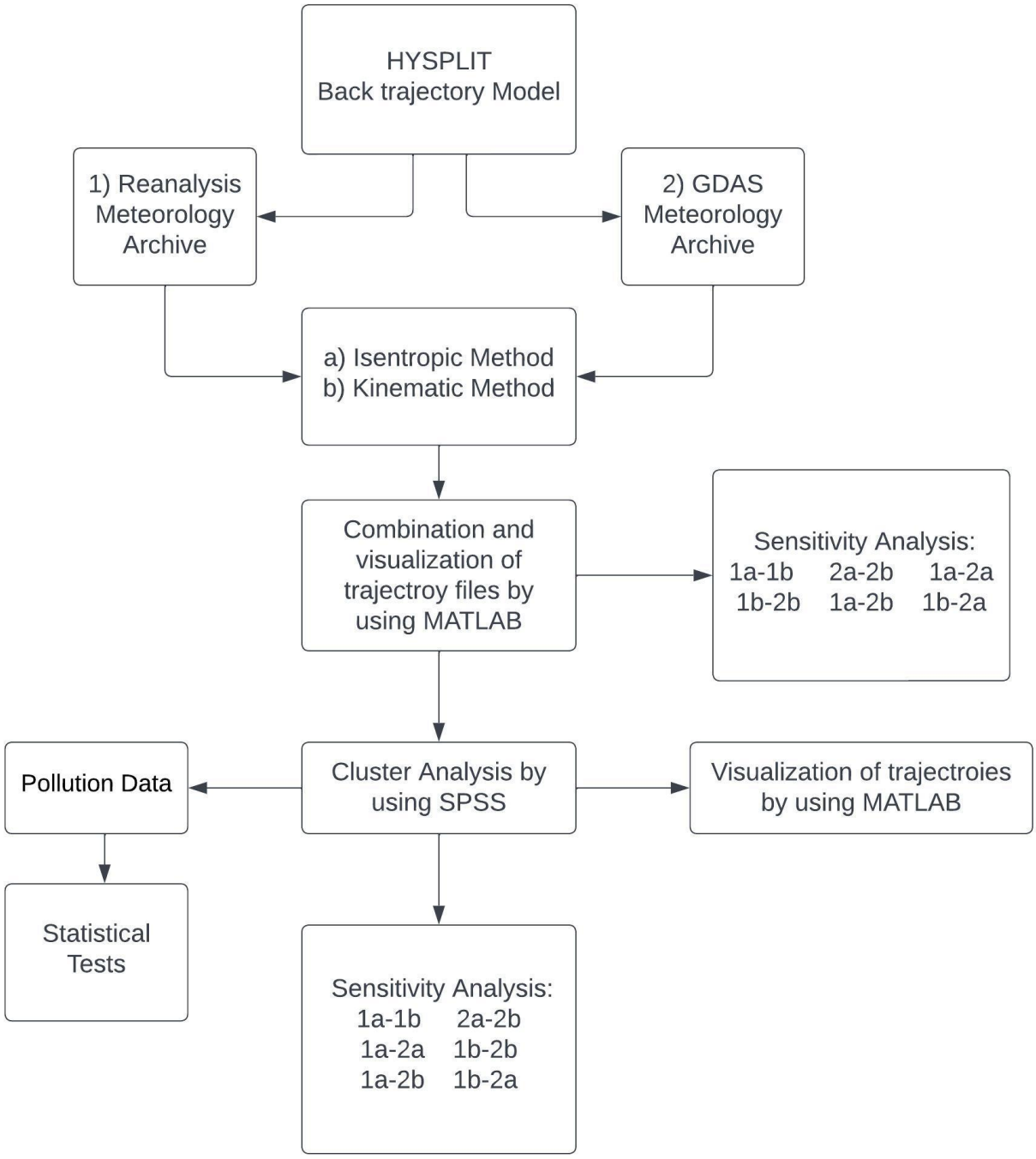


Figure 5. Methodology of the study

3.2.1. Sensitivity Analysis

In this study, sensitivity analysis was applied to the back trajectories. Previous research has utilized slightly different statistical metrics of trajectory sensitivity than the ones presented here (Harris et al., 2005; Rolph & Draxler, 1990; Stohl & Wotawa, 1995).

To evaluate the impact of the model inputs, sensitivity analyses were conducted on the trajectories. The focus of these analyses was to assess the differences resulting from variations in the model inputs. In order to determine these differences, the absolute horizontal transport deviation (AHTD), the absolute vertical transport deviation (AVTD), and the relative horizontal transport deviation (RHTD) were calculated. These parameters were calculated using the following equations (Stohl & Wotawa, 1995):

$$\mathrm{AHTD(t)} = \frac{1}{\mathrm{N}}\sum_{\mathrm{n=1}}^{\mathrm{N}}\{[\mathrm{X_n(t)} - \mathrm{x_n(t)}]^2 + [\mathrm{Y_n(t)} - \mathrm{y_n(t)}]^2\}^{1/2} \qquad \text{(Equation 2)}$$

$$\mathrm{AVTD(t)} = \frac{1}{\mathrm{N}}\sum_{\mathrm{n=1}}^{\mathrm{N}}|\mathrm{Z_n(t)} - \mathrm{z_n(t)}| \qquad \text{(Equation 3)}$$

$$\mathrm{RHTD(t)} = \frac{1}{\mathrm{N}}\sum_{\mathrm{n=1}}^{\mathrm{N}}\frac{\{[\mathrm{X_n(t)} - \mathrm{x_n(t)}]^2 + [\mathrm{Y_n(t)} - \mathrm{y_n(t)}]^2\}^{1/2}}{\mathrm{AL_n(t)}} \qquad \text{(Equation 4)}$$

$$\mathrm{AL_n(t)} = \frac{1}{2}\sum_{\mathrm{t_i=2}}^{\mathrm{t}}\langle\{[\mathrm{X_n(t_i)} - \mathrm{X_n(t_{i-1})}]^2 + [\mathrm{Y_n(t_i)} - \mathrm{Y_n(t_{i-1})}]^2\}^{1/2} + \{[\mathrm{x_n(t_i)} - \mathrm{x_n(t_{i-1})}]^2 + [\mathrm{y_n(t_i)} - \mathrm{y_n(t_{i-1})}]^2\}^{1/2}\rangle \qquad \text{(Equation 5)}$$

In this equations, N is the total number of back trajectories, X,Y,Z is the location of the first set of back trajectories; x,y,z is the location of the second set of back trajectories; ALn(t) represents the average length of the two back trajectories compared (Stohl & Wotawa, 1995). In this study, 1461

back trajectories were simulated for each data set, therefore the N value given in Equations 1,2 and 3 was taken as 1461.

The formula given in Equation 2 (also in Equations 4 and 5) is actually the Euclidean length. Considering that the Earth is spherical, it would be incorrect to calculate the distance between two locations directly by the Euclidean length. When the studies on back trajectory analyses were examined, it was seen that the distance calculated in Equation 2 were replaced by the distance calculated by the Haversine formula (Cabello et al., 2008; Markou & Kassomenos, 2010). The Haversine formula is utilized to compute the shortest distance between two points with known latitude and longitude values on the earth's geolocation. The Haversine formula introduces a small error of approximately 0.3% assuming the Earth is a perfect sphere, making it a suitable choice for distance calculations in trajectory analyses. Therefore, the distance between two locations in Equations 2, 4 and 5 were calculated using the Haversine formula in Equation 6:

$$a = \sin^2((\varphi B - \varphi A)/2) + \cos \varphi A * \cos \varphi B * \sin^2((\lambda B - \lambda A)/2)$$
$$c = 2 * \operatorname{atan2}(\sqrt{a}, \sqrt{(1-a)})$$
$$d = R \cdot c \qquad \text{(Equation 6)}$$

In this equation, "φ" and "λ" represent latitude and longitude (in radians), respectively. "B" represents the second position and A represents the first position. "R" represents the mean radius of the earth (6371 km taken). "d" is the distance (also called as the Great Circle Distance (in km)) (Stohl & Wotawa, 1995).

These parameters allow researchers to quantify the influence of different meteorological data sets and vertical velocity methods on the simulated trajectories. These formulas were applied to all the trajectory data sets (in Table *4*) using MATLAB with a script that I created. The script that used to calculate sensitivity analysis parameters is in **APPENDIX C.**

3.2.2. Cluster Analysis

Different techniques are used in cluster analysis, the most used among these techniques is the k-means technique. Therefore, in this study, cluster analysis was performed with the k-means technique.

For cluster analysis, the non-hierarchical k-means technique was performed by using IBM SPSS Statics 23 (IBM, 2023). Different methods are used to cluster trajectories. To select the optimum number of clusters, the percent change in total-root-mean-square-deviation (TRMSD) between clusters is evaluated (Dorling et al., 1992b). Besides, total spatial variance (TSV) technique is used to cluster trajectories.(Draxler et al., 1998; Stein et al., 2015). In this study, TRMSD is used. The threshold is taken as 5% change in TRMSD (Dorling et al., 1992b).

When performing k-means cluster analysis, it is necessary to select seeds. A seed is basically a starting cluster centroid. Seeds are either determined based on foreknowledge or selected at random. The robust solution method was applied in order to identify stable centroid positions by Kassomenos et al. (2010). From a randomly selected data set (GDAS1-Isentropic) among four data sets, the first 50 trajectories were randomly selected and used as seeds. To ensure the stability and consistency of the results, cluster analysis was performed multiple times with varying numbers of seeds. The number of seeds was progressively decreased from 50 to 40, 30, 20, and finally 10. The purpose of this approach was to observe if the results would remain unchanged even with a reduced number of seeds. The impact of seed selection on the outcome of the cluster analysis was monitored. The robust solution includes determining the point at which further reduction in the number of seeds would no longer affect the results significantly. Once a stage where the choice of seeds no longer influenced the outcomes was reached, it was decided that the cluster analysis results obtained with 50 seeds would be used for further analysis and interpretation. Figure 6 shows the results of k-means analysis applied to a data set selected as an example (GDAS1-model vertical velocity), with different number of seeds. As seen in Figure 6, no difference was observed in the optimum number of clusters as the number of seeds decreased. This result is the same for other data sets. By conducting this seed selection analysis, the robustness and reliability of k-means cluster analysis results were ensured. Only latitude and longitude were used in the cluster analysis, but not height.

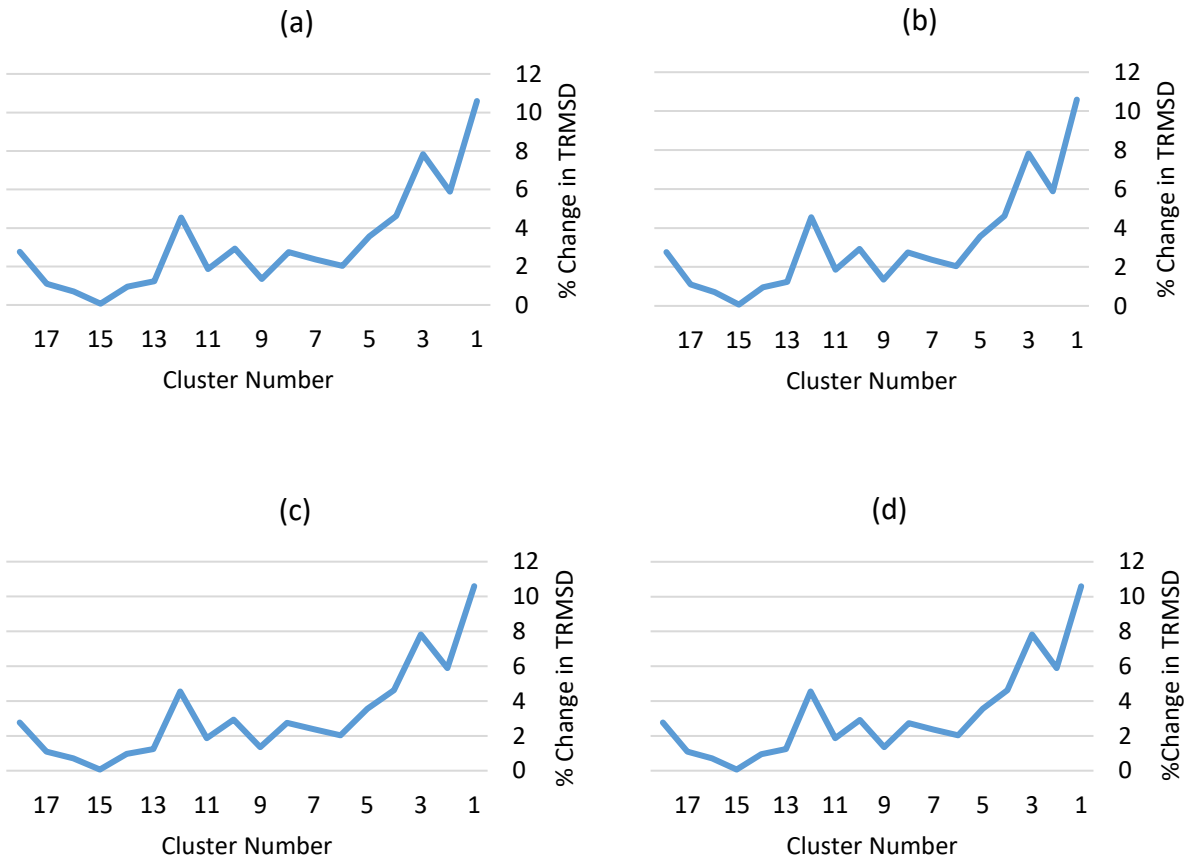


Figure 6. K-means cluster analysis GDAS1-Model vertical velocity TRMSD results (a) 50 seeds, (b) 40 seeds, (c) 30 seeds, (d) 20 seeds

3.2.2.1. Similarity Algorithm

Clustering analysis was applied utilizing the k-means technique on the back trajectories calculated with different data sets. The similarity algorithm was utilized to determine the similarities of clusters. The similarity index was used to check how similar, that is, the same, the back trajectories represented by the cluster centers were in each data set.

Similarity algorithm (SA) is a statistical tool used to quantify the degree of similarity between two variables or data sets. The similarity between the two cluster is calculated with Equation 7, where

"n_i'" represents the number of identical trajectories in corresponding clusters (Cui, Song, Zhong, et al., 2021).

$$S_A = \frac{\sum_{i=1}^{K} n_i'}{n} \times 100\% \qquad \text{(Equation 7)}$$

3.2.2.2. Pollution Data

In order to understand how the back trajectories obtained with different inputs affect the source-receptor relationship, it is necessary to combine the trajectories with pollution data. Back trajectories are generally used to identify sources of long-distance transported pollutants (Borge et al., 2007a; Cabello et al., 2008; Cape et al., 2000; Sirois & Bottenheim, 1995; Y. Q. Wang et al., 2004). Therefore, in this study, a secondary pollutant, namely sulfate (SO_4^{-2}), is chosen to relate the atmospheric transport pathways (i.e. clusters) and source regions. In the receptor site, daily particulate matter (PM) samples were collected between 2010 and 2013 years using a Gent-PM_{10} stacked filter unit (SFU), which collects coarse (10 μm>d>2.5 μm, $PM_{10–2.5}$) and fine ($PM_{2.5}$) fractions separately. Sulfate concentration in all collected samples were analyzed by Dionex DX-120 Model Ion Chromatography (IC). The details of the PM sampling and analysis can be found in Genç Tokgöz (2023). In this study, sulfate concentration in PM_{10} i.e. sum of the fine and coarse fraction, is used to investigate the influence of back trajectories generated with different input variables on the interpretation of source-receptor relations.

3.2.2.3. Statistical Tests

The Kruskal-Wallis test and median test are used in this study.

3.2.2.3.1. Median Test

In this study, the Median test was used to test whether there is a seasonal (summer-winter) difference between comparisons consisting of different data sets. The median test investigates the equality of the medians on a given variable between two or more populations.

3.2.2.3.2. Kruskal-Wallis Test

In this study, the Kruskal-Wallis test was employed to the combined pollution data. The application of the Kruskal-Wallis test in this context aimed to assess test whether there was a difference between medians of clusters of back trajectories simulated with different model inputs.

The null hypothesis of the Kruskal-Wallis test assumes that the medians of all the groups, represented by the cluster centers, are equal. This implies that there are no systematic differences among the groups in terms of their central tendency. The alternative hypothesis, on the other hand, posits that at least one group's population median is different from the population median of at least one other group. By setting up these null and alternative hypotheses, the Kruskal-Wallis test allows for a statistical examination of whether there are meaningful differences in the medians of the groups.

4. RESULTS AND DISCUSSION

In this chapter, our different back trajectory data sets were compared with sensitivity analysis. The variation of sensitivity parameters with respect to time and season were explained. One of the trajectory statistics methods, namely Cluster Analysis were applied to the back trajectory data sets. The variation of transport pattern and the variation in the interpretation of the source-receptor relationship with respect to the different trajectory data sets were discussed.

4.1.Trajectory Comparison

Sensitivity analysis was conducted to quantify the differences between trajectories when different meteorology archives and vertical velocity methods were used to run the HYSPLIT model. As discussed in Chapter 3 (Methods), the most common meteorological archives are NCEP/NCAR Reanalysis (denoted by 1) and GDAS1(denoted by 2) while the most common vertical velocity methods are isentropic (denoted by a) and model vertical velocity method (denoted by b).

Back trajectories simulated by 1a, 1b, 2a, and 2b data sets (Figure *7*) were compared with each other by means of the sensitivity parameters: RHTD, AHTD, and AVTD. These differences were quantified for first 24-hr (i.e. 24-hr back trajectories), second 24-hr (48-hr back trajectories, third 24-hr (72-hr back trajectories), and fourth 24-hr (96-hr back trajectories) to examine the time evolution of the differences, and the results are presented in Table 5 to Table 8.

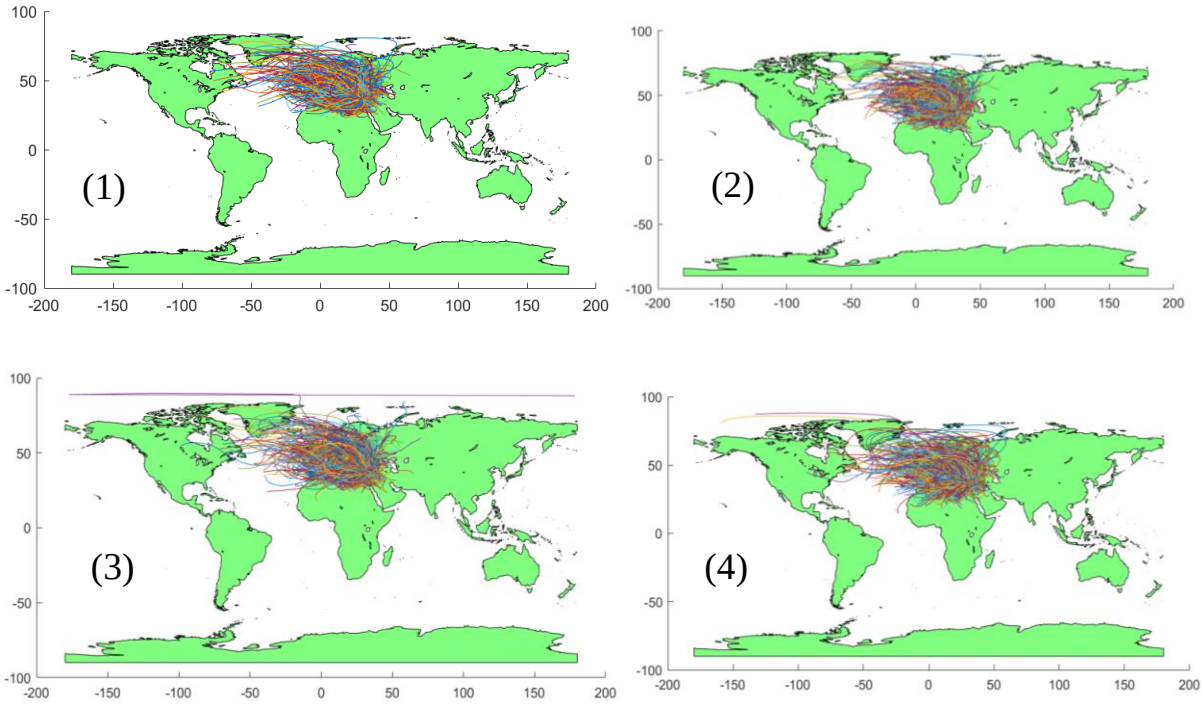


Figure 7. Back trajectories in each data set (1) 1a, (2) 1b, (3) 2a, (4) 2b

To investigate the effects of long-range transport of pollutants usually 3 to 5 days back trajectories are used (Abdalmogith & Harrison, 2005; Cabello et al., 2008; Moody et al., 1995; Sirois & Bottenheim, 1995; Y. Q. Wang et al., 2004). In the literature, the sensitivities of trajectories are generally reported up to 96-hr (4 days). Therefore, in this study the results of sensitivity analysis of 96-hr back trajectories will be discussed in more detail to provide comparison with the literature.

The results of the sensitivity analysis which is conducted at 96-hr is presented in Table 5. The sensitivity of trajectories to meteorological input (i.e. meteorological archive data set) for isentropic and model vertical velocity trajectories are 38% and 40%, respectively (comparison 1a-2a and 1b-2b). Harris et al. (2005) reported similar magnitudes (35% for isentropic and 40% for model vertical velocity trajectories) using NCEP/NCAR Reanalysis and ERA-40 data sets. Table 5 shows horizontal (AHTD), and vertical (AVTD) differences are greater in model vertical velocity trajectories than in isentropic trajectories. The relative increase in RHTD, AHTD, and AVTD statistics of model vertical velocity trajectories to isentropic trajectories are 7%, 25%, and 34%, respectively. Higher magnitudes of sensitivity statistics in model vertical velocity trajectories

indicate that although both isentropic and model vertical velocity trajectories are sensitive to the meteorological input, model vertical velocity trajectories are more sensitive, particularly in the vertical component of the trajectory.

The differences between trajectories concerning the vertical motion method are examined by comparing isentropic and model vertical velocity trajectories using NCEP/NCAR Reanalysis and GDAS1 data sets (comparison 1a-1b and 2a-2b). When NCEP/NCAR Reanalysis data is used, the differences between isentropic and model vertical velocity trajectories are less (RHTD value of 25%) than those generated with GDAS1 archive (RHTD value of 45%). The lowest horizontal deviation (AHTD value of 652 km) is calculated between isentropic and model vertical velocity trajectories when NCEP/NCAR Reanalysis archive is used. Harris et al. (2005) compared isentropic and model vertical velocity trajectories using NCEP/NCAR Reanalysis archive and reported similar values for all sensitivity statistics (29% (25% in current study), 705 km (652 km in current study), 1384 m (1262 m in current study) for RHTD, AHTD, and AVTD, respectively). When the GDAS1 archive is used, the horizontal and vertical deviations between isentropic and model vertical velocity trajectories become greater. The relative increase in all sensitivity statistics of isentropic versus model vertical velocity trajectories with GDAS1 to isentropic versus model vertical velocity trajectories with NCEP/NCAR Reanalysis is significantly high (79%, 93%, and 65% for RHTD, AHTD, and AVTD, respectively). This significant increase in sensitivity statistics indicates that trajectories simulated by the GDAS1 archive are significantly sensitive to the vertical motion method. It is not wrong to state that trajectories are robust concerning the vertical motion method when NCEP/NCAR Reanalysis archive is used.

Trajectories generated with different meteorological archives and vertical motion methods are also compared to determine sensitivity to both meteorological archives and vertical transport methods (1a-2b,1b-2a). All sensitivity statistics are high (42-45%, 1067-1240 km, 1558-1933 m for RHTD, AHTD, and AVTD, respectively) when both meteorological and vertical motion methods differ. The sensitivity statistics of GDAS1 to vertical transport method (2a-2b comparison) are higher than the 1b-2a comparison, indicating that differences in the meteorological archive and vertical transport method did not merge directly. Some differences cancel out because the summation of

the differences in comparisons 1a-2a,1b-2b and 1a-1b, 2a-2b resulted in bigger values than 1a-2b and 1b-2a.

Table 5. Back trajectory Comparison at 96-hr results

Comparison	Description	RHTD, %	AHTD, km	AVTD, m
1a-1b	Sensitivity of the NCEP/NCAR Reanalysis archive to the velocity method	24.8	652	1262
2a-2b	Sensitivity of the GDAS1 archive to the velocity method	44.5	1258	2077
1a-2a	Sensitivity to meteorology archive	37.8	940	1221
1b-2b		40.4	1171	1632
1a-2b	Sensitivity to both meteorology archive and velocity method	44.7	1240	1933
1b-2a		41.8	1067	1558

Table 6. Back trajectory Comparison at 72-hr Results

Comparison	Description	RHTD, %	AHTD, km	AVTD, m
1a-1b	Sensitivity of the NCEP/NCAR Reanalysis archive to the velocity method	17.3	390	969
2a-2b	Sensitivity of the GDAS archive to the velocity method	33.6	844	1797
1a-2a	Sensitivity to meteorology archive	29.9	639	1017
1b-2b		30.4	780	1418
1a-2b	Sensitivity to both meteorology archive and velocity method	32.9	830	1671
1b-2a		31.1	701	1279

Table 7. Back trajectory Comparison at 48-hr Results

Comparison	Description	RHTD, %	AHTD, km	AVTD, m
1a-1b	Sensitivity of the NCEP/NCAR Reanalysis archive to the velocity method	10.3	209	728
2a-2b	Sensitivity of the GDAS archive to the velocity method	21.2	480	1449
1a-2a	Sensitivity to meteorology archive	22.5	403	778
1b-2b		21.7	452	1134
1a-2b	Sensitivity to both meteorology archive and velocity method	22.9	481	1314
1b-2a		22	418	952

Table 8. Back trajectory Comparison at 24-hr Results

Comparison	Description	RHTD, %	AHTD, km	AVTD, m
1a-1b	Sensitivity of the NCEP/NCAR Reanalysis archive to the velocity method	3.9	76	477
2a-2b	Sensitivity of the GDAS archive to the velocity method	10.4	193	1091
1a-2a	Sensitivity to meteorology archive	12.9	199	556
1b-2b		12.5	206	846
1a-2b	Sensitivity to both meteorology archive and velocity method	13.13	224	970
1b-2a		12.8	197	667

Figure 8 and Figure 9 illustrates the distribution of sensitivity statistics for a time period of up to 96-hr, represented as box and whisker plots. The points above the box whisker plot are outliers. The bottom line of the box whisker plot (Q0) means it is the dataset's minimum data point, omitting outliers. The upper line of the box whisker plot (Maximum (Q4) means it is the highest data point in the data set, omitting outliers. The median (Q2) is the medium value in the data set. The line inside the boxes represents the median. The first quartile is shown as the box's top line (Q1), the bottom of the data set. The box’s bottom line (Q3) is the median of the upper half of the data set.

Consistent with the earlier findings, the horizontal deviations observed in the 1a-1b comparison exhibit a smaller range and are generally less pronounced. This suggests that the variations in horizontal positions between the trajectories in this comparison are relatively minor. However, when considering the distribution of vertical deviations, it becomes evident that all comparisons, except for 1a-1b and 1a-2a, exhibit a broader range. The notably lower vertical distribution observed in the 1a-1b and 1a-2a comparisons indicates that isentropic trajectories and trajectories simulated using the NCEP/NCAR Reanalysis archive are less prone to vertical deviation, indicating less sensitivity to the vertical motion method. The lower sensitivity of isentropic trajectories to the meteorological archive can be ascribed to the assumption to calculate the vertical position of the air mass. Isentropic trajectories are determined based on the assumption that an air mass is limited to an isentropic surface, meaning it follows a constant potential temperature. However, this assumption is only valid in dry adiabatic conditions; therefore, there will be uncertainties in the actual atmospheric conditions due to moisture (Stohl & Wotawa, 1995). By

employing this assumption, isentropic trajectory simulations effectively constrain the amount of noise or variability in the vertical velocity, resulting in lower sensitivity statistics for vertical positions.

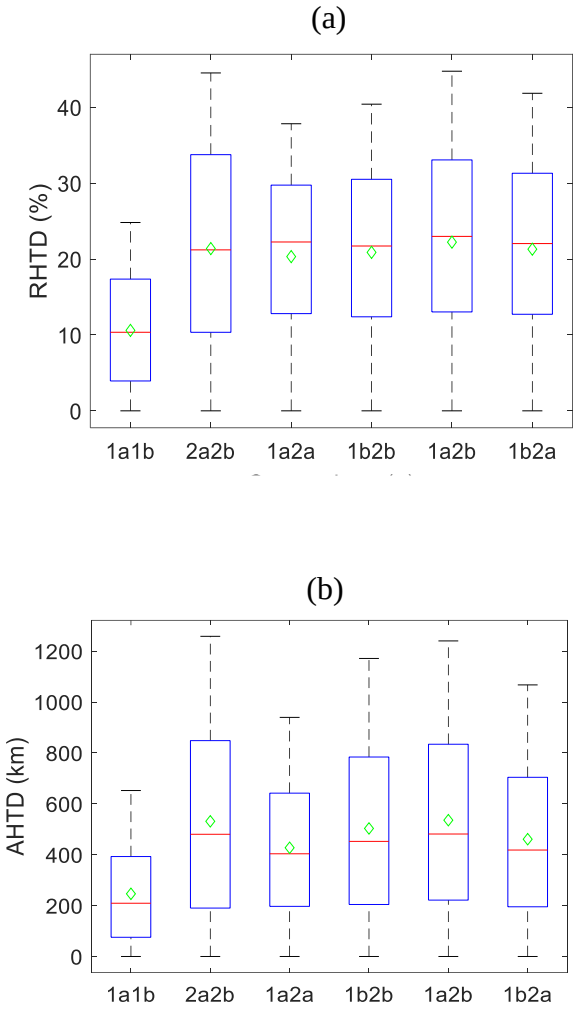


Figure 8. Distributions of trajectories up to 96-hr: (a) RHTD, (b) AHTD

(c)

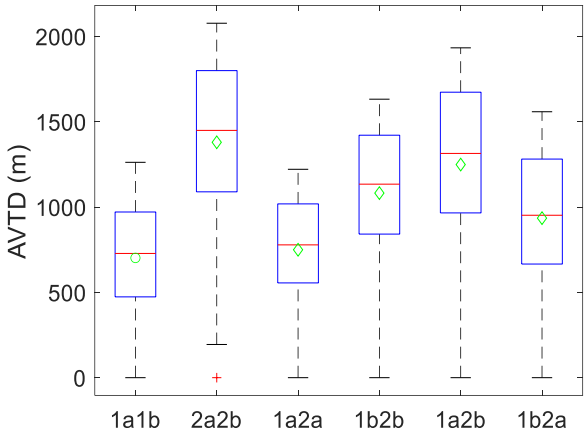


Figure 9. Distributions of trajectories up to 96-hr: (c) AVTD

The average elevation and length of 96-hr back trajectories are also calculated for each data set and presented in Table 9. The vertical position of trajectories is nearly similar in isentropic trajectories (1a and 2a). Model vertical velocity trajectories reach greater elevations compared to isentropic trajectories. The difference is quite noticeable for model vertical velocity trajectories simulated with GDAS1 than NCEP/NCAR Reanalysis archive (Figure 10). The higher temporal and vertical resolution of the GDAS1 archive compared to the NCEP/NCAR Reanalysis archive may better represent the vertical position of the air parcel in model vertical velocity trajectories (NOAA, 2023). The NCEP/NCAR Reanalysis archive's lower resolution might restrict the air parcel's vertical position. Consistent with the vertical position, model vertical velocity trajectories simulated by the GDAS1 archive travelled the longest distance (2952 km) among the average distance travelled by each data set given in Table 9. When a trajectory attains high elevation, it means there are high wind speeds. When high wind speeds accompany an air parcel, it travels faster both in the horizontal and vertical direction and will travel long distances.

Table 9. Average Trajectory Elevation and Length (mean values)

Data set	Elevation (m)		Length (km)	
	Isentropic (a)	Model Vertical Velocity (b)	Isentropic (a)	Model Vertical Velocity (b)
NCEP/NCAR Reanalysis (1)	2290	2264	2710	2637
GDAS1(2)	2238	2551	2632	2952

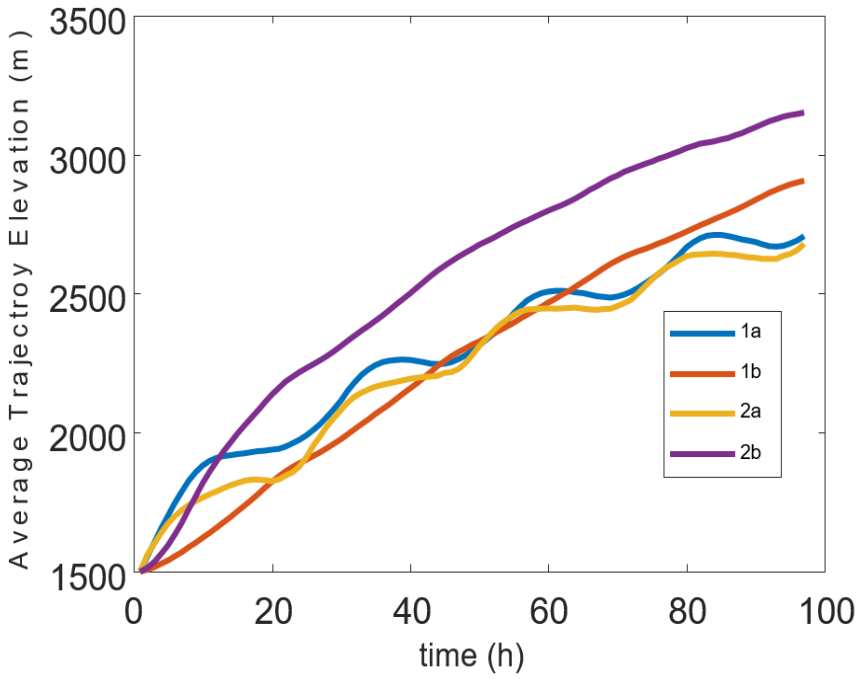


Figure 10. Average trajectory elevations

4.1.1. Time Evolution of Sensitivity Parameters

Sensitivity statistics were calculated hourly from time zero to 96-hr back in time, and results are presented in Figure 11, Figure 12, and Figure 13. It is well known that the uncertainty of a trajectory increases as it travels back in time. Consistent with this statement, all deviation statistics (i.e., RHTD, AHTD, and AVTD) for each comparison are found to increase with the total travel time, i.e. 96-hr back trajectories are more sensitive than preceding 72-hr, 48-hr, and 24-hr back trajectories (Figure 11, Figure 12, and Figure 13). The lowest horizontal deviation parameters

(RHTD and AHTD values) are calculated for 1a-1b comparison and confirm our previous statement of the robustness of trajectories generated with the NCEP/NCAR Reanalysis archive. Time variation of vertical deviation statistics (i.e. AVTD) is immediate and large for the first 6 h, and then continues to increase but at a slower rate. The lowest vertical deviations are calculated for 1a-1b and 1a-2a comparisons.

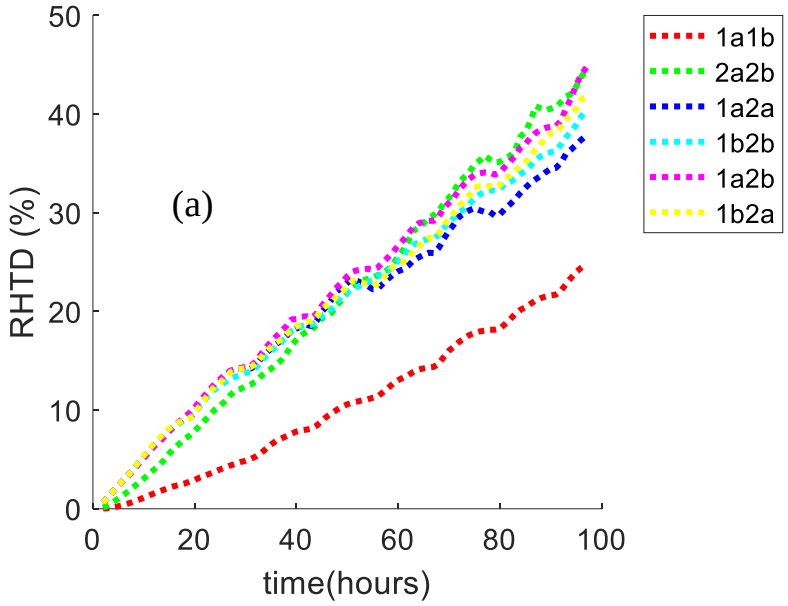


Figure 11. Time Evaluation of Sensitivity Parameters: (a) RHTD

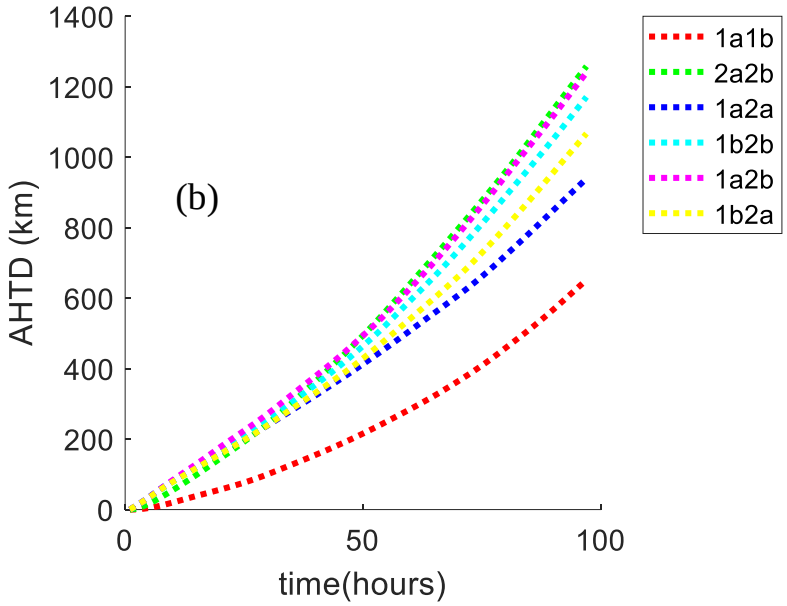


Figure 12.Time Evaluation of Sensitivity Parameters: (b) AHTD

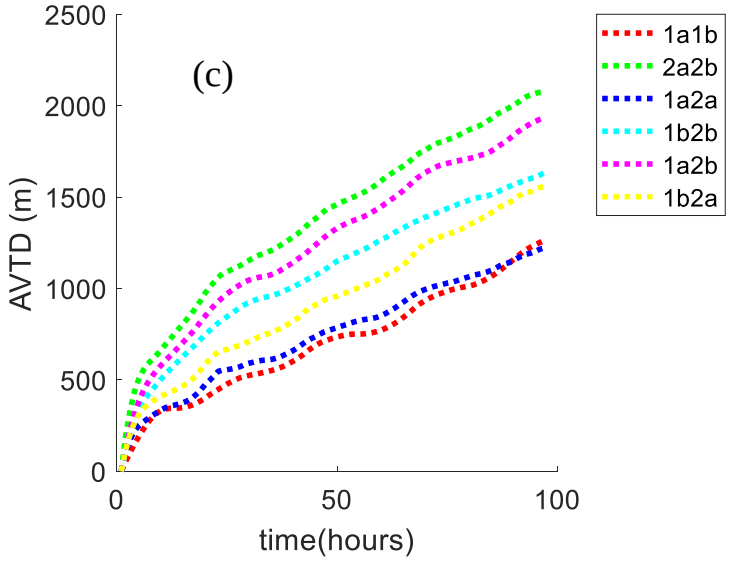


Figure 13. Time Evaluation of Sensitivity Parameters: (c) AVTD

4.1.2. Seasonal Variation of Sensitivity Parameters

The seasonal variations of sensitivity parameters are also examined considering the colder months with high precipitation represents the winter season (from November to April) and hotter and dryer months represents the summer season (from May to October). The seasonality of sensitivity parameters up to 96-hr were tested by the Median test at 95% confidence levels for each comparison and the results are reported in Table 10. In general, all sensitivity parameters showed statistically significant difference between seasons, except the RHTD parameter for comparison 2a-2b and 1b-2a, AHTD parameter of 1b-2b and 1a-2b comparison, and the AVTD parameter of 1b-2a comparisons.

Table 10. Statistical summary of the seasonal variations of the sensitivity parameters

Sensitivity Parameters	Comparison	Winter		Summer		Median Test
		Mean	Median	Mean	Median	
RHTD (%)	1a-1b	11.6	10.6	9.5	9.2	Reject
	2a-2b	21.5	20.6	21.3	20.9	Retain
	1b-2b	19.6	19.2	22.3	22.3	Reject
	1a-2a	19.1	19.5	21.4	22.5	Reject
	1a-2b	21.2	20.4	23.4	23.3	Reject
	1b-2a	20.8	20.5	21.8	22.2	Retain
AHTD (km)	1a-1b	310	263	184	152	Reject
	2a-2b	590	532	474	422	Reject
	1b-2b	532	478	475	425	Retain
	1a-2a	463	430	393	366	Reject
	1a-2b	577	513	495	442	Retain
	1b-2a	513	459	401	358	Reject
AVTD (m)	1a-1b	777	759	629	643	Reject
	2a-2b	590	532	474	422	Reject
	1b-2b	925	900	1236	1298	Reject
	1a-2a	729	713	769	806	Reject
	1a-2b	1141	1156	1355	1426	Reject
	1b-2a	935	923	912	929	Retain

* The significance level is 0.05

4.2. Cluster Analysis

In this study, k-means clustering analysis was applied to 1a, 1b, 2a, and 2b data sets. Sensitivity analysis was performed on the cluster centers determined as a result of k-means analysis. The similarity index of cluster centers was calculated. Finally, statistical tests were carried out on the cluster center data combined with the pollution data.

4.2.1. K-means Cluster Analysis

Based on the TRMSD outcomes of the clustering analysis conducted on all the data sets, it was found that when the cluster number is decreased from 4 to 3, the TRMSD value increases significantly. K-means cluster analysis 50-seed TRMSD results were presented in Figure 14. Therefore, the optimal number of clusters that could be chosen as a common solution was 4. However, as presented in Figure 15, Figure 16, Figure 17, and Figure 18, the direction of cluster centers did not match with each other when optimum cluster number is 4. The directions of the cluster centers of data set 1a are different from other data sets. It is crucial that the directions of the cluster centers must be matched in the analyses to be made. Consequently, it was decided to consider the optimum number of clusters as 5 instead.

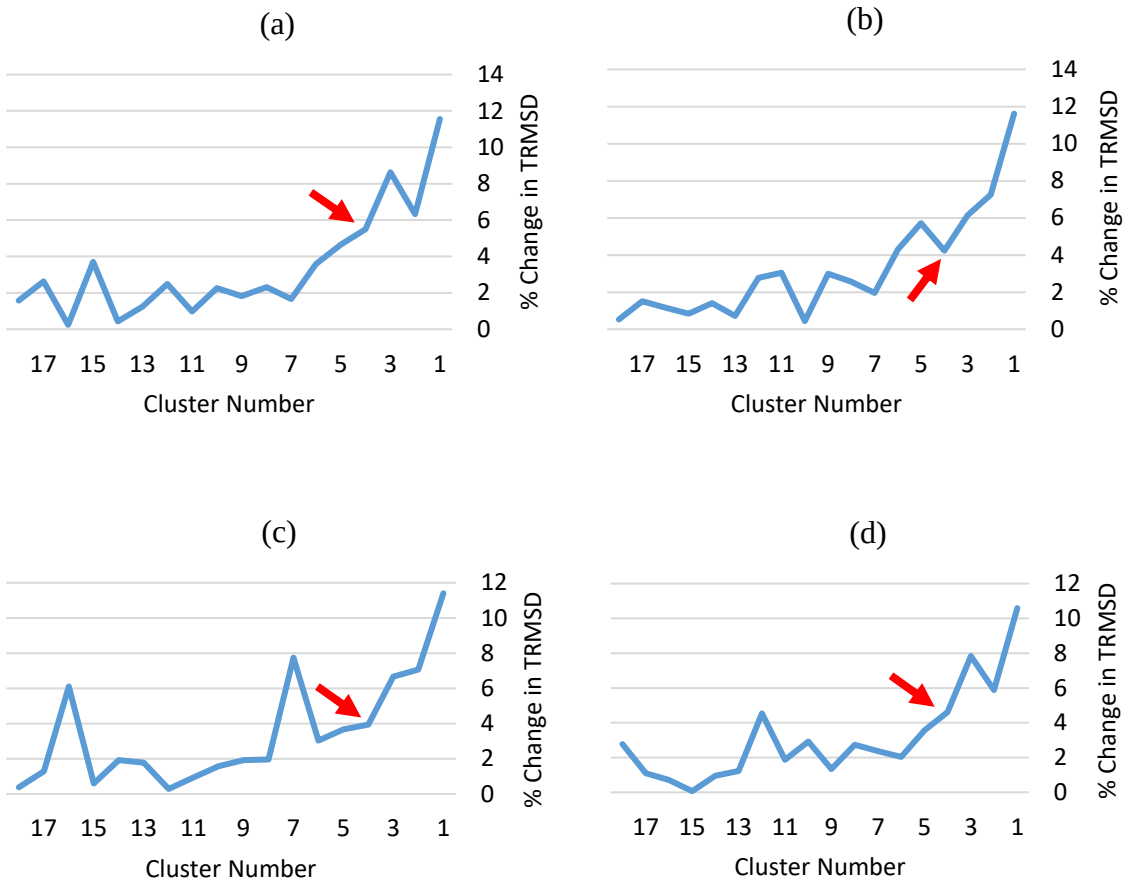


Figure 14. K-means cluster analysis 50-Seed TRMSD results (a) 1a, (b) 1b, (c) 2a, (d) 2b

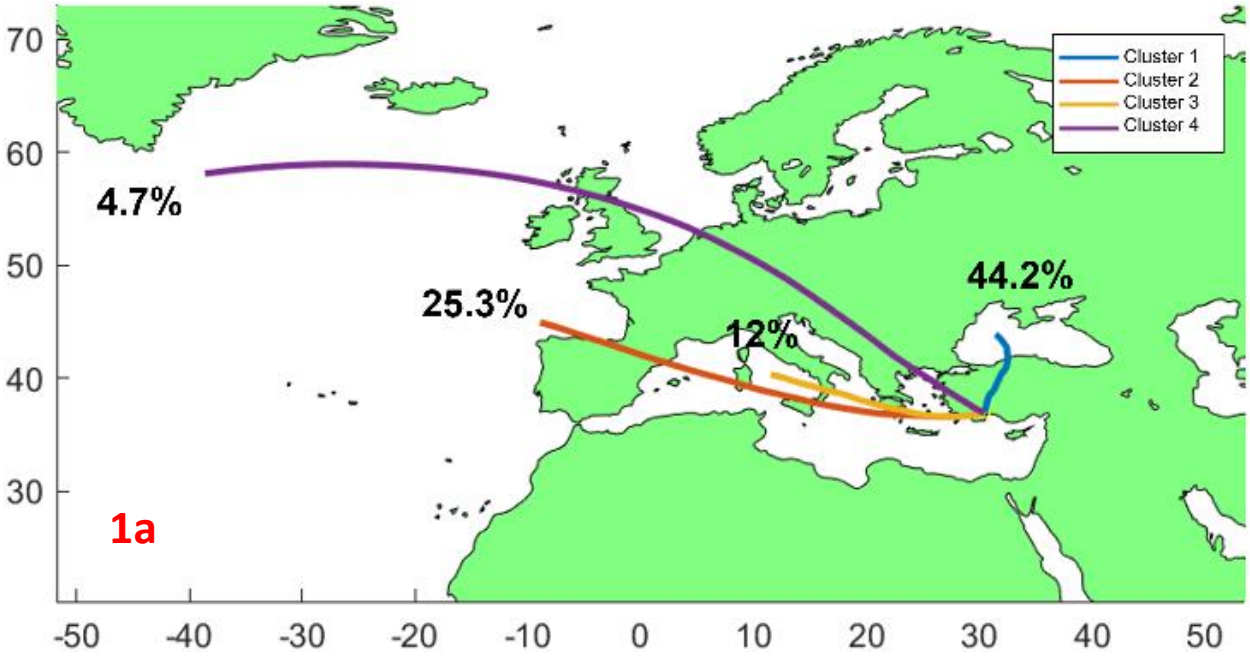


Figure 15. 1a data set when optimum number of cluster is 4

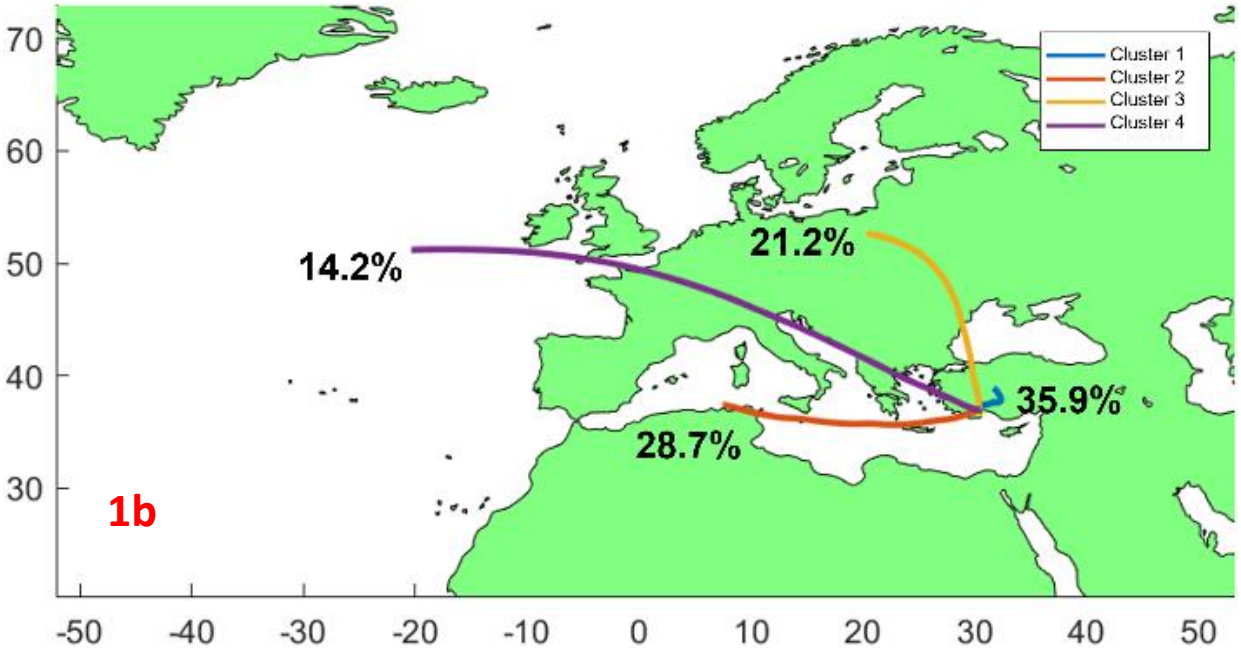


Figure 16. 1b data set when optimum number of cluster is 4

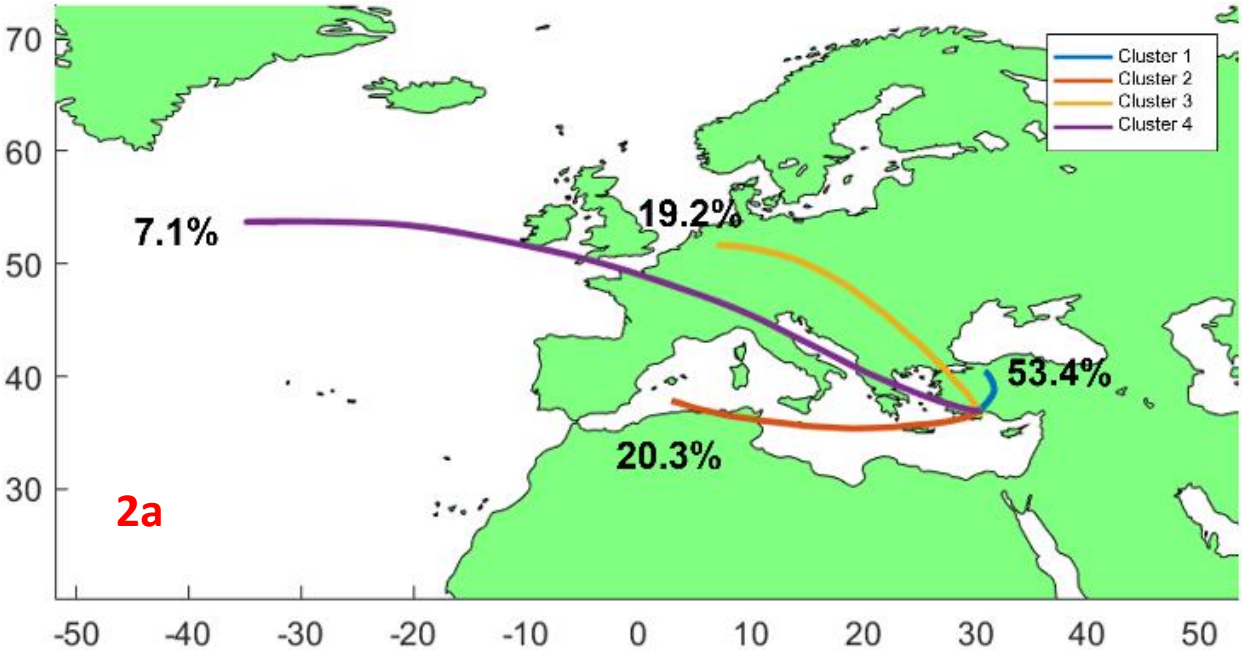


Figure 17. 2a data set when optimum number of cluster is 4

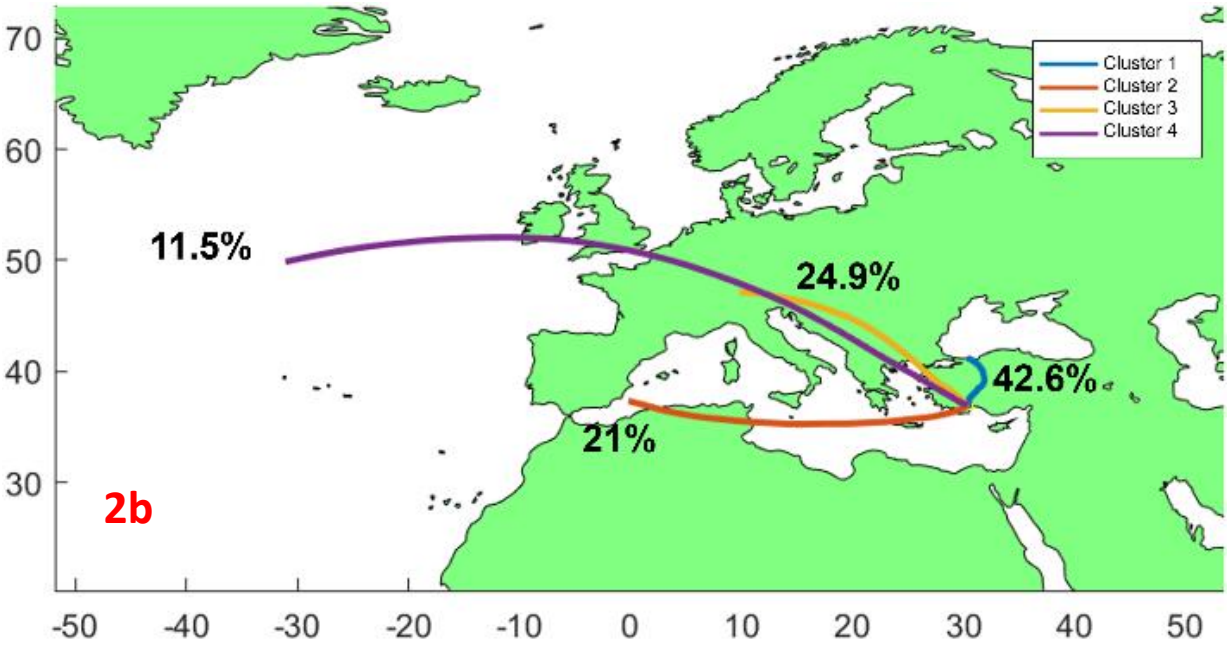


Figure 18. 2b data set when optimum number of cluster is 4

To compare the cluster centroids in each data set, optimum number of clusters are selected as 5 in this study The resulting cluster centroids (cluster means) are presented in Figure 19, Figure 20, Figure 21, and Figure 22. As presented from Figure 19 to Figure 22, Cluster 1 represents the air movements from Northeast (NE). The air mass movements from Northeast can be classified as slow moving (NE-slow) as in 1b and 2a data sets, and moderately moving (NE-mod) as in 1a and 2b data sets.

Slow air movements from Southwest (SW-slow) are represented by Cluster 2, while moderate air flows from the Southwest (SW-mod) are represented by Cluster 3.

The fastest air movements are from Northwest (NW-long) and classified in Cluster 4. Finally, Cluster 5 represents the moderate flow air masses from Northwest (NW-mod).

The cluster centroids obtained by k-means technique generated quite similar circulation pattern for each data set (Figure 19, Figure 20, Figure 21, Figure 22). There are some differences in curvature and length of the cluster centroids, however we can conclude that even different input variables were used to simulate back trajectories, the resulting centroids are quite similar.

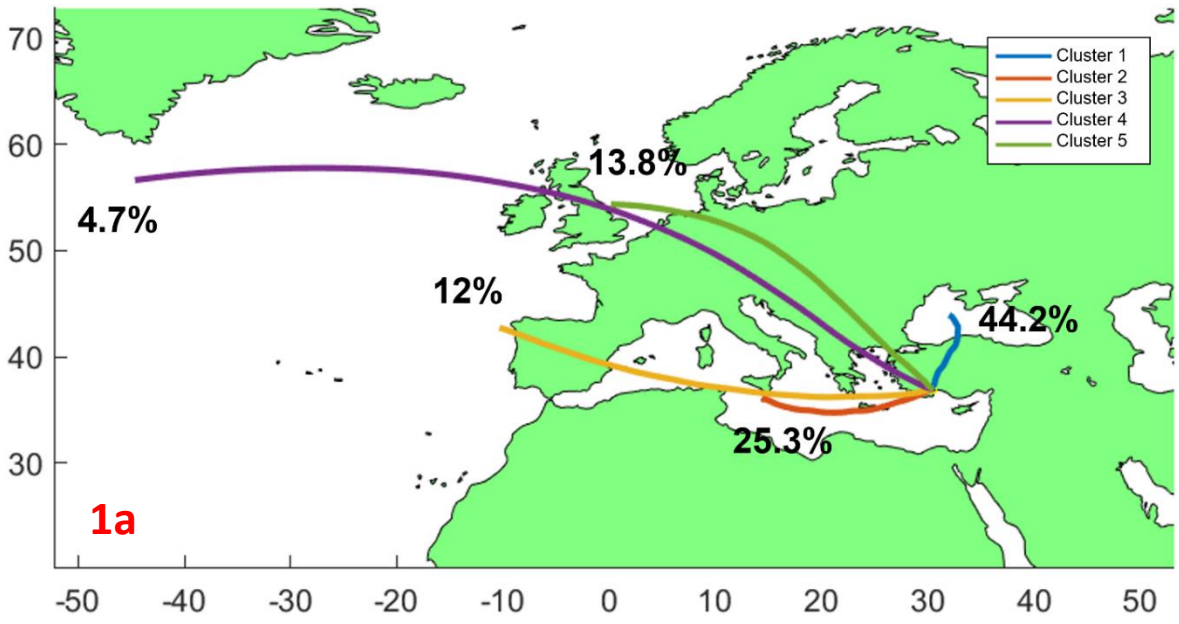


Figure 19. 1a data set when optimum number of cluster is 5

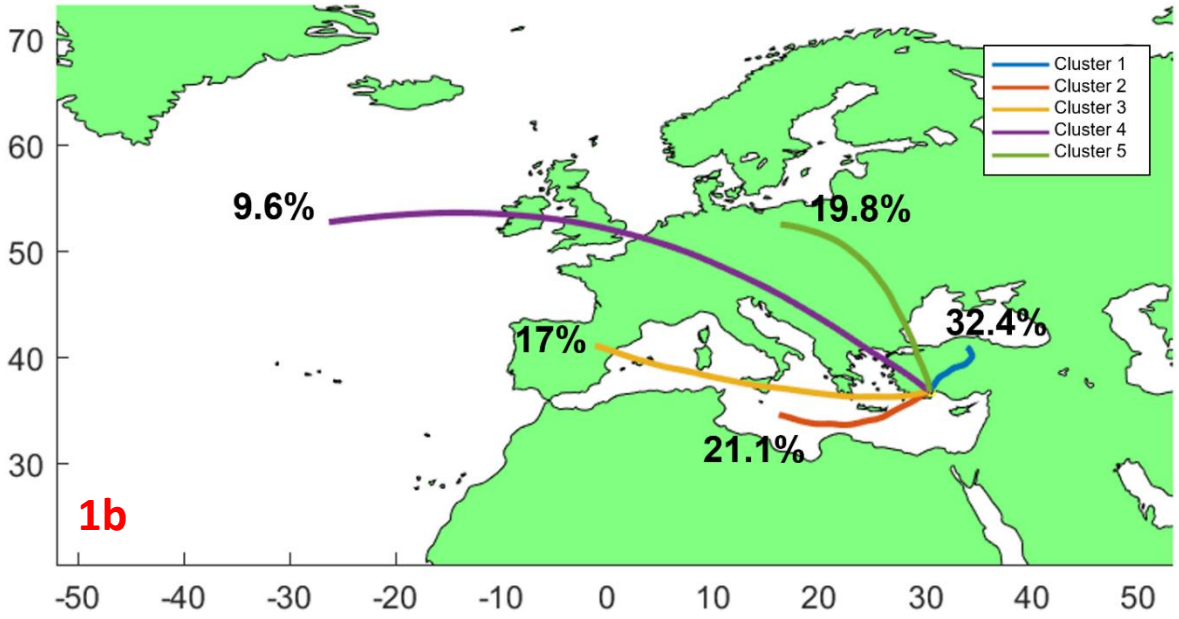


Figure 20. 1b data set when optimum number of cluster is 5

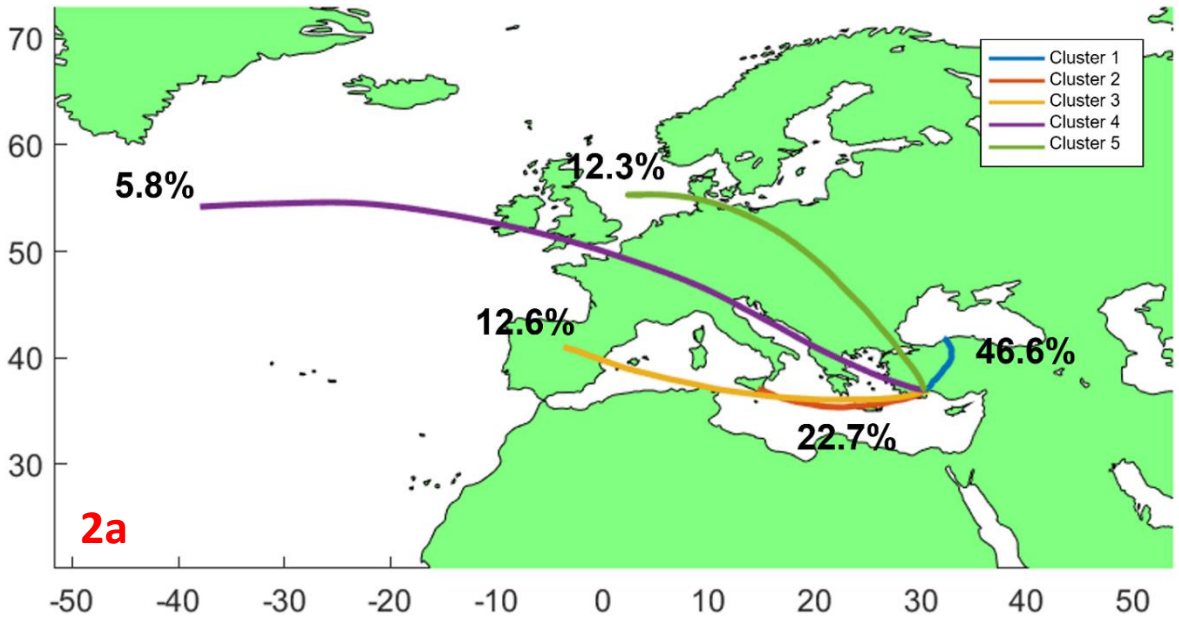


Figure 21. 2a data set when optimum number of cluster is 5

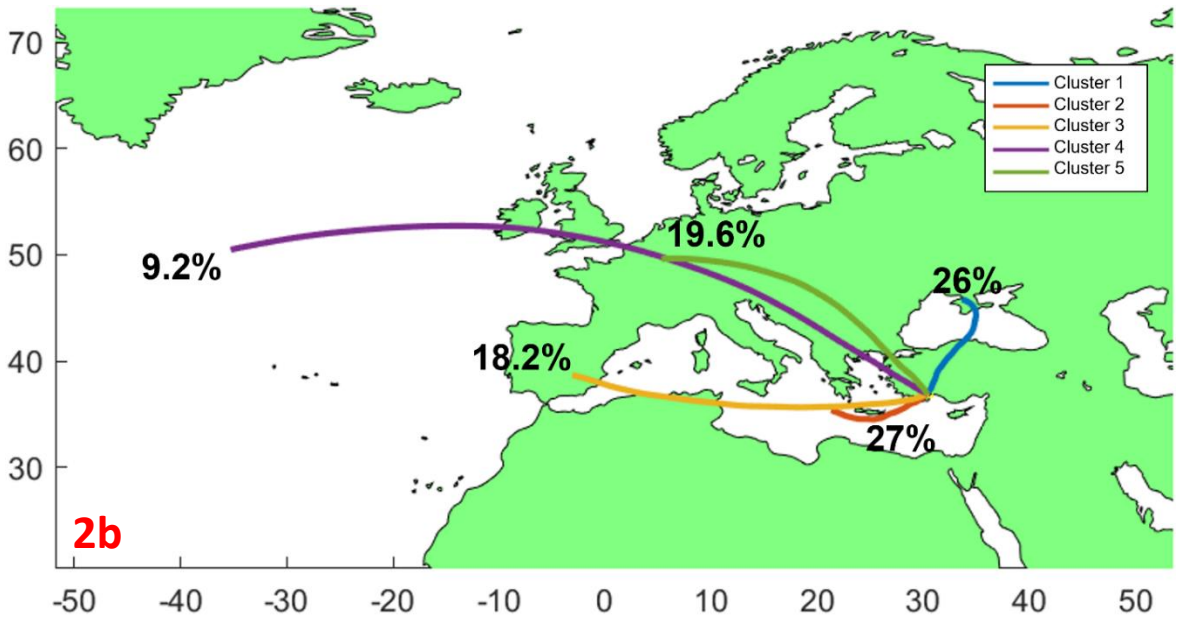


Figure 22. 2b data set when optimum number of cluster is 5

4.2.2. Sensitivity of Cluster Analysis to the HYSPLIT Model Inputs

The sensitivity analysis is applied to cluster centroids to determine the differences in transport patterns when trajectories are generated from different meteorological archives and vertical velocity methods. The results are presented in Table 11. Note that only horizontal deviation statistics (i.e. RHTD and AHTD) were presented in Table 11 because the pressure or height of trajectories were not used in k-means cluster analysis.

Table 11. Sensitivity analysis of cluster centroids

96h	Description	RHTD, %					AHTD, km				
		Cluster ID					Cluster ID				
		1	2	3	4	5	1	2	3	4	5
1a-1b	Sensitivity of the NCEP/NCAR Reanalysis archive to the velocity method	37	12	23	24	33	356	227	773	1249	1084
2a-2b	Sensitivity of the GDAS archive to the velocity method	38	36	9.8	8.6	10	456	627	263	450	638
1a-2a	Sensitivity to meteorology archive	23	6.2	18	10	4	230	111	581	502	252
1b-2b		45	27	12	12	26	532	476	321	671	835
1a-2b	Sensitivity to both meteorology archive and velocity method	20	38	24	16	17	246	631	756	920	633
1b-2a		19	16	7.5	17	15	174	301	210	786	873

Considering the sensitivity analysis results for Cluster 1, slow and moderate air movements from the Northeast (NE-slow, NE-mod), the results 1a-2a and 1b-2b are remarkable. 1b-2b comparison was higher than all other comparisons. Based on the difference between 1a-2a and 1b-2b comparisons, it can be said that the model vertical velocity method for Cluster 1 is more influenced by the meteorological archive than the isentropic velocity method. Therefore, for Cluster 1, the model vertical velocity method is quite sensitive to the meteorological archive.

For Cluster 2, slow-moving air masses from Southwest (SW-slow), show that comparison 1a-2a has the lowest values. This shows that parameter a, that is, the isentropic velocity method, is the least affected parameter from the meteorology archive. This implies that the model vertical velocity method is more sensitive to the meteorological archive than the isentropic velocity method. When the results of 1a-1b and 2a-2b comparisons are examined, it is seen that the GDAS1 meteorology archive is more sensitive to the vertical velocity method than the NCEP/NCAR

Reanalysis meteorology archive. In addition, comparison 1a-2b has the highest values. Compared to the 1b-2a comparison, it is seen that the 1a-2b combination has the results most affected by the model inputs.

When analyzing the results for Cluster 3, moderate air flows from the Southwest (SW-mod), the highest RHTD values are calculated for 1a-1b and 1a-2b comparisons. This indicates that NCEP/NCAR Reanalysis archive is more sensitive to the vertical velocity method. Longer trajectories from SW are represented in Cluster 3 when NCEP/NCAR Reanalysis archive is used with isentropic velocity method (1a). Actually, this statement can be corroborated with the average length of the trajectories (Table 9). The average length of back trajectories generated from 1a and 1b datasets are calculated as 2710 and 2637 km, respectively.

When Table 11 is analyzed for Cluster 4, the fastest air movements from Northwest (NW-long), a significant difference is seen between the results 1a-1b and 2a-2b. The sensitivity values between these two comparisons vary significantly (i.e. 8.6% to 24%). Specifically, the comparison with the highest sensitivity value is 1a-1b, indicating a strong sensitivity of the NCEP/NCAR Reanalysis meteorology archive to the vertical velocity method within this cluster. On the other hand, the 2a-2b comparison has the lowest sensitivity, suggesting a relatively lower sensitivity of the GDAS1 meteorology archive to the vertical velocity method. It appears that the two meteorology archives differ substantially when considering vertical velocity. The NCEP/NCAR Reanalysis meteorology archive demonstrates a higher sensitivity, indicating that changes or uncertainties in the vertical velocity have a more pronounced impact on trajectories derived from 1a-1b data set within Cluster 4.

The highest deviation statistics for Cluster 5, the moderate flow air masses from Northwest (NW-mod), belong to the comparison 1a-1b (RHTD value of 33%), similar to the results of Cluster 4. In other words, the NCEP/NCAR Reanalysis meteorology archive is highly sensitive to the vertical velocity method. This finding suggests that variations in the vertical velocity had a significant impact on the resulting trajectories for Cluster 5 when NCEP/NCAR Reanalysis archive is used. In contrast, the lowest results were noted in the 1a-2a comparison. When comparing the 1a-2a and 1b-2b comparisons, it becomes apparent that the isentropic velocity method is less affected by the

meteorology archive compared to the model vertical velocity method. In other words, changes in the meteorological input had a relatively smaller influence on trajectories generated using the isentropic velocity method, as indicated by the lower sensitivity values in the mentioned comparisons.

The percentage of trajectories corresponding to each cluster center is shown in the Figure 19, Figure 20, Figure 21, and Figure 22, and also presented in Table 12 for each trajectory dataset. It is obvious that the percentage of trajectories in each cluster shows variation for each dataset. For example, the percentage of back trajectories classified in Cluster 1 (NE) are 44% for 1a dataset and 26% for 2b dataset.

Table 12. The percentage of back trajectories in each cluster center

Data set	**Cluster ID**				
	1	**2**	**3**	**4**	**5**
	NE	SW-slow	SW-mod	NW-fast	NW-mod
1a	44.2	25.3	12	4.7	13.8
1b	32.4	21.1	17	9.6	19.8
2a	46.6	22.7	12.6	5.8	12.3
2b	26	27	18.2	9.2	19.6

As Southwest-slow (Cluster 2) and Southwest-mod (Cluster 3) corresponds to the Southwest direction, the sum of these two clusters is accepted as Southwest (SW), and the sum of NW–fast (Cluster 4) and NW-mod (Cluster 5) are accepted as NW and the resulting percentages are presented in Table xx.

Table 13. The percentage of back trajectories in the main transport directions

Data set	**NE**	**SW**	**NW**
1a	44.2	37.3	18.5
1b	32.4	38.1	29.4
2a	46.6	35.3	18.1
2b	26	45.2	28.8

It’s clearly seen from Table 13 that if the model vertical velocity method is selected, the percentages of NE air flows are decreased (i.e, 44% to 32% for the NCEP/NCAR Reanalysis archive and from 47% to 26% for the GDAS1 archive). The decrease in percentages is 27% and 44% for the NCEP/NCAR Reanalysis and GDAS1 archive, respectively.

Contrary, trajectories from the NW increases when the model vertical velocity method is chosen. The increase in the percentages of trajectories is 59% for both meteorological archives. For the trajectories from the SW, the change in the velocity method did not influence the percentages of trajectories for the NCEP/NCAR Reanalysis archive. However, GDAS1 archive showed 28% increase (from 35% to 45%) in SW trajectories when model vertical velocity is chosen.

For the isentropic trajectories, meteorological archive selection did not influence the percentages of trajectories (i.e. 5% increase in NE, 5% decrease in SW, and 2% decrease in NW trajectories). However, model vertical velocity method changed the percentages of trajectories in NE and SW directions (i.e. 20% decrease in NE, 19% increase in SW) when GDAS1 archive is used.

As a result, we may say that back trajectories simulated with GDAS1 archive and model vertical velocity (2b dataset) are more likely classified into westerly direction (i.e. SW and NW).

To examine whether a trajectory arriving in our receptor site is classified in the same cluster when trajectories are simulated with different datasets, percentages of the same trajectories in each cluster is calculated using the similarity index and presented in Table 14. There is no definite threshold value for the similarity index value. Kassomenos et al. (2010a) used 70% and Cui, Song, & Zhong (2021) used 85% as the threshold value for similarity index.

According to the Table 14 in most instances, the clusters were unable to maintain a high level of similarity. The highest similarity index results (64-83%) belong to Cluster 1 which is slow-moving and moderately moving air movements originating from Northeast. The high similarity index value in Cluster 1 indicates that the back trajectories have a significant degree of similarity in terms of their paths and locations. This suggests that the air masses followed similar paths and experienced similar atmospheric conditions even different input datasets are used.

The lowest similarity index values (24-63%) are calculated for Cluster 5 which represents the moderate flow air masses from Northwest (NW-mod). The lowest similarity index of Cluster 5 suggests that the back trajectories have divergent paths and locations. This indicates less consistency among the trajectories, suggesting that the air masses followed distinct paths and experienced different atmospheric conditions when different input datasets are used. It may indicate the influence of atmospheric disturbances, or local-scale factors that resulted in the divergence of the trajectories.

For all clusters, the highest similarity index values are calculated for 1a-2a comparison (63-83%). Therefore, it can be concluded that clusters generated from isentropic back trajectories are less sensitive to the meteorological archive.

Table 14. Similarity Index Results

Similarity %	**Cluster ID**				
	1	**2**	**3**	**4**	**5**
1a-1b	79	72	62	55	40
2a-2b	64	45	56	51	47
1a-2a	83	63	65	63	63
1b-2b	66	53	59	61	52
1a-2b	66	50	52	54	48
1b-2a	72	54	60	51	24

4.2.3. The Influence of Cluster Analysis on The Interpretation of The Pollution Data

In order to investigate the influence of back trajectories generated with different input variables on the interpretation of source-receptor relations, pollution data is used. Sulfate, a secondary pollutant, is selected as pollutant because its long-range transport is most probable. The statistical summary of sulfate data is presented in Table 15.

Table 15. Descriptive statistics of SO_4^{-2} data

Count	460
Average*	3603
Std*	2249
Median*	2989
Min*	51.5
Max*	17567

* Concentrations are reported in ng m^{-3}.

The Kruskal-Wallis test was performed to specify whether the median sulfate concentration between clusters per data set differs significantly (Table 16). Sulfate concentrations in each cluster center per data set are presented in Figure 23 as a box-whisker plot.

Among the four data set, significant differences in sulfate median concentrations are reported only for clusters generated from GDAS1 trajectories (both isentropic (2a) and model vertical (2b)). If a researcher used Reanalysis archive, he/she would conclude that there were no significant differences in sulfate concentration between clusters for both 1a and 1b data sets. However, if the researcher used GDAS1 archive, then the interpretation of the pollution data would be different. Considering the 2a data set (GDAS1-isentropic), the highest sulfate concentrations are due to short air flows from Northeast (Cluster 1) and the lowest concentrations are observed with the fastest air movements from Northwest (Cluster 4). If 2b data set (GDAS1-model vertical) were used, then the highest concentrations would be reported for moderate flow air masses from Northwest (Cluster 5) and the lowest concentrations would be reported for the moderate air flows from the Southwest (Cluster 3).

The differences observed in the source-receptor interpretation of pollution data indicates that GDAS1 trajectories are sensitive to the vertical motion method and the source regions may differ depending on the vertical motion method.

Table 16. Median concentrations of SO_4^{-2} (ng m^{-3}) and K-W test result

SO_4^{-2} (ng m^{-3})	**Cluster ID**					**K-W Test**
	1	**2**	**3**	**4**	**5**	
1a	3420	2540	2646	2617	3058	Retained
1b	3226	2485	2649	2718	3445	Retained
2a	3625	2511	2658	1967	2858	Rejected
2b	3285	2993	2521	2529	3392	Rejected

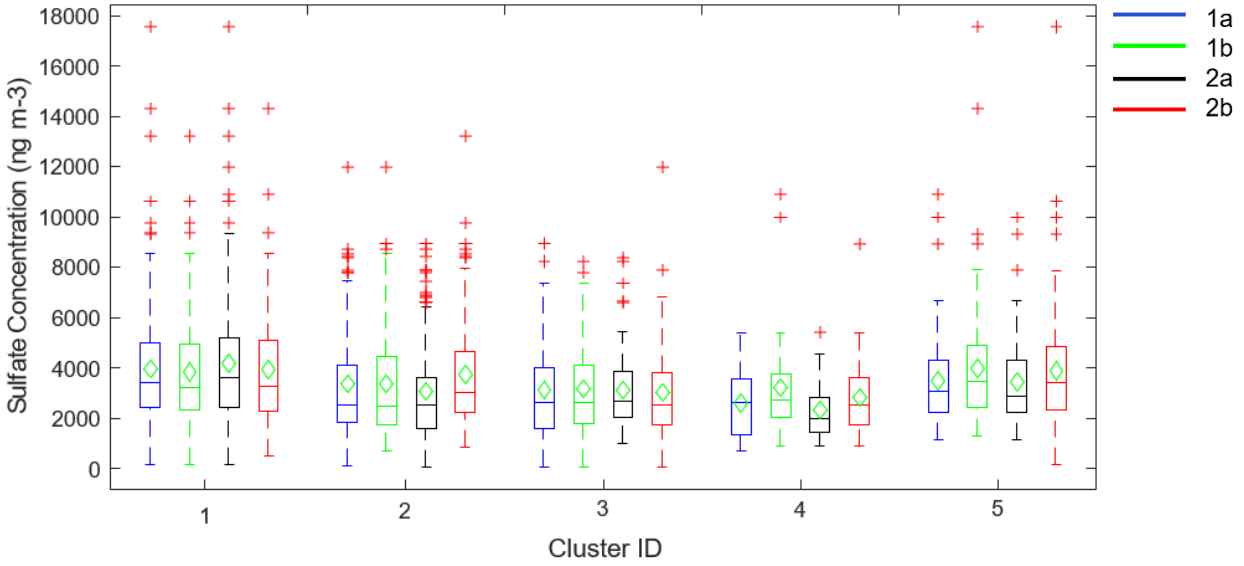


Figure 23. Sulfate concentrations in each cluster center per data set

5. CONCLUSION

This study conducts a sensitivity analysis for a rural area in the Mediterranean part of Turkey to identify differences between HYSPLIT back trajectories generated by the most frequently used meteorological archive (NCEP/NCAR Reanalysis and GDAS1) and vertical motion methods (isentropic and model vertical velocity).

For this particular receptor site, back trajectories are found to be sensitive to the meteorological archive (38-40%), vertical motion method (25-44%), and both meteorological archive and vertical motion method (42-45%). The trajectory uncertainties (either isentropic or model vertical velocity) are within the typical range (around 20% of distance travelled) when NCEP/NCAR Reanalysis archive is used. In contrast, the deviations become larger than the typical trajectory error with the GDAS1 archive. Air quality researchers working in this particular receptor site might need to inspect either isentropic or model vertical velocity trajectories when NCEP/NCAR Reanalysis archive is used. However, inspecting both isentropic and model vertical velocity trajectories with the GDAS1 archive is strongly recommended.

The back trajectories simulated with the most frequently used meteorological archive (NCEP/NCAR Reanalysis and GDAS1) and vertical motion methods (isentropic and model vertical velocity) are classified into five cluster centers using the k-means cluster technique. The percentage of trajectories in each cluster center showed variation for each dataset, however, the resulting centroids showed quite similar circulation pattern. The air movements from Northeast (NE) were represented by Cluster 1. Slow air movements from Southwest (SW-slow) were represented by Cluster 2. Moderate air flows from the Southwest (SW-mod) were represented by Cluster 3. The fastest air movements from Northwest (NW-long) were represented by Cluster 4. Moderate flow air masses from Northwest (NW-mod) were represented by Cluster 5.

The sensitivity of cluster centers to the input data were also investigated. The model vertical velocity method for Cluster 1 was found to be more influenced by the meteorological archive than the isentropic velocity method. For Cluster 2, GDAS1 meteorology archive was found to be more

sensitive to the vertical velocity method. The NCEP/NCAR Reanalysis meteorology archive was found to be more sensitive to the vertical velocity method for Cluster 3, Cluster 4 and Cluster 5.

Cluster Analysis results also showed that back trajectories simulated with GDAS1 archive and model vertical velocity (2b dataset) were more likely classified into westerly direction (i.e. SW and NW). In general, clusters generated from isentropic back trajectories were less sensitive to the meteorological archives, hence indicated the highest similarity index values (63-83%).

Secondary air pollutant, sulfate, was used to track the potential source direction of pollution affecting the receptor site. Cluster centers generated from the NCEP/NCAR Reanalysis archive (1a and 1b dataset) did not show significant differences in sulfate concentration, however, significant differences in sulfate concentration were found when GDAS1 archive used (2a and 2b dataset). The cluster centers highlighted as the potential source direction of sulfate were different for 2a (GDAS1-isentropic) and 2b (GDAS-model vertical velocity) datasets, most probably due to large deviations in horizontal and vertical profiles of GDAS1 trajectories, hence causing discrepancies in the source-receptor interpretation.

As a result, in this particular receptor site, to simulate the back trajectories HYSPLIT model is recommended to run using the GDAS1 archive with both isentropic and model vertical velocity as the source regions were found to be different depending on the vertical motion method.

In general, it is known that back trajectory uncertainties are somewhat specific to location because of the influence of meteorology and topography. In each location or region different meteorological and geographical regimes exist, therefore, HYSPLIT model is recommended to run with different data inputs for more reliable simulations of atmospheric transport. Utilizing different meteorology archives provides a broader perspective on atmospheric conditions. Different meteorological models have varying strengths and weaknesses, so combining their outputs can lead to a more nuanced understanding of the complex atmospheric processes that HYSPLIT aims to simulate.

The NCEP/NCAR Reanalysis archive has the longest meteorological archive covering the globe and most of the previous studies were conducted using the NCEP/NCAR Reanalysis archive. However, the NCEP/NCAR Reanalysis archive has low resolution, and may not interpret the trajectory path accurately. When available, the meteorological archives having high resolution (i.e. horizontal, vertical, time) should be preferred because uncertainties (i.e. interpolation and its associate errors) are minimized in the dense meteorological archives.

Another important criterion for choosing the meteorological archive should be the existence of vertical velocity field. If available in the meteorological archive, vertical velocity from the meteorological archive should be used to run the HYSPLIT model. The vertical velocity calculated from the isentropic and/or isobaric vertical velocity methods (i.e. approaches) may not represent the vertical motion of the air masses therefore these two vertical velocity methods should be preferred only if there are specific reasons. Excessive reliance on the vertical velocity calculated from assumptions (for example isentropic and isobaric back trajectories) may result in misinterpretation of the source-receptor relations and should be avoided. Of course, needless to say that, whenever possible multiple inputs should be used to drive the HYSPLIT model to get confidence about the air parcel pathways.

5.1. Recommendations for Future Research

The following can be recommended for future studies within the scope of the results obtained from this study.

- Back trajectories also differ depending on the arrival height and these differences can be examined with sensitivity analysis.

- It is recommended to conduct studies on the effect of synoptic-scale meteorological parameters (pressure, wind, etc.) on the back trajectories generated from different meteorological archive and vertical velocity method. The synoptic situations when back trajectories agree with each other and the synoptic situations when back trajectories differ from each other should be determined.

- Sensitivity analysis should be conducted using the other trajectory analysis methods such as the Potential Source Contribution Function (PSCF) so that the differences in source regions can be examined quantitatively.

- Episodic events such as Saharan dust intrusions can be examined to identify which model inputs are more appropriate to simulate the back trajectories and to define the synoptic meteorological conditions causing such episodic events.

- Back trajectories generated using different arrival time can also be investigated to determine the influence of arrival time.

- The sensitivity due to clustering technique could also be investigated.

REFERENCES

Abdalmogith, S. S., & Harrison, R. M. (2005). The use of trajectory cluster analysis to examine the long-range transport of secondary inorganic aerosol in the UK. *Atmospheric Environment*, *39*(35), 6686–6695. https://doi.org/10.1016/J.ATMOSENV.2005.07.059

Ağaç, K. (2016). *İstanbul Kağıthane Bölgesinde Yüksek PM10 Konsantrasyonlarının Meteorolojik Olarak İncelenmesi*.

Alan, Ş., Sarışahin, T., Acar Şahin, A., Kaplan, A., & Pınar, N. M. (2019). An assessment of ragweed pollen and allergen loads in an uninvaded area in the Western Black Sea region of Turkey. *Aerobiologia 2019 36:2*, *36*(2), 183–195. https://doi.org/10.1007/S10453-019-09620-Z

Al-Momani, I. F., Güllü, G., Ölmez, I., Eler, Ü., Örtel, E., Şirin, G., & Tuncel, G. (1997). Chemical composition of eastern Mediterranean aerosol and precipitation: Indications of long-range transport. *Pure and Applied Chemistry*, *69*(1), 41–46. https://doi.org/10.1351/PAC199769010041/MACHINEREADABLECITATION/RIS

Alp, K., & Hanedar, A. O. (2009). Determination of transport processes of nocturnal ozone in Istanbul atmosphere. *International Journal of Environment and Pollution*, *39*(3/4), 213. https://doi.org/10.1504/IJEP.2009.028686

Anil, I., Alagha, O., & Karaca, F. (2017). Effects of transport patterns on chemical composition of sequential rain samples: trajectory clustering and principal component analysis approach. *Air Quality, Atmosphere & Health 2017 10:10*, *10*(10), 1193–1206. https://doi.org/10.1007/S11869-017-0504-X

ARL - Global Reanalysis Data Archive. (2023). https://www.ready.noaa.gov/gbl_reanalysis.php

Asaf, D., Pedersen, D., Peleg, M., Matveev, V., & Luria, M. (2008). Evaluation of background levels of air pollutants over Israel. *Atmospheric Environment*, *42*(36), 8453–8463. https://doi.org/10.1016/J.ATMOSENV.2008.08.011

Ashbaugh, L. L., Malm, W. C., & Sadeh, W. Z. (1985). A residence time probability analysis of sulfur concentrations at grand Canyon National Park. *Atmospheric Environment (1967)*, *19*(8), 1263–1270. https://doi.org/10.1016/0004-6981(85)90256-2

Baid, U., Talbar, S., & Talbar, S. (2016). *Comparative Study of K-means, Gaussian Mixture Model, Fuzzy C-means algorithms for Brain Tumor Segmentation*. *137*, 583–588. https://doi.org/10.2991/ICCASP-16.2017.85

Baker, J. (2010). A cluster analysis of long range air transport pathways and associated pollutant concentrations within the UK. *Atmospheric Environment*, *44*(4), 563–571. https://doi.org/10.1016/J.ATMOSENV.2009.10.030

Balcılar, İ. (2018). Temporal variations of Eastern Black Sea aerosol. *GLOBAL NEST JOURNAL*, *20*(1), 115–121.

Baltaci, H., Arslan, H., & Akkoyunlu, B. O. (2022). High PM10 source regions and their influence on respiratory diseases in Canakkale, Turkey. *International Journal of Environmental Science and Technology*, *19*(2), 797–806. https://doi.org/10.1007/S13762-020-02914-7/METRICS

Baltaci, H., & Ezber, Y. (2022). Characterization of atmospheric mechanisms that cause the transport of Arabian dust particles to the southeastern region of Turkey. *Environmental Science*

and Pollution Research, *29*(15), 22771–22784. https://doi.org/10.1007/S11356-021-17526-Y/METRICS

Baltaci, H., Ozgen Alemdar, C. S., & Akkoyunlu, B. O. (2020). Background atmospheric conditions of high PM10 concentrations in Istanbul, Turkey. *Atmospheric Pollution Research*, *11*(9), 1524–1534. https://doi.org/10.1016/J.APR.2020.06.020

Borge, R., Lumbreras, J., Vardoulakis, S., Kassomenos, P., & Rodríguez, E. (2007a). Analysis of long-range transport influences on urban PM10 using two-stage atmospheric trajectory clusters. *Atmospheric Environment*, *41*(21), 4434–4450. https://doi.org/10.1016/J.ATMOSENV.2007.01.053

Borge, R., Lumbreras, J., Vardoulakis, S., Kassomenos, P., & Rodríguez, E. (2007b). Analysis of long-range transport influences on urban PM10 using two-stage atmospheric trajectory clusters. *Atmospheric Environment*, *41*(21), 4434–4450. https://doi.org/10.1016/j.atmosenv.2007.01.053

Bowman, K. P., Lin, J. C., Stohl, A., Draxler, R., Konopka, P., Andrews, A., & Brunner, D. (2013). Input Data Requirements for Lagrangian Trajectory Models. *Bulletin of the American Meteorological Society*, *94*(7), 1051–1058. https://doi.org/10.1175/BAMS-D-12-00076.1

Cabello, M., Orza, J. A. G., Galiano, V., & Ruiz, G. (2008). Influence of meteorological input data on backtrajectory cluster analysis – a seven-year study for southeastern Spain. *Advances in Science and Research*, *2*(1), 65–70. https://doi.org/10.5194/ASR-2-65-2008

Cape, J. N., Methven, J., & Hudson, L. E. (2000). The use of trajectory cluster analysis to interpret trace gas measurements at Mace Head, Ireland. *Atmospheric Environment*, *34*(22), 3651–3663. https://doi.org/10.1016/S1352-2310(00)00098-4

Çapraz, Ö., & Deniz, A. (2021). Particulate matter (PM10 and PM2.5) concentrations during a Saharan dust episode in Istanbul. *Air Quality, Atmosphere and Health*, *14*(1), 109–116. https://doi.org/10.1007/S11869-020-00917-4/METRICS

Celenk, S. (2019). Detection of reactive allergens in long-distance transported pollen grains: Evidence from Ambrosia. *Atmospheric Environment*, *209*, 212–219. https://doi.org/10.1016/J.ATMOSENV.2019.04.040

Chan, C. Y., Wong, K. H., Li, Y. S., Chan, L. Y., & Zheng, X. D. (2006). The effects of Southeast Asia fire activities on tropospheric ozone, trace gases and aerosols at a remote site over the Tibetan Plateau of Southwest China. *Tellus B*, *58*(4), 310–318. https://doi.org/10.1111/J.1600-0889.2006.00187.X

Chen, Y., Ebenstein, A., Greenstone, M., & Li, H. (2013). Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(32), 12936–12941. https://doi.org/10.1073/PNAS.1300018110

Cheng, I., Zhang, L., Blanchard, P., Dalziel, J., Tordon, R., Huang, J., & Holsen, T. M. (2013). Comparisons of mercury sources and atmospheric mercury processes between a coastal and inland site. *Journal of Geophysical Research: Atmospheres*, *118*(5), 2434–2443. https://doi.org/10.1002/JGRD.50169

Compton, J. E., Harrison, J. A., Dennis, R. L., Greaver, T. L., Hill, B. H., Jordan, S. J., Walker, H., & Campbell, H. V. (2011). Ecosystem services altered by human changes in the nitrogen cycle: a new perspective for US decision making. *Ecology Letters*, *14*(8), 804–815. https://doi.org/10.1111/J.1461-0248.2011.01631.X

Cooter, E. J., Rea, A., Bruins, R., Schwede, D., & Dennis, R. (2013). The role of the atmosphere in the provision of ecosystem services. *Science of The Total Environment*, *448*, 197–208. https://doi.org/10.1016/J.SCITOTENV.2012.07.077

Cui, L., Song, X., & Zhong, G. (2021). Comparative Analysis of Three Methods for HYSPLIT Atmospheric Trajectories Clustering. *Atmosphere 2021, Vol. 12, Page 698*, *12*(6), 698. https://doi.org/10.3390/ATMOS12060698

Cui, L., Song, X., Zhong, G., Xing, J., Kelly, J., Zhao, J., Zhang, Y., & Zhu, Y. (2021). Comparative Analysis of Three Methods for HYSPLIT Atmospheric Trajectories Clustering. *Atmosphere 2021, Vol. 12, Page 698*, *12*(6), 698. https://doi.org/10.3390/ATMOS12060698

Dinçer, F., Elbir, T., & Müezzinoğlu, A. (2003). Hysplit Back-Trajectory Modeli ile İzmir ve Çevresindeki Hava Kirletici Kaynakların Bölgenin Hava Kalitesine Etkilerinin Analizi. *Yanma ve Hava Kirliliği Kontrolü VI. Ulusal Sempozyumu*. http://hkadtmk.org/Bildiriler/HKK-2003/bildiriler/dincer.pdf

Dogan, G., Gullu, G., Karakas, D., & Tuncel, G. (2010). Comparison of Source Regions Affecting SO4 2- and NO3 - Concentrations at the Eastern Mediterranean and Black Sea Atmospheres. *Current Analytical Chemistry*, *6*(1), 66–71. https://www.academia.edu/5329487/Comparison_of_Source_Regions_Affecting_SO4_2_and_NO3_Concentrations_at_the_Eastern_Mediterranean_and_Black_Sea_Atmospheres

Dogan Rastgeldi, T., & Atbinici, M. (2022). TEMPORAL AND SPATIAL ANALYSIS OF PM10 AND SO2 CONCENTRATION WITH THE USE OF GIS IN SOUTHEASTERN ANATOLIA REGION CITIES OF TURKEY (2010-2020). *Applied Ecology and Environmental Research*, *20*(5), 4079–4093. https://doi.org/10.15666/AEER/2005_40794093

Dogan, T. R., Yilmaz, D., & Yalcin, S. (2021). Gamma ray characterization of the albedo of atmospheric dust from Southeast Anatolia, Turkey. *Https://Doi.Org/10.1080/10739149.2021.1915797*, *49*(6), 604–615. https://doi.org/10.1080/10739149.2021.1915797

Dorling, S. R., Davies, T. D., & Pierce, C. E. (1992a). Cluster analysis: A technique for estimating the synoptic meteorological controls on air and precipitation chemistry—Method and applications. *Atmospheric Environment. Part A. General Topics*, *26*(14), 2575–2581. https://doi.org/10.1016/0960-1686(92)90110-7

Dorling, S. R., Davies, T. D., & Pierce, C. E. (1992b). Cluster analysis: A technique for estimating the synoptic meteorological controls on air and precipitation chemistry—Method and applications. *Atmospheric Environment. Part A. General Topics*, *26*(14), 2575–2581. https://doi.org/10.1016/0960-1686(92)90110-7

Dörter, M., Sağırlı, E., Karakaş, D., & Yenisoy-Karakaş, S. (2020). Investigation of washing mechanisms in volume-based fractional rain samples in high altitude semirural site by determining polycyclic aromatic hydrocarbons, elemental carbon, and organic carbon. *Polycyclic Aromatic Compounds*, *40*(1), 179–193. https://doi.org/10.1080/10406638.2018.1545134

Draxler, R. R. (1996). Boundary layer isentropic and kinematic trajectories during the August 1993 North Atlantic Regional Experiment Intensive. *Journal of Geophysical Research: Atmospheres*, *101*(D22), 29255–29268. https://doi.org/10.1029/95JD03760

Draxler, R. R., Spring, S., Maryland, U. S. A., & Hess, G. D. (1998). An Overview of the HYSPLIT_4 Modelling System for Trajectories, Dispersion, and Deposition. *Australian Meteorological Magazine*, *47*, 295–308.

ECMWF. (2023). *ECMWF | Parameter details*. https://codes.ecmwf.int/grib/param-db/?id=135

Eşsiz, M., & Acar, Z. (2023). Synoptic analysis of the January 2004 snowstorm: Example of Çanakkale. *International Journal of Engineering and Geosciences*, *8*(1), 11–18. https://doi.org/10.26833/ijeg.988115

European Environment Agency. (2017, February 14). *LRTAP - European Environment Agency*. https://www.eea.europa.eu/help/glossary/eea-glossary/long-range-transport-of-air-pollutants

Feingold, G., Eberhard, W. L., Veron, D. E., & Previdi, M. (2003). First measurements of the Twomey indirect effect using ground-based remote sensors. *Geophysical Research Letters*, *30*(6). https://doi.org/10.1029/2002GL016633

Fenger, J. (2002). Urban air quality. *Developments in Environmental Science*, *1*(C), 1–52. https://doi.org/10.1016/S1474-8177(02)80004-3

Ferrarese, S., Longhetto, A., Cassardo, C., Apadula, F., Bertoni, D., Giraud, C., & Gotti, A. (2002). A study of seasonal and yearly modulation of carbon dioxide sources and sinks, with a particular attention to the Boreal Atlantic Ocean. *Atmospheric Environment*, *36*(35), 5517–5526. https://doi.org/10.1016/S1352-2310(02)00669-6

Gebhart, K. A., Schichtel, B. A., & Barna, M. G. (2005). Directional biases in back trajectories caused by model and input data. *Journal of the Air & Waste Management Association (1995)*, *55*(11), 1649–1662. https://doi.org/10.1080/10473289.2005.10464758

Genç Tokgöz, D. D. (2013). *Temporal variation in aerosol composition at Northwestern Turkey*. Middle East Technical University.

Genç Tokgöz, D. D. (2023). *Characterization of size segregated PM10 at a rural site in the Eastern Mediterranean. Manuscript in preparation*.

Genç Tokgöz, D. D., & Tuncel, G. (2008). Investigation of Influence of Synoptic Scale Atmospheric Movements on Observed Particulate Matter Over Northwestern Turkey. *National Air Pollution and Control Conference*, 741–749.

Genç Tokgöz, D. D., & Tuncel, G. (2011). Cluster Analysis of Back Trajectories Arriving in Northwestern Turkey and Chemical Composition of Aerosols. *Ninth National Environmental Engineering Conference*.

Google Earth. (2023). https://earth.google.com/web/@39.46024535,37.94818859,1360.72262502a,2502663.57000768d,35y,0.67989728h,0t,0r

Govender, P., & Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric Pollution Research*, *11*(1), 40–56. https://doi.org/10.1016/J.APR.2019.09.009

Gozzo, L. F., Rocha, R. P. da, Gozzo, L. F., & Rocha, R. P. da. (2013). Air-Sea Interaction Processes Influencing the Development of A Shapiro-Keyser Type Cyclone over the Subtropical South Atlantic Ocean. *PApGe*, *170*(5), 917–934. https://doi.org/10.1007/S00024-012-0584-3

Güllü, G., Doğan, G., & Tuncel, G. (2005). Atmospheric trace element and major ion concentrations over the eastern Mediterranean Sea: Identification of anthropogenic source regions. *Atmospheric Environment*, *39*(34), 6376–6387. https://doi.org/10.1016/J.ATMOSENV.2005.07.031

Güllü, G. H., Ölmez, I., Aygün, S., & Tuncel, G. (1998). Atmospheric trace element concentrations over the eastern Mediterranean Sea: Factors affecting temporal variability. *Journal of Geophysical Research: Atmospheres*, *103*(D17), 21943–21954. https://doi.org/10.1029/98JD01358

Güllü, G. H., Ölmez, I., & Tuncel, G. (2000). Temporal variability of atmospheric trace element concentrations over the eastern Mediterranean Sea. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *55*(7), 1135–1150. https://doi.org/10.1016/S0584-8547(00)00206-8

Güllü, G., Ölmez, I., & Tuncel, G. (2004). Source apportionment of trace elements in the Eastern Mediterranean atmosphere. *Journal of Radioanalytical and Nuclear Chemistry*, *259*(1), 165–171. https://doi.org/10.1023/B:JRNC.0000015823.02927.4A/METRICS

Günaydin, G. C., & Tuncel, G. (2003). Sources Regions Affecting Chemical Composition of Aerosols and Precipitation in the Eastern Mediterranean Atmosphere Determined Using Trajectory Statistics. *Air Pollution Processes in Regional Scale*, 121–133. https://doi.org/10.1007/978-94-007-1071-9_13

Harris, J. M., Draxler, R. R., & Oltmans, S. J. (2005). Trajectory model sensitivity to differences in input data and vertical transport method. *Journal of Geophysical Research: Atmospheres*, *110*(D14), 1–8. https://doi.org/10.1029/2004JD005750

Harris, J. M., & Kahl, J. D. (1990). A descriptive atmospheric transport climatology for the Mauna Loa Observatory, using clustered trajectories. *Journal of Geophysical Research: Atmospheres*, *95*(D9), 13651–13667. https://doi.org/10.1029/JD095ID09P13651

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, *28*(1), 100. https://doi.org/10.2307/2346830

Hondula, D. M., Sitka, L., Davis, R. E., Knight, D. B., Gawtry, S. D., Deaton, M. L., Lee, T. R., Normile, C. P., & Stenger, P. J. (2010). A back-trajectory and air mass climatology for the Northern Shenandoah Valley, USA. *International Journal of Climatology*, *30*(4), 569–581. https://doi.org/10.1002/JOC.1896

Horlock, J. H. (1967). A Course in Thermodynamics. By J. KESTIN. Blaisdell, 1966. 615 pp. *Journal of Fluid Mechanics*, *29*(3), 622–623. https://doi.org/10.1017/S0022112067211065

HYSPLIT Trajectory Model Configuration. (2007). https://www.arl.noaa.gov/documents/workshop/NAQC2007/HTML_Docs/trajmetd.html

IBM. (2023). *IBM SPSS Statistics 23*. https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-23

Im, U., Incecik, S., Guler, M., Tek, A., Topcu, S., Unal, Y. S., Yenigun, O., Kindap, T., Odman, M. T., & Tayanc, M. (2013). Analysis of surface ozone and nitrogen oxides at urban, semi-rural and rural sites in Istanbul, Turkey. *Science of The Total Environment*, *443*, 920–931. https://doi.org/10.1016/J.SCITOTENV.2012.11.048

Im, U., Markakis, K., Koçak, M., Gerasopoulos, E., Daskalakis, N., Mihalopoulos, N., Poupkou, A., Kindap, T., Unal, A., & Kanakidou, M. (2012). Summertime aerosol chemical composition in the Eastern Mediterranean and its sensitivity to temperature. *Atmospheric Environment*, *50*, 164–173. https://doi.org/10.1016/J.ATMOSENV.2011.12.044

İm, U., Tayanç, M., & Yenigün, O. (2006). Analysis of major photochemical pollutants with meteorological factors for high ozone days in Istanbul, Turkey. *Water, Air, and Soil Pollution*, *175*(1–4), 335–359. https://doi.org/10.1007/S11270-006-9142-X/METRICS

Im, U., Tayanç, M., & Yenigün, O. (2008). Interaction patterns of major photochemical pollutants in Istanbul, Turkey. *Atmospheric Research*, *89*(4), 382–390. https://doi.org/10.1016/J.ATMOSRES.2008.03.015

Kahl, J. D., Harris, J. M., Herbert, G. A., & Olson, M. P. (1989). Intercomparison of three long-range trajectory models applied to Arctic haze. *Tellus B: Chemical and Physical Meteorology*, *41 B*(5), 524–536. https://doi.org/10.3402/TELLUSB.V41I5.15109

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, B., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., … Joseph, D. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *BAMS*, *77*(3), 437–472. https://doi.org/10.1175/1520-0477(1996)077

Karaca, F., Alagha, O., Ertürk, F., Yilmaz, Y. Z., & Özkara, T. (2008). Seasonal Variation of Source Contributions to Atmospheric Fine and Coarse Particles at Suburban Area in Istanbul, Turkey. *Https://Home.Liebertpub.Com/Ees*, *25*(5), 767–781. https://doi.org/10.1089/EES.2007.0170

Karaca, F., Anil, I., & Alagha, O. (2009). Long-range potential source contributions of episodic aerosol events to PM10 profile of a megacity. *Atmospheric Environment*, *43*(36), 5713–5722. https://doi.org/10.1016/J.ATMOSENV.2009.08.005

Karaca, F., & Camci, F. (2010). Distant source contributions to PM10 profile evaluated by SOM based cluster analysis of air mass trajectory sets. *Atmospheric Environment*, *44*(7), 892–899. https://doi.org/10.1016/J.ATMOSENV.2009.12.006

Kasparoglu, S., Incecik, S., & Topcu, S. (2018). Spatial and temporal variation of O3, NO and NO2 concentrations at rural and urban sites in Marmara Region of Turkey. *Atmospheric Pollution Research*, *9*(6), 1009–1020. https://doi.org/10.1016/J.APR.2018.03.005

Kassomenos, P., Vardoulakis, S., Borge, R., Lumbreras, J., Papaloukas, C., & Karakitsios, S. (2010a). Comparison of statistical clustering techniques for the classification of modelled atmospheric trajectories. *Theoretical and Applied Climatology*, *102*(1), 1–12. https://doi.org/10.1007/S00704-009-0233-7/METRICS

Kassomenos, P., Vardoulakis, S., Borge, R., Lumbreras, J., Papaloukas, C., & Karakitsios, S. (2010b). Comparison of statistical clustering techniques for the classification of modelled atmospheric trajectories. *Theoretical and Applied Climatology*, *102*(1), 1–12. https://doi.org/10.1007/S00704-009-0233-7/METRICS

Kilic, M., & Kilic, S. (2023). Ionic Compositions of Sequential Rainfall Samples as Source Signatures of Forest Fire Emissions. *Chromatographia*, *86*(2), 153–165. https://doi.org/10.1007/S10337-023-04233-8/METRICS

Kilic, M., & Pamukoglu, M. Y. (2023). Characterization of water-insoluble particulate matters in sequential rain samples collected by a novel automatic sampler in Antalya, Turkey. *Atmospheric Pollution Research*, *14*(4), 101722. https://doi.org/10.1016/J.APR.2023.101722

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., & Fiorino, M. (2001). The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society*, 247–267. https://opensky.ucar.edu/islandora/object/articles%3A10212/

Knippertz, P., & Wernli, H. (2010). A Lagrangian Climatology of Tropical Moisture Exports to the Northern Hemispheric Extratropics. *Journal of Climate*, *23*(4), 987–1003. https://doi.org/10.1175/2009JCLI3333.1

Koçak, M., Kubilay, N., Herut, B., & Nimmo, M. (2007). Trace metal solid state speciation in aerosols of the northern Levantine Basin, East Mediterranean. *Journal of Atmospheric Chemistry*, *56*(3), 239–257. https://doi.org/10.1007/S10874-006-9053-7/METRICS

Koçak, M., Kubilay, N., & Mihalopoulos, N. (2004). Ionic composition of lower tropospheric aerosols at a Northeastern Mediterranean site: implications regarding sources and long-range transport. *Atmospheric Environment*, *38*(14), 2067–2077. https://doi.org/10.1016/J.ATMOSENV.2004.01.030

Koçak, M., Mihalopoulos, N., & Kubilay, N. (2007a). Chemical composition of the fine and coarse fraction of aerosols in the northeastern Mediterranean. *Atmospheric Environment*, *41*(34), 7351–7368. https://doi.org/10.1016/J.ATMOSENV.2007.05.011

Koçak, M., Mihalopoulos, N., & Kubilay, N. (2007b). Contributions of natural sources to high PM10 and PM2.5 events in the eastern Mediterranean. *Atmospheric Environment*, *41*(18), 3806–3818. https://doi.org/10.1016/J.ATMOSENV.2007.01.009

Koçak, M., Mihalopoulos, N., & Kubilay, N. (2009a). Origin and source regions of PM10 in the Eastern Mediterranean atmosphere. *Atmospheric Research*, *92*(4), 464–474. https://doi.org/10.1016/J.ATMOSRES.2009.01.005

Koçak, M., Mihalopoulos, N., & Kubilay, N. (2009b). Origin and source regions of PM10 in the Eastern Mediterranean atmosphere. *Atmospheric Research*, *92*(4), 464–474. https://doi.org/10.1016/J.ATMOSRES.2009.01.005

Koçak, M., Theodosi, C., Zarmpas, P., Séguret, M. J. M., Herut, B., Kallos, G., Mihalopoulos, N., Kubilay, N., & Nimmo, M. (2012). Influence of mineral dust transport on the chemical composition and physical properties of the Eastern Mediterranean aerosol. *Atmospheric Environment*, *57*, 266–277. https://doi.org/10.1016/J.ATMOSENV.2012.04.006

Kohonen, T. (1990). The Self-Organizing Map. *Proceedings of the IEEE*, *78*(9), 1464–1480. https://doi.org/10.1109/5.58325

Kubilay, N., Nickovic, S., Moulin, C., & Dulac, F. (2000). An illustration of the transport and deposition of mineral dust onto the eastern Mediterranean. *Atmospheric Environment*, *34*(8), 1293–1303. https://doi.org/10.1016/S1352-2310(99)00179-X

Kubilay, N., Oguz, T., Koçak, M., & Torres, O. (2005). Ground-based assessment of Total Ozone Mapping Spectrometer (TOMS) data for dust transport over the northeastern Mediterranean. *Global Biogeochemical Cycles*, *19*(1), 1–9. https://doi.org/10.1029/2004GB002370

Kuzu, S. L., Saral, A., Demir, S., Summak, G., & Demir, G. (2013). A detailed investigation of ambient aerosol composition and size distribution in an urban atmosphere. *Environmental Science and Pollution Research*, *20*(4), 2556–2568. https://doi.org/10.1007/S11356-012-1149-9/METRICS

Lyons, T. J., & Scott, W. D. (1990). Principles of air pollution meteorology. In *Principles of air pollution meteorology*. Belhaven Press, Pinter. https://doi.org/10.2307/635293

Ma, Y. F., Du, B. Y., Wang, Q., Hu, Q. Q., Bian, Y. S., Wang, M. B., & Jin, S. Y. (2019). Analysis of the atmospheric pollution transport pathways and sources in Shenyang, based on the HYSPLIT model. *IOP Conference Series: Earth and Environmental Science*, *351*(1), 012030. https://doi.org/10.1088/1755-1315/351/1/012030

MacDonald, A. M., Anlauf, K. G., Leaitch, W. R., Chan, E., & Tarasick, D. W. (2011). Interannual variability of ozone and carbon monoxide at the Whistler high elevation site: 2002-2006. *Atmospheric Chemistry and Physics*, *11*(22), 11431–11446. https://doi.org/10.5194/ACP-11-11431-2011

Mahapatra, P. S., Ray, S., Das, N., Mohanty, A., Ramulu, T. S., Das, T., Chaudhury, G. R., & Das, S. N. (2013). Urban air-quality assessment and source apportionment studies for Bhubaneshwar, Odisha. *Theoretical and Applied Climatology*, *112*(1–2), 243–251. https://doi.org/10.1007/S00704-012-0732-9/METRICS

Markou, M. T., & Kassomenos, P. (2010). Cluster analysis of five years of back trajectories arriving in Athens, Greece. *Atmospheric Research*, *98*(2–4), 438–457. https://doi.org/10.1016/J.ATMOSRES.2010.08.006

Martin, D., Bergametti, G., & Strauss, B. (1990). On the use of the synoptic vertical velocity in trajectory model: Validation by geochemical tracers. *Atmospheric Environment. Part A. General Topics*, *24*(8), 2059–2069. https://doi.org/10.1016/0960-1686(90)90240-N

Mayer, H. (1999). Air pollution in cities. *Atmospheric Environment*, *33*(24–25), 4029–4037. https://doi.org/10.1016/S1352-2310(99)00144-2

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M. A., Arbel, T., Avants, B. B., Ayache, N., Buendia, P., Collins, D. L., Cordier, N., … Van Leemput, K. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, *34*(10), 1993–2024. https://doi.org/10.1109/TMI.2014.2377694

Moody, J. L., Oltmans, S. J., Levy, H., Merrill, J. T., Moody, J. L., Oltmans, S. J., Levy, H., & Merrill, J. T. (1995). Transport climatology of tropospheric ozone: Bermuda, 1988-1991. *JGR*, *100*(D4), 7179–7194. https://doi.org/10.1029/94JD02830

Mutlu, A. (2020). Air quality impact of particulate matter (PM10) releases from an industrial source. *Environmental Monitoring and Assessment*, *192*(8), 1–17. https://doi.org/10.1007/S10661-020-08508-7/METRICS

Mylona, S. (1996). Sulphur dioxide emissions in Europe 1880–1991 and their effect on sulphur concentrations and depositions. *Tellus B*, *48*(5), 662–689. https://doi.org/10.1034/J.1600-0889.1996.T01-2-00005.X

NCEP/NCAR Global Reanalysis Products, 1948-continuing - Dataset - DASH Search - Production. (2023). https://data.ucar.edu/dataset/ncep-ncar-global-reanalysis-products-1948-continuing

NOAA. (2007). *Trajectory Vertical Motion*. https://www.ready.noaa.gov/documents/Tutorial/html/traj_vert.html

NOAA. (2011). *Air Resources Laboratory HYSPLIT Model Research*. www.arl.noaa.gov/HYSPLIT_info.php

NOAA. (2012). *GDAS0.5*. https://www.ready.noaa.gov/data/archives/gdas0p5/readme_gdas0p5_info.txt

NOAA. (2019a). *Air Resources Laboratory - GDAS Data Archive*. https://www.ready.noaa.gov/gdas1.php

NOAA. (2019b). *GFS0.25*. https://www.ready.noaa.gov/data/archives/gfs0p25/readme_gfs0p25_info.txt

NOAA. (2019c). *NOAA - Air Resources Laboratory*. https://www.ready.noaa.gov/faq_md3.php

NOAA. (2023). *READY - Gridded Data Archives*. https://www.ready.noaa.gov/archives.php

NOAA - Air Resources Laboratory -. (2019). https://www.ready.noaa.gov/faq_md3.php

NOAA Air Resources Laboratory. (2007). *HYSPLIT Vertical Motion Options*. https://www.arl.noaa.gov/documents/workshop/NAQC2007/HTML_Docs/trajvert.html

Oğuz, K., & Dündar, C. (2014). Analysis of Dust Transport Event via Remote Sensing and Numerical Forecast Model. *İzmir Aegean Geographical Journal*, *23*(2), 53–64. http://disc.sci.gsfc.nasa.gov/giovanni

One-Dimensional Isentropic Flow. (2006). http://brennen.caltech.edu/fluidbook/basicfluiddynamics/compressibleflow/onedimensionalisentropicflow.pdf

Ortiz-Amezcua, P., Guerrero-Rascado, J. L., Granados-Muñoz, M. J., Bravo-Aranda, J. A., & Alados-Arboledas, L. (2014). Characterization of atmospheric aerosols for a long range transport of biomass burning particles from canadian forest fires over the southern iberian peninsula in july 2013. *Optica Pura y Aplicada*, *47*(1), 43–49. https://doi.org/10.7149/OPA.47.1.43

Oruc, I. (2022). Transport routes and potential source areas of PM10 in Kirklareli, Turkey. *Environmental Monitoring and Assessment*, *194*(2), 1–14. https://doi.org/10.1007/S10661-022-09772-5/METRICS

Özdemir, E. T. (2019). Investigations of a Southerly Non-Convective High Wind Event in Turkey and Effects on PM10 Values: A Case Study on April 18, 2012. *Pure and Applied Geophysics*, *176*(10), 4599–4622. https://doi.org/10.1007/S00024-019-02240-1/METRICS

Ozdemir, H., Mertoglu, B., Demir, G., Deniz, A., & Toros, H. (2012). Case study of PM pollution in playgrounds in Istanbul. *Theoretical and Applied Climatology*, *108*(3–4), 553–562. https://doi.org/10.1007/S00704-011-0543-4/METRICS

Öztürk, F. (2009). *Investigation of short and long term trends in the eastern Mediterranean aerosol composition*. Middle East Technical University.

Öztürk, F., Zararsz, A., Dutkiewicz, V. A., Husain, L., Hopke, P. K., & Tuncel, G. (2012). Temporal variations and sources of Eastern Mediterranean aerosols based on a 9-year observation. *Atmospheric Environment*, *61*, 463–475. https://doi.org/10.1016/J.ATMOSENV.2012.07.051

Papayannis, A., Mamouri, R. E., Amiridis, V., Giannakaki, E., Veselovskii, I., Kokkalis, P., Tsaknakis, G., Balis, D., Kristiansen, N. I., Stohl, A., Korenskiy, M., Allakhverdiev, K., Huseyinoglu, M. F., & Baykara, T. (2012). Optical properties and vertical extension of aged ash layers over the Eastern Mediterranean as observed by Raman lidars during the Eyjafjallajökull eruption in May 2010. *Atmospheric Environment*, *48*, 56–65. https://doi.org/10.1016/J.ATMOSENV.2011.08.037

Peña, J. M., Lozano, J. A., & Larrañaga, P. (1999). An empirical comparison of four initialization methods for the K-Means algorithm. *Pattern Recognition Letters*, *20*(10), 1027–1040. https://doi.org/10.1016/S0167-8655(99)00069-0

Piñero-García, F., Ferro-García, M. A., Chham, E., Cobos-Díaz, M., & González-Rodelas, P. (2015). A cluster analysis of back trajectories to study the behaviour of radioactive aerosols in the south-east of Spain. *Journal of Environmental Radioactivity*, *147*, 142–152. https://doi.org/10.1016/J.JENVRAD.2015.05.029

Qin, N., Kong, X. Z., Zhu, Y., He, W., He, Q. S., Yang, B., Ou-Yang, H. L., Liu, W. X., Wang, Q. M., & Xu, F. L. (2012). Distributions, sources, and backward trajectories of atmospheric polycyclic aromatic hydrocarbons at Lake Small Baiyangdian, Northern China. *TheScientificWorldJournal*, *2012*. https://doi.org/10.1100/2012/416321

Rastgeldi Dogan, T., & Yalcin, S. P. (2020, December 3). *The Atmospheric Transported Desert Dust Over Sanliurfa (Turkey) And Its Structural Properties*. Sigma Journal of Engineering and Natural Sciences. https://www.researchgate.net/publication/365960890_THE_ATMOSPHERIC_TRANSPORTED_DESERT_DUST_OVER_SANLIURFA_TURKEY_AND_ITS_STRUCTURAL_PROPERTIES

Reanalysis Data Sources. (2007). https://www.cpc.ncep.noaa.gov/products/wesley/data.html

REVIHAAP Project. (2013). https://apps.who.int/iris/handle/10665/341712

Rolph, G. D., & Draxler, R. R. (1990). Sensitivity of Three-Dimensional Trajectories to the Spatial and Temporal Densities of the Wind Field. *JApMe*, *29*(10), 1043–1054. https://doi.org/10.1175/1520-0450(1990)029

Sari, D., İncecik, S., & Ozkurt, N. (2016). Surface ozone levels in the forest and vegetation areas of the Biga Peninsula, Turkey. *Science of The Total Environment*, *571*, 1284–1297. https://doi.org/10.1016/J.SCITOTENV.2016.07.168

Sari, D., Incecik, S., & Ozkurt, N. (2020). Analysis of surface ozone episodes using WRF-HYSPLIT model at Biga Peninsula in the Marmara region of Turkey. *Atmospheric Pollution Research*, *11*(12), 2361–2378. https://doi.org/10.1016/J.APR.2020.09.018

Saunders, R. O., Scotty, E., & Kahl, J. D. W. (2013). The sensitivity of single air parcel trajectory calculations to starting elevation. *The Science of the Total Environment*, *463–464*, 229–236. https://doi.org/10.1016/J.SCITOTENV.2013.06.007

Sciare, J., Bardouki, H., Moulin, C., & Mihalopoulos, N. (2003). Aerosol sources and their Contribution to the chemical composition of aerosols in the Eastern Mediterranean Sea during summertime. *Atmospheric Chemistry and Physics*, *3*(1), 291–302. https://doi.org/10.5194/ACP-3-291-2003

Sciare, J., Oikonomou, K., Cachier, H., Mihalopoulos, N., Andreae, M. O., Maenhaut, W., & Sarda-Estève, R. (2005). Aerosol mass closure and reconstruction of the light scattering coefficient over the Eastern Mediterranean sea during the MINOS campaign. *Atmospheric Chemistry and Physics*, *5*(8), 2253–2265. https://doi.org/10.5194/ACP-5-2253-2005

Seibert, P., Kromp-Kolb, H., Baltensperger, U., Jost, D. T., & Schwikowski, M. (1994). Trajectory Analysis of High-Alpine Air Pollution Data. *Air Pollution Modeling and Its Application X*, 595–596. https://doi.org/10.1007/978-1-4615-1817-4_65

Shannon, W. D. (2007). Cluster Analysis. *Handbook of Statistics*, *27*, 342–366. https://doi.org/10.1016/S0169-7161(07)27011-7

Sirois, A., & Bottenheim, J. W. (1995). Use of backward trajectories to interpret the 5-year record of PAN and O3 ambient air concentrations at Kejimkujik National Park, Nova Scotia. *Journal of Geophysical Research: Atmospheres*, *100*(D2), 2867–2881. https://doi.org/10.1029/94JD02951

Srinivas, C. V., Venkatesan, R., Baskaran, R., Rajagopal, V., & Venkatraman, B. (2012). Regional scale atmospheric dispersion simulation of accidental releases of radionuclides from Fukushima Dai-ichi reactor. *Atmospheric Environment*, *61*, 66–84. https://doi.org/10.1016/J.ATMOSENV.2012.06.082

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., & Ngan, F. (2015). NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System. *Bulletin of the American Meteorological Society*, *96*(12), 2059–2077. https://doi.org/10.1175/BAMS-D-14-00110.1

Stohl, A. (1998). Computation, accuracy and applications of trajectories—A review and bibliography. *Atmospheric Environment*, *32*(6), 947–966. https://doi.org/10.1016/S1352-2310(97)00457-3

Stohl, A., & Seibert, P. (1998). Accuracy of trajectories as determined from the conservation of meteorological tracers. *Quarterly Journal of the Royal Meteorological Society*, *124*(549), 1465–1484. https://doi.org/10.1002/QJ.49712454907

Stohl, A., & Wotawa, G. (1995, October). *Interpolation errors in wind fields as a function of spatial and temporal resolution and their impact on different types of kinematic trajectories*. https://folk.nilu.no/~andreas/publications/4.pdf

Sturman, A., & Zawar-Reza, P. (2002). Application of back-trajectory techniques to the delimitation of urban clean air zones. *Atmospheric Environment*, *36*(20), 3339–3350. https://doi.org/10.1016/S1352-2310(02)00253-4

Su, L., Yuan, Z., Fung, J. C. H., & Lau, A. K. H. (2015). A comparison of HYSPLIT backward trajectories generated from two GDAS datasets. *The Science of the Total Environment*, *506–507*, 527–537. https://doi.org/10.1016/J.SCITOTENV.2014.11.072

Tawadrou, A., & Katsabani, P. (2005). Prediction of surface blast patterns in limestone quarries using artificial neural networks. *Fragblast*, *9*(4), 233–242. https://doi.org/10.1080/13855140600761863

Theodosi, C., Im, U., Bougiatioti, A., Zarmpas, P., Yenigun, O., & Mihalopoulos, N. (2010). Aerosol chemical composition over Istanbul. *Science of The Total Environment*, *408*(12), 2482–2491. https://doi.org/10.1016/J.SCITOTENV.2010.02.039

Topuz, M., & Karabulut, M. (2017). Hatay Çöl Tozu Taşınımı Değerlendirmesi. *Journal of Turkish Studies*, *12*(Volume 12 Issue 3), 565–580. https://doi.org/10.7827/TURKISHSTUDIES.11575

Toros, H., Geertsema, G., & Cats, G. (2014). Evaluation of the HIRLAM and HARMONIE Numerical Weather Prediction Models during an Air Pollution Episode over Greater Istanbul Area. *CLEAN-SOIL AIR WATER*, *42*(7), 863–870. https://doi.org/10.1002/CLEN.201200306

Tošić, I., & Unkašević, M. (2013). Extreme daily precipitation in Belgrade and their links with the prevailing directions of the air trajectories. *Theoretical and Applied Climatology*, *111*(1–2), 97–107. https://doi.org/10.1007/S00704-012-0647-5/METRICS

Türküm, A., Pekey, B., Pekey, H., & Tuncel, G. (2008). Comparison of sources affecting chemical compositions of aerosol and rainwater at different locations in Turkey. *Atmospheric Research*, *89*(4), 306–314. https://doi.org/10.1016/J.ATMOSRES.2008.03.011

Ünal, A. (2016). *Diyarbakır İli İçin 2015 Yılında Çöl Tozları Taşınımının Bsc-dream8b Modeli İle Araştırılması*. http://hdl.handle.net/11527/15243

Uygur, N., Karaca, F., & Alagha, O. (2010). Prediction of sources of metal pollution in rainwater in Istanbul, Turkey using factor analysis and long-range transport models. *Atmospheric Research*, *95*(1), 55–64. https://doi.org/10.1016/J.ATMOSRES.2009.08.007

Uygur, N., & Saral, A. (2013). Evaluation of the effect of Marmara Sea on the characterization of the coastal line atmosphere in İstanbul. *Sigma Journal of Engineering and Natural Sciences*, *31*(3), 429-. https://sigma.yildiz.edu.tr/article/779

Vitali, L., Righini, G., Piersanti, A., Cremona, G., Pace, G., & Ciancarella, L. (2017). M-TraCE: a new tool for high-resolution computation and statistical elaboration of backward trajectories on the Italian domain. *Meteorology and Atmospheric Physics*, *129*(6), 629–643. https://doi.org/10.1007/S00703-016-0491-8/METRICS

Wang, L., Liu, Z., Sun, Y., Ji, D., & Wang, Y. (2015). Long-range transport and regional sources of PM2.5 in Beijing based on long-term observations from 2005 to 2010. *Atmospheric Research*, *157*, 37–48. https://doi.org/10.1016/J.ATMOSRES.2014.12.003

Wang, Y. Q., Zhang, X. Y., Arimoto, R., Cao, J. J., & Shen, Z. X. (2004). The transport pathways and sources of PM10 pollution in Beijing during spring 2001, 2002 and 2003. *Geophysical Research Letters*, *31*(14). https://doi.org/10.1029/2004GL019732

Yavuz, V., Özen, C., Çapraz, Ö., Özdemir, E. T., Deniz, A., Akbayır, İ., & Temur, H. (2022). Analysing of atmospheric conditions and their effects on air quality in Istanbul using SODAR and CEILOMETER. *Environmental Science and Pollution Research International*, *29*(11), 16213–16232. https://doi.org/10.1007/S11356-021-16958-W

Zemmer, F., Karaca, F., & Ozkaragoz, F. (2012). Ragweed pollen observed in Turkey: Detection of sources using back trajectory models. *Science of The Total Environment*, *430*, 101–108. https://doi.org/10.1016/J.SCITOTENV.2012.04.067

Zhao, N., Wang, G., Li, G., Lang, J., & Zhang, H. (2020). Air pollution episodes during the COVID-19 outbreak in the Beijing–Tianjin–Hebei region of China: An insight into the transport pathways and source distribution. *Environmental Pollution*, *267*, 115617. https://doi.org/10.1016/J.ENVPOL.2020.115617

Zhou, L., Hopke, P. K., & Liu, W. (2004). Comparison of two trajectory based models for locating particle sources for two rural New York sites. *Atmospheric Environment*, *38*(13), 1955–1963. https://doi.org/10.1016/J.ATMOSENV.2003.12.034

APPENDICES

APPENDIX A

NCEP/NCAR GLOBAL REANALYSIS DATA SET DETAILS

Pressure Level Data

2.5 degree latitude-longitude global grid

144x73 points from 90N-90S, 0E-357.5E

1/1/1948 - present with output every 6 hours

Levels (hPa): 1000,925,850,700,600,500,400,300,250,200, 150,100,70,50,30,20,10

Surface or near the surface (.995 sigma level) winds and temperature

Precipitation

Model Type: LAT-LON

Vert Coord: 2

Numb X pt: 144

Numb Y pt: 73

Numb Levels: 18

Sfc Variables: 5 PRSS T02M U10M V10M TPP6

Upper Levels: 6 HGTS TEMP UWND VWND WWND RELH

Sigma Level Data (CONUS extract for the DATEM archive, not available)

The spectral coefficients on 28 model sigma surfaces were processed to obtain required fields 4 per day on a global Gaussian grid of 1.875 degree resolution. A regional sub-grid covering the continental US and Canada was extracted.

Current USA spatial domain: 21.9N 127.5W to 60.0N 52.5W

Output every 6 hours

Levels: .995,.982,.964,.943,.916,.884,.846,.801,.751,.694,.633,
.568,.502,.436,.372,.312,.258,.210,.168,.133,.103,.078, .058,.042,.029,.018,.010,.003

MODEL TYPE: MERCATOR

VERT COORD: 1

POLE LAT: 21.904

POLE LON: -127.5

REF LAT: 21.904

REF LON: -127.5

REF GRID: 136.5

ORIENTATION: 0.

CONE ANGLE: 0.

SYNC X: 1.

SYNC Y: 1.

SYNC LAT: 21.904

SYNC LON: -127.5

NUMB X: 57

NUMB Y: 41

NUMB LEVELS: 29

SFC VARIABLES: 01 PRSS

UPPER LEVELS: 05 TEMP SPHU UWND VWND WWND (*ARL - Global Reanalysis Data Archive*, 2023)

APPENDIX B

GDAS1 DATA SET DETAILS

Table 17. Meteorological Fields contained in the GDAS Archive. For accumulation/average fields, 6-h acc/avg at 00, 06, 12, 18 UTC (NOAA, 2019a)

Field	Units	Label	Data Order
Pressure at surface	hPa	PRSS	S1
Pressure reduced to mean sea level	hPa	MSLP	S2
Accumulated precipitation (6 h accumulation)	m	TPP6	S3
u-component of momentum flux (3- or 6-h average)	N/m2	UMOF	S4
v-component of momentum flux (3- or 6-h average)	N/m2	VMOF	S5
Sensible heat net flux at surface (3- or 6-h average)	W/m2	SHTF	S6
Downward short wave radiation flux (3- or 6-h average)	W/m2	DSWF	S7
Relative Humidity at 2m AGL	%	RH2M	S8
U-component of wind at 10 m AGL	m/s	U10M	S9
V-component of wind at 10 m AGL	m/s	V10M	S10
Temperature at 2m AGL	K	TO2M	S11
Total cloud cover (3- or 6-h average)	%	TCLD	S12
Geopotential height	gpm*	SHGT	S13
Convective available potential energy	J/Kg	CAPE	S14
Convective inhibition	J/kg	CINH	S15
Standard lifted index	K	LISD	S16
Best 4-layer lifted index	K	LIB4	S17
Planetary boundary layer height	m	PBLH	S18
Temperature at surface	K	TMPS	S19
Accumulated convective precipitation (6 h accumulation)	m	CPP6**	S20
Volumetric soil moisture content	frac.	SOLM	S21
Categorial snow (yes=1, no=0) (3- or 6-h average)		CSNO	S22
Categorial ice (yes=1, no=0) (3- or 6-h average)		CICE	S23
Categorial freezing rain (yes=1, no=0) (3- or 6-h average)		CFZR	S24
Categorial rain (yes=1, no=0) (3- or 6-h average)		CRAI	S25
Latent heat net flux at surface (3- or 6-h average)	W/m2	LHTF	S26
Low cloud cover (3- or 6-h average)	%	LCLD	S27
Middle cloud cover (3- or 6-h average)	%	MCLD	S28
High cloud cover (3- or 6-h average)	%	HCLD	S29
Geopotential height	gpm*	HGTS	U1

Temperature	K	TEMP	U2
U-component of wind with respect to grid	m/s	UWND	U3
V-component of wind with respect to grid	m/s	VWND	U4
Pressure vertical velocity	hPa/s	WWND	U5
Relative humidity	%	RELH	U6

APPENDIX C

SENSITIVITY ANALYSIS SCRIPT

```
% Two data set will be compared
% set 1 and set2 are inputs

% set1 and set 2 includes lat in degree, lon in degree
% and height in meters

% t=0,1,2,.....96 hour therefore i=1:97
% n=1:1461; 1461 daily back trajectories in each data
%set

% Create separate matrix  for lat in radians, lon in
%radians and height in meters

% set A and set B contain the data sets used. By
%changing them to reise, revert, gdasise, gdasvert; 6
%comparisons can be made by using 4 data sets

setA=transpose(reise);


Lat_set1=setA(1:97,:)*pi()/180;% latitutes matrix in
radians
Lon_set1=setA(98:194,:)*pi()/180;% longititutes matrix
in radians
Height_set1=setA(195:291,:);% height matrix in meters

setB=transpose(gdasvert);


Lat_set2=setB(1:97,:)*pi()/180;% latitutes matrix in
radians;
Lon_set2=setB(98:194,:)*pi()/180;% longititutes matrix
in radians
Height_set2=setB(195:291,:);% height matrix in meters

%HAVERSINE FORMULA
% d is equal to Great Circle Distance in km
% R : earth radius (6371 km)
% AHTD is equal to sum of d divided by total number of
%trajectories
```

```
% actaully AHTD is equal to average d at each time
interval
earth_radius= 6371; % in km

delta_phi= Lat_set2-Lat_set1;% difference btwn
latitudes in radians
delta_lambda= Lon_set2-Lon_set1;% difference btwn
longitudes in radian

a=sin(delta_phi/2).*sin(delta_phi/2)+cos(Lat_set1).*cos
(Lat_set2).*sin(delta_lambda/2).*sin(delta_lambda/2);
c=2.*atan2(sqrt(a),sqrt(1-a));
d=earth_radius*c; % in km

for i=1:97
AHTD (i,:)= mean (d(i,:));
end

%average lenght of two trajectories that has been
%compared can be calculated as below

% average lenght for set 1:
for k=1:1461
    for i=2:97
    delta_phi_set1(i,k)=Lat_set1(i,k)-Lat_set1(i-1,k);%
difference btwn latitudes of individual trajectories at
time i and i-1 in radian
delta_lambda_set1(i,k)= Lon_set1(i,k)-Lon_set1(i-1,k);%
difference btwn longitudes of individual trajectories
at time i and i-1 in radian

a1(i,k)=sin(delta_phi_set1(i,k)/2).*sin(delta_phi_set1(
i,k)/2)+cos(Lat_set1(i,k)).*cos(Lat_set1(i-
1,k)).*sin(delta_lambda_set1(i,k)/2).*sin(delta_lambda_
set1(i,k)/2);

end
end

for k=1:1461
    for i=2:97

    sqrt_a1(i,k)=real(sqrt(a1(i,k)));
    c1(i,k)=2.*atan2(sqrt_a1(i,k),sqrt(1-a1(i,k)));
```

```
    d1(i,k)=earth_radius*c1(i,k);% t=0,1,2,3.....96 if
t=72 is needed, it means d1(73,:)
end
end

% average lenght for set 2:
for k=1:1461
   for i=2:97
    delta_phi_set2(i,k)=Lat_set2(i,k)-Lat_set2(i-1,k);%
difference btwn latitudes of individual trajectories at
time i and i-1 in radian
delta_lambda_set2(i,k)= Lon_set2(i,k)-Lon_set2(i-1,k);%
difference btwn longitudes of individual trajectories
at time i and i-1 in radian

a2(i,k)=sin(delta_phi_set2(i,k)/2).*sin(delta_phi_set2(
i,k)/2)+cos(Lat_set2(i,k)).*cos(Lat_set2(i-
1,k)).*sin(delta_lambda_set2(i,k)/2).*sin(delta_lambda_
set2(i,k)/2);


end
end
for k=1:1461
  for  i=2:97

    sqrt_a2(i,k)=real(sqrt(a2(i,k)));
    c2(i,k)=2.*atan2(sqrt_a2(i,k),sqrt(1-a2(i,k)));
    d2(i,k)=earth_radius*c2(i,k);% t=0,1,2,3.....96 if
t=72 is needed, it means d1(73,:)
end
end

L_1b2a=(d1+d2)/2;% AVERAGE LENGHT

% Relative horizontal transport deviation can be
%calculated:

for i=1:97
    for k=1:1461
dummy1(i,k) = (d(i,k)./L(i,k));
    end
end

for i=1:97
```

```
RHTD (i,:)= mean (dummy1(i,:));
end

 % ABSOLUTE VERTICAL TRANSPORT DEVIATION can be
%calculated:

dummy2=abs(Height_set2(:,:)-Height_set1(:,:));
for i=1:97
AVTD (i,:)= mean (dummy2(i,:));% in meters
end

% 96 hour results
t1=97
t2=73
t3=49
t4=25
RHTD_96h=RHTD(t1)
AHTD_96h=AHTD(t1)
AVTD_96h=AVTD(t1)
% 72 hour results
RHTD_72h=RHTD(t2)
AHTD_72h=AHTD(t2)
AVTD_72h=AVTD(t2)
%48 hour results
RHTD_48h=RHTD(t3)
AHTD_48h=AHTD(t3)
AVTD_48h=AVTD(t3)
%24 hour results
RHTD_24h=RHTD(t4)
AHTD_24h=AHTD(t4)
AVTD_24h=AVTD(t4)
```

APPENDIX D

VISUALIZATION SCRIPT OF ALL TRAJECTORIES

```
% x and y are the coordinates of the starting point
% (station)

% lat and lon are files containing the coordinates of the
% all trajectories. This scripts includes reise (1a) data
% set visualization.

% Since it is 1461 days of trajectory data, i increases
%from 1 to 1461.

f1=figure
geoshow('landareas.shp', 'FaceColor', [0.5 1.0 0.5]);
y=36.9700;
x=30.4338;
line(x,y,'marker','square','markersize',4,'color','y')
hold on
lat=reise_all_lat;
lon=reise_all_lon;

for i = 1:1461
  plot(lon{i,:},lat{i,:},'LineWidth', 0.5)
hold on;
end
```

APPENDIX E

CLUSTER VISUALIZATION SCRIPT

```
% x and y are the coordinates of the starting point
% (station)

% lat1 and lon1 are files containing the coordinates of
% the all trajectories. This scripts includes reise (1a)
% data set visualization.

% Since it is 5 cluster center data, i increases from 1
% to 5.



f1=figure
geoshow('landareas.shp', 'FaceColor', [0.5 1.0 0.5]);
y=36.9700;
x=30.4338;
line(x,y,'marker','square','markersize',4,'color','r')
text(x,y,' Station','vertical','top');
hold on
lat1=lat_reise5;
lon1=lon_reise5;

for i = 1:5
  plot(lon1{i,:},lat1{i,:}, 'k-', 'LineWidth', 1.5)
hold on;
end
```

APPENDIX F

PUBLICATIONS DERIVED FROM THESIS

1. Determination of Sensitivity of Back Trajectories by Using Reanalysis and GDAS Meteorology Archive

Doğrusever F., Genç Tokgöz D. D.

1st International Conference on Applied Engineering and Natural Sciences, Konya, Turkey, 1 - 03 November 2021, pp.475.

Determination of Sensitivity of Back Trajectories by Using Reanalysis and GDAS Meteorology Archive

Firdevs Doğrusever*, D. Deniz G. Tokgöz[2]

[1]*Department of Environmental Engineering/Hacettepe University, Turkey*
ORCID ID 0000-0002-5501-9804
[2]*Department of Environmental Engineering/Hacettepe University, Turkey*
ORCID ID 0000-0001-7904-2497
(firdevs.dogrusever@hacettepe.edu.tr)

Abstract – Determination of back trajectories is carried out by using the HYSPLIT model, which is widely used by air quality researchers to locate the source of a pollutant. There are several meteorology archives and vertical velocity methods that the model offers to its users.

Previous researches have shown that two different meteorology archives, GDAS (Global Data Assimilation System) and Reanalysis, and two different vertical velocity methods, model vertical velocity and isentropic are mostly used in the studies performed in Turkey. It is not certain that which meteorology archive or vertical velocity method gives better results. Back trajectories calculated for the same region can be quite different due to the model inputs. Even the fact that some studies do not even need to give any information about the model inputs reveals the necessity of doing a study on this subject. Therefore, the differences in the model inputs should be analyzed and it is necessary to highlight the difference.

To date, no study has been conducted in Turkey. The most important contribution of the study is numerically presenting the results of two different model inputs for the first time in the Mediterranean Region which is the most studied region in Turkey.

In this study, to identify the effects of different model inputs on results, HYSPLIT back trajectories have been visualized by using MATLAB, cluster analyses were performed by using SPSS, and then clusters were visualized by using MATLAB.

Consequently, the study focused on a non-consensus practice that which meteorology archive and vertical velocity methods should be used among practitioners. In this way, researchers can reach a consensus if it is determined that the model inputs have a significant effect on the results.

Keywords – Back trajectory; Clustering; HYSPLIT; GDAS; Reanalysis; Air quality modeling

2. Sensitivity of HYSPLIT Back Trajectory Clustering to Meteorological Inputs

GENÇ TOKGÖZ D. D., Doğrusever F.

10th International Symposium on Atmospheric Sciences (ATMOS22), İstanbul, Turkey, 18 October 2022.

Sensitivity of HYSPLIT Back Trajectory Clustering to Meteorological Inputs

D. Deniz Genç Tokgöz[1], Firdevs Doğrusever[2]

[1]Hacettepe University Faculty of Engineering, Department of Environmental Engineering, Ankara, Turkey

denizgenctokgoz@hacettepe.edu.tr; firdevs.dogrusever@hacettepe.edu.tr

ABSTRACT

In this study, 96-hr back trajectories for a rural area, located on the Eastern Mediterranean coast of Turkey, were simulated by the HYSPLIT model (web-version). Two different meteorological archives (Reanalysis and GDAS1) and two different vertical transport velocity options (isentropic and model vertical velocity) were chosen for the simulation of daily back trajectories between 2010 and 2013 years. These model inputs were selected as they are the most common input variables. Cluster Analysis by SPSS (k-means technique) was applied for each back trajectory data set to classify them into similar groups (clusters). Based on their speed, back trajectories in each data set were classified into 3 clusters as fast moving, mean, and slow-moving air mass movements. Clusters in each data set were compared with the corresponding Cluster in the other data set. Generated clusters were not significantly different although they indicated higher sensitivity to the vertical velocity method than the meteorological archive.

***Keywords*:** Cluster Analysis, k-means technique, sensitivity analysis, SPSS, MATLAB.

1. INTRODUCTION

Due to adverse impact of air pollution on human health (WHO 2013), ecosystem (Bytnerowicz, Omasa, and Paoletti 2007), and materials (Watt et al. 2009), ambient air quality standards have been getting stricter. In order to develop effective abatement strategies, it is essential to identify the pollutants and their sources.

Air masses may travel over the different geographical locations (e.g., sea, deserts, forests, etc) and pollution sources (i.e. cities, industrial facilities, etc) before arriving the receptor site, hence they can be loaded with natural and anthropogenic pollutants (Perez et al., 2017). To determine the contribution of local pollution, the history of air masses need to be investigated. Back trajectory calculations are the common tool for simulation of air masses. A back trajectory simulates the paths of air masses before intercepting the receptor site. Although there are large numbers of software programs that can be used to simulate back trajectories, the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model is widely used. HYSPLIT model requires meteorological data as input data and variety of parameters such as vertical transport velocity model, arriving time, arrival height and etc. The sensitivity to meteorological input data set (archives) has been computed (Su et al., 2015; Harris et al., 2005) and found to have significant influence (20-40%). Those studies also indicate that trajectories vary depending on the geographical properties of the receptor sites. Therefore, it is suggested to use different meteorological archives and input parameters to investigate the sensitivity of trajectories to input variables for each receptor site.

Typically, a back trajectory has an error of 20% of the distance travelled (Stohl, 1998). In air pollution studies, large number of trajectories over a long period of time are used in statistical analysis to increase the accuracy of the trajectory analysis.

Cluster Analysis (CA) is a multivariate statistical technique that is used to determine the atmospheric transport pattern and to examine the influence of different atmospheric transport pattern on observed chemical composition at the receptor. Many researchers have used CA to examine the relation between synoptic scale transport patterns and atmospheric pollution (Dorling et al., 1992; Cape et al., 2000, Borge et al., 2007, Cabello et al., 2008; Markou and Kassomenos, 2010). The criterion used in CA is to split a large trajectory data set into a number of groups based on trajectory transport speed and direction, simultaneously (Abdalmogith and Harrison, 2005,

Brankov et al., 1998). The produced groups in CA are called clusters. Members of each cluster have similar trajectory length and curvature, while distinct clusters represent different synoptic regimes. When coupled with aerosol chemical composition, CA can be a good tool to examine the effect of synoptic scale atmospheric patterns on observed chemical composition; hence will aid to establish source-receptor relationship. The uncertainty in individual trajectory is high and increases with the total travel distance (Stohl and Koffi, 1998). In CA large sets of trajectories are used, therefore the accuracy of trajectory analysis is improved.

In this study, for a receptor site located in the Eastern Mediterranean Turkey, back trajectories are simulated using the input data which are most widely used. Simulated back trajectories are classified by CA and then generated clusters are compared to determine the sensitivities to the most widely used input variables.

2. MATERIAL AND METHOD

In this study, four-days back trajectories (96 h), extending four days back in time and arriving at a rural site in the Eastern Mediterranean (360 58' 12" N and 300 26' 2" E) at 12:00 UTC (Coordinate Universal Time), at altitudes 1500 m from the surface were simulated for each day in the years 2010, 2011, 2012 and 2013. Trajectory arrival height may have different effects on trajectories (Cabello et al., 2008; Yang et al.,2017). The long-range transport studies in the Eastern Mediterranean region usually chooses 1500 m arrival height as the most representative transport layer (Dayan et al., 2017). In this study only 1500 m arrival height back trajectories were simulated as the objective is to determine the sensitivity of meteorological input data and vertical transport method on the trajectories.

Two different meteorological data achieves (Reanalysis and GDAS1) and two different vertical velocity models (isentropic and model vertical)) were used as model input. These input variables are chosen as these are the most widely used input variables. For each trajectory data set 1461 daily back trajectories (365 days/year * 4 years + 1 day for leap year =1461 daily back trajectories) were simulated. The details of trajectory data set simulated in this study is given in Table 1.

Table 1. Trajectory data set (96 h) simulated in this study

Symbol of data	Data Description		Total number of back trajectories
	Meteorology Archive	Vertical Velocity Method	
1a	Reanalysis meteorology archive	Isentropic	1461
1b	Reanalysis meteorology archive	Model vertical velocity	1461
2a	GDAS1 meteorology archive	Isentropic	1461
2b	GDAS1 meteorology archive	Model vertical velocity	1461

Non-hierarchical k-means technique in IBM SPSS Statics 23 was used for cluster analysis of these data sets.

Sensitivity analysis was performed for the generated clusters to determine the differences caused by the model inputs. For sensitivity analysis, the absolute horizontal transport deviation (AHTD), the absolute vertical transport deviation (AVTD) and the relative horizontal transport deviation (RHTD) were calculated using the below equations:

$$AHTD(t) = \frac{1}{N}\sum_{n=1}^{N}\{[X_n(t) - x_n(t)]^2 + [Y_n(t) - y_n(t)]^2\}^{1/2} \quad (1)$$

$$AVTD(t) = \frac{1}{N}\sum_{n=1}^{N}|Z_n(t) - z_n(t)| \quad (2)$$

$$RHTD(t) = \frac{1}{N}\sum_{n=1}^{N}\frac{\{[X_n(t)-x_n(t)]^2+[Y_n(t)-y_n(t)]^2\}^{1/2}}{AL_n(t)} \quad (3)$$

$$AL_n(t) = \frac{1}{2}\sum_{t_i=2}^{t}\langle \begin{matrix}\{[X_n(t_i) - X_n(t_{i-1})]^2 + [Y_n(t_i) - Y_n(t_{i-1})]^2\}^{1/2} + \\ \{[x_n(t_i) - x_n(t_{i-1})]^2 + [y_n(t_i) - y_n(t_{i-1})]^2\}^{1/2}\end{matrix}\rangle \quad (4)$$

In this equations, N is the total number of back trajectories, X,Y,Z is the location of the first set of back trajectories; x,y,z is the location of the second set of back trajectories; ALn(t) represents the average length of the two back trajectories compared. In this study, cluster centroids were compared, therefore the N value given in Equations 1,2 and 3 was taken as 1.

The formula given in Equation 1 (also in Equations 3 and 4) is actually the Euclidean length. Considering that the Earth is spherical, it would be incorrect to calculate the distance between two locations directly by the Euclidean length. When the studies on back trajectory analyses were examined, it was seen that the distance calculated in Equation 1 were replaced by the distance calculated by the Haversine formula (Markou and Kassomenos, 2010; Cabello at al., 2008). The Haversine formula is used to calculate the shortest distance between two points with known latitude and longitude values on the earth's geolocation. The Haversine formula assumes that the earth is a perfect sphere and the shortest distance between two points is calculated with 0.3% error. Therefore, the distance between two locations in Equations 1, 3 and 4 were calculated using the Haversine formula given below:

$$a = sin^2((\varphi B - \varphi A)/2) + cos\,\varphi A * cos\,\varphi B * sin^2((\lambda B - \lambda A)/2)$$
$$c = 2 * atan2(\sqrt{a}, \sqrt{(1-a)})$$
$$d = R \cdot c \qquad (5)$$

In this equation, φ and λ represent latitude and longitude (in radians), respectively. B represents the second position and A represents the first position. R represents the mean radius of the earth (6371 km taken). d is the distance (also called as the Great Circle Distance (in km)).

3. RESULTS

There is no definitively accepted method to determine the optimum number of clusters in cluster analysis. In general, depending on the decrease in the number of clusters, that is, when the number of clusters is reduced from the number of k clusters to the number of k-1 clusters, the percentage change in the total root mean square deviation (TRMSD) (5% change is considered significant) is used as the criterion (Dorling et al., 1992).

In this study, the optimum number of clusters is found to be different for each data set. It is not possible to match the clusters in one data set with the clusters in the other data set when the number

of clusters is different. In order to compare the cluster in each data set with the corresponding cluster in the other data set, the optimum cluster number was selected as 3 (Figure 1).

Generated Cluster in each data set were compared with the corresponding Cluster in the other data set. Comparisons were designed to determine the sensitivity to the vertical transport velocity (1a-1b, 2a-2b), meteorological archive (1a-2a,1b-2b), and both meteorological archive and vertical transport velocity (1a-2b,1b-2a). The sensitivity parameters: absolute horizontal transport deviation (AHTD), absolute vertical transport deviation (AVTD) and the relative horizontal transport deviation (RHTD) were calculated for each comparison. Table 2-4 presents AHTD, RHTD and AVTD statistics (median values) calculated for Cluster 1 (C1), Cluster 2 (C2), and Cluster 3 (C3), respectively.

Table 2. C1-Comparison of fast-moving air movements

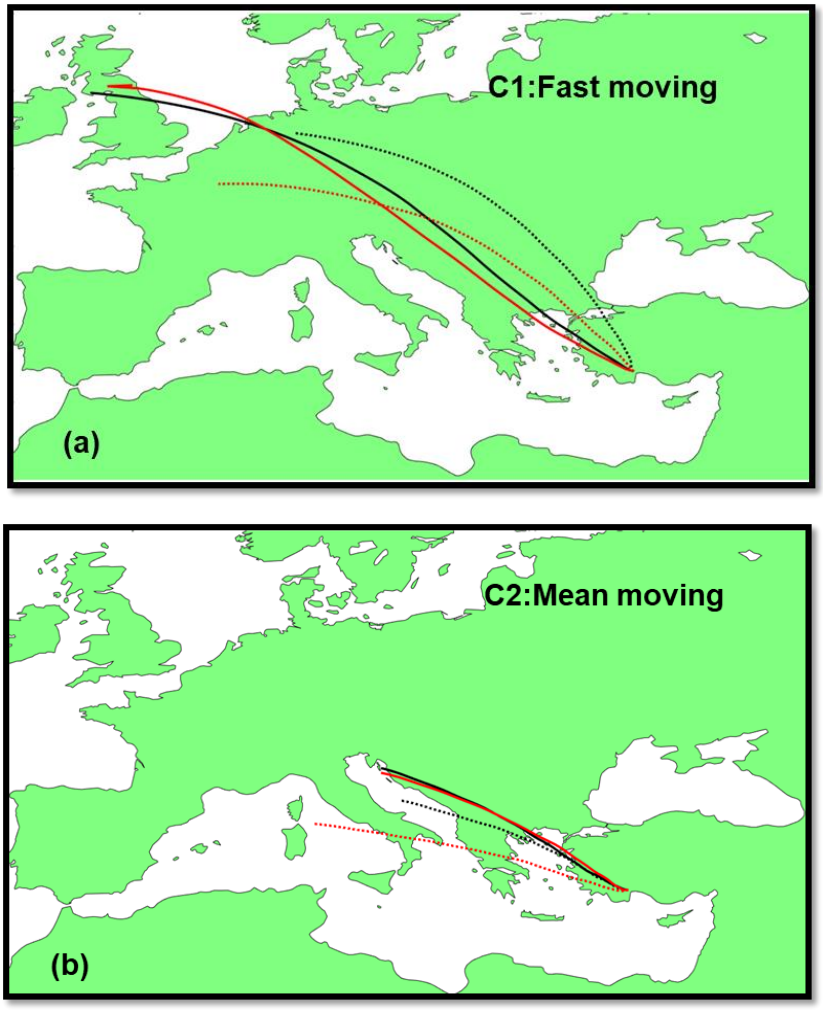
Comparison	**Description**	**RHTD, %**	**AHTD, km**	**AVTD, m**
1a-1b	Sensitivity of the Reanalysis archive to the velocity method	16.7	575	297
2a-2b	Sensitivity of the GDAS1 archive to the velocity method	15.1	571	272
1a-2a	Sensitivity to meteorology archive	1.8	65	519
1b-2b		6.7	195	900
1a-2b	Sensitivity to both meteorology archive and velocity method	14.9	518	599
1b-2a		17.4	653	17

Table 3. C2-Comparison of mean moving air movements.

Comparison	**Description**	**RHTD, %**	**AHTD, km**	**AVTD, m**
1a-1b	Sensitivity of the Reanalysis archive to the velocity method	7	114	113
2a-2b	Sensitivity of the GDAS1 archive to the velocity method	12.2	236	89
1a-2a	Sensitivity to meteorology archive	1.7	28	87
1b-2b		13.4	244	137
1a-2b	Sensitivity to both meteorology archive and velocity method	11	207	34
1b-2a		5.4	92	157

Table 4. C3-Comparison of slow moving air movements

Comparison	Description	RHTD, %	AHTD, km	AVTD, m
1a-1b	Sensitivity of the Reanalysis archive to the velocity method	24.1	237	108
2a-2b	Sensitivity of the GDAS1 archive to the velocity method	14.6	125	340
1a-2a	Sensitivity to meteorology archive	4.9	35	95
1b-2b		15.5	179	452
1a-2b	Sensitivity to both meteorology archive and velocity method	12.8	115	431
1b-2a		27.8	266	106



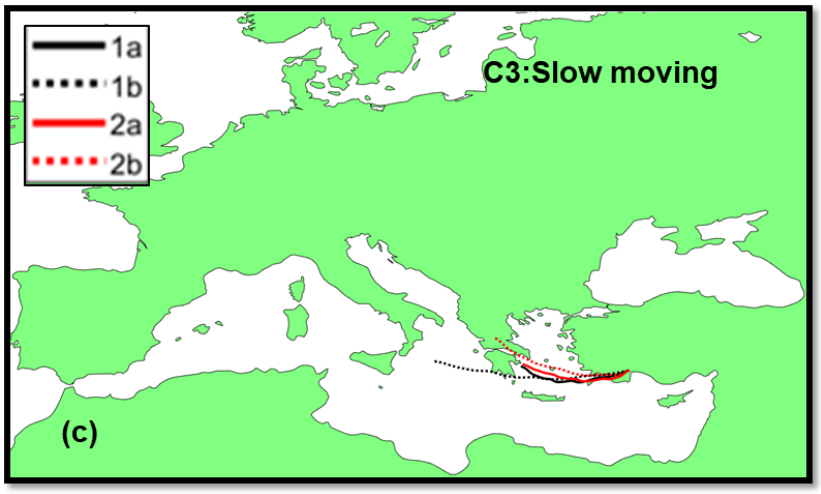


Figure 1. Variation of clusters with respect to data set (a) C1-Fast moving, (b) C2-Mean moving, (c) C3-slow moving (1a: Reanalysis meteorology archive with isentropic velocity method, 1b: Reanalysis meteorology archive with model vertical velocity method, 2a: GDAS1 meteorology archive with isentropic velocity method, 2b: GDAS1 meteorology archive with model vertical velocity method)

4. DISCUSSION

Figure clearly shows that back trajectories were classified according to their transport speed and direction. Cluster 1 (C1) represents air movements with the highest velocity and named as “fast moving air movements”. Cluster 3 (C3) represents slow air movements and named as “slow moving air movements”. Air movements that are faster than Cluster 3 but slower than Cluster 1 are defined as Cluster 2 (C2) and called as “mean moving air movements”.

The elevation of the clusters presented in Figure 2 also confirms the above classification with respect to trajectory speeds. Trajectories move faster at high elevations but their speed decreases

as they lose elevation. The elevation of C1, C2, and C3 was between 1500-6220 m, 1498-3092 m, and 677 -1750 m, respectively.

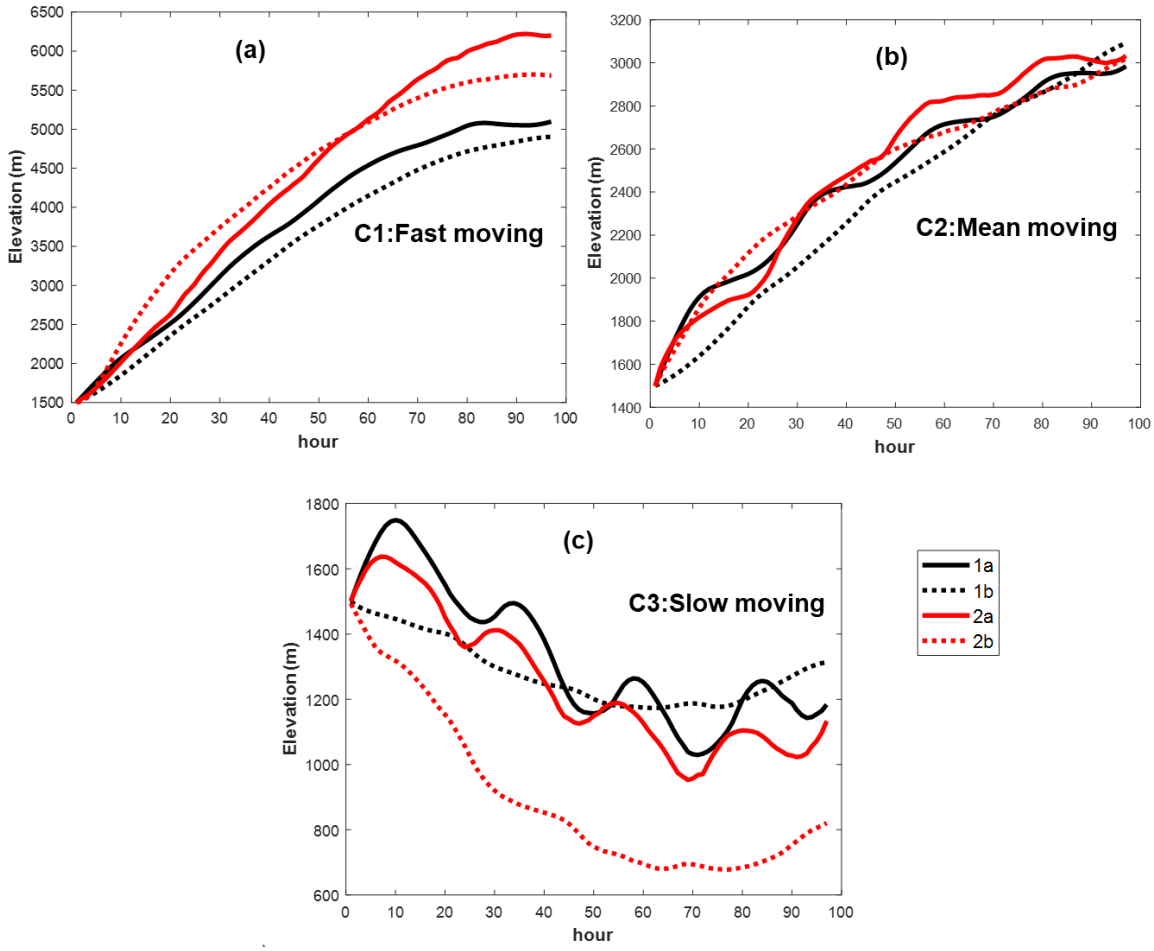


Figure 2. Average elevation of clusters (a) C1-Fast moving, (b) C2-Mean moving, (c) C3-Slow moving (1a: Reanalysis meteorology archive with isentropic velocity method, 1b: Reanalysis meteorology archive with model vertical velocity method, 2a: GDAS1 meteorology archive with isentropic velocity method, 2b: GDAS1 meteorology archive with model vertical velocity method)

Among the statistics reported in Tables 2-4, AVTD were found to be less informative for comparison of clusters. Therefore, AHTD and RHTD values were considered when evaluating the sensitivity of clusters to the meteorological archive and vertical velocity method.

AHTD and RHTD parameters indicated that clusters are more sensitive to vertical transport method. Clusters generated from model vertical velocity back trajectories were found to be slightly different than those obtained by isentropic back trajectories (Figure 1). These differences were calculated as 15-17% for C1 (Fast moving air movements), 7-12% for C2 (Mean moving air movements) and 15-24% for C3 (Slow moving air movements). The sensitivity to meteorological archive was calculated between 2% and 15.5%. The difference due to meteorological archive for C1, C2 and C3 were 2-7%, 2-13% and 5-16%, respectively.

The sensitivities of clusters (C1, C2, and C3) generated from isentropic back trajectories (1a and 2a data sets) were calculated between 2-5%. Clusters generated from model vertical velocity back trajectories (1b and 2b) indicated 7-16% sensitivity to meteorological archive. The sensitivities of clusters from model vertical velocity back trajectories were nearly three times higher than those generated from isentropic back trajectories. This indicates that clusters generated from model vertical velocity back trajectories are more sensitive to meteorological archive.

The sensitivity of meteorological archive to vertical velocity method was calculated by comparing 1a-1b, and 2a-2b data sets. Reanalysis meteorological archive's sensitivity to vertical velocity model is calculated between 7 and 24%. Medium moving trajectories, represented by C2, indicated lowest sensitivity to vertical velocity model while Fast moving (C1) and slow moving trajectories (C3) indicated higher sensitivities (17 and 24%, respectively). Sensitivity of GDAS1 archive to velocity method is nearly constant in all clusters (12-15%).

C1 (Table 2) showed the highest sensitivity to vertical transport method (15-17%) and lowest sensitivity to the meteorological archive (2-7%). When evaluating the sensitivity to both meteorological archive and vertical transport velocity together, similar values close to the sensitivity of the vertical velocity method were found. This shows that vertical velocity method has a dominant effect on fast moving air movements.

C2 (Table 3) has the highest sensitivity to GDAS1 meteorological archive (13%) when model vertical velocity method is used. Isentropic trajectories generated quite similar clusters (1a-2a) with 2% RHTD value. This indicates that medium moving trajectories are not sensitive to meteorological archive when isentropic velocity model is chosen. Although the difference is not significant, the sensitivity of GDAS1 meteorological archive to vertical velocity model is high (12%).

C3 (Table 4), representing the slow moving air movements, showed the highest sensitivity to model vertical velocity method when Reanalysis archive is used (28%). The sensitivity of GDAS1 and Reanalysis archive to velocity method is 15 and 24%, respectively. Isentropic trajectories generated quite similar clusters (5%) while clusters from model vertical velocity trajectories were a bit different (15%).

Sensitivity of each data set to generated clusters were also calculated for preceding 24, 48, and 72-hr back trajectories (not presented here). Sensitivities were found to increase with the total travel time i.e. 96-hr back trajectories were more sensitive than preceding back trajectories.

In summary, the results of cluster analysis of back trajectories, simulated using Reanalysis and GDAS1 meteorological archives and isentropic and model vertical velocity methods, indicated that clusters are not significantly different (Figure 1). Although the length and curvature of clusters may show variation, direction of clusters are quite similar with the corresponding clusters.

ACKNOWLEDGEMENTS

This work was supported by Hacettepe University Scientific Research Projects Coordination Unit. Project Number: 18766, 2022.

REFERENCES

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

Brankov, E., Rao, S. T. & Porter, P. S. (1998). *A trajectory-clustering-correlation methodology for examining the long-range transport of air pollutants*, Atmospheric Environment, 32, 1525-1534.

Borge, R., Lumbreras, J., Vardoulakis, S., Kassomenos, P. & Rodriguez, E. (2007). *Analysis of long-range transport influences on urban PM10 using two-stage atmospheric trajectory clusters*, Atmospheric Environment, 41**,** 4434-4450.

Bytnerowicz, A., K. Omasa, and E. Paoletti. (2007). *Integrated effects of air pollution and climate change on forests: A northern hemisphere perspective*, Environmental Pollution 147 (3):438-45. doi: 10.1016/j.envpol.2006.08.028.

Cabello, M., Orza, J. A. G., Galiano, V., & Ruiz, G. (2008). *Influence of meteorological input data on backtrajectory cluster analysis – a seven-year study for southeastern Spain*, Advances in Science and Research, 2(1), 65–70.

Cape, J. N., Methven, J. & Hudson, L. E. (2000). *The use of trajectory cluster analysis to interpret trace gas measurements at Mace Head, Ireland,* Atmospheric Environment, 34**,** 3651-3663.

Dayan, U., P. Ricaud, R. Zbinden and F. Dulac (2017), *Atmospheric pollution over the eastern Mediterranean during summer – a review*, Atmos. Chem. Phys. 17(21), 13233-13263.

Dorling, S. R., Davies, T. D., & Pierce, C. E. (1992). *Cluster analysis: A technique for estimating the synoptic meteorological controls on air and precipitation chemistry—Method and applications,* Atmospheric Environment. Part A. General Topics, 26(14), 2575–2581.

Harris, J. M., Draxler, R. R. & Oltmans, S. J. (2005). *Trajectory model sensitivity to differences in input data and vertical transport method,* J. Geophys. Res.: Atmos., 110.

Markou, M. T., & Kassomenos, P. (2010). *Cluster analysis of five years of back trajectories arriving in Athens, Greece,* Atmospheric Research, 98(2–4), 438–457.

Perez, I.A., Sanchez, M. L., Garcia1, M. A., Pardo, N.(2017). *Boundaries of air mass trajectory clustering: key points and applications,* Int. J. Environ. Sci. Technol.14:653–662.

San José, R., Stohl, A., Karatzas, K., Bohler, T., James, P., & Pérez, J. L. (2005). *A modelling study of an extraordinary night time ozone episode over Madrid domain,* Environmental Modelling & Software, 20(5), 587–593.

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., & Ngan, F. (2015). *NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System*, Bulletin of the American Meteorological Society, 96(12), 2059–2077.

Stohl, A. & Koffi, N. E. (1998). *Evaluation of trajectories calculated from ECMWF data against constant volume balloon flights during ETEX,* Atmospheric Environment, 32, 4151-4156.

Stohl, A. (1998). *Computation, accuracy and applications of trajectories - A review and bibliography,* Atmospheric Environment, 32, 947-966.

Su, L., Yuan, Z., Fung, J. C. H. & Lau, A. K. H. (2015). *A comparison of HYSPLIT backward trajectories generated from two GDAS datasets*, Science of the Total Environment, 506-507, 527-537.

Watt, J., J. Tidblad, R. Hamilton, and V. Kucera. (2009). *The effects of air pollution on cultural heritag*e, The Effects of Air Pollution on Cultural Heritage.

WHO. 2013. *Review of Evidence on Health Aspects of Air Pollution-REVIHAAP Project*, Technical Report. Bonn, Germany.

Yang, W., G. Wang and C. Bi (2017), *Analysis of long-range transport effects on PM2.5 during a short severe haze in Beijing, China*, Aerosol and Air Quality Research 17(6), 1510-1522.

3. Comparison of HYSPLIT Back Trajectories Simulated with Different Meteorological Inputs

Doğrusever F., Genç Tokgöz D. D.

10th International Conference on Environmental Management, Engineering, Planning and Economics (CEMEPE) & SECOTOX Conference, Skaithos Island, Greece, 05 -09 June 2023

Comparison of HYSPLIT Back Trajectories Simulated with Different Meteorological Inputs

F. Doğrusever[1], D. D. G. Tokgöz*[1]

[1]Department of Environmental Engineering, Hacettepe University, Ankara, 06800, Turkey

Corresponding author: E-mail: denizgenctokgoz@hacettepe.edu.tr, Tel +90 312 297 7800 (135), Fax: +90 312 2992053

Abstract

The Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model (web-version) was used to simulate hourly 96-hr back trajectories arriving at a rural site in the Eastern Mediterranean, at altitudes 1500 m from the surface, for each day between 2010 and 2013 years. Two meteorological data archives (NCEP/NCAR Reanalysis and GDAS1) and vertical velocity methods (isentropic and model vertical velocity) were used as model inputs as these are the most widely used input variables. The sensitivity of trajectories to model inputs was measured by the absolute horizontal transport deviation (AHTD), the absolute vertical transport deviation (AVTD) and the relative horizontal transport deviation (RHTD) statistics. Both the meteorological archive and vertical transport method significantly influenced the trajectories. Trajectories simulated by NCEP/NCAR Reanalysis archive were less sensitive to the vertical transport method than trajectories simulated by the GDAS1 archive. Back trajectories simulated for the same region could be very different from each other due to differences in model inputs. Since trajectory uncertainties could be specific to location, multiple inputs must be used to simulate trajectories to examine the air mass pathway.

Keywords: Air quality modelling, Backward trajectory, Long-range transport, single air parcel trajectory.