ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

# DATA ANALYSIS AND SIMULATION APPLICATIONS ON EUROPEAN AIR TRAFFIC MODELLING AND SPATIOTEMPORAL GRID EMISSION MODELLING

M.Sc. THESIS

Yiğit Bekir Kaya

Department of Aeronautics and Astronautics

Aeronautics Engineering Programme

Thesis Advisor: Assoc. Prof. Dr. Gökhan İnalhan

MAY 2015

**ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY**

# DATA ANALYSIS AND SIMULATION APPLICATIONS ON EUROPEAN AIR TRAFFIC MODELLING AND SPATIOTEMPORAL GRID EMISSION MODELLING

**M.Sc. THESIS**

**Yiğit Bekir Kaya**
**511131142**

**Department of Aeronautics and Astronautics**

**Aeronautics Engineering Programme**

**Thesis Advisor: Assoc. Prof. Dr. Gökhan İnalhan**

**MAY 2015**

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

# AVRUPA HAVA TRAFİĞİ VE UZAY-ZAMANSAL GRİD SALINIM MODELLEMEDE VERİ ANALİZİ VE SİMULASYONU UYGALAMALARI

**YÜKSEK LİSANS TEZİ**

**Yiğit Bekir Kaya**
**511131142**

**Uçak ve Uzay Mühendisliği Anabilim Dalı**

**Uçak Mühendisliği Programı**

**Tez Danışmanı: Doç. Dr. Gökhan İnalhan**

**MAYIS 2015**

**Yiğit Bekir Kaya**, a **M.Sc.** student of ITU **Graduate School of Science Engineering and Technology** 511131142, successfully defended the **thesis** entitled "**DATA ANALYSIS AND SIMULATION APPLICATIONS ON EUROPEAN AIR TRAFFIC MODELLING AND SPATIOTEMPORAL GRID EMISSION MODELLING**", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

**Thesis Advisor :** **Assoc. Prof. Dr. Gökhan İnalhan** ..............................
İstanbul Technical University

**Jury Members :** **Prof. Dr. İbrahim Özkol** .............................
İstanbul Technical University

**Dr. Nazım Kemal Üre** ..............................
İstanbul Technical University

**Date of Submission : 4 May 2015**
**Date of Defense : 22 May 2015**

***To my beloved mother,***

## FOREWORD

I want to thank my family, my advisor Mr. Inalhan, my laboratory mates, and Mr. Unal from Eurasia Institute for their continous support. I also want to thank Utku Eren for bringing the research to the state that I took up where he left off.
It has been an honor for me to do my research in Control and Avionics Laboratory since 2013.

May 2015

Yiğit Bekir Kaya
Computer Engineering BSc. &
Aeronautics Engineering BSc. and
MSc. Student

## TABLE OF CONTENTS

**FOREWORD** .......... **ix**
**TABLE OF CONTENTS** .......... **xi**
**ABBREVIATIONS** .......... **xiii**
**LIST OF TABLES** .......... **xiv**
**LIST OF FIGURES** .......... **xv**
**SUMMARY** .......... **xvii**
**ÖZET** .......... **xx**
**1. INTRODUCTION** .......... **1**
1.1 Purpose of Thesis .......... 1
1.2 Literature Review .......... 1
1.2.1 Air traffic literature review .......... 1
1.2.2 Emissions literature review .......... 2
**2. DATA SPECIFICATIONS AND PREPARATION** .......... **5**
2.1 ALL_FT+ Data Fields, Used Fields .......... 5
2.2 BigData Management Tool .......... 7
2.3 Data Flaws and Fixes .......... 9
2.3.1 The same consecutive airspaces flaw .......... 10
2.3.2 Not connected consecutive airspaces flaw .......... 11
2.3.3 Airspaces with zero elapsed time flaw .......... 11
2.3.4 Inconsistent planned and actual profiles flaw .......... 12
2.3.5 Scripts used for filtering data set .......... 14
**3. EUROPEAN AIR TRAFFIC FLOW MODEL ANALYSIS** .......... **17**
3.1 Airport Classification .......... 17
3.1.1 Airport classification according to air traffic flow rate .......... 21
3.1.2 Airport classification according to flight durations .......... 24
3.2 Capacity Calculation .......... 26
3.3 Capacity Normalisation .......... 27
3.4 Off-block Time Normalisation .......... 28
3.5 Flight Duration Normalisation .......... 28
3.6 Aircraft Tracking .......... 28
3.7 Simulation Algorithm .......... 30
3.7.1 Simulate .......... 30
3.7.2 SearchSlotDeparture and SearchSlotArrival .......... 31
3.7.3 SearchSlot and SearchSlotInTimeWindow .......... 32
3.7.4 Preprocess function .......... 33
3.7.4.1 Group_Airport_by_Region and Calculate_Nominal_Capacities .......... 34
3.7.4.2 AssignFlightTimes, GetWindowTime, Normalise_Flight_Durations, Normalise_Flight_Times .......... 35
3.7.5 Variables, symbols, and function definitions .......... 36
**4. EMISSION DISTRIBUTION AS 4D GRIDS OVER MARMARA REGION** .......... **39**
4.1 Impact of Aircraft Emissions on Local Air Quality. .......... 39
4.2 Data Preparation for Emission .......... 40
4.3 Emission Modelling .......... 42
4.3.1 Flight mode determination .......... 43
**5. RESULTS AND CONCLUSIONS** .......... **47**

5.1 Data Understanding .... 47
5.2 Initial Simulation Results .... 47
5.3 Emission Grid Results .... 52
5.3.1 Day of week emission distribution .... 54
5.3.2 Monthly emission distribution .... 56
5.3.3 Hourly emission distribution .... 58
5.3.4 Emission distribution by altitude .... 60
**REFERENCES .... 62**
**APPENDICES .... 64**
APPENDIX A: Total Flow Chart .... 64
APPENDIX B: Case Study on Unexpected Arrival Delays at Frankfurt Airport .. 65
Case 1: Arrival demand over capacity on EDDP airport .... 65
Case 2 and 3: Arrival demand over capacity on LGIR airport .... 65
APPENDIX C: ALL_FT+ Detailed Fields .... 67
**CURRICULUM VITAE .... 71**
**OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS : .... 72**

## ABBREVIATIONS

| | |
|---|---|
| **ACC** | Area Control Center |
| **ANSP** | Air Navigation Service Provider |
| **AOBT** | Actual Off-Block Time |
| **ASM** | Airspace Management Planning Charts |
| **ATC** | Air Traffic Control |
| **ATCO** | Air Traffic Control Officer |
| **ATFM** | Air Traffic Flow Management |
| **ATM** | Air Traffic Management |
| **ATS** | Air Traffic Service |
| **ATFCM** | Air Traffic Flow and Capacity Management |
| **AUA** | Air Traffic Control Unit Airspace |
| **CFMU** | Central Flow Management Unit |
| **CMAQ** | Community Multi-scale Air Quality Model |
| **COBT** | Calculated Off-Block Time |
| **CPR** | Correlated Position Report |
| **CTFM** | Current Tactical Flight Model |
| **CTOT** | Calculated Take-Off Time |
| **DDR** | Demand Data Repository |
| **ECAC** | European Civil Aviation Conference |
| **EOBT** | Estimated Off-Block Time |
| **ETA** | Estimated Time of Arrival |
| **ETFMS** | Enhanced Tactical Flow Management System |
| **FIR** | Flight Information Region |
| **FOA** | First Order Approximation |
| **FPL** | Filed Flight Plan |
| **FTFM** | Filed Tactical Flight Model |
| **IBT** | In-Block Time |
| **IFPS** | Integrated Initial Flight Plan Processing System |
| **IOBT** | Initial Off-Block Time |
| **LOBT** | Last Off-Block Time |
| **NEVAC** | Network Estimation and Visualization of ACC Capacity |
| **RTFM** | Regulated Tactical Flight Model |
| **SAAM** | System for Airspace Analysis at Macroscopic Level SESAR |
| **SMOKE** | Sparse Matrix Operator Kernel Emission |
| **SRS** | Schedule Reference Service |
| **STAR** | Standard Terminal Arrival Route |

**LIST OF TABLES**

**Page**

**Table 2.1 :** Fields Used in Simulation ... 6
**Table 2.2 :** Airspace Profile Fields ... 6
**Table 2.3 :** Delay Calculation Methods ... 7
**Table 3.1 :** The busiest airports in Europe, November 2011 ... 17
**Table 3.2 :** The busiest airports in Europe, July 2011 ... 18
**Table 3.3 :** The Simulate Algorithm ... 31
**Table 3.4 :** SearchSlotDeparture Algorithm ... 31
**Table 3.5 :** SearchSlot Algorithm ... 32
**Table 3.6 :** SearchSlotInTimeWindow Algorithm ... 32
**Table 3.7 :** Preprocess Algorithm ... 34
**Table 3.8 :** Group_Airports_by_Region Algorithm ... 34
**Table 3.9 :** Calculate_Nominal_Capacities Algorithm ... 35
**Table 3.10 :** AssignFlightTimes Algorithm ... 35
**Table 3.11 :** GetTimeWindowTime Algorithm ... 35
**Table 3.12 :** Normalise_Flight_Durations Algorithm ... 36
**Table 3.13 :** The busiest airports in Europe, November 2011 ... 36
**Table 3.14 :** The busiest airports in Europe, November 2011 ... 36
**Table 5.1 :** Total Mass of the Polutants in Tonnes ... 52
**Table 5.2 :** Pollutant Emission Correlation Score with Day of Week. ... 56
**Table 5.3 :** Relative Pollution Factors for Each Pollutant and Phase of Flight. ... 56
**Table 5.4 :** Pollutant Emission Correlation Score with Hour of the Day. ... 59
**Table A.1 :** Sample from DDR2 data. ... 65
**Table A.2 :** Arrival Demand in EDDP Airport. ... 65
**Table A.3 :** Sample from DDR2 data. ... 66
**Table A.4 :** Arrival Demand in LGIR Airport. ... 66
**Table A.5 :** Capacity Calculation with Different Methods. ... 66
**Table A.6 :** ALL_FT+ Detailed Fields ... 67

## LIST OF FIGURES

**Page**

**Figure 2.1 :** BigData Management Tool ........................................................ 9
**Figure 2.2 :** An example for same consecutive airspace problem for planned profile. ........................................................ 10
**Figure 2.3 :** An example for same consecutive airspace problem for actual profile. 10
**Figure 2.4 :** An example for not connected consecutive airspaces for planned profile. ........................................................ 11
**Figure 2.5 :** An example of Airspaces with Zero Elapsed Time for planned data.... 11
**Figure 2.6 :** The Airspaces in FIR level that belong to the flawed data for Inconsistent Planned and Actual Profiles Problem. ........................................ 13
**Figure 2.7 :** An example of category three, flawed data. .................................. 14
**Figure 3.1 :** Flow distribtuion in July 2011.......................................................... 18
**Figure 3.2 :** Flow distribtuion in November 2011 ............................................... 19
**Figure 3.3 :** Connectivity graphs of major airports, July 2011. ............................ 20
**Figure 3.4 :** Connectivity graphs of major airports, November 2011. .................... 20
**Figure 3.5 :** Movement distribution across Europe, July 2011. ............................ 21
**Figure 3.6 :** Movement distribution across Europe, November 2011. .................... 22
**Figure 3.7 :** Air traffic volume ratios across Europe, July 2011. ........................... 23
**Figure 3.8 :** Air traffic volume ratios across Europe, November 2011................... 23
**Figure 3.9 :** Flight Distribution across Europe, July 2011. .................................... 24
**Figure 3.10 :** Flight Distribution across Europe, November 2011. ......................... 24
**Figure 3.11 :** Flight Distribution of London Heathrow Airport, July 2011. ............ 25
**Figure 3.12 :** Flight Distribution of Marco Polo Airport, July 2011........................ 25
**Figure 3.13 :** Franfurk Airport Arrival Demand Profile, July 2011......................... 27
**Figure 3.14 :** Simulation Algortithm Flow Chart.................................................... 38
**Figure 4.1 :** Marmara Region Map......................................................................... 39
**Figure 4.2 :** Regular Flight with a Small Grade Example....................................... 43
**Figure 4.3 :** Graded Flight Example........................................................................ 44
**Figure 5.1 :** Connection Graph of Busiest Airports. ............................................... 48
**Figure 5.2 :** Delay and Connection Graph of Europe Region. ................................ 49
**Figure 5.3 :** Frankfurt Departure Delay Profile, July 2011. .................................... 49
**Figure 5.4 :** Frankfurt Departure Delay Profile, Nobember 2011........................... 50
**Figure 5.5 :** Frankfurt Arrival Delay Profile, July 2011. ........................................ 50
**Figure 5.6 :** Frankfurt Arrival Delay Profile, Nobember 2011. .............................. 51
**Figure 5.7 :** Live Simulation European Region Only. ............................................ 51
**Figure 5.8 :** Live Simulation, Delays Occuring. .................................................... 52
**Figure 5.9 :** Total Emission in Tonnes for Marmara Region, 2011 (low resolution). ........................................................ 53
**Figure 5.10 :** Total Emission in Tonnes for Marmara Region, 2011 (high resolution). ........................................................ 53
**Figure 5.11 :** HC Total Emission in Tonnes for Each Day of Week, 2011. ............ 55

**Figure 5.12 :** NOx Total Emission in Tonnes for Each Day of Week, 2011. ........... 55
**Figure 5.13 :** CO Total Emission in Tonnes for Each Day of Week, 2011. ............. 56
**Figure 5.14 :** CO Total Emission in Tonnes for Each Month of Year, 2011. ........... 57
**Figure 5.15 :** HC Total Emission in Tonnes for Each Month of Year, 2011. ........... 57
**Figure 5.16 :** NOx Total Emission in Tonnes for Each Month of Year, 2011. ......... 58
**Figure 5.17 :** NOx Total Emission in Tonnes for Each Hour of Day, 2011. ............ 58
**Figure 5.18 :** HC Total Emission in Tonnes for Each Hour of Day, 2011................ 59
**Figure 5.19 :** CO Total Emission in Tonnes for Each Hour of Day, 2011................ 59
**Figure 5.20 :** HC Total Emission in Tonnes for Each Altitude................................ 60
**Figure 5.21 :** CO Total Emission in Tonnes for Each Altitude................................ 60
**Figure 5.22 :** NOx Total Emission in Tonnes for Each Altitude. ............................ 61
**Figure A.1 :** Life Cycle of the Application Code Structure .................................... 64

**DATA ANALYSIS AND SIMULATION APPLICATIONS ON EUROPEAN AIR TRAFFIC MODELLING AND SPATIOTEMPORAL GRID EMISSION MODELLING**

## SUMMARY

With its intrinsic complexity, rapidly growing demand and almost saturated infrastructures, Air Transport Management is one of the most challenging fields of the near future. To be able to respond the need, there are many conducted researches and initiatives as Single European Sky ATM Research (SESAR) programme in Europe and its US counterpart NextGen. Resilience2050.eu is one of the projects carried out in Europe with the purpose of analyzing the current air traffic system's deficiencies in terms of resilience ability. Aimed to design disruption and perturbation adaptive ATM concepts of future beyond SESAR within the boundaries of safety, this research project analyzes the current system dynamics focusing on the propagation of unexpected and undesired events through the whole ATM system which underlies the theme of this thesis.

Macro analysis of the European Air Traffic Network system play a key role in pinpointing the elements and events that drive the system which are crucial for a resilient structure. Specifically the air transportation network across the airports, airspaces, subspaces and its segmentation define the structure of the flow network. In addition, scheduled flights, their densities across this network, and corresponding flight patterns define the nature of the flow across the network. The analyses provide an insight about the transportation network system's dynamics and serve to the construction of more accurate models which will enable the application of optimization problems into real world conditions. Accordingly, they form an essential part for the purpose of building a resilient structure and have an utter importance to construct reliable and robust air traffic management and control infrastructures.

The analyses in this project have been conducted with two types of data as ALL_FT+ trajectory data and DDR capacity data. Each file of the trajectory data gives the all flights' trajectories with point and airspace profiles over European Air Traffic Network whereas DDR capacity data provides the declared capacities for different entities of air transportation system such as airports, airspaces etc. Among from various flight models and radar data in ALL_FT+ structure, Planned and Actual Data have been extracted with some assumptions. The planned data structure represents the prior determined trajectory that is intended to be flown whereas the actual data is the actual flown trajectory. Since the ALL_FT+ data has a complex structure with many flaws, series of fixing and filtering algorithms have been developed to preclude the wrong computations which may conclude with false deductions.

The three parameters that create the infrastructure of analyses are capacity, traffic and delays. As it was mentioned, capacity values are obtained via DDR capacity data and the traffic and delays are extracted from ALL_FT+ data. The first leg of the

analyses is to construct the network graph of the European Airspace where each node in the graph may represent airports. This graph does not only depict the interactions between airports, but also provide a framework to work on with traffic data. Each branch (edge) of the graph stands for reciprocal traffic flows between the related nodes. The complete connectivity graph is generated by the ALL_FT+ data itself thanks to continuous airspace profiles of each flight . The analysis is conducted bidirectional in each branches of the graph and the time interval of the whole analysis has been chosen as a month to see daily and weekly pattern and their variations based on flight plans of each day. Additionally, the findings that obtained through the traffic flow graphs regarding the magnitude of sectors also have been confirmed via monitoring the number aircrafts per hour in each sector.

As a third parameter of the analyses, delays and their propagation across the network graph have been investigated. Delay is not only a solid merit of quality of service but also a key element which defines resilience in the Air Traffic Network. Individual delays for each phases that all flights experience have been calculated with seeking the differences between planned and actual data's elapsed times for each phase where the possible reasons for these differences may be unexpected events such as bad weather conditions or strikes. Aggregating these individual delays in terms of airspaces yields the delay distributions of airspaces over the European airspace network. Moreover classifying these delays in hourly intervals will depict the propagation of delays, and an example of this approach has been conducted over Frankfurt area.

As a secondary approach, instead of using airspace centered analyses, network of airports have been investigated which shifts the focus to the propagation of delays between airport pairs. The approach introduces the examination of airport delays into analyses, hence the delays of Actual Off-Block Times and Taxi phases have been calculated and status of airports are monitored in terms of arrival/departure traffics and declared capacities. Incorporating the airports into the analyses provided the push-back delays and their evolutions through each flight's path which resulted with a complete track of delays.

The analyses have been conducted in different perspectives and results of each are given with comments. The events, conditions and procedures that drive the air transportation network system have been identified to get deeper understanding and the comparisons of potential modelling techniques for European Air Transportation Network have been presented holistically.

In addition to resilience of the European Air Traffic, emissions and air quality effects of the flights over the Marmara region have been investigated.

In the last decade, air traffic has increased dramatically with a significant increase in emissions. Our goal is to quantify the impact of aircraft emissions on regional air quality, especially in regards to HC, NOx, and CO, and ozone. Here the focus is on Marmara Region, which is the busiest region in Turkey in terms of air traffic.

First, aircraft HC, NOx, and CO emissions are estimated based on the Smoke Number (SN) by using the ‘‘first order’’ method. The Emissions and Dispersion Modeling System (EDMS) is used for gaseous species. HC, NOx, and CO emissions are estimated once based on the characteristic SN and a second time using the mode-specific SN. Further, aircraft emissions are processed in two ways: (1) allocating the emissions at the airport itself, and (2) by accounting for flight paths, mode, and plume rise.

When the more conservative emission estimates are used (i.e, the characteristic SN estimates allocated to the region.

**AVRUPA HAVA TRAFİĞİ VEUZAY-ZAMANSAL GRİD SALINIM MODELLEMEDE VERİ ANALİZİ VE SİMULASYONU UYGALAMALARI**

## ÖZET

Oldukça karmaşık yapısı, kendisine duyulan talebin hızlı artışı ve hızla tükenmekte olan altyapı kapasiteleri ile Havayolu Taşımacılığı Yönetimi yakın geleceğin en zorlu alanlarından biri halini almıştır. Bu alanda oluşan ihtiyacı karşılayabilmek amacı ile çok uluslu araştırmalar gerçekleştirilmekte ve gerek Avrupa'da (SESAR) gerekse Amerika Birleşik Devletlerinde (NextGen) çeşitli toplu girişim hareketlerinde bulunulmaktadır. Bu amaçla, Avrupa Birliği bünyesinde gerçekleştirilen projelerden biri olan Resilience2050.eu mevcut durumdaki hava trafik sisteminde bulunan ve sistemin beklenmedik olaylara direncini ciddi derecede etkileyen eksiklileri belirlemeyi hedef edinmiştir. Nihai amacı güvenlik sınırları çerçevesinde, aksaklıklara ve istenmedik karışıklıklara dayanıklı, geleceğin Hava Trafik Yönetimi konseptlerini geliştirmek olan bu projenin temelinde mevcut sistemin dinamiklerini ve beklenmeyen/istenmeyen olayların sistem içerisinde nasıl bir etki yaratarak devam ettiklerini inceleyen analizler bulunmaktadır. Bahsi geçen analizler bu tezin odak noktasını oluşturmaktadır.

Mevcut sistemin daha kararlı, oluşabilecek beklenmedik etkilere karşı daha dayanıklı ve verimli hale getirilme amacını taşıyan süreçte, Avrupa Hava Trafiği Ağı'nın makro analizi, hava trafik sistemlerinin alt bileşenlerinin ve sistemin davranışlarını oluşturan olayların belirlenmesinde çok büyük bir önem taşımaktadır. Hava taşımacılığı ağındaki havalimanları, havaalanları ve bu alanların segmentleri, bu ağı oluşturan temel elemanlar olarak nitelendirilirken, uçuş planları ve bu uçuşların ağ elemanları arasındaki yoğunluklarının oluşturduğu kalıplar ise trafik akışının karakteristiğini meydana getirmektedir. Bu bağlamda trafik ağı sisteminin dinamiklerini inceleyen makro analizler, bu sistemi daha yüksek doğrulukla temsil edebilecek modellerin oluşturulması için bir adım olmakta ve sistemi daha verimli kılmak amacı ile uygulanabilecek optimizasyon problemlerini gerçek sistem karateristikleri ile buluşturan bir köprü görevi görmektedir. Bu nedenlerden dolayı geleceğin daha güvenilir ve dayanıklı Hava Trafik Yönetimi ve Kontrol alt sistemlerinin oluşturulmasında, makro analizler oldukça temel ve önemli bir yer tutmaktadır.

Bu proje kapsamında yürütülen makro analizlerde ALL_FT+ ve DDR Capacity olmak üzere iki farklı data yapısı kullanılmıştır. Temel olarak ALL_FT+ verileri uçağın rotalarını verirken, DDR Capacity verisi trafik ağının elemanları olan havalimanları ve havaalanlarına ilişkin altyapıların belirlenen kapasitelerini içermektedir. ALL_FT+ verileri günlük olarak dosyalanırken, her uçuş için 7 adet uçuş modeli ve bir adet radar ölçümleri veri alt grubu içermesi planlanmıştır ancak bu 8 adet veri alt grubu her uçuş için mevcut değildir. Ayrıca tüm bu 8 grup için de nokta bazlı ve havaalanı bazlı olmak üzere iki farklı data profili bulunmakta ve bu iki profil aynı veriye iki farklı bakış açısı kazandırmaktadır. Nokta bazlı profilde uçuşun rotası noktalar dizisi olarak ifade edilir iken, havaalanı bazlı profilde uçuşun rotası, içerisinden geçtiği hava alanlarının giriş ve çıkış noktaları olarak ifade edilmektedir. Dahası, tüm bu 8 grubun dışında her uçuş verisinin başına eklenmiş ve yaklaşık 70 veri alanından oluşan bir ortak veri alanı bulunmakta ve bu alan, uçağın park alanından ilk harekete geçmesi planlanan ve geçtiği zamanlar, kalkış ve iniş havalimanları, uçuş numarası, uçuşun iptal olup olmadığı ve uçuşa dair bazı prosedür

mesajları gibi bazı önemli bilgiler içermektedir. DDR capacity verisi ise yaklaşık olarak bir aylık veriler halinde depolanmış ve Avrupa Hava Sahası içerisinde bulunan havalimanlarının saatlik iniş/kalkış kapasitelerini ve kontrol havaalanlarının saatlik uçak barındırma kapasitelerini içermektedir.

ALL_FT+ datasının içerdiği aynı uçuşa ait bu 8 adet veri alt grubu içerisinden tezin temasını oluşturacak analizleri gerçekleştirmek amacı ile 3 adet veri grubu seçilmiştir. Birinci alt veri grubu (FTFM) planlanan uçuş datasını, ikinci alt grup (CTFM) planlanan uçuş datası ile radar verilerinin füzyonundan oluşan uçuş modelini, üçüncü alt grup (CPF-REF) ise radar verilerini içermektedir. Bu bağlamda birinci grup “Planlanan Uçuş Verisi” olarak atanırken mevcut olması durumunda üçüncü grup “Gerçek Uçuş Verisi” olarak kullanılmaktadır. Üçüncü grubun data içerisinde bulunmaması durumunda ise ikinci grup olan CTFM verisi gerçek uçuş verisi olarak kabul edilmektedir. Birçok geniş çaplı veri yapılarında olduğu gibi ALL_FT+ verisi de bazı hatalı veriler içermektedir. Bu hatalı verilerin olduğu gibi kullanılması özellikle gecikme analizlerinde oldukça yanlış sonuçlar vermektedir. Bu nedenle bir takım filtremele ve düzenleme algoritmaları yazılmış ve analizlerden önce veriler bu algoritmalardan geçirilerek yanlış çıkarımlara neden olabilecek sonuçların giderilmesi sağlanmıştır.

Kapasite, trafik ve gecikme parametreleri, bu çalışmada yapılan analizlerin alt yapısını oluşturmaktadır. Bu parametrelerin yalnız veya birbirleri ile birlikte incelenmeleri sonucu sistem davranışlarına dair bir takım sonuçlar elde edilmektedir. Daha önce de bahsedildiği gibi, parametrelerden biri olan kapasite DDR Capacity verisinin işlenmesi ile doğrudan elde edilmektedir. Trafik ve gecikme değerleri ise ALL_FT+ datası üzerinden hesaplanmaktadır. İlk işlem olarak, Avrupa havasahası üzerinde düğüm noktaları havalimanlarını ya da havaalanlarını temsil eden bağlantı grafı oluşturulmuştur. Veri yapısı içerisinde havasahaları çeşitli boyutlarda temsil edilmektedir. Bu bölgelerden en çok kullanılanları boyutlarına göre büyükten küçüğe “Uçuş Bilgi Bölgesi (FIR)”, “Hava Trafik Kontrolü Birim Havaalanı (AUA)” ve “Temel Havaalanı (ES)” olarak sıralanabilir. Bu çoklu temsil, analizin istenilen düzeyde ve detayda gerçekleştirilebilmesine olanak tanımaktadır.

Üçüncü ve son parametre olan gecikme değerleri, hizmet kalitesinin en önemli ölçütlerinden biri olmakla birlikte, hava trafik ağının beklenmeyen ve/veya istenmeyen değişikliklere ne kadar dayanıklı olduğunun en temel göstergesidir. Bu bağlamda gecikmeler ve bu gecikmelerin sistemde nasıl dağıldığına dair analizler yapılmıştır. İlk olarak data içerisinde bulunan tüm uçuşların, uçuşun her fazı için yaşadıkları gecikme değerleri planlanan ve gerçekleşen uçuş dataları arasında yaşanan farklar ile hesaplanmıştır. Tüm uçuşların tüm havaalanlarında yaşadığı gecikmeler hesaplandıktan sonra bu gecikmeler ilgili havaalanı başlığı altında tekrar organize edilerek, havaalanlarının gecikme üretme karakteristiği çıkarılmıştır. Bu işlemin bir günlük veri üzerinde gerçekleştirilmesi üzerine, o gün için Avrupa havasahası ağı üzerinde problemli bölgelerin görüntülenmesi sağlanmıştır. Benzer işlemin saatlik olarak sınıflandırılması bir gün içerisinde gecikme karakteristiklerinin saatlik olarak nasıl değiştiği sonucunu vermektedir. Dolayısı ile hava trafik ağı sisteminde herhangi bir noktada oluşan aşırı gecikme üretimi sorununun sistem içerisinde bir sonraki saatte ne gibi bir etki yarattığı incelenebilmektedir. Bu durum Frankfurt havalimani üzerinde örneklendirilmiş ve sonuçlar ilgili bölümde verilmiştir.

Havaalanı bazlı analizlerin gerçekleştirilmesi sonucu, hava trafiği dinamiğini modelleme amacına hizmet edecek daha farklı yaklaşımların denenmesi yoluna gidilmiştir. Burada amaç sistem içerisinde gerçekleşen olayları en iyi şekilde temsil edebilecek ve olayların yarattığı etkinin yayılımını neden sonuç ilişkisi içerisinde daha net gösterebilecek modelleri oluşturma metodolojisini araştırmaktır. Bu nedenle ikinci bir yaklaşım olarak, havaalanı bazlı analizler kullanmak yerine, havalimanlarının ağından oluşan ve gecikmelerin yayılımını havalimanı ağı üzerinde inceleyen analizler gerçekleştirilmiştir. Burada bağlantı graflarının düğümleri havalimanlarını temsil ederken gecikmeler havalimanı çiftleri arasında incelenmiştir. Bu nedenle, ilk etapta havalimanı analizleri gerçekleştirilmiştir. Uçağın ilk hareket ettiği zaman olarak tanımlanan “Gerçek Hareket Zamanı”nda yaşanan gecikmeler ve uçağın taksi süresince yaşadığı gecikmeler de analiz edilmiş ve bu analizlerin sonucunda elde edilen bulgular havalimanına inen ve havalimanından kalkan uçakların saatlik sayıları ve ilgili havalimanı için ilan edilen kapasite değerleri ile birlikte değerlendirilmiştir. Havalimanlarının durumunun da modelleme süreci içerisine katılması ile hesaplanan yeni gecikme değerleri sistemde gerçekleşen olaylar hakkında daha fazla bilgi verirken uçuşun tüm fazlarında gerçekleşen gecikmelerin ve bu gecikmelerin ne nedenle oluştuğu hakkında da varsayımlarda bulunma fırsatı tanımaktadır. Ayrıca bu sayede gecikmeler uçuşun başlangıcından sonuna kadar tamamen izlenebilmektedir

Bu çalışmada Avrupa Havasahası üzerinde gerçekleşen uçuşların ve yer operasyonlarının oluşturduğu sistemin modellenmesine ön ayak olacak analizler gerçekleştirilmiş ve modelleme için önerilerde bulunulmuştur. Analizler temel olarak iki farklı bakış açısı ile gerçekleştirilmiş ve iki yaklaşımın da avantaj ve dezavantajları belirtilmiştir. Hava Trafik sistemini süren ve etkileyen olaylar analizlerde görülen sonuçlar ile bağdaştırılmış ve bu sonuçların bir bütün olarak değerlendirilmesi ile modelleme yöntemi hakkında önerilerde bulunulmuştur.

Avrupa Hava Trafiği resilience konusu dışında, uçuşların oluşturduğu salınımların getirdiği Marmara bölgesindeki hava kalitesine etkisi de incelenmiştir.

En son on yılda hava trafiği oldukça artmıştır. Bu da salınımların miktarını oldukça arttırmıştır. Bu tezdeki amacımız bölgesel hava kalitesini sayısallaştırımaktır. Özellikle HC, NOx ve COler için. Dolayısıyla Türkiye’nin hava trafiği açısından en meşgul olan bölgesi üzerinde yoğunlaştık.

İlk olarak havaaracından salınan HC, NOx ve CO Duman Sayısı (SN) birinci mertebeden yöntemi ile tahmin edilir. Salınım ve Dağılma Modelleme (EDMS) gaz halindeki maddeler için kullanılır. HC, NOx, and CO salınımları ilk olarak karakteristik duman sayılarıyla ikinci olarak da modele özel katsayıları ile tahmin edilir.

Ayrıca hava aracı emisyonları iki farklı yöntemle daha işlenir: salınımları havalimanına atamak yada uçuş yolları, fazı ve dumanın yüksekliğine gore hesaplanır (yani duman sayıları bölgeye atanır).

# 1. INTRODUCTION

This study is about data analysis and simulation applications on European air traffic and spatiotemporal grid emission modeling using the same EUROCONTROL ALLFT+ and DDR2 data provided by the PRISME in Resilience2050.eu project. Resilience2050.eu is a collaborative project that was funded through the FP7 AAT Call 5, topic AAT.2012.6.2- 4: Identifying new design principles fostering safety, agility, and resilience for ATM (EUROCONTROL, 2014).

## 1.1 Purpose of Thesis

This thesis aims two things. Firstly, creating a nominal model where the delaying factors other than scheduling are eliminated and running simulations to determine the effect of other disturbances separately. Secondly, creating a spatiotemporal emission model generated by the aircrafts, partitioned in grids.

## 1.2 Literature Review

Since this study consists of two applications of the same data there are two literature reviews for each project.

### 1.2.1 Air traffic literature review

In 2010 Global Air Transport deals with 2.4 billion passengers, 43 million tonnes of cargo, 32 million jobs, just 1 accident for every 1.4 million flights, %2 of global carbon emissions and $545 billion in revenue (Bisignani, 2010). With its coverage over the whole world, it provides connections between various regions where each region has peculiar characteristics and procedures. Notwithstanding the differences between regions, any event that concludes with a significant effect in any of these regions may have concrete impact on other. This worldwide transport phenomena with its entities such as aircrafts, airports and airspaces as well as its regional infrastructures and their independencies, is an extremely complex and active system. This current dynamic air traffic system consistently enlarges with the growing demand. In fact, various analyses and forecasts assume that the growth of air traffic will be approximately 5% per year where the financial crisis, epidemics etc. usually show only a temporary impact (AIRBUS).

The ability to recover quickly from any disruption or unexpected event is a crucial and difficult to attain feature in complex sociotechnical systems such as Air Traffic System where large number of interacting human operators and technical systems, functioning over various regions under different organizations and procedures, have to manage air traffic system with certain safety and efficiency even in the case of uncertainty and disturbances. Although procedures and regulations tend to specify working processes in ATM to a considerable extent, the flexibility and system oversight of human operators are essential for efficient and safe operations in normal and more rare conditions (FAA/EUROCONTROL, 2010). Hence, the role of human operators in resilient ATM structure is vital, and construction of more automated and adaptive ATM structure of future is contingent upon good comprehension of the current human-invoked resilience. Even though the roles and responsibilities of humans will change with the advancement of current system towards to more automated ATM, intelligence, perception and flexibility of human operators will be the fundamental source for resilience of ATM system in the foreseeable future.

This study is the next phase of the previous conference paper presented on ICRAT (Kaya, Inalhan, 2014) and the master thesis (Koyuncu, Eren, Inalhan, 2013).

### 1.2.2 Emissions literature review

Emissions inventories have long been noted for being one of the most, if not the most, uncertain aspect of air quality modeling (e.g., Sawyer et al., 2000). This uncertainty inhibits accurate air quality modeling (e.g., Hanna et al., 1998), effective air quality management, and detailed understanding of the mechanisms impacting the formation and fate of particulate matter in the atmosphere. For example, in modeling studies, inaccurate emissions can lead to either poor model performance or, worse, to the introduction of compensatory errors being introduced (NARSTO, 2000). Many air quality management programs, such as trading, control strategy assessment and permitting depend on accurate knowledge of source emissions rates. Understanding the formation and transport of pollutants requires knowing the properties and rates of source emissions.

Emissions from aircraft can be estimated based on the number of landing and takeoff cycles (LTO). Aircraft engines emit pollutants at different rates during the various phases of operation, such as: idling, taxing, takeoff, climbing, cruising, and approach for landing. Different emissions estimates must be employed for commercial air carrier, air taxi, general aviation, and military aircraft. The Emissions and Dispersion

Modeling System (EDMS) Version 4.01 (FAA, 2001) is widely used for estimating emissions from aircraft and ground support equipment (GSE). This model can calculate emissions for volatile organic compounds (VOCs), nitrogen oxides (NOX), carbon monoxide (CO), and sulfur oxides (SOX). Currently, EDMS does to calculate particulate matter (PM) emissions, coarse or fine, for aircraft although it estimates $PM_{2.5}$ emissions for ground support equipment (FAA, 2001). This is expected to change as of Version 4.3.

## 2. DATA SPECIFICATIONS AND PREPARATION

### 2.1 ALL_FT+ Data Fields, Used Fields

In this study ALL_FT+ from PRISME has been used as data source for calculating delays. ALL_FT+ has information about flights over Europe airspace between 1st of March and 30th of November 2011.

The ALL_FT+ data analyzed includes airspace profile for every flight including airspace entity (FIR, ES, ERSA, AUA) entry time, geographical location of the entry point, exit time and geographical location of the exit point. In our previous work (Url-1, Url-2), by using this data, we have shown a data analytic synthesis of the network structure and the design of the delay propagation model. Specifically, for the same type airspace entities, flight reports continue sequentially (e.g. the exit time of a FIR is the entry time of the following FIR, and the exit point of a FIR is the entry point of the following FIR). Hence, it is possible to create continuous flight segments for every flight. The collective hulls of the flight entry and exit points correspond to the airspace geographical coverage

Continuous flight segments enable to create connectivity and traffic flow graphs by determining connected airspaces within the flight trajectory data (ALL_FT+ data). For every type of airspace entities (i.e. NAS, FIR, AUA, ES, ERSA etc.), directed flow graphs have been generated through ALL_FT+ data. Specifically the interconnectivity of the airspaces is inferred from identification of the flight airspace boundary crossings. As such The European AUA (ATC Unit Airspace) and The European ES (Elementary Air Space) connectivity and flow graph as calculated from the ALL_FT+ data.

ALL_FT+ data has 143 fields corresponding to a single flight, some of them are just numeric or logical, and some of them are compound fields containing information about point-to-point or airspace-to-airspace trajectory of a flight. In this study airspace to airspace trajectory is more meaningful since the airspaces used for a flight does not vary significantly and gives a bigger picture about how a flight cruises and arrives en-route.

Raw data is stored in text based symbol separated files. Every line represents a single flight whereas fields corresponding to the flight are separated with other symbols. Having data stored unprocessed also have advantage to represent the data in desired format including different number of data points because of different filters used for each format. The fields used in the simulation can be seen in Table 2.1. The detailed description of the Airspace Profile is depicted in Table 2.2.

**Table 2.1 :** Fields Used in Simulation.

| # | Field Name | Type | Comment |
|---|---|---|---|
| 1 | DepAir | String | Departure Airport |
| 2 | ArrAir | String | Arrival Airport |
| 3 | AOBT | DateTime | Actual Off-Block Time |
| 4 | IOBT | DateTime | Initial Off-Block Time |
| 5 | AirID | ID | Unique ID for each flight |
| 6 | AirOp | String | Airline |
| 7 | AirMode | String | Aircraft Model |
| 8 | FlStat | String | Flight Status |
| 9 | FTFM_AS | AirspaceProfile | Planned Profile |
| 10 | CTFM_AS | AirspaceProfile | Fused Profile |
| 11 | CPF_AS | AirspaceProfile | Actual Profile |

**Table 2.2 :** Airspace Profile Fields.

| # | Field Name | Type | Comment |
|---|---|---|---|
| 1 | ASTyp | String | Airspace Type, should be FIR |
| 2 | EntTime | Timestamp | Entry time to specified airspace |
| 3 | ExtTime | Timestamp | Exit time to specified airspace |
| 4 | ASID | ID | Unique ID of airspace |
| 5 | EntAlt | Integer | Entry flight level in 1/100 feet to AS |
| 6 | ExtAlt | Integer | Exit flight level in 1/100 feet to AS |
| 7 | EntGeo | Geo location | Latitude, longitude of entry to AS |
| 8 | ExtGeo | Geo location | Latitude, longitude of exit from AS |
| 9 | EntDist | Integer | Entry distance to beginning |
| 10 | ExtDist | Integer | Exit distance to beginning |

Whole raw data has been filtered by analyzed route and also type of airspace, i.e. being only Flight Information Region (FIR). Also only entry and exit times, time windows, and airspace ids are used for delay calculations with the help of Table 2.3.

**Table 2.3 :** Delay Calculation Methods.

| **Type** | **Formula** |
|---|---|
| I. *Delay up to push-back phase* | AOBT-EOBT |
| II. *Delay at taxi to take-off* | (AAP[0].EntTime -AOBT) - (PAP[0].EntTime - EOBT) |
| III. *Delay at take-off to TMA exit* | (AAP[0].ExtTime - AAP[0].EntTime) - (PAP[0].ExtTime - PAP[0].EntTime) |
| **Delay generation at sectors** | |
| IV. *Delay at en-route* | (AAP[-1].EntTime - AAP[0].ExtTime) - (PAP[-1].EntTime - PAP[0].ExtTime) |
| **Delay generation at arrival airport** | |
| V. *Delay at TMA entrance to touch-down* | (AAP[-1]. ExtTime - AAP[-1]. EntTime) - (PAP[-1]. ExtTime - PAP[-1]. EntTime) |

In Table 2.3, AAP stands for Actual Airspace Profile, whereas PAP stands for Predicted Airspace Profile. Also -1 indexing means the last trajectory in the list. Departure delay is calculated as sum of three departure delay types for higher level of representation. After calculating these delays, data set has been generated for every day and time step. Summer season is used as the analysis data set. This data set consists of 94 days and for every day there are 48 data points, i.e. approximately 4500 data points.

### 2.2 BigData Management Tool

The BigData Management Tool (BMT, see figure 2.1) converts unstructured text file to managed and efficient database. It allows BigData to be consumed efficiently by applications processing BigData provided. Just including a driver to database allows access to labeled data.

BMT can process any BigData which is in text format separated by symbols and other extensions can be built for other possibilities. In our problem ALLFT+ is a sample data source for BMT, i.e. it can process other types of BigData sources too. In order to parse a new data source, a small Python script should be written for parsing text data into a hierarchical array for each element (in ALLFT+ this element is a flight).

1. BMT Stores data in MongoDB because of the properties of MongoDB below:
    a) Schemaless design: There is no strict scheme or table structure for storing data. Also data is stored in documents in JSON (JavaScript Object Notation) corresponding to object oriented class structure making it easier to be consumed by client applications.
    b) Flexible, distributed: MongoDB can be distributed among cluster to multiple machine almost seamlessly. Only a little administration process is required compared to relational database administration load.
    c) Highly efficient and available: MongoDB performs highly efficient analysis on BigData and has efficient availability options eliminating chance of failure drastically
2. Reduces the size of data stored
    a) Up to 5-6 times (with extensions the size of data can be further reduced up to 60-140 times with some trade-off issues)
    b) Reducing size and distributed structure of MongoDB makes data processing ultra faster depending on the scale of the data
3. Allows selective parsing, i.e. eliminates redundant information processing by selecting only the field required for the analysis prior to creating database entry boosting storage, memory, and computational efficiency.
4. Assigns labels to features in order to efficiently manage data in client applications, i.e. labels are assigned with simple description language specialised for BMT (a JSON structure)
5. Transforms, filters and scales (e.g. normalizes) data while parsing saving a lot of database access time
6. Filtering an entry before whole fields are processed increases parsing speed considerably

7. BMT eliminates client application's parsing overhead at every run since the data is already parsed the analysis can be performed only by accessing created MongoDB database
8. By utilising Python rather than MATLAB the parsing time of a single day of ALLFT+ data is dropped from 5 hours to 14 minutes at the same specs
   a) It can be further reduced by distributing database across shards and utilising supercomputers
9. Python is free and highly supported not only by science but also development community
   a) Python has a very lightweight infrastructure and comes bundled with many of the UNIX distributions (such as OSX)
   b) Python is open-source and allows extensions to developed programs efficiently

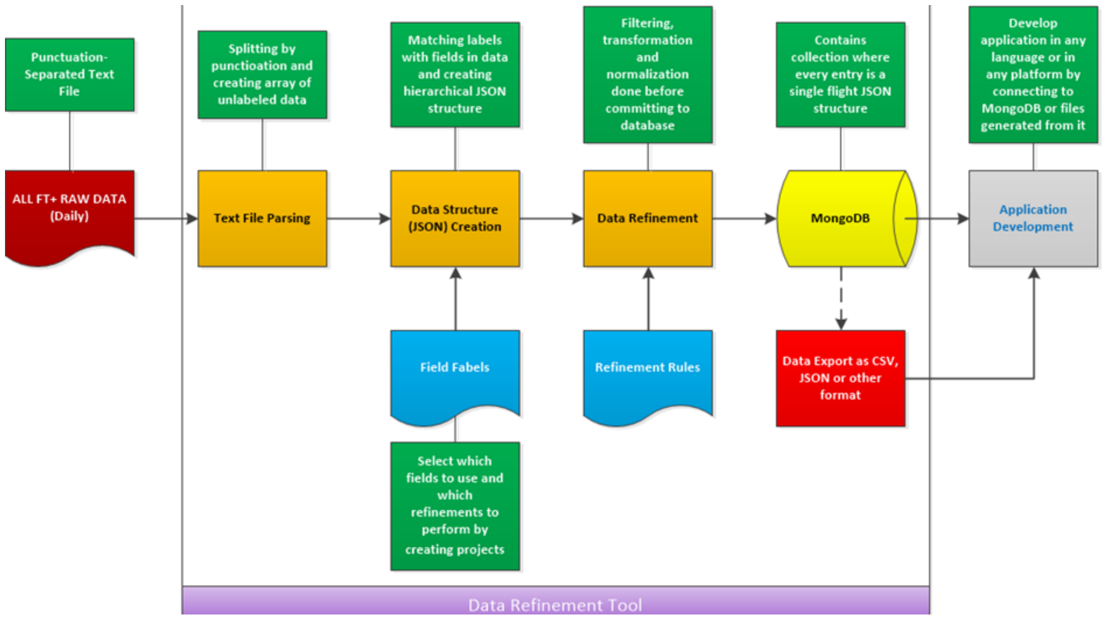


**Figure 2.1 :** BigData Management Tool.

### 2.3 Data Flaws and Fixes

Using the planned and actual profile data as is resulted in anomalies during simulations and delay calculations because of various flaws in the flight data. Previously there were four major flaws detected, being:

1. The Same Consecutive Airspaces Flaw
2. Not Connected Consecutive Airspaces Flaw

3. Airspaces with Zero Elapsed Time Flaw

4. Inconsistent Planned and Actual Profiles Flaw

In addition to those flaws new ones are detected by further investigating the flight data:

1. Flight Entry Duplication Flaw

2. Flights with Wrong Tail Numbers Flaw

3. Flights with Crucial Information Missing Flaw

Each of these problems are explained in details with examples from the data and their proposed solutions for this project is provided as follows.

### 2.3.1 The same consecutive airspaces flaw

In the airspace profile sections of both Planned and Actual data, the same airspaces exist consecutively. An example of this occurrence is displayed below.

```
Flight12491;ESTL;ESTL;UNY434;20110301100000
100500:ESAAFIR:102325:FIR:560507N0131225E:563200N0134100E:1:60:0:58
105325:ESAAFIR:111315:FIR:563200N0134100E:555923N0140603E:60:60:59:124
113315:ESAAFIR:115345:FIR:555923N0140603E:560459N0131204E:60:6:125:182
120345:ESAAFIR:120405:FIR:560459N0131204E:560507N0131225E:4:1:183:184
```

**Figure 2.2 :** An example for same consecutive airspace problem for planned profile.

The airspace profile in FIR level of Flight12491 for planned data is given in figure 2.2. The example clearly shows that, even if there is a geographic coordinate connection between the lines, the exit times of each line are not equal to the entry times of the next line. Despite the fact that aircraft follows a continuous route, there are time gaps between the lines. The actual data of the same flight is also given in figure 2.3:

```
Flight12491;ESTL;ESTL;UNY434;20110301100000
095200:ESAAFIR:100120:FIR:560507N0131225E:563200N0134100E:1:60:0:58
100410:ESAAFIR:102820:FIR:563200N0134100E:555923N0140603E:60:44:59:124
110610:ESAAFIR:112455:FIR:555923N0140603E:560459N0131204E:40:7:125:182
113455:ESAAFIR:113515:FIR:560459N0131204E:560507N0131225E:4:1:183:184
```

**Figure 2.3 :** An example for same consecutive airspace problem for actual profile.

The actual airspace profiles also suffer from the same problem. For this problem, the same sectors are reduced to one and entry and exit times of this single airspace are taken as the entry time of the first line and the exit time of the last time. There are

161 this type of cases for planned data and 1395 cases for the actual data in March 1, 2011.

### 2.3.2 Not connected consecutive airspaces flaw

This problem is similar to the same consecutive airspace problem, except it occurs in different airspaces. An example of this situation is given below.

In figure 2.4 a huge time gap between EDGGFIR and EGTTUIR airspaces this gap also shows itself in geographic entry exit coordinates and flight levels. There are 97 cases like this one in planned profile and 1442 cases in actual profiles for March 1, 2011 data. For this problem, in order to maintain the connectivity of the airspaces the mean of the first airspaces exit and second airspaces entry times is calculated and the mean is assigned to entry and exit times. The same procedure is also utilized for both coordinates and flight levels.

```
Flight5398;ETAD;KHOP;RCH573;20110301052000
053000:EDGGFIR:054455:FIR:495835N0064154E:505026N0064139E:12:220:0:134
062540:EGTTUIR:064320:FIR:532900N0002406E:550007N0013913W:320:320:654:871
064320:EGPXUIR:072725:FIR:550007N0013913W:560000N0100000W:320:320:871:1414
072725:NOTA:075355:FIR:560000N0100000W:570000N0150000W:320:320:1414:1740
075355:EGGXFIR:091325:FIR:570000N0150000W:590000N0300000W:320:320:1740:2649
091325:CZZZFIR:095705:FIR:590000N0300000W:600036N0402347W:320:360:2649:3245
095705:BGGLFIR:104305:FIR:600036N0402347W:594424N0513413W:360:360:3245:3871
104305:CZZZFIR:112050:FIR:594424N0513413W:580130N0595542W:360:360:3871:4386
112050:CCCCFIR:142410:FIR:580130N0595542W:414637N0831920W:360:360:4386:6829
142410:KKKKFIR:152145:FIR:414637N0831920W:364018N0872936W:360:0:6829:7497
```

**Figure 2.4 :** An example for not connected consecutive airspaces for planned profile.

### 2.3.3 Airspaces with zero elapsed time flaw

In some cases, there are airspaces with zero elapsed time (entry and exit times are the same) in the airspace sequences. An example of this situation is illustrated in figure 2.5.

```
Flight7879;LEBL;EBBR;VLG8992;20110301171000
172500:LECBFIR:173420:FIR:411749N0020442E:415439N0013545E:0:245:0:100
173420:LECBUIR:173955:FIR:415439N0013545E:423658N0012901E:245:300:100:179
173955:LFFFUIR:185050:FIR:423658N0012901E:503032N0032621E:300:196:179:1136
185050:EBBUFIR:190240:FIR:503032N0032621E:505405N0042904E:196:2:1136:1232
185050:EBURUIR:185050:FIR:503032N0032621E:503032N0032621E:196:196:1136:1136
```

**Figure 2.5 :** An example of Airspaces with Zero Elapsed Time for planned data.

In the last line of the figure 2.5, EBURUIR airspace has the same entry and exit times. The situation is same for the coordinates and the flight levels. This type of error not only alters the connectivity of airspace sequences but also creates singularities in the delay density calculations (delay density is the delay generated for a unit elapsed time in regarding airspace). This type of cases are filtered and removed from the airspace sequences to handle the singularities and maintain the connectivity. There are 107 such cases in the planned data and 43 in the actual data.

### 2.3.4 Inconsistent planned and actual profiles flaw

The calculations of delay generations involve the processing of each airspace in airspace profiles of both planned and actual data. During these calculations extreme results appeared in delay generations which orientated the focus of the study to further examination of planned and actual data. After an extensive assessment, the planned and actual data separated into five different categories based on their airspaces. These categories for March 1, 2011 are presented with the number of flights they contain as follows. Out of 29661 flights:

1. Airspace sequences of Planned and Actual data are exactly the same (18769 flights)
2. Only the first and the last airspaces of Planned and Actual data are the same (5661 flights)
3. Only the first airspaces of Planned and Actual data are the same (2804 flights)
4. Only the last airspaces of Planned and Actual data are the same (2187 flights)
5. Neither the first nor the last airspaces of Planned and Actual data are the same (240 flights)

The first category is the easiest one to calculate sector delays because planned and actual data are perfectly matched. Since the second category involves changes in the plan and the flight occurred via different path, delay generations of each sector are calculated with some assumptions. At this point, delays for different airspaces (airspaces that belongs to a different path and the airspaces that the separation from the original course begins and ends) calculated holistically and the total delay is shared between those different airspaces with the proportion of their elapsed times over the total elapsed time. This assumption for the delay calculation necessitates the time connectivity of the consecutive airspaces for a correct partition of the total delay.

The first and the second categories can be interpreted as domestic flights over the European Airspace. On the other hand, most of the third category can be classified as the flights start from somewhere in the European airspace and end at another continent. Similarly the most of the flights in the fourth category can be interpreted as the flight coming from another continent to Europe. Since CPF-REF data has no coverage outside the European airspace, the incomplete actual data in these flights make sense. The same logic applies for the fifth category, which can be explained as transit flights.

However the last three categories has also many flights with flawed data. In order to see if these flawed data are peculiar to some region, airspaces of those flights are plotted as it is given in figure 2.6. The figure clearly shows that the problem is related to the data itself rather than any region over European airspace. Utilization of these three categories in delay calculations result with incorrect solutions. Because of that reason they are not included into delay calculation process, but they are processed in traffic calculations since the aircrafts exist in the actual data. Consequently, for the delay calculations around %75 of the total data is utilized whereas the traffic calculations have been carried out with %90 of the data.

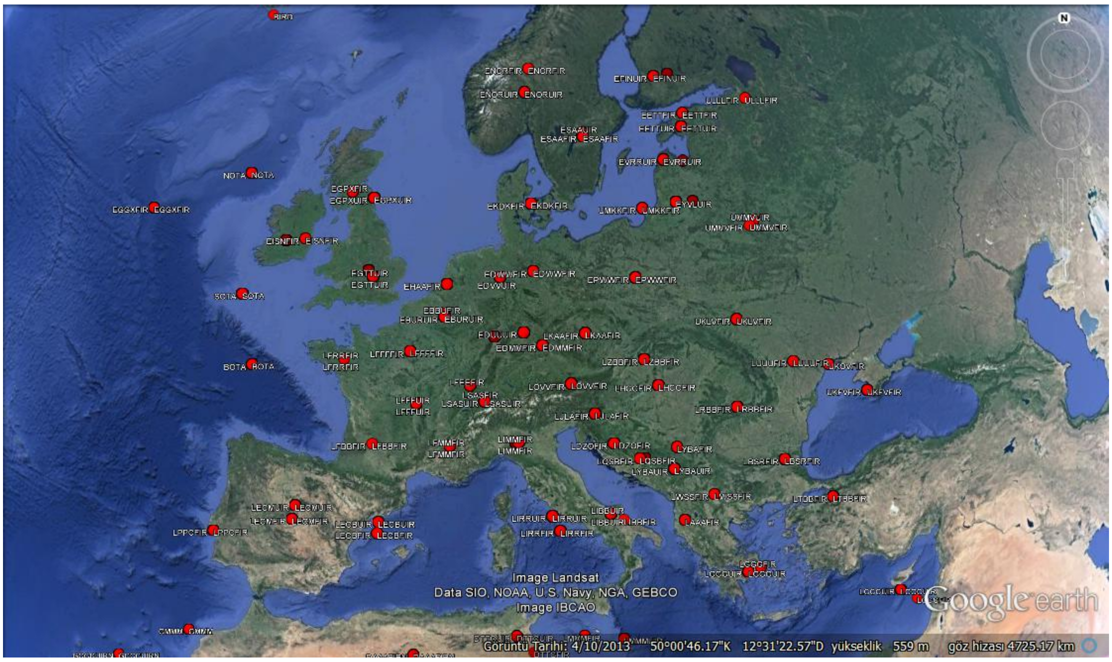


**Figure 2.6 :** The Airspaces in FIR level that belong to the flawed data for Inconsistent Planned and Actual Profiles Problem.

Figure 2.7 demonstrates the category three flawed data. Same colors indicated the common airspaces and the times elapsed in each sector for both planned and actual data is presented just below the airspace name. As it can be seen from the figure, both planned and actual data has same first airspaces, however, after the second airspace separation begins and flight ends in different airspaces.

Also note that, both LIMMFIR and LSASFIR are the FIR regions of European airspace. Therefore, it is a crystal-clear fact that, this category three data is actually a flawed data and using these kind of flights with flawed data will eventually add wrong results to the delay generation data pool.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *PLANNED* | EGTTFIR 965 | EBURUIR 1465 | EDVVUIR 60 | EDUUUIR 2450 | LOVVFIR 295 | LIMMUIR 965 | LIMMFIR 700 | | |
| *ACTUAL* | EGTTFIR 835 | EBURUIR 1350 | EDUUUIR 770 | LFFFUIR 140 | EDUUUIR 535 | LSASUIR 10 | EDUUUIR 45 | LSASUIR 500 | LSASFIR 455 |

**Figure 2.7 :** An example of category three, flawed data.

### 2.3.5 Scripts used for filtering data set

In order to apply the filters for duplication, wrong tail number, and crucial information missing flaws Python and JavaScript scripts are developed.

**duplicate_filter.js:** This filter detectes the flight entries with the same departure, arrival airport, callsign and actual off block time (day/month/year) and removes the redundancies.

**grouped_by_callsign.js:** This script groups the flight entries with the same same departure, arrival airport, callsign and actual off block time. It also provides the count of entries which are in the same group.

**remove_wrong_tail.py:** This filter removes the flights in the same group expect ones with “TE” states being “Terminated” (This code uses grouped_by_callsign.js’ output).

**find_anomalies.js:** This script detects flight entries without FTFM Airspace Profles, callsigns and initial off-block time or flights having “ZZZZ” as departure or arrival and saves it to anomalies collection in the database.

**remove_anomalies.py:** This code removes the flights in the all flights collection stored in anomalies collection in the database.

**get_scheduled.js:** This script filters the flight with flight type “S”, i.e. scheduled flights and stores those flights to scheduled collection in the database.

**get_europe.js:** This script filter the flights with departure, arrival pairs starting with L-L, L-E, E-L, E-E initials and saves to europe collection in the database. This way the flights in Europe has been extracted.

**code_sharing.js:** This script groups the scheduled flights with departure, arrival, registration mark and initial off block time. This group denotes the flights using codesharing. This flights with more than one instances saved to code_sharing collection in the database.

**remove_code_sharing.py:** This code takes copies of codeshared flights from code_sharing collection and removes them from the all flights collection in the database.

**remove_wrong_same_deparrcall.py:** This code groups the flights with the same departure, arrival airport and callsign for each day. The groups having more than one flights are filtered out and checked if the pattern is present during the month. If there is no pattern for these flights, this code removes them. If there is a pattern, the analysis goes further and groups are divided into subgroups (the flights with close inial off-block times in the same day are grouped together) and by adding suffixes such as _1 and _2 to the callsign these flights are split.

# 3. EUROPEAN AIR TRAFFIC FLOW MODEL ANALYSIS

## 3.1 Airport Classification

The busiest airports in 2011 can be observed on Table 1 and Table 2. The order of busiest airports stays nearly constant with minor changes. However the movements per day is higher in summer season compared to winter season. Hence, one can expect higher delays in summer season than winter season under the assumption that the actual capacities are not regulated according to seasons.

Moreover, with seasonality the main delay generators, i.e. major airports, change suggesting there should be separate models for each season. Examining air traffic flow between regions and air traffic flow generation and absorption graphs of busiest airports for each season can further support this hypothesis of needing to use separate models for each season.

**Table 3.1 :** The busiest airports in Europe, November 2011.

| Rank | Airport | Movements/Day |
|---|---|---|
| 1 | Frankfurt | 1296 |
| 2 | Charles de Gaulle | 1288 |
| 3 | Heathrow | 1242 |
| 4 | Schiphol | 1076 |
| 5 | Madrid | 1073 |
| 6 | Munich | 1058 |
| 7 | Ataturk | 855 |
| 8 | Leonardo da Vinci | 788 |
| 9 | Barcelona | 735 |
| 10 | Vienna | 689 |

**Table 3.2 :** The busiest airports in Europe, July 2011.

| Rank | Airport | Movements/Day |
|---|---|---|
| 1 | Charles de Gaulle | 1500 |
| 2 | Frankfurt | 1380 |
| 3 | Heathrow | 1360 |
| 4 | Schiphol | 1306 |
| 5 | Madrid | 1229 |
| 6 | Munich | 1171 |
| 7 | Leonardo da Vinci | 1002 |
| 8 | Ataturk | 944 |
| 9 | Barcelona | 926 |
| 10 | Gatwick | 792 |

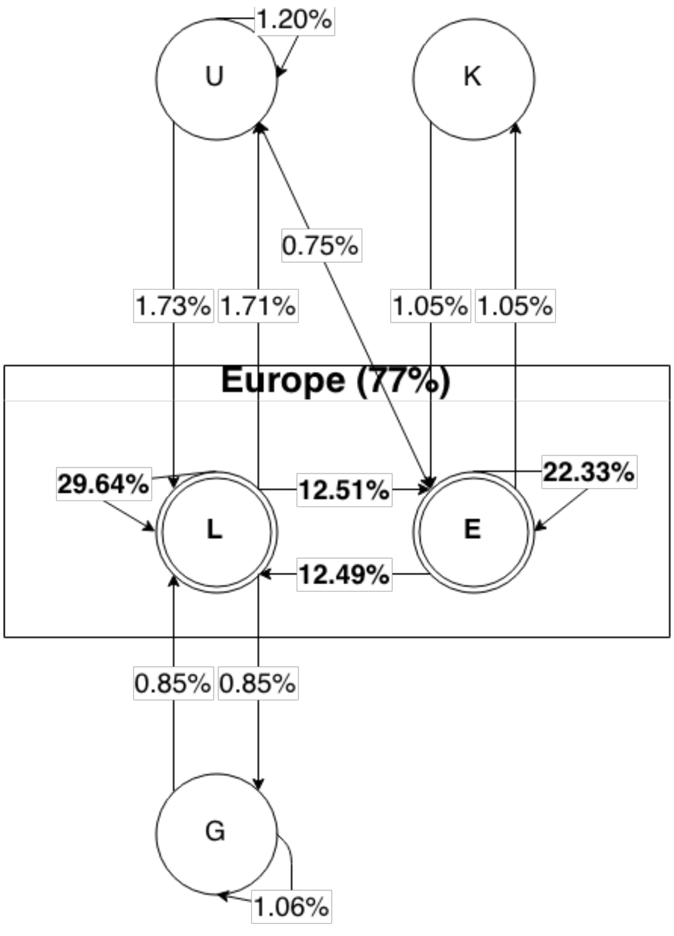


**Figure 3.1 :** Flow distribtuion in July 2011.

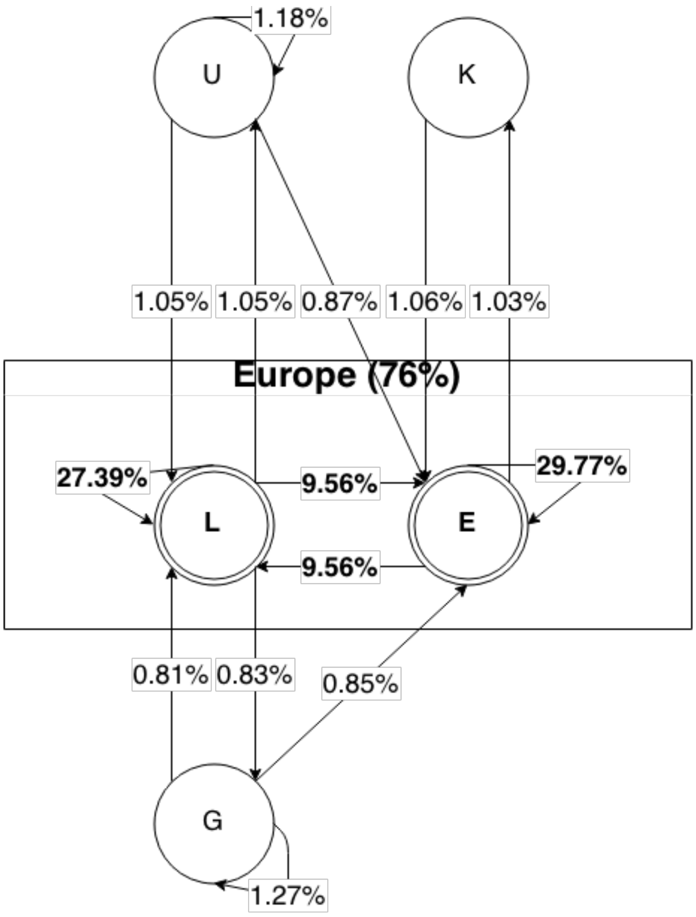


**Figure 3.2 :** Flow distribtuion in November 2011.

In figures 1 and 2 air traffic flows are presented as weighted directed graphs. Weights on the edges represent the percentage of air traffic flowing from a region to another denoted by ICAO region codes with respect to all traffic volume in July and November 2011.

While building the flow distribution graphs only regions with strongest connections have been selected. Connections with less than 0.7% are not considered as significant in order to simplify the model.

It has been observed that air traffic flux varies from one season to another. Since there is not a big variation in total air traffic flow in Europe from season to season, Europe can be assumed as a closed system, i.e. while total flow is constant, rate distribution shifts from one region to another, e.g. air traffic flow between Northern and Southern Europe shifts to inner air traffic flow at Northern Europe in winter season.

Since the ALL_FT+ data consists mainly of European flight data, it is more accurate to focus on Europe region. By implementing a model focused primarily on Europe will represent a stronger model with at least 73% actual flow coverage.

This graph structure also suggests that the traffic sensitivity to delays changes from season to season depending on air traffic flow rates. However, investigating the connections in airport-to-airport level will make presentation of delay sensitivities more accurate.

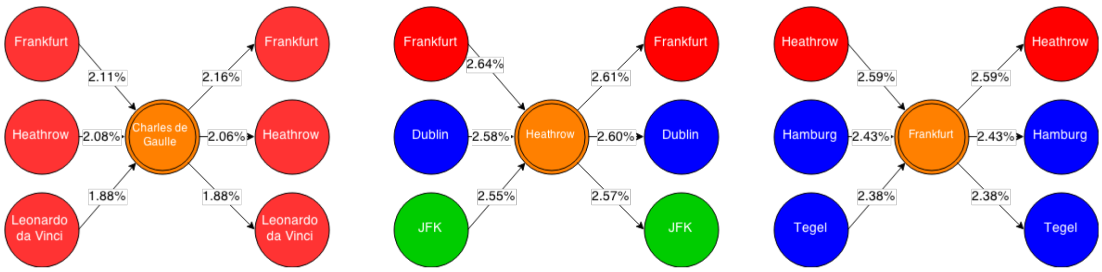


**Figure 3.3 :** Connectivity graphs of major airports, July 2011.

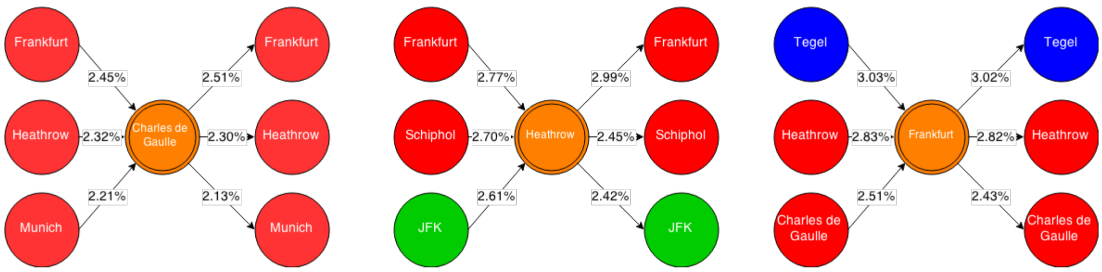


**Figure 3.4 :** Connectivity graphs of major airports, November 2011.

In Figure 3 and 4 connectivity graphs of the airports with most movements are demonstrated. Orange airports are the investigated airports, red airports are the busiest airports in Europe, blue airports are airports in Europe that are not in top 10 busiest airports in Europe and green airports are the airports outside Europe.

The color-coding comes in handy because having a connection with a red airport causes more delays because of high traffic volume and having a connection with blue airport causes rather less delays. Having a connection with green airports does not provide an accurate picture caused by lack of data outside the Europe region.

By only looking at color coding one can predict that the delay sensitivity of Charles de Gaulle is higher than Frankfurt airport. Also the delay sensitivity of Frankfurt increases from July to November because of the shift of airport connections.

Charles de Gaulle, Heathrow, and Frankfurt have different connectivity characteristics, so one can predict that the delay trends will be different from airport to airport and season to season.

Seasonality has an impact on connection sensitivities. The links get stronger in winter than in summer, so the delay propagation rate is expected to be higher in winter.

### 3.1.1 Airport classification according to air traffic flow rate

Both figures 5 and 6 suggest that about **82-83%** of all airports in Europe have air traffic flow rate less than one movement per hour averaged about one month and this rate reduces at higher flow rates.

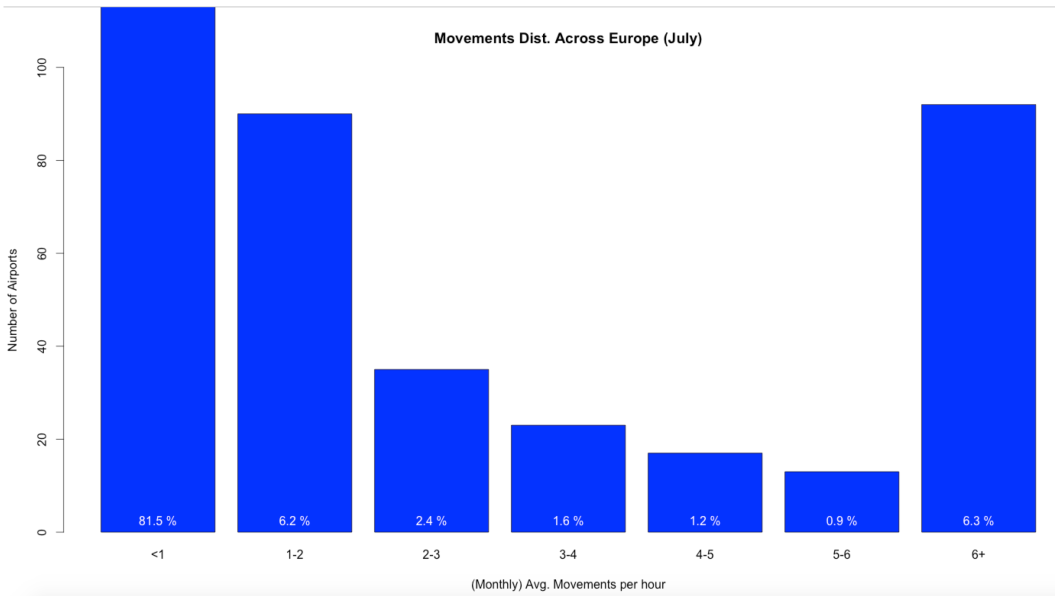


**Figure 3.5 :** Movement distribution across Europe, July 2011.

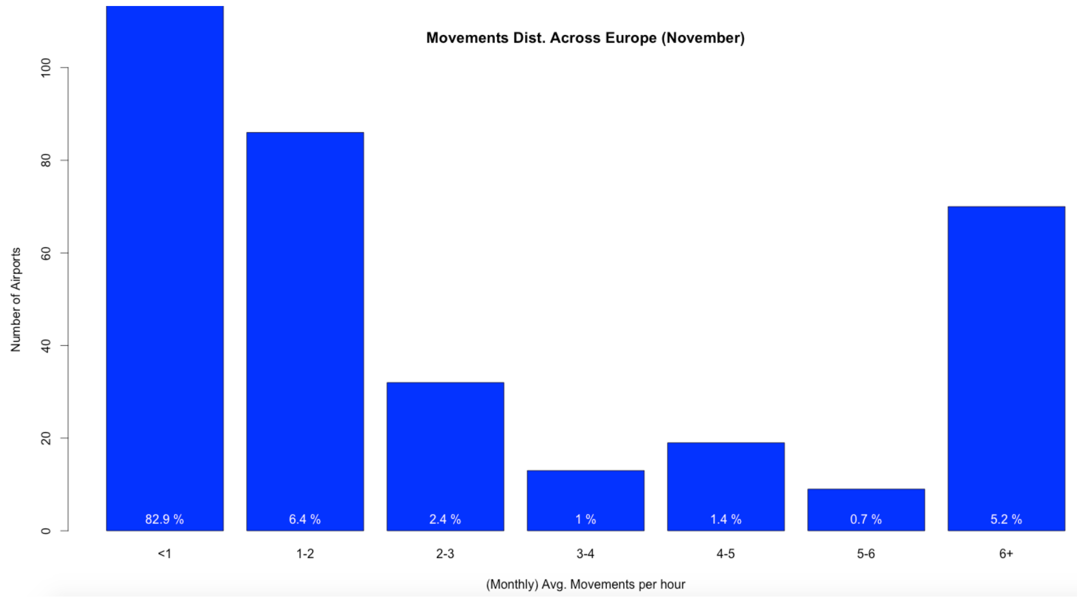


**Figure 3.6 :** Movement distribution across Europe, November 2011.

Since major airports are defined as "airports having one or more movements per 15 minutes" ratio of European major airports is only **7-8%** among all **1458** airports in Europe. However, the massive rate of air traffic flow is generated and by major airports being **72-76%** as seen on Figure 3.7 and 3.8.

Because of this distribution a simplification model has been proposed. Minor airports are grouped by their *IATA* regions and modeled as aggregated airports. Capacity of flights from those aggregated airports is determined from their cumulated demand with the same capacity calculation procedures used for major airports. With this simplification total number of airports has been reduced from **2057** to **304** consisting of **122 European** major airports and **182** aggregated airports (e.g. LFXX, ENXX, EGXX).

In the simulation, non-European airports and aggregated airports are assumed to have infinite capacity, i.e. their capacity is set to a large number.

Also the distribution of departure movements and arrival movements are very similar thus number of *total movements* is sufficient for representing the movement distribution across Europe.

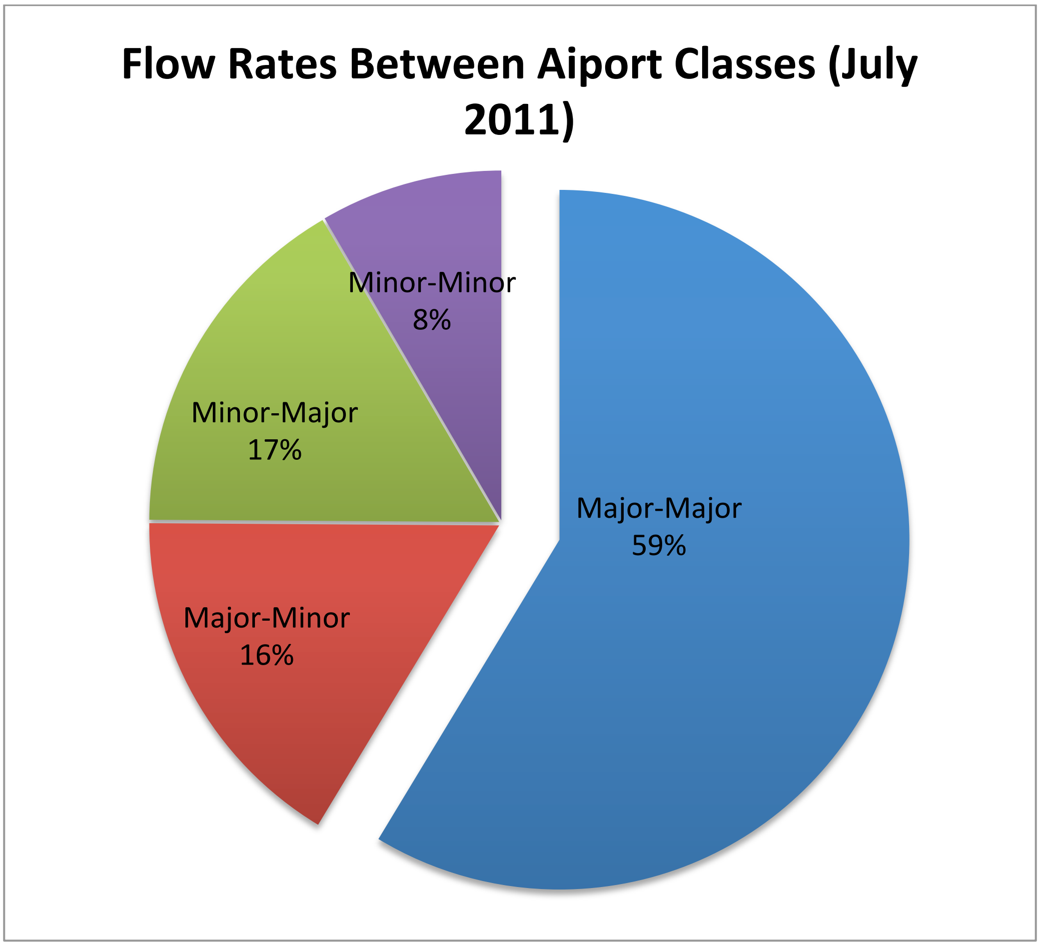


**Figure 3.7 :** Air traffic volume ratios across Europe, July 2011.

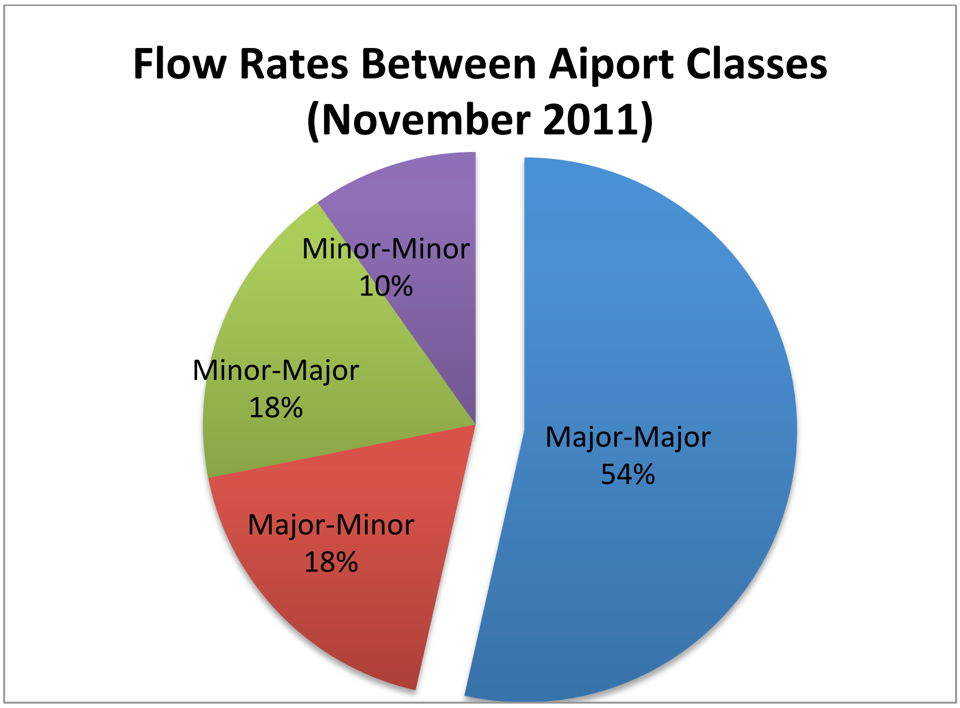


**Figure 3.8 :** Air traffic volume ratios across Europe, November 2011.

### 3.1.2 Airport classification according to flight durations

Figures 3.9 and 3.10 depict that average flight duration in Europe is distributed normally with a left skew. Also most of the flights take between **1 and 2 hours**. This distribution changes slightly with seasonality so it can be modeled as single normal distribution independent from seasonality.

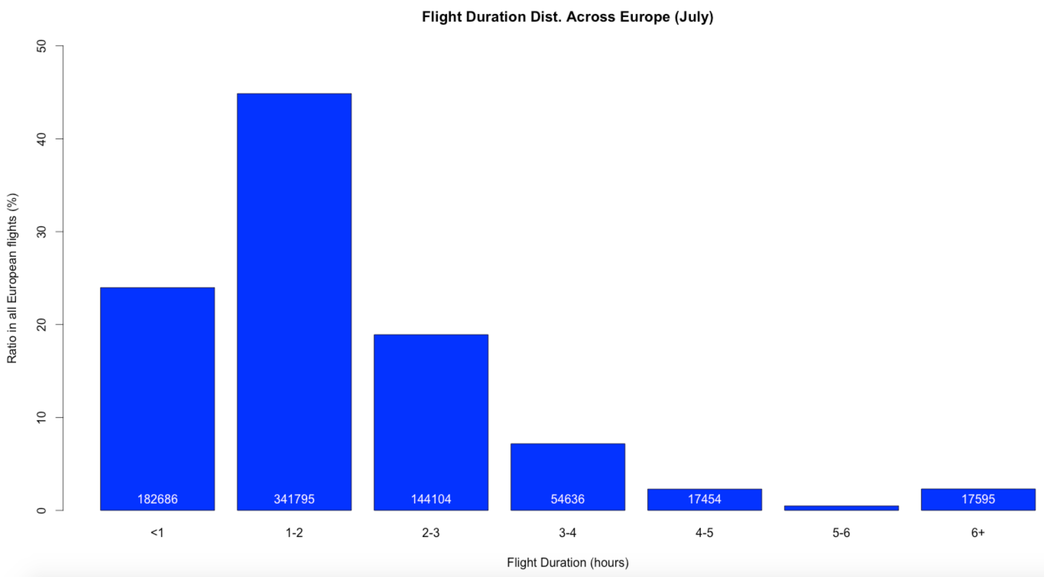


**Figure 3.9 :** Flight Distribution across Europe, July 2011.

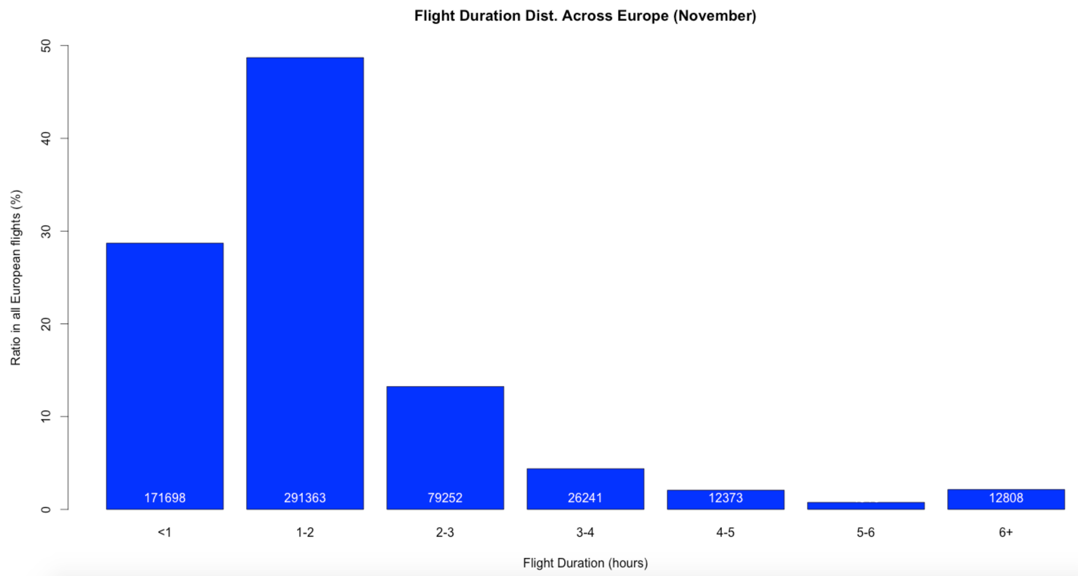


**Figure 3.10 :** Flight Distribution across Europe, November 2011.

Figure 3.11 and 3.12 clearly show that *Marco Polo Airport* is very similar to average Europe distribution in figure 8, while *Heathrow Airport* has a different distribution. *Heathrow* is a major hub, thus most of the flights are long distance flights and consequently *Heathrow*'s delay characteristics are expected to be different from *Marco Polo*'s.

Moreover, the mean flight duration of European flights are **143** minutes in July and **138** minutes in November. Marco Polo has a mean duration of **97** minutes while Heathrow has **243** minutes in November.

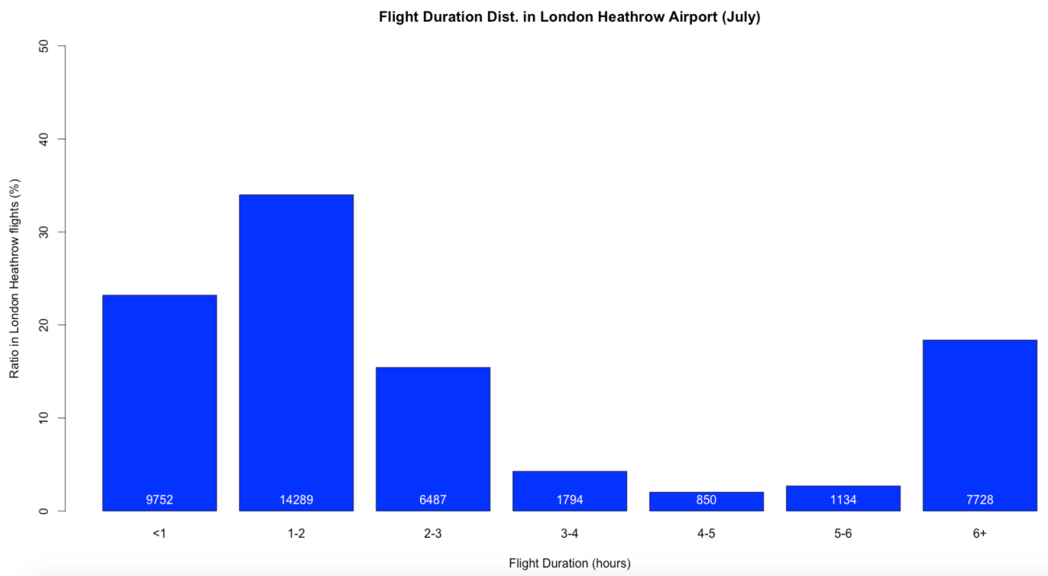


**Figure 3.11 :** Flight Distribution of London Heathrow Airport, July 2011.

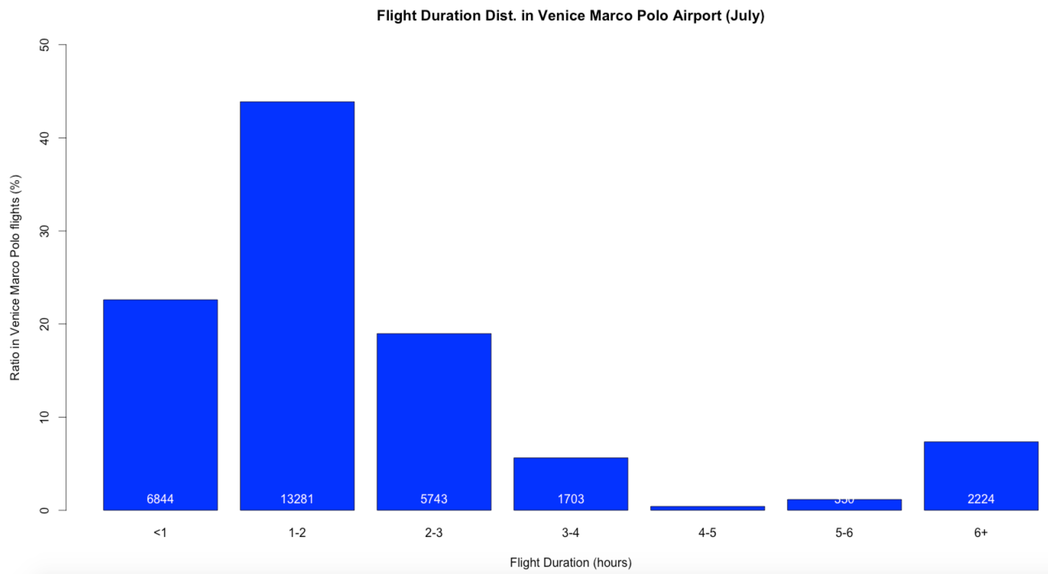


**Figure 3.12 :** Flight Distribution of Marco Polo Airport, July 2011.

### 3.2 Capacity Calculation

Python and JavaScript scripts have been developed for calculating capacities using declared capacites from DDR2 data and the actual demand of a specific airport.

**calculate_demand.py:** This code calculates flight durations by utilizing the initial off-block time and the last point of FTFM airspace profile and also determines departure and arrival time windows. Both departure and arrival demand lists are updated by one for each flight. These lists are saved to database as departure_demand and arrival_demand.

**group_airports_by_regions.py:** This code classifies airports by calculating total demand by adding airport's entry at the departude_demand and arrival_demand collections in the database. If total demand is less than 1 the airport is considered as "minor" airport and the airport is grouped with its region denoted by the first two letters of the ICAO code. By adding XX to end of this two letters, new airport name is generate (e.g. LTAC could become LTXX). After that these demand are grouped together and grouped_dep_demand and grouped_arr_demand collections are created in the database.

**ddr_parse.py:** DDR2 capacity file gets parsed and saved to the database as ddr2 collection

**grouped_ddr2.js:** This script groups the capacity declerations with the "A", "D" categories by their category, entry and exit times and determines a time frame for declerad capacities. This way multiple declarations of the same airport can be captured. This results are saved to ddr2_aggregated collection in the database.

**calculate_capacities.py:** This code is used for calculating ultimate capacities of airports by using grouped_dep_demand, grouped_arr_demand, and ddr2_aggregated collections in the database.

This process is applied for both departure and arrival. For both departure and arrival if an aggregated airport name ends with XX, i.e. if it is a minor airport, the capacity is determined as infinity. Also if an aiport is a major airport, ddr2 data and demand data are compared. If capacity is 999, i.e. infinity, 99% capacity is selected, if it is not 999, then the maximum of declared and 95% capacity is selected as capacity.

Finally, these capacity values are stored in dep_capacities and arr_capacities collections in the database.

What's more, it is observerd that airports can shift their departure and arrival capacities when there is a congesition and being at full capacaity. However, this situation is not considered at the algorithm.

### 3.3 Capacity Normalisation

In Figure 13 arrival movements per 15 minutes (i.e. time window) has been demonstrated. In the figure declared capacity is only **11**, while nominal capacity (above 95%) is **14** per 15 minutes.

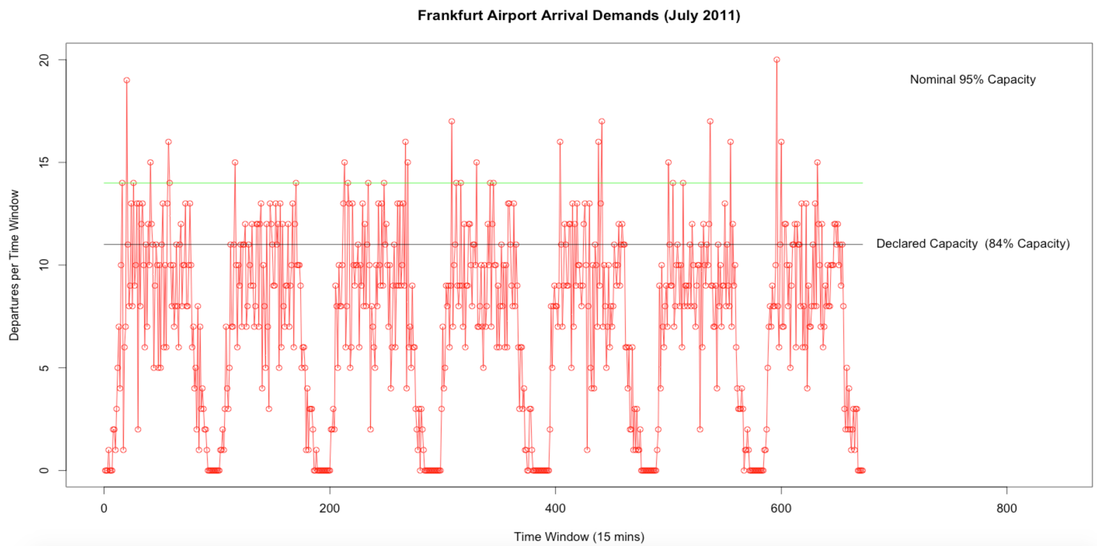


**Figure 3.13 :** Franfurk Airport Arrival Demand Profile, July 2011.

According to *DDR2 capacity declarations* dataset *Frankfurt Airport*'s declared departure capacity is infinity and arrival capacity is **45** per hour, i.e. **11** per 15 minutes. Since infinity does not provide information about the capacity, the capacity of a major airport is set to 99% nominal capacity and capacity of minor airport is set to a large number as depicted before.

In theory the utilization rate ρ should not be more than one that is demand should not exceed capacity. With declared capacity, demand exceeds capacity about **16% (~2.5σ)** of the time not indicating a good assumption for the actual capacity value. Thus, a nominal value has been calculated for reducing capacity exceeding rate to **5% (~3σ)** for being acceptable.

### 3.4 Off-block Time Normalisation

Initial off-block times should be normalised for the sake of nominal simulation. For these purpose Python and JavaScript scripts have been developed.

**route_times.js:** The scheduled flights with the same departure, arrival airports and callsigns are grouped in route_times collection in the database. This way if the same callsign is used for different origin destination pairs then this callsign becomes seperated.

**normalize_flight_times.py:** This code inspects every flight in the route_times collection and list the flights with the same initial off-block time. The miniumum initial off-block time is selected as the initial off-block time for every flight in the group. If there are flights with the same initial off-block time in the group, they are removed. The resulting groups are stored in normalised_times collection in the databsase.

**unwind_normalized_flights.js:** This script does the inverse of grouping to the initial off-block times in normalized_times and stores it as a flat list in the normalised_flight_times collection in the database.

### 3.5 Flight Duration Normalisation

Another entity that should normalised for simulation is flight duration. For the sake of simulation same flight should have same flight duration every time.

**group_flight_duration.js:** This script groups the scheduled flights with the same departure, arrival airport and callsign and calculates the flight duration by subtracting the initial off-block time from the last points in FTFM airspace profile. Later, these durations are stored in the durations collection in the database.

**normalize_flight_duration.py:** This code calculates the flight durations as described before. After that the flight durations with the statistical minum will be selected from the durations collection in the database, i.e. if there is an outlier it should not count as the mimimum flight time.

### 3.6 Aircraft Tracking

Another approach is to track every single aircrafts. This approach makes the simulation closer to actual process because of the connected flights flown with the same aircraft.

There are six Python scripts developed. The sq_main is the main script that acts as the framework of the process. The subprocesses are split into sq_1, sq_2, sq_3, and sq_4 scripts and called from the sq_main script.

Main script is designed for tracking the flights day by day by checking schedule list. For each day intial off-block times are sorted from lower to higher and grouped by their registration marks. After that anomalies and aircrafts with no registration marks re detected and removed.

First, the disconnections in flights that are grouped are detected by checking if the first flights arrival is not equal to second connected flights's departure airport and empty flights are inserted to this gaps.

Sq_1 fills the gaps with the proper flights by checking the historical data. If there is only one gap or two sequential gaps, then this script determines the proper aircrafts with the registration marks and fills them with those.

If there are more than two gaps, Sq_2 is exectued. 12-13 gaps can be observed in a connected flight. There is also the case that two different groups might use the same registration mark leading to be grouped as a single group causing ambiguties.

1. ENCN:EKCH:5:5
2. ENTO:EKCH:5:10
3. EKCH:ENCN:7:20,
4. EKCH:ENTO:8:20
5. ENCN:EKCH:8:35

The flights above (departure airport, arrival airport, departure time) have the same registration mark, however it can be observed that 1,3, and 5 are one group and 2 and 4 are another group. Another case is there are 17 flights in a single group but actually it is wrongly addition of a 8 and a 9 group. When this problem is solved by Sq_2 script, some of the disconnetions disappear.

Further, another problem there are flight having only one disconnection, however it is not possible to place other flights to this gap because the initial off-block times of flight before and after the gap are not consistent. This problem can be solved by removing the flights before and after the gap.

Further, sq_4 groups the flights with no registration marks and places them to the gaps and detects the gaps that can be fixed and rest of the aircrafts are marked with null_k (k=1,2,3..).

Finally, sq_fixed_null_id.py places the flights with null_k registartion marks to corresponding gaps and irrelevant groups are registered with corrected marks. At the end, Sq_fixed_register_at_sc.py fixes the scheduled list.

### 3.7 Simulation Algorithm

In this section pseudo-code of simulation algorithm will be presented. The algorithm consists of various components. It has a two-way recursive algorithm. Flow chart of the whole algorithm can be seen on Figure 3.14.

#### 3.7.1 Simulate

The simulation starts with preprocessing of airports, demands, queues, and slots. The step of simulation is 15 minutes and simulation takes place between times $T_{start}$ and $T_{end}$. At every time step, an airport A is picked from all airports and its departure queue Q and its slot $S_{dep}$ are selected for management.

Up until there is neither flight in queue Q nor there is no place in slot left all flights at time t and in departing from airport A are managed. At every management step a flight has been picked (Dequeue) from the beginning of the queue Q. After that the arrival slot $S_{arr}$ and duration D of the flight has been picked from the list $S_{arr,all}$ and $D_{all}$ between A and B, respectively.. Consequently *SearchSlotDeparture* recursive function starts with the current time window and flight. Recursive function runs until it assigns the flight F to departure and arrival slots or fails to assign if t reaches to $T_{end}$.

After assignment is complete with or without success, the same process repeats for every flight in the queue or until slots are full and for all airports for the particular time window t. If there are still flights left in the queue these flights are assigned to the next time window of the queue of the next time window of the airport. When all assignments are complete for the time window t, t is set to next time window by *EnqueueAll* function.

**Table 3.3 :** The Simulate Algorithm.

| **Simulate($T_{start}$, $T_{end}$, $A_{all}$')** |
|---|
| Preprocess() |
| $t_{window}$ <- 15 mins, t <- $T_{start}$ |
| **while** t <= $T_{end}$ **do** |
|     **foreach** A **in** $A_{all}$ **do** |
|         Q <- $Q_{dep}$(A, t), $S_{dep}$ <- $S_{dep,all}$(A), |
|         **while** Length(Q) > 0 **or** Length($S_{dep}$(t)) > 0 **do** |
|             F <- Dequeue(Q), B <- Destination(F), $S_{arr}$ <- $S_{arr,all}$(B), D <- $D_{all}$(A, B) |
|             *SearchSlotDeparture*(IOBT(F)) |
|         **if** Length(Q) > 0 **then** |
|             EnqueueAll($Q_{dep}$(A, t + $t_{window}$), Q) |
|                 t <- t + $t_{window}$ |

### 3.7.2 SearchSlotDeparture and SearchSlotArrival

First of all, calculate $t_{tw}$, start time of the scheduled time $t_{sched}$'s time window. Start searching for a free slot starting from scheduled time in departure slot and continue until a slot is found. Then, if the time $t_{OBT}$ is found, start searching for arrival slot for flight departing at time $t_{OBT}$.

**Table 3.4 :** SearchSlotDeparture Algorithm.

| **SearchSlotDeparture($t_{sched}$)** |
|---|
| $t_{tw}$ <- GetTimeWindowTime($t_{sched}$) |
| $t_{OBT}$ <- *SearchSlot*($S_{dep}$, $t_{tw}$, $t_{sched}$) |
| **if** Found($t_{OBT}$) **then** |
|     *SearchSlotArrival*($t_{OBT}$) |

Start with calculating time of arrival for given off-block time by adding the duration. Similarly, calculate the start of the time window and search for a slot in arrival. If slot is found, search for slot in departure slots by calculating the new departure time $t_{OBT}'$.

If there is a slot in exactly $t_{TOA}$, then $t_{OBT}'$ will be equal to $t_{OBT}$. If there is a slot free in the time window then assign the flight to the first available slot in the time window starting from the $t_{OBT}'$. Similarly, if there is a slot in exactly $t_{TOA}$, then $t_{OBT}''$ will be also equal to $t_{OBT}$.

If there is no slot left in the time window, continue searching by returning to the *SearchSlotDeparture* function with $t_{OBT}'$.

### 3.7.3 SearchSlot and SearchSlotInTimeWindow

First, search for a slot in time window. If there is a place return it. Otherwise search slot in next time window starting from the beginning of the time window. If $t_{tw}'$ reach $T_{end}$ return *NotFound* value returning **false** for *Found* function.

**Table 3.5 :** SearchSlot Algorithm.

| **SearchSlot(S, $t_{tw}$, $t_{slot}$)** |
|---|
| $t_{slot}'$ <- *SearchSlotInTimeWindow*(S, $t_{tw}$, $t_{slot}$) |
| **if** Found($t_{slot}'$) **then** |
|   **return** $t_{slot}'$ |
| $t_{tw}'$ <- $t_{tw}$ +$t_{window}$ |
| **if** $t_{tw}'$ <= $T_{end}$ **then** |
|   **return** SearchSlot(S, $t_{tw}'$, $t_{tw}'$) |
| **return** *NotFound* |

Segment time window into C equal length slots. Check if the slot is free at the specific time $t_{slot}$. If it is free return it, otherwise check from start of the time window until $t_{slot}$ and if there are still no slot available search from next time of $t_{slot}$ until $t_{window}$. If there are no slots available in the time window return *NotFound*.

**Table 3.6 :** SearchSlotInTimeWindow Algorithm.

| **SearchSlotInTimeWindow(S, $t_{tw}$, $t_{slot}$)** |
|---|
| C <- Length(S($t_{tw}$)) |
| $t_{step}$ <- $t_{window}$ / C |

```
t_slot' <- Round((t_slot - t_tw) / t_step)
if Available(S(t_tw)(t_slot')) then
        return t_slot'
t_slot' <- 0
while t_slot' < t_slot do
        if Available(S(t_tw)(t_slot')) then
                return t_slot'
        t_slot' <- t_slot' + t_step
t_slot' <- t_slot + t_step
while t_slot' < t_window do
        if Available(S(t_tw)(t_slot')) then
                return t_slot'
return NotFound
```

### 3.7.4 Preprocess function

First of all scheduled flights are picked from the data by filtering the type of flight. If the flight type is scheduled the flight is included in simulation.

Minor airports are grouped with *Group_Airports_by_Region* and merged with major airports. After grouping airports departure and arrival capacities are calculated. Since the same flights can have different flight times the minimum duration is picked as nominal time for simulation.

Since there are special events in the data the schedule data is not completely normalized. Because of that, days closer to normal picked for every day in a week from a month and a normalized demands week has been created.

After that departure queues are populated with normalized flights according to normalized scheduled off block times.

Finally, slots for every airport and every time window are allocated and set free for departure and arrival.

**Table 3.7 :** Preprocess Algorithm.

| **Preprocess()** |
|---|
| $F_{all}$ <- *Pick_Scheduled_Flights()* |
| $A_{all}$ <- *Group_Airports_by_Region*() |
| ($C_{dep,all}$, $C_{arr,all}$) <- *Calculate_Nominal_Capacities*() |
| $D_{all}$ <- *Normalise_Flight_Durations*() |
| $F_{grouped}$ <- *Normalise_Flight_Times()* |
| $Q_{dep}$ <- *Populate_Departure_Queues*() |
| ($S_{dep.all}$, $S_{arr,all}$) <- *Allocate_Slots_Free*() |

### 3.7.4.1 Group_Airport_by_Region and Calculate_Nominal_Capacities

Major airports are kept as is, while minor airports (movements per time window < 1) are grouped as aggregated airports at their region R.

**Table 3.8 :** Group_Airports_by_Region Algorithm.

```
Group_Airports_by_Region()
foreach A in A_all' do
        if MovementsPerHour(A) >= 4 then
                Append(A_all, A)
        else
                R <- Region(A)
                Append(A_reg(R), A)
AppendAll(A_all, A_reg)
return A_all
```

95% confidence value of movements in airport for both departure and arrival has been calculated as

$$C = m + 1.96\Sigma \tag{3.1}$$

If the statistical upper level is less then actual (sample) 95% capacity, capacity is increased by one in each iteration, until the actual 95% level is met. Usually the stastical and sample upper levels are equal with some exceptions.

**Table 3.9 :** Calculate_Nominal_Capacities Algorithm.

```
Calculate_Nominal_Capacities()

foreach A in A_all do
        f <- Count(Movements(A)), m <- mean(f), Σ <- sd(f)
        C <- m + 1.96 Σ
        while Count(Movements (A) < C) / Count(f) < 0.95 do
                C <- C + 1
        (C_dep,C_arr) <- C
        (C_dep,all, C_arr,all) <- Append((C_dep,C_arr))
        return (C_dep,all, C_arr,all)
```

### 3.7.4.2 AssignFlightTimes, GetWindowTime, Normalise_Flight_Durations, Normalise_Flight_Times

If slot is found successfully, $t_{TOA}$ and $t_{OBT}$ are assigned as Actual Take-off Time and Actual Off-Block Time of F respectively. After that flight F is assigned to departure and arrival slots at these times.

**Table 3.10 :** AssignFlightTimes Algorithm.

```
AssignFlightTimes(t_TOA, t_OBT)

ATOA(F) <- t_TOA
AOBT(F) <- t_OBT
S_dep(t_OBT) <- F
S_arr(t_TOA) <- F
```

A time is rounded up to be a lower factor of $t_{window}$.

**Table 3.11 :** GetTimeWindowTime Algorithm.

```
GetTimeWindowTime(t)

return (t // t_window) * t_window
```

For each source-destination-callsign triplet collect the durations by subtracting scheduled off block time from scheduled time of arrival and set the duration of that triplet as the statistical minimum for the set collected.

Statistical minimum is the least element in the sample, which is higher than the lowest end of the normal curve, i.e. this element is the smallest element in a set, which is not an outlier.

**Table 3.12 :** Normalise_Flight_Durations Algorithm.

| **Normalise_Flight_Durations()** |
|---|
| **for all** F: |
| A <- Source(F) |
| B <- Destination(F) |
| CS <- Callsign(F) |
| D <- STOA(F)-SOBT(F) |
| Append($D_{A,B,CS}$, D) |
| **for all** $D_{A,B,CS}$ <- statistical_min($D_{A,B}$) |
| **return** $D_{A,B,CS}$ |

Minimum of each source-destination-callsign triplet's IOBT is set as all of the flights' IOBT.

### 3.7.5 Variables, symbols, and function definitions

**Table 3.13 :** The busiest airports in Europe, November 2011.

| Variable | Description |
|---|---|
| $t_{window}$ | Time Window Width (e.g. 15 minutes) |
| $T_{start}$, $T_{end}$ | Start and End Times |
| $A_{all}'$, $A_{all}$ | All Airports, All Normalized Airports |
| $A_{reg}$ | Regional Airports |
| $C_{dep,all}$, $C_{arr,all}$ | All Departure/Arrival Capacities |
| $Q_{dep}$ | Departure Queue |
| $S_{dep.all}$, $S_{arr,all}$ | All Departure/Arrival Slots |
| f | Departure/Arrival Frequency |
| m | Departure/Arrival Mean Frequency |
| Σ | Departure/Arrival Standard Deviation |
| $D_{all}$ | All Flight Durations Between OD Pairs |
| $t_{step}$ | Time Step |
| $t_{slot}$ | Calculated Slot Time |
| $t_{OBT}$ | Off Block Time |
| $t_{TOA}$ | Time of Arrival |

**Table 3.14 :** The busiest airports in Europe, November 2011.

| Symbols | Description |
|---|---|

| | |
|---|---|
| a <- 5 | Assignment Operator, e.g. 5 is assigned to a |
| 6 // 4 | // Integer Division Operator, e.g. 6 // 4 = 1 |
| *NotFound* | There is no free space left in time window |
| $D_{all}$(A, B) | Duration of flight between airports A and B |
| Found() | Return **False** if *NotFound* otherwise **True** |
| Enqueue(Q, A) | Insert a to the end of the queue Q |
| A <- Dequeue(Q) | Pick the first element from the top of the queue Q |
| EnqueueAll(Q, $A_{list}$) | Insert all elements A in $A_{list}$ to the Q in the same order |
| R <- Region(A) | Get region code from the airport, e.g. if A is LTBA then R is LTXX |
| IOBT(F) | Initial (Scheduled) Off Block Time of flight F |
| AOBT(F) | Actual Off Block Time of flight F |
| ATOA(F) | Actual Time of Arrival of flight F |
| Length(L) | Number of elements in a list L |
| Append(L, A) | Insert A to the end of the list L |
| AppendAll(L, $A_{list}$) | Insert all A in $A_{list}$ to the end of the list L |
| Available(S($t_{tw}$)($t_{slot}$)) | Checks if slot at index $t_{slot}$ time window $t_{tw}$ is empty |
| {} | Empty Dictionary (i.e. a <- {}, a[3] <- 4 => a = [3: 4]) |

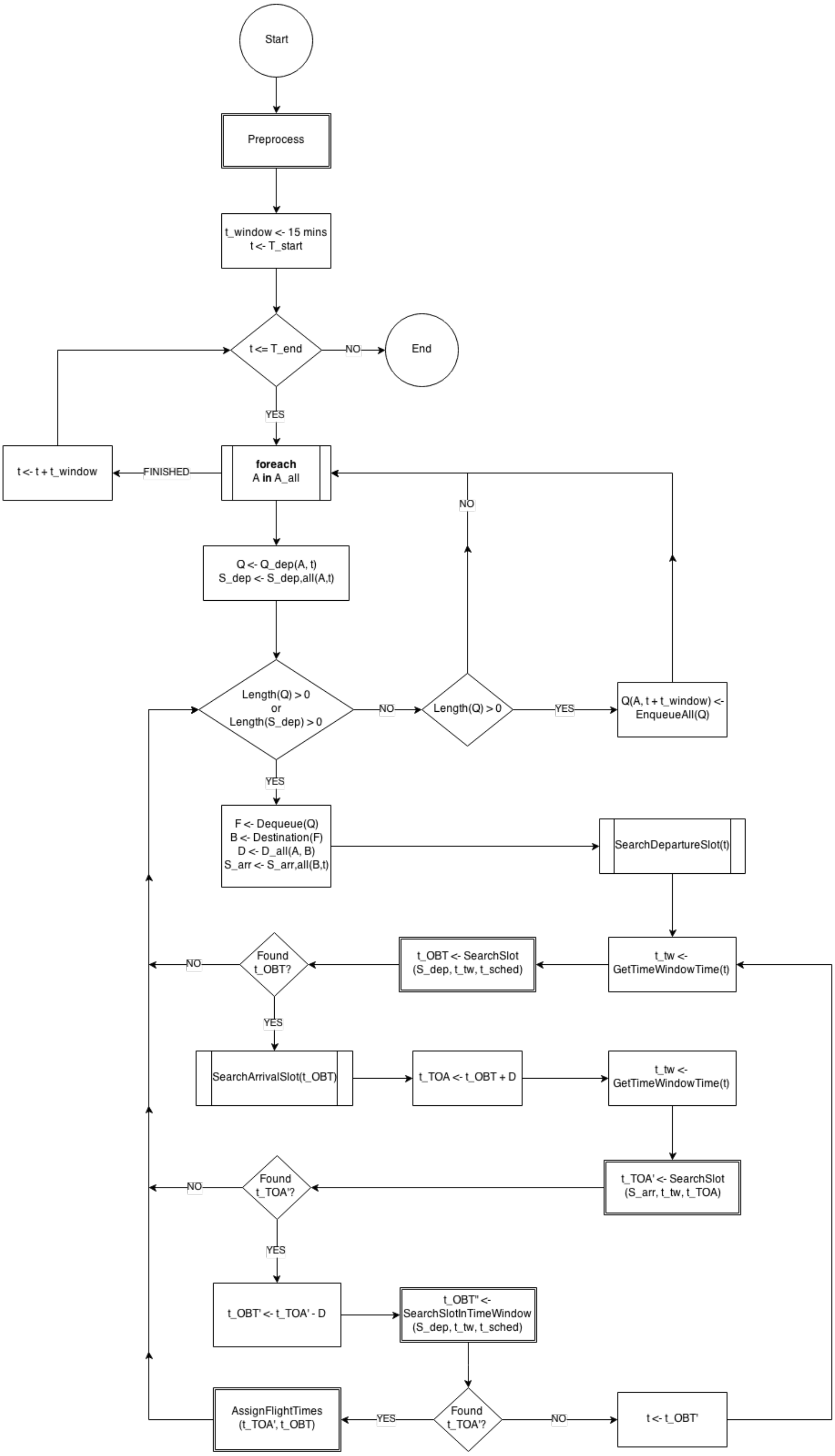


**Figure 3.14 :** Simulation Algortithm Flow Chart.

## 4. EMISSION DISTRIBUTION AS 4D GRIDS OVER MARMARA REGION

A detailed modeling approach was used to quantify the impact of aircrafts on air quality from the whole Mediterranean area, which is located in Turkey (Fig. 4.1). First, a detailed inventory was developed for aircraft and other emissions. Then, air quality simulations were performed to relate these emissions to regional air quality around this region using the Community Multi-scale Air Quality Model (CMAQ) (Byun and Ching, 1999). Whole 2011 was selected as the focus episode because of the data provided by the EUROCONTROL. This episode has a number of days and grids characteristics of high air pollution levels.

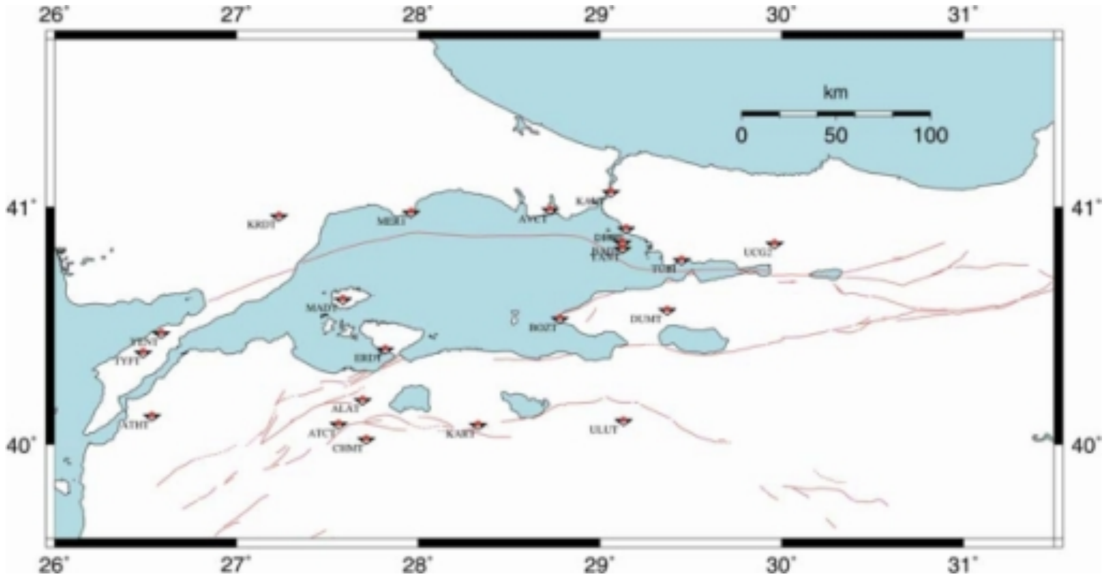


**Figure 4.1 :** Marmara Region Map.

For this purpose, historical flight data has been used.

### 4.1 Impact of Aircraft Emissions on Local Air Quality.

The effect of airports on local air quality is of growing concern. In many great cities of the world, emissions from automobiles, power plants, refineries, and other major sources are being steadily reduced. As a result, emissions from aircraft operations, which have stayed the same or increased, are a growing contributor to air pollution, especially in the immediate vicinity of airports. Assessment of the potential impact of airports on the health of nearby residents requires estimates of the contribution of airport operations to pollutant concentrations. This can be done by either traditional source oriented air quality modeling or by receptor modeling methods, both of which face significant challenges. Applying standard air quality models to aircraft operations is difficult for several reasons. The emission rates of pollutants from aircraft turbine and piston engines are not well known, and the emissions vary greatly during takeoff, landing, and taxiing, being a maximum during taxi and takeoff (Popp et al., 1999). During takeoff emissions are not a point or line source, but follow a curved path of varying height that is difficult to model realistically. Furthermore, many airports are on the coast with complex winds from lake or sea breezes. Receptor modeling is an alternative to source-oriented models. It uses chemical composition of the emissions to distinguish between sources. In the case of airports, this task is usually made difficult by the presence of a high volume of diesel vehicle emissions delivering goods and people to and from the airport. Jet fuel is very similar to diesel fuel (Spicer et al., 1992), making it difficult to distinguish aircraft from diesel emissions using ordinary receptor modeling methods. The variation of concentrations of pollutants with wind direction and speed is potentially a way of separating out the mix of sources around a large airport. Henry et al. (2002) have shown that nonparametric regression of pollutant concentrations on wind direction is an accurate way to determine the direction of nearby sources. The present paper extends this method to nonparametric regression of hourly pollutant concentrations on two variables, wind direction and wind speed. Nonparametric regression is used to estimate the average concentration of a pollutant such as sulfur dioxide ($SO_2$) as a function of wind direction and speed. As will be shown below, wind speed is useful in distinguishing ground level emissions from elevated emissions like aircraft. In this way, the contribution of airport operations can be distinguished from vehicular traffic and other sources.

**4.2 Data Preparation for Emission**

As part of the Fall Line Air Quality Study (FAQS), an emissions inventory was prepared for each source category in Georgia (Unal et al., 2003). In this work VOC, $PM_{2.5}$ emissions for commercial aircraft were calculated using EDMS Version 4.01.

The literature is reviewed to improve the estimates of fine NOx, CO, and HC emissions from aircraft. Recently, the Federal Aviation Administration (FAA) has developed a first- order approximation (FOA) method (Wayson et al., 2003) for estimating $PM_{2.5}$ emissions from aircraft. In this method a relationship was developed that relates $PM_{2.5}$ emissions to Smoke Number (SN) and fuel flow rate (FF) as follows:

$$EI = 0.6\times(SN)^{1.8}\times(FF) \tag{4.1}$$

Where EI is the emission index in mg of $PM_{2.5}$ emitted per second, SN the Smoke Number, and FF the fuel flow rate in kg s-1.

SN is a dimensionless number that identifies the smoke level and is determined by means of the loss of reflectance of a filter used to trap smoke particles from a prescribed mass of exhaust per unit area of filter (Wayson et al., 2003). The data leading to the FOA formula were collected by probes placed at fixed distances behind the engines. It is possible that emissions that are in the vapor phase at the probe orifice may condense into particles farther away. If this is the case, the amount of particles measured by the probe would be less than the total amount of particles caused by the aircraft engine. Therefore FOA may lead to an under- estimation of HC, NOx, and CO emissions from aircraft. An important part of applying Eq. (1) is to find the correct SN and FF for individual engines in each operational mode (i.e., idling, takeoff, climb-out, and approach). Here the International Civil Aviation Organization (ICAO) database was utilized to determine the SN and FF for each aircraft and engine type as available. In the ICAO database different statistics, such as average, standard deviation and characteristic value are provided for SN. Characteristic value for SN is the mean of the values of all the engines tested and corrected to the reference standard engine and reference ambient conditions divided by the coefficient corresponding to the number of engines tested as explained by the ICAO manual (Wayson et al., 2003). We have estimated total PM emissions at Marmara Region using two different methods. In the first method we utilized only the characteristic value SN. For aircraft, which do not have the corresponding characteristic value in the ICAO database, we utilized a database average value. HC, NOx, and CO emissions calculated with this method.

In the second method we utilized the mode specific SN values. For most aircraft SN is only measured for the takeoff mode. We developed a statistical relation, through linear regression, between the takeoff mode and other modes using the data for the engines tested in other modes, and utilized this relation to estimate the SN for modes other than takeoff.

For aircraft, which do not have the corresponding takeoff SN in the ICAO database, we utilized the database average value. It is estimated that the highest contributor to total emissions is the climb-out mode with 65 percent. Takeoff and idling modes come after climb-out with 17 and 12 percent, respectively. The approach mode makes about 6 percent of the total emissions.

It should be noted that Eq. (4.1) provides aircraft emissions per LTO.

For our assessment of airport-related air quality impacts, we prepared four different sets of emissions inventories; these are: (i) without aircraft emissions; (ii) with aircraft emissions estimated using the characteristic value method; (iii) with aircraft emissions estimated using the mode specific method; and (iv) with mode specific aircraft emissions but without GSE emissions.

### 4.3 Emission Modelling

An air quality model like CMAQ needs hourly, gridded, and speciated emissions. Here Sparse Matrix Operator Kernel Emission (SMOKE) (CEP, 2003) is used for spatiotemporal distribution and speciation. Historically SMOKE treated aircraft emissions as point sources and did not distribute them spatially. However, recently SMOKE started to give the user an option to spatially allocate airport-emissions to grid cells based upon a ‘‘point location’’ (CEP, 2003; Strum et al., 2004). Furthermore, SMOKE uses a default temporal profile for all aircraft source category code (SCC) types. In order to better resolve and distribute emissions from aircraft, we developed a new emissions processing framework. This framework involves the following parts: temporal distribution; three-dimensional (3-D) spatial distribution; and speciation.

Temporal distribution: Hourly emissions profiles were developed from activity profiles for Hartsfield–Jackson airport (Nissalke, 2003). There were three different temporal profiles: monthly; weekly; and diurnal. Some of these temporal profiles differ significantly from default temporal profiles used in SMOKE.

Spatial distribution: Emissions from aircraft are distributed in 3-D space especially in takeoff, climb- out, and approach modes. However, they are generally put into the first layer of the air quality model in a single grid cell that coincides with the airport location. Here, emissions were distributed using two different methods. In the first method emissions were injected into the three first-layer cells. In the second method, actual aircraft location data were used to distribute the emissions horizontally and vertically. The location data consisted of 3-D coordinates typical for different aircraft types during takeoff and landing in 2011.

### 4.3.1 Flight mode determination

Flight profiles (see in Figure 4.2 and 4.3) are inspected to determine the flight modes of each flight. Some extra algorithms are developed for special cases. After determining the flight modes the pollutants are calculated respectively.

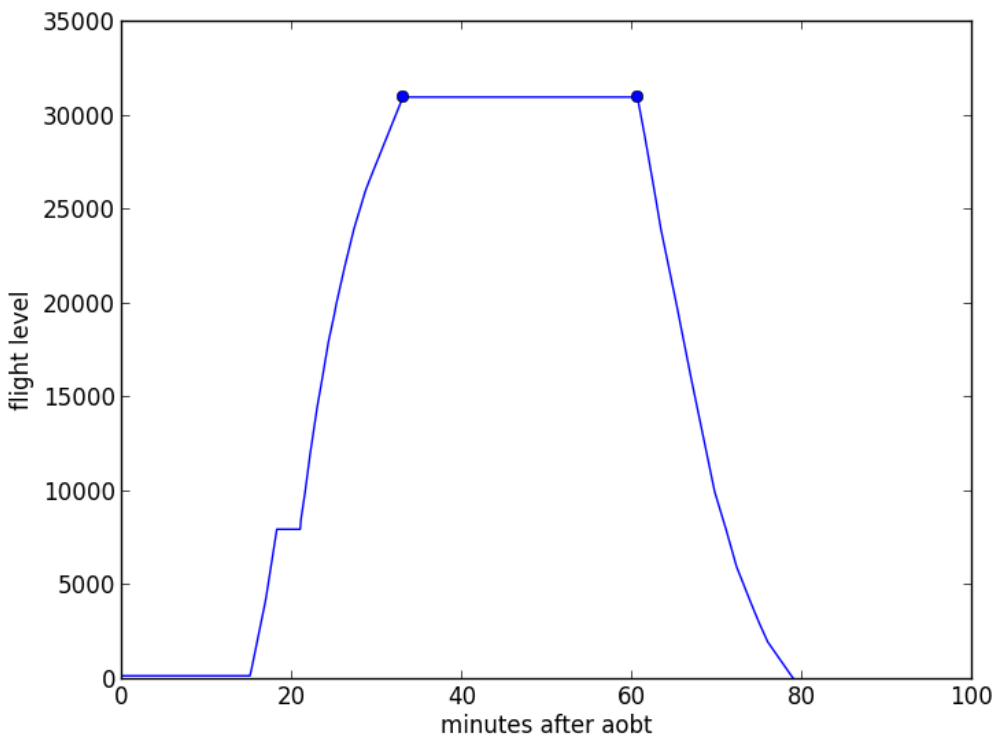


**Figure 4.2 :** Regular Flight with a Small Grade Example.

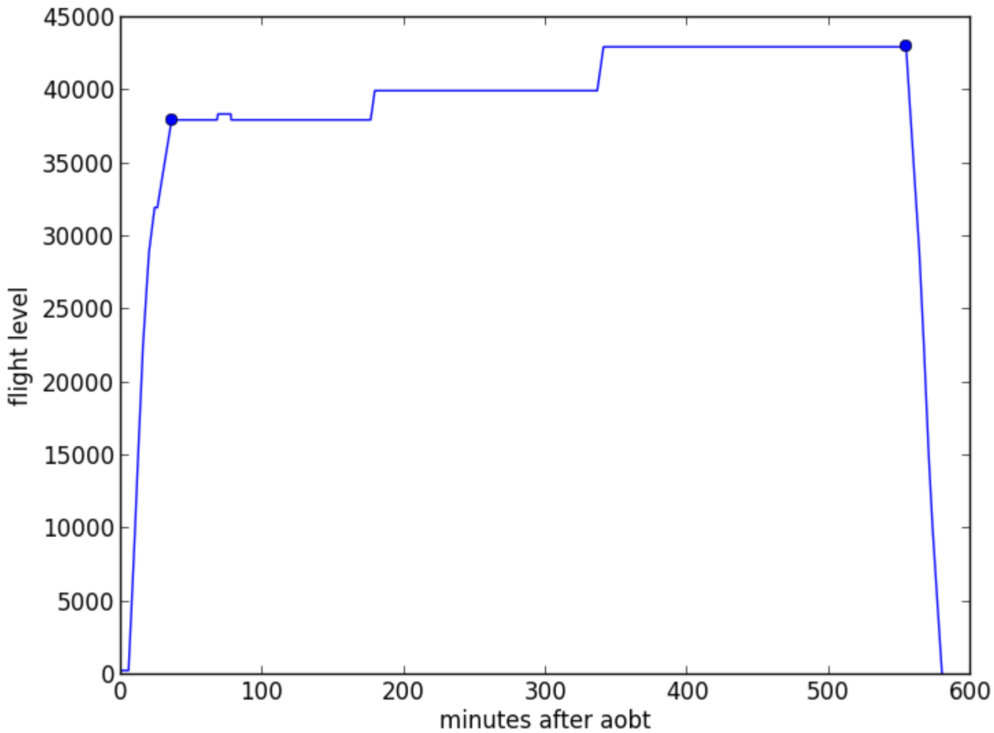


**Figure 4.3 :** Graded Flight Example.

Highest flight level and minimum cruise level have been determined for flight mode calculation. In order to calculate minimum cruise level a tolerance constant has been used. This tolerance constant has been selected by experimenting. The formula for minimum cruise level is as follows:

$$\frac{(z_{max} \times t_{duration} \times C_{tolerance})}{100 \times 100} \tag{4.1}$$

It has been observed that the cruise altitude is related with the duration of the flight and the maximum flight level. After cruise level is determined the first highest point is selected as the starting of cruise and the last higher points is selected as cruise end. After determinin the cruise start and end points (shown with dots on figure 4.2 and 4.3) other phases are determined relatively: Zero to cruise as ascending, cruise start to end as cruise and cruise to zero  as descending. After determining this phases these connected with the actual flight modes, Taxi In, Take-off, Climb, Cruise, Descend, Landing, Taxi Out. After determining this phases these are mapped to Idle, Take-Off, Climb and Approach.

In the end, Taxi-In and Taxi-Out are classified as Idle, Take-off as Take-off, Ascending, Descending and Landing as Climb (or Land) and Cruise as Cruise.

# 5. RESULTS AND CONCLUSIONS

## 5.1 Data Understanding

Within the scope of the European Air Traffic Network's Macro analysis, ALL_FT+ and DDR data structures and their crucial features have been completely revealed after extensive research. ALL_FT+ data has been employed to extract the information of daily flight routes and their crossed airspace profiles whereas DDR capacity data has utilized to observe the capacity declarations for ATC unit airspaces and arrival/departure flow rates of airports. Since the ultimate goal of the analyses is to observe the dynamics of delay propagation, FTFM, CTFM and CPF-REF data profiles have been extracted from eight different data structures of ALL_FT+ data. Two different profiles of each flight model: point profile and airspace profile, have been processed and analyzed simultaneously. Besides extracting these three models, auxiliary parameters as departure, arrival airports and Off-Block Times also have been sorted out in the contemplation of clarifying all phases of each flight. The attempt of using Off-Block Times also provided an insight about airports' status and delays generated by ground procedures in macro analyses, which is also a subsidiary information for micro analyses.

## 5.2 Initial Simulation Results

In the figure 5.1 connection graphs of busiest airport are demonstrated. Red airports denote the major airport whereas blue airports denote the major airports that are not in the top ten busiest airports and connected to those airports. Frankfurt Airport (EDDF) is one of the most connected nodes, thus it is studied closer.

Figure 5.2 shows how European airports are connected and how their delay distributions are. The simulation results have similar results in terms of delay distribution, however since only the delays caused by capacity managements are calculated the results vary.

Departure and arrival delay characteristics of Frankfurt Airport (EDDF) in July and November, 2011 are presented in figures 5.3, 5.4, 5.5, and 5.6. There are some outlier delays occurring three times in a month. These delays are caused by different reasons that are listed in the case study in the appendix section.

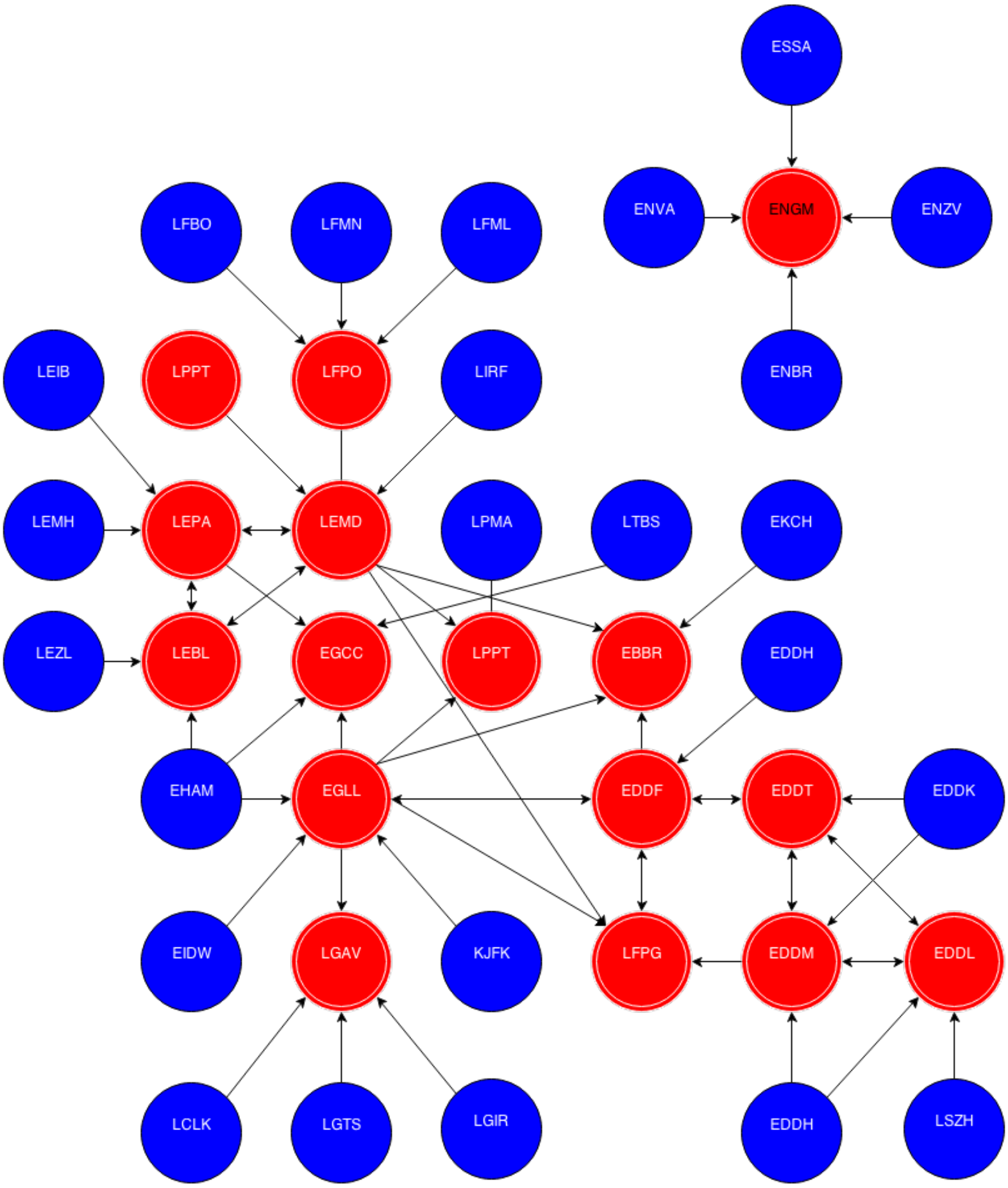


**Figure 5.1 :** Connection Graph of Busiest Airports.


**Figure 5.2 :** Delay and Connection Graph of Europe Region.

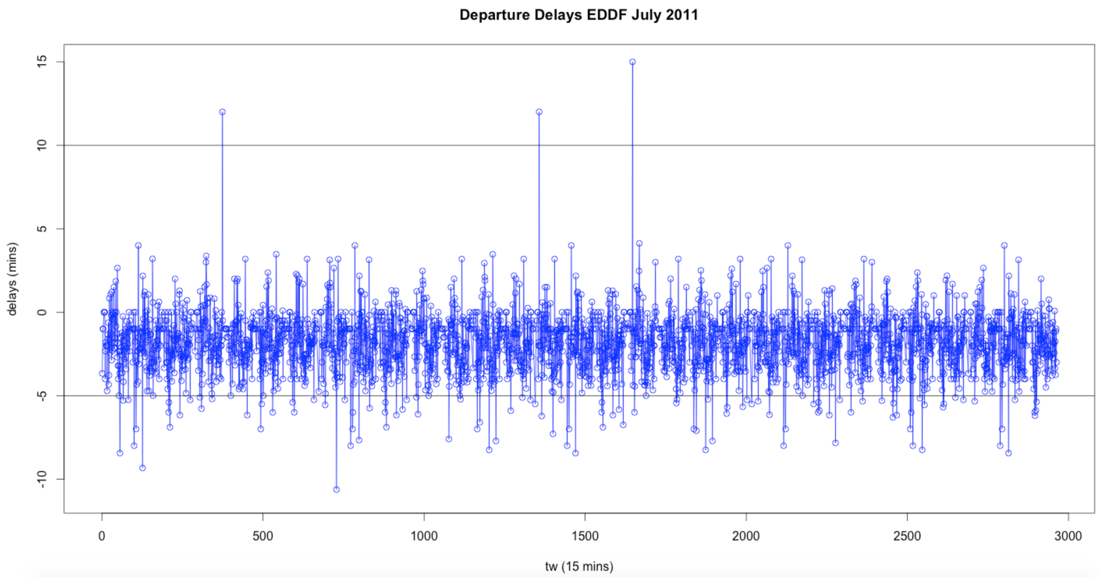


**Figure 5.3 :** Frankfurt Departure Delay Profile, July 2011.

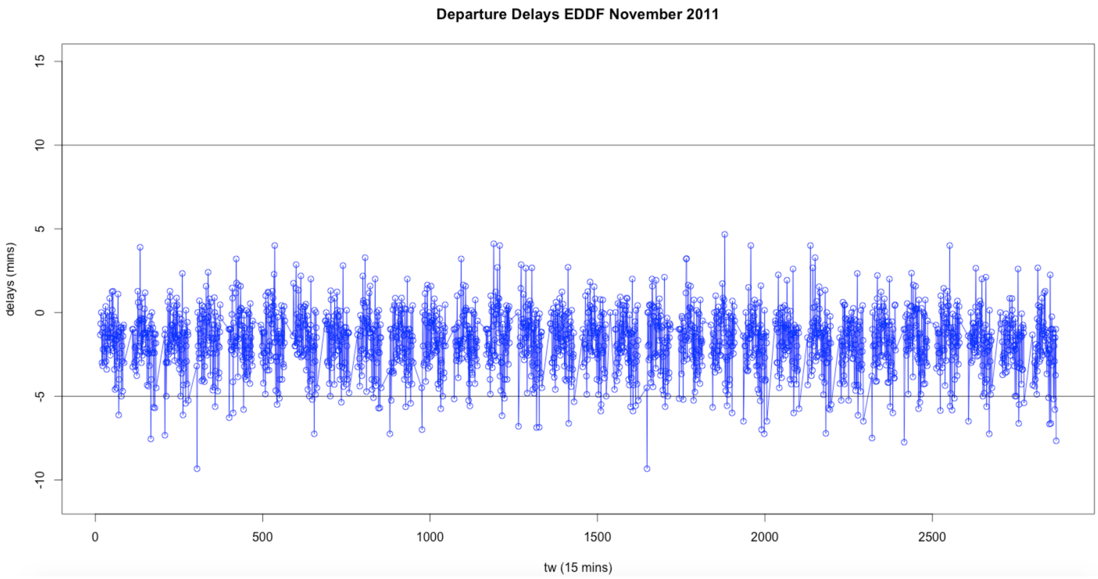


**Figure 5.4 :** Frankfurt Departure Delay Profile, Nobember 2011.

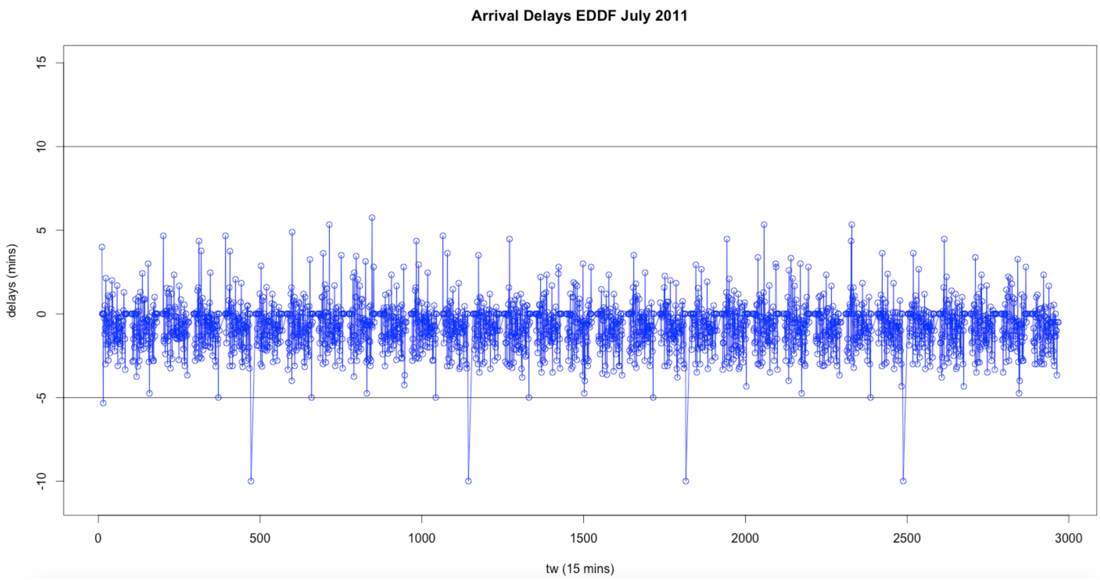


**Figure 5.5 :** Frankfurt Arrival Delay Profile, July 2011.

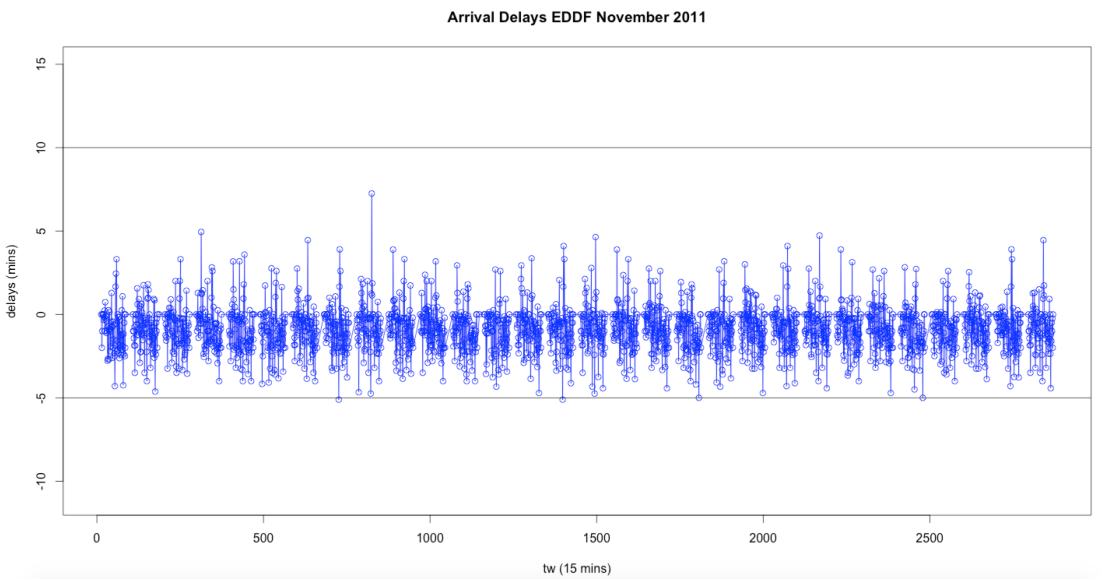


**Figure 5.6 :** Frankfurt Arrival Delay Profile, Nobember 2011.

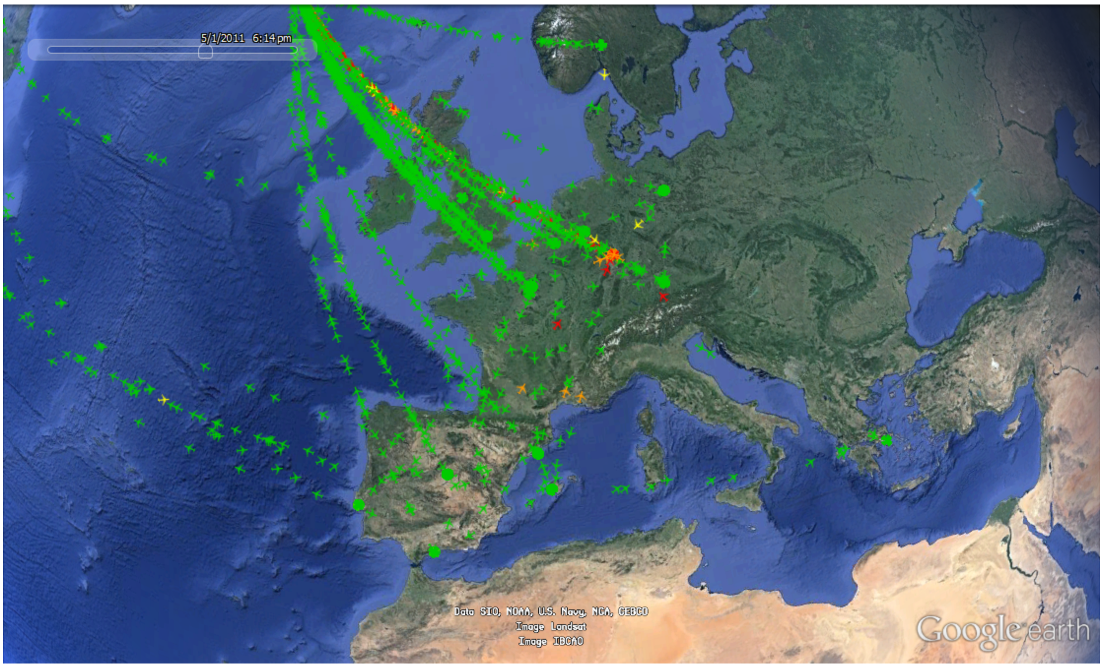


**Figure 5.7 :** Live Simulation European Region Only.

A live simulation has been developed and run in Google Earh shown in Figure 5.7 and 5.8. These animations show the propagation of the delay from one airport to another and how the delay model behaves like epidemic distribution.

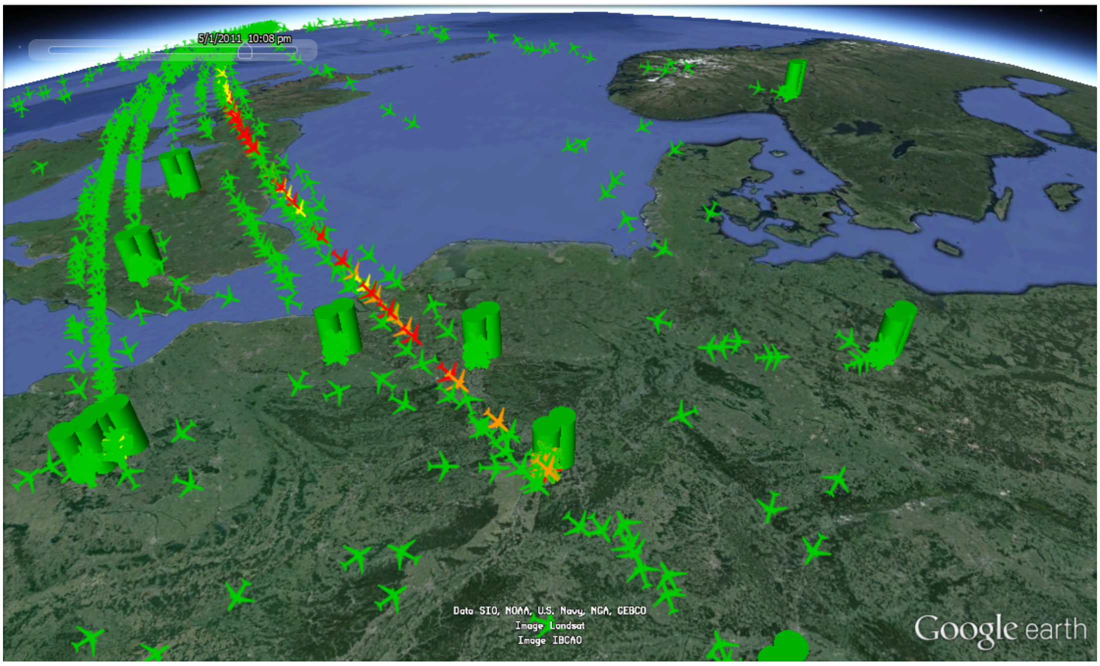


**Figure 5.8 :** Live Simulation, Delays Occuring.

### 5.3 Emission Grid Results

All of the data provided in ALL_FT+ for year 2011 has been used as BigData application and the cumulative results have been calculated using bining for each grid. The data is organized as 4D grids with the dimesions latitude, longitude, altitude and the time windows which aircraft passes by. Since there are 9 months of data starting from March ending at November, results are reflecting data expect the Winter. In table 5.1 total mass of each polutant studied can be seen tabulated in tonnes.

**Table 5.1 :** Total Mass of the Polutants in Tonnes

| Pollutant | Total Mass in Tonnes |
|---|---|
| HC | 231.3918 |
| NOx | 13300.67 |
| CO | 1405.744 |

According to figure 5.9 the region with the highest emission footprint is Ataturk Airport region. The rates slowly decay on the flight routes. There is an emission density on the region where flights routes are intersecting. This can be observed in more detail in figure 5.10 where emission paths are connected.

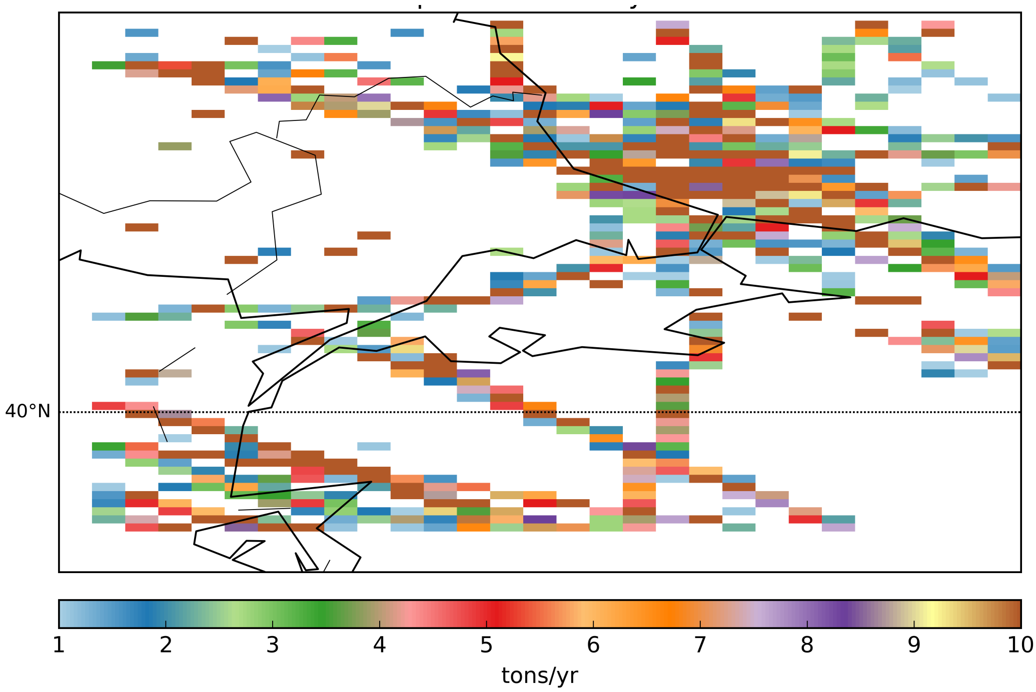


**Figure 5.9 :** Total Emission in Tonnes for Marmara Region, 2011 (low resolution).

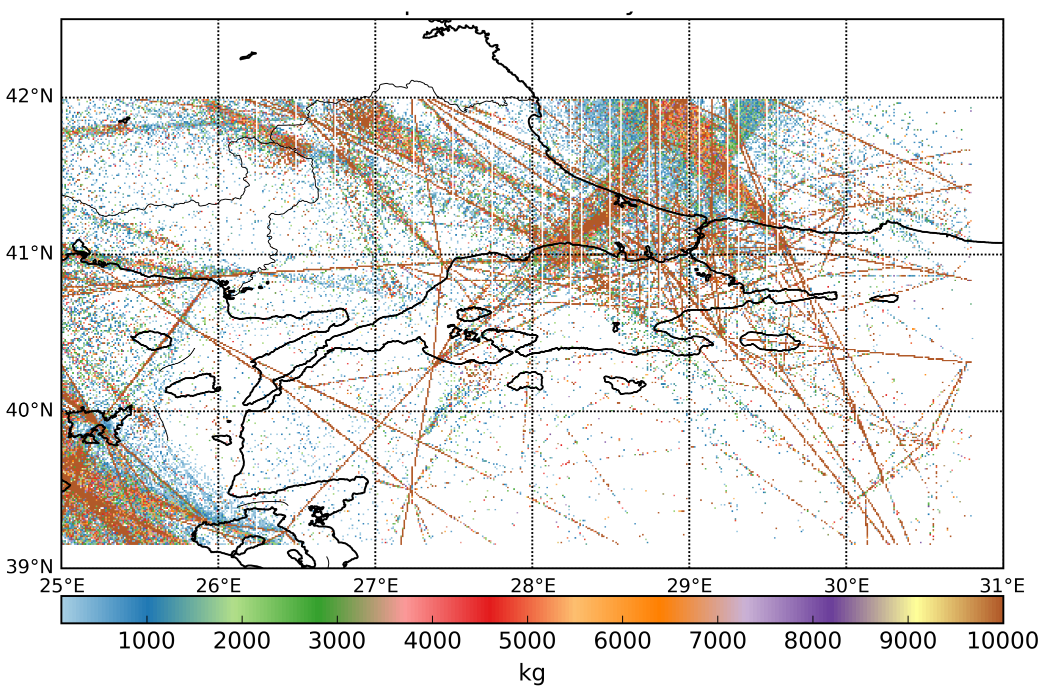


**Figure 5.10 :** Total Emission in Tonnes for Marmara Region, 2011 (high resolution).

There are fields in the data enabling categorization and aggregation of emissions in different aspects. In data there are:

- Month
- Day of Week
- Hour of Day
- Altitude
- Longitude and Latitude
- Origin, Destination Airports
- Airline Carrier
- Type of Flight
- Etc.

However, we focused mainly on temporal (month, day of week, hour of day) and spatial (altitude, longitud, and latitude) distributions of emissions generated by the aircrafts using the flight data over the Marmara region.

### 5.3.1 Day of week emission distribution

Firstly our main focus was investigating how day of week effects the distribution of each pollutant. In figure 5.11, 5.12, and 5.13 the trends can be observed as highly correlated with the day of week. This result can be confirmed by checking each pollutant's correlation value with the day of week (see Table 5.2).

While HC has the highest correlation its trend resembles the trend of CO emssion distribution. However, NOx behaves differently. This behaviour is observed at all features because it is related with the emission coefficients in Table 5.3. The coefficients are divided by the takeoff coefficients in order to emphasize the characteristic differences between NOx and other pollutants. Because HC and CO Take Off to Idle Ratio is lower while NOx has a higher Take Off to Idle Ratio.

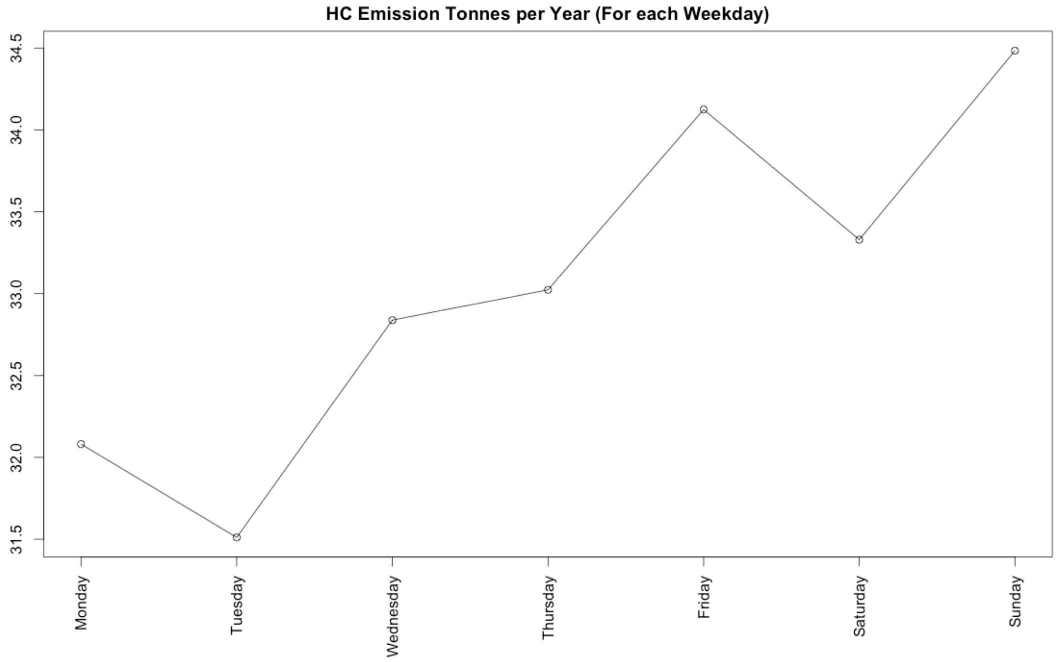


**Figure 5.11 :** HC Total Emission in Tonnes for Each Day of Week, 2011.

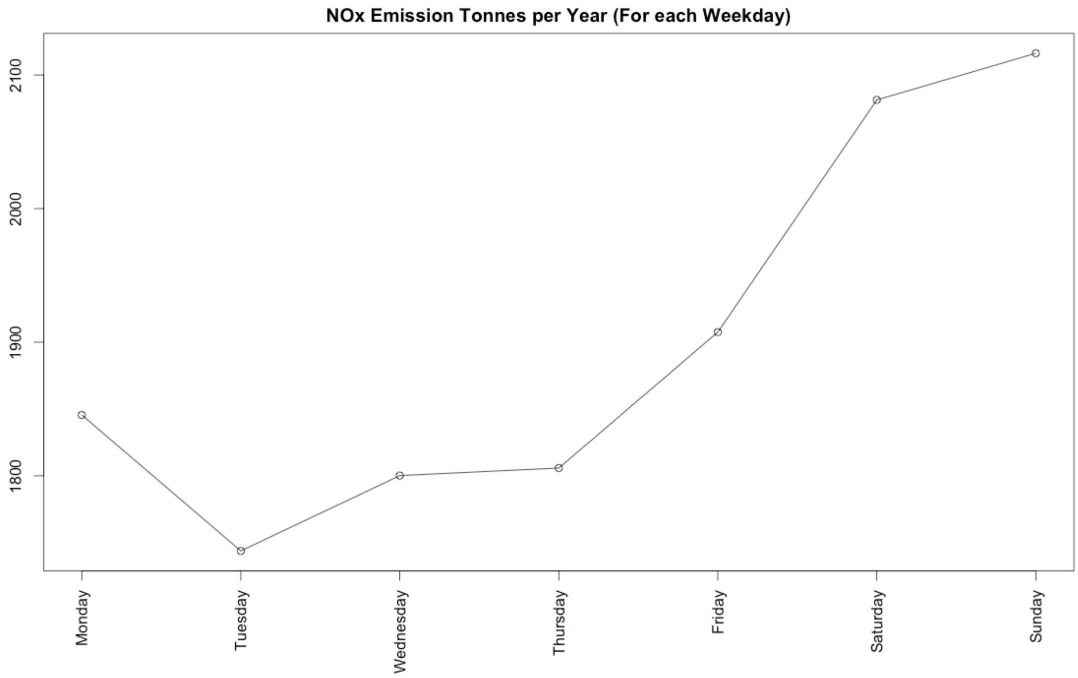


**Figure 5.12 :** NOx Total Emission in Tonnes for Each Day of Week, 2011.

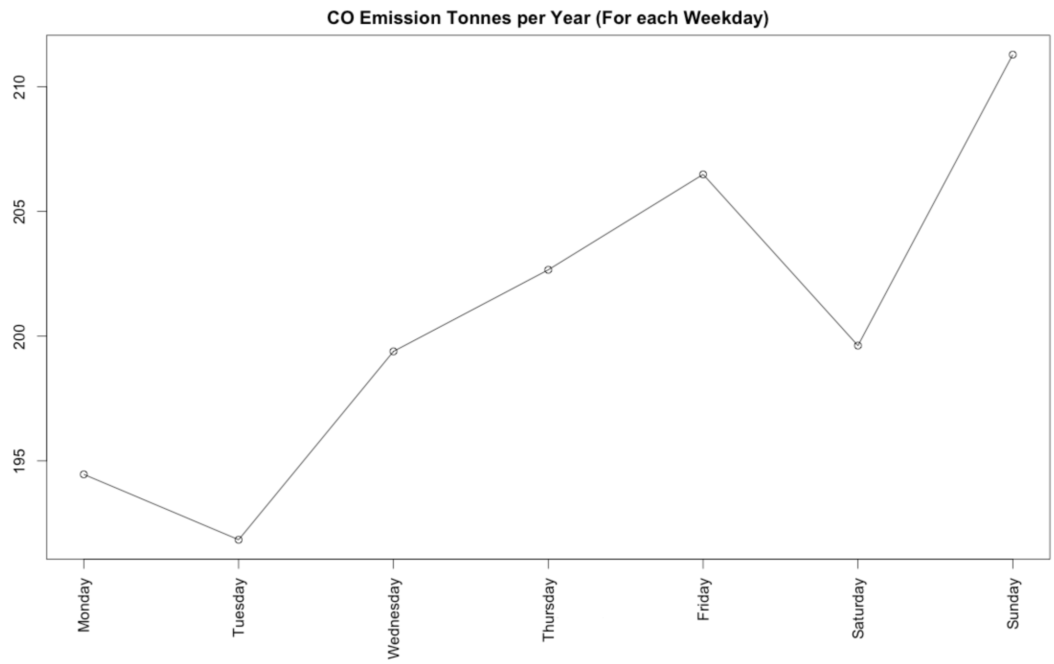


**Figure 5.13 :** CO Total Emission in Tonnes for Each Day of Week, 2011.

**Table 5.2 :** Pollutant Emission Correlation Score with Day of Week.

| Pollutant | Correlation with Weekday |
|---|---|
| HC | 0.8896148 |
| NOx | 0.8492917 |
| CO | 0.8417995 |

**Table 5.3 :** Relative Pollution Factors for Each Pollutant and Phase of Flight.

| Pollutant | Climb/Takeoff | Approach/Takeoff | Idle/Takeoff |
|---|---|---|---|
| HC | 1.142857143 | 10.5 | 100.5714286 |
| NOx | 0.809824903 | 0.382782101 | 52.19512195 |
| CO | 1.317073171 | 12.29268293 | 0.178988327 |

### 5.3.2 Monthly emission distribution

Another important factor in categorizing emission distribution is month. There are significant differences in emmissions between seasons. In Figure 5.14, 5.15, and 5.16 it can be observed that the emission constantly climbes up from March through July and statys high until October and dropbs again. However, this behaviours is a little different in NOx. For NOx emissions grow until August and drops again in October.

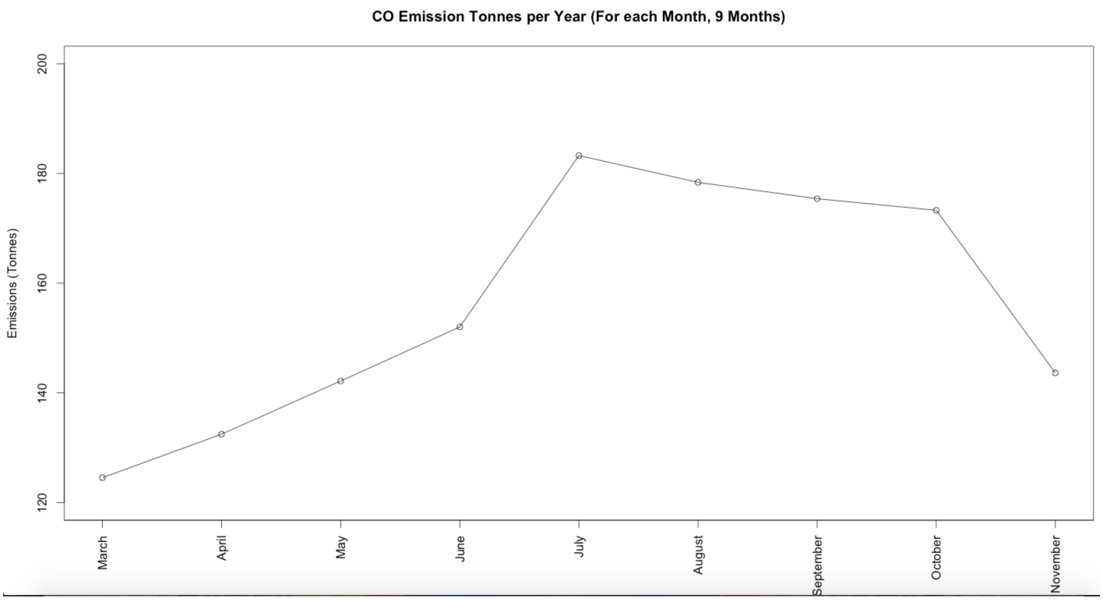


**Figure 5.14 :** CO Total Emission in Tonnes for Each Month of Year, 2011.

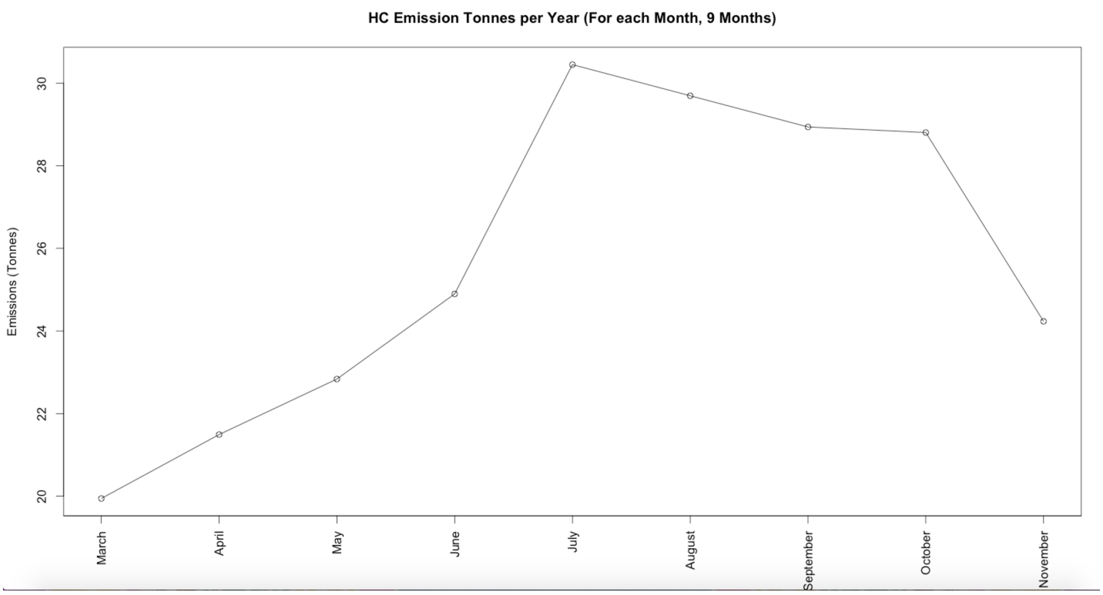


**Figure 5.15 :** HC Total Emission in Tonnes for Each Month of Year, 2011.

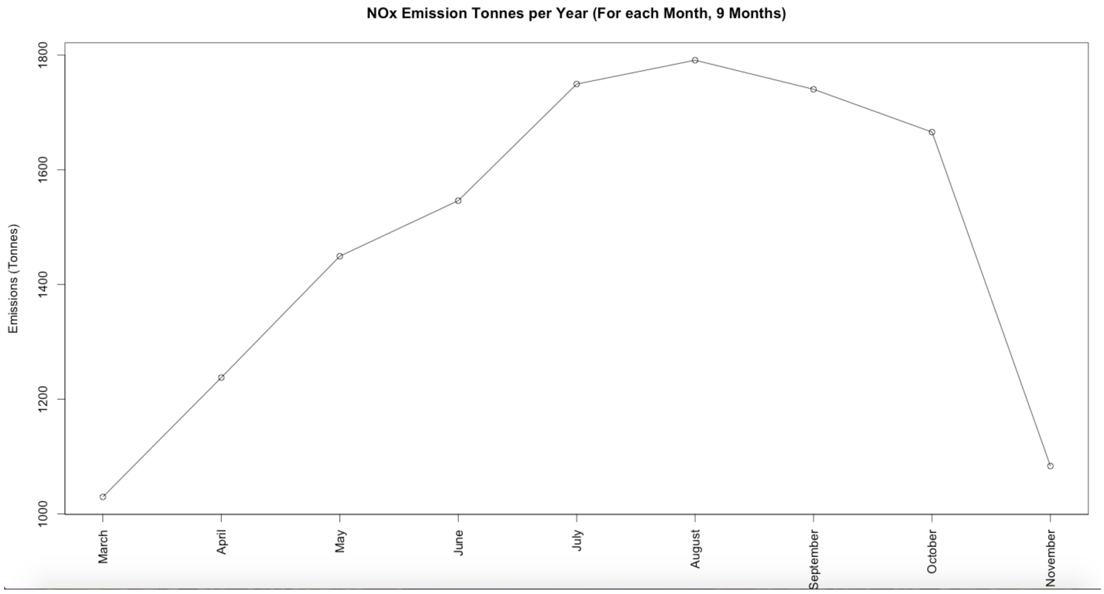


**Figure 5.16 :** NOx Total Emission in Tonnes for Each Month of Year, 2011.

### 5.3.3 Hourly emission distribution

Another highly correlated factor of emissions are the hour of the day. Since there is a higher flight rate at the 8-9 pm, there is a higher emission rate. Correlation rates can be seen in Table 5.4. This can be interpereted as there is a nearly a linear correlation between NOx and hour of day.

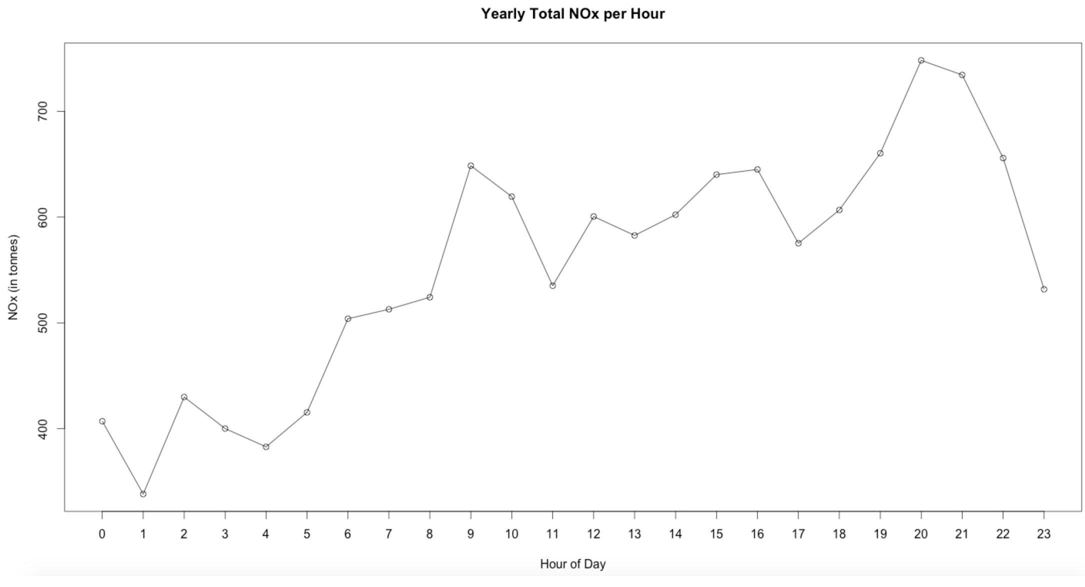


**Figure 5.17 :** NOx Total Emission in Tonnes for Each Hour of Day, 2011.

**Table 5.4 :** Pollutant Emission Correlation Score with Hour of the Day.

| Pollutant | Correlation with Hour of the Day |
|---|---|
| HC | 0.6067262 |
| NOx | 0.8307334 |
| CO | 0.5738456 |

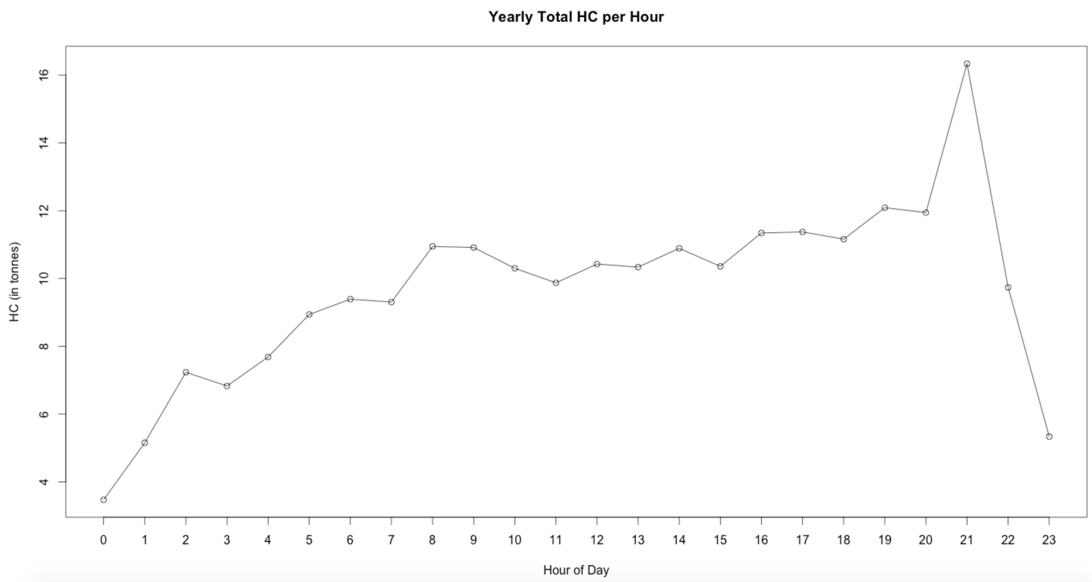


**Figure 5.18 :** HC Total Emission in Tonnes for Each Hour of Day, 2011.

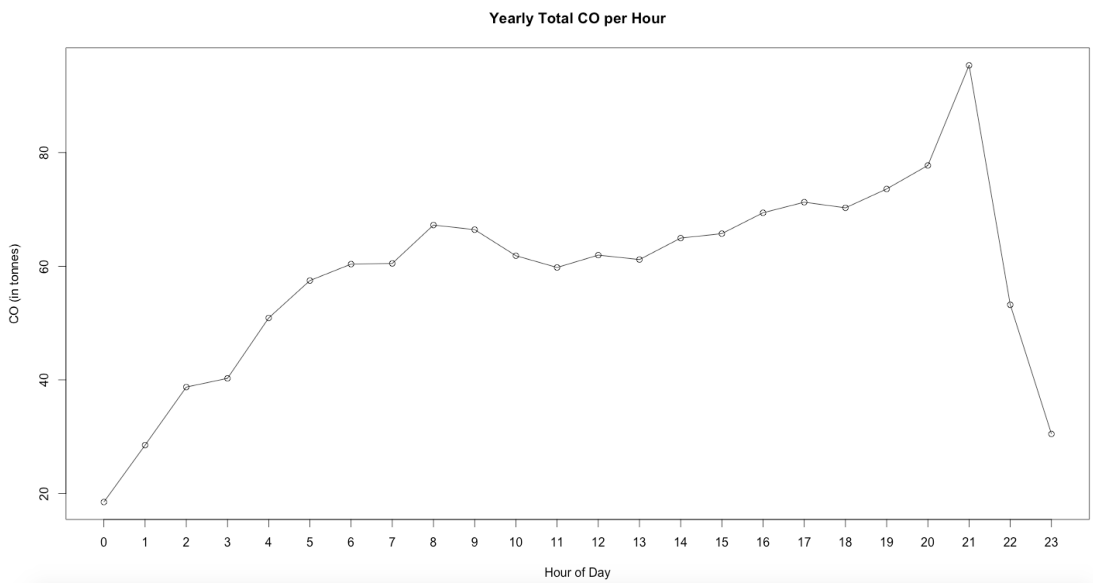


**Figure 5.19 :** CO Total Emission in Tonnes for Each Hour of Day, 2011.

### 5.3.4 Emission distribution by altitude

Another spatial categorization is altitude. There is a high emission peak at between 200 and 300 feet and between 10000 and 20000 feet. However, NOx has only a peak at between 10000 and 20000 feet. Because climb and idle coeeficients are relatively closer to each other in NOx (see Table 5.3).

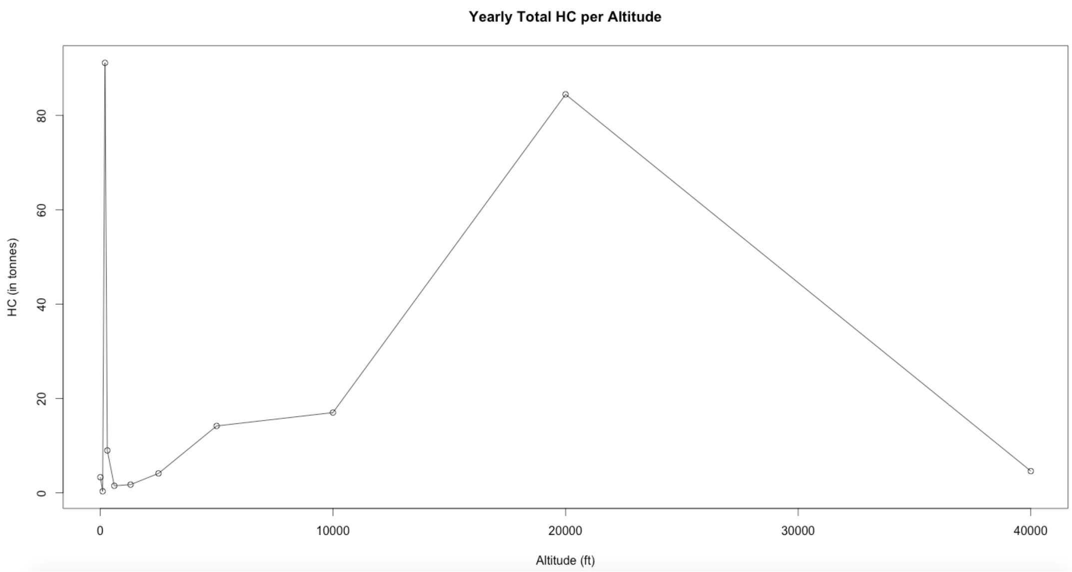


**Figure 5.20 :** HC Total Emission in Tonnes for Each Altitude.

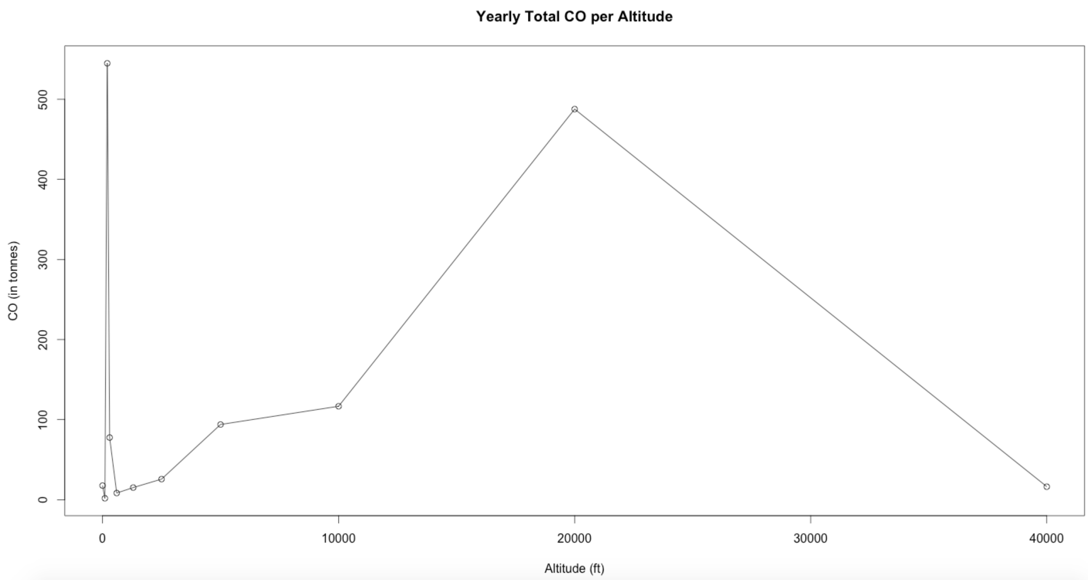


**Figure 5.21 :** CO Total Emission in Tonnes for Each Altitude.

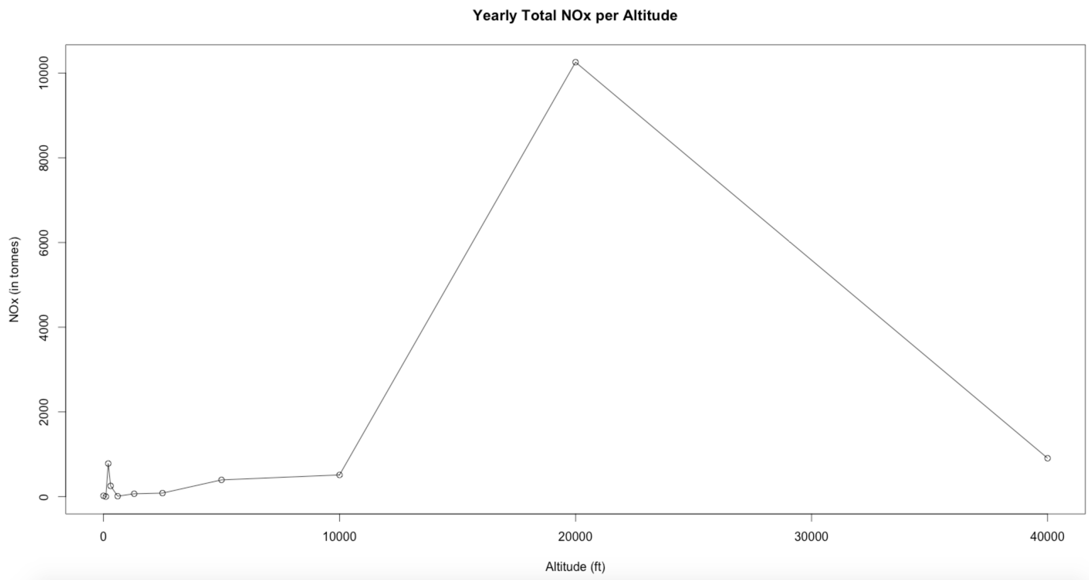


**Figure 5.22 :** NOx Total Emission in Tonnes for Each Altitude.

In conclusion, month, day of week, hour of the day, longitude, latitude, and altitude have a very high correlation with each pollutant. A prediction or simulation model can be modeled using these parameters as features for prediction or simulation

## REFERENCES


**AIRBUS**, Flying smart, thinking big: Global Market Forecast 2009-2028, http://www.slideshare.net/xiongjianvlsi/air-bus-global-market-forecast

**FAA/Eurocontrol (2010).** Human performance in air traffic management safety: A white paper, *FAA/Eurocontrol Action Plan 15*.

**Bisignani, G.** (2010). State of the Air Transport Industry, 66*th IATA Annual General Meeting and World Air Transport Summit*, Berlin, Germany.

**Byun, D.W., Ching, J.K.S., (1999).** Science Algorithms of the EPA Models 3 community multiscale air quality (CMAQ) modeling system. Washington, US EPA.

**CEP, (2003).** Sparse matrix operator kernel emissions modeling system (SMOKE). User Manual, Carolina Environmental Program.

**EUROCONTROL.** (2014). D4.1 Concept for the new, efficient Air Traffic System, *Resilience2050.eu*

**FAA, (2001).** Emissions and dispersion modeling system (EDMS) 4.01. Federal Aviation Administration, Washington DC.

**Hanna, S.R., Chang, J.C., Fernau, M.E., (1998).** Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables. Atmospheric Environment 32, 3619–3628.

**Henry, R.C., Chang, Y.-S., Spiegelman, C.H., (2002).** Locating nearby sources of air pollution by nonparametric regression of atmospheric concentrations on wind direction. Atmo- spheric Environment 36, 2237–2244.

**Kaya Y. B., Inalhan G., (2014).** A Dynamic Bayesian Belief Network Approach for Modelling the ATM Network Delays. *International Conference on Research in Air Transportation*, May 26-30, 2014 Istanbul, Turkey.

**Koyuncu E., Eren U. and Inalhan G., (2013).** Data Analytic Synthesis and Stochastic Modeling of the European ATM Network Flow Model, *Satellite Meeting - Complexity Science and Transportation Systems '13 (ECCS'13)*, September 18th, 2013, Barcelona, Spain

**NARSTO, (2000).** An assessment of tropospheric ozone pollu- tion. Richland, WA, NARSTO.

**Nissalke, T., (2003).** Personal communication.

**Popp, P.J., Bishop, G.A., Stedman, D.H., (1999).** Method for commercial aircraft nitric oxide emission measurements. Environmental Science and Technology 33, 1542–1544.

**Sawyer, R.F., Harley, R.A., Cadle, S.H., (2000).** Mobile sources critical review: NARSTO assessment. Atmospheric Envir- onment 34 (12–14), 2161–2281.

**Spicer, C.W., Holdren, M.W., Smith, D.L., Hughes, D.P., Smith, M.D., (1992).** Chemical composition of exhaust from aircraft turbine engines. Journal of Engineering for Gas Turbines and Power-Transactions of the ASME 114, 111–117.

**Strum, M., Driver, L., Mason, R., Smith, J.D., Billings, R., (2004).** Spatially allocating airport-related emissions for air quality modeling: new approach and new data. Presented at 13th International Emission Inventory Conference, Clear Water, FL.

**Unal, A., Tian, D., Hu, Y., Russell, A.G., (2003).** 2000 Emissions inventory for fall line air quality study (FAQS). Atlanta, GA, Prepared for Georgia Department of Natural Re- sources, Environmental Protection Division.

**Url-1**, ATC Fast Time Simulator and Air Traffic Optimizer (AirTOp), http://www.airtopsoft.com/products.html

**Url-2**, Total Airspace and Airport Modeler (TAAM), http://ww1.jeppesen.com/industry-solutions/aviation/government/total-airspace-airport-modeler.jsp

**Wayson, R.L., Fleming, G.G., Kim, B., (2003).** Status report on proposed methodology to characterize jet/gas turbine engine particulate matter emissions. US Department of Transportation, Federal Aviation Administration.

# APPENDICES

## APPENDIX A: Total Flow Chart

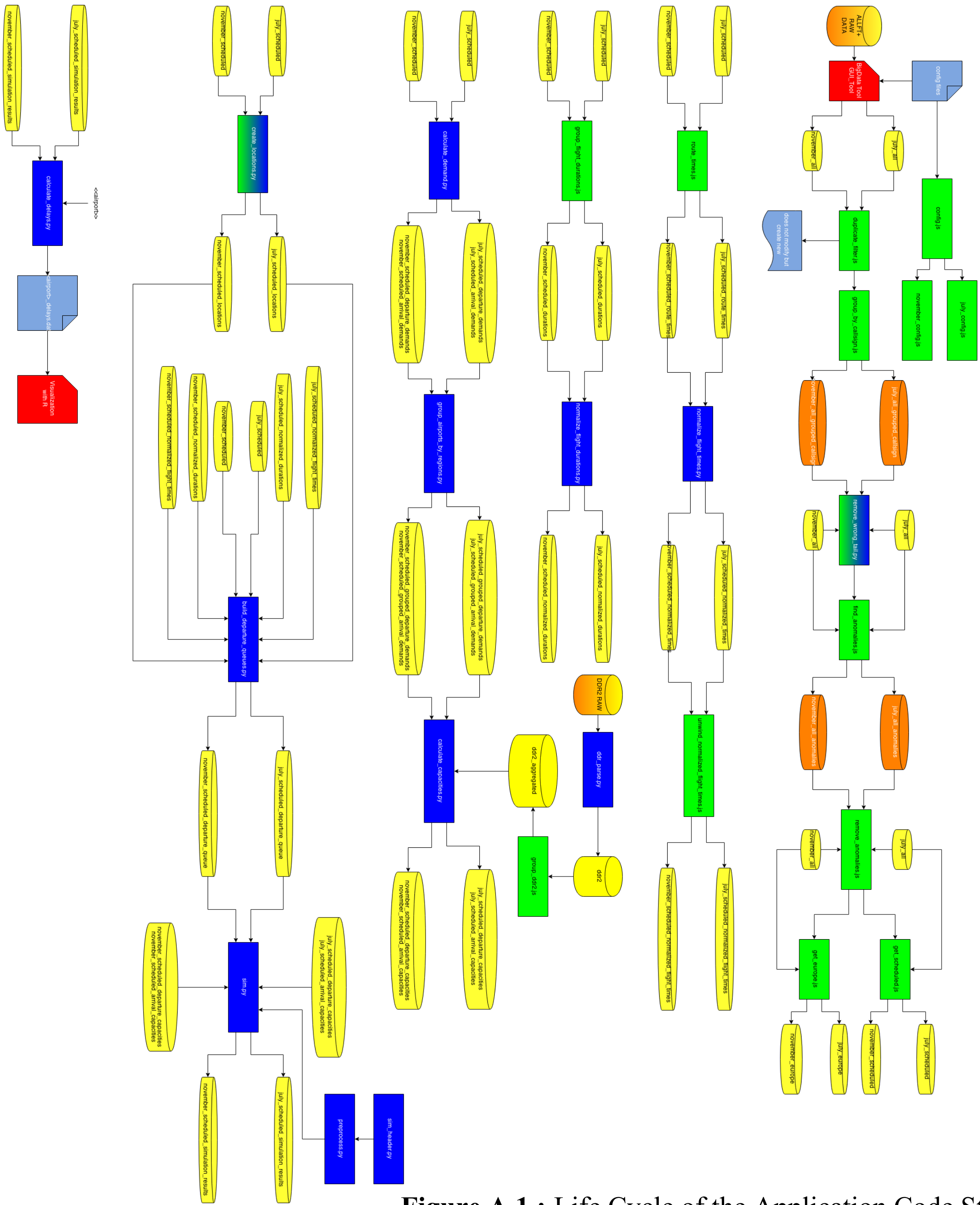

**Figure A.1 :** Life Cycle of the Application Code Structure

## APPENDIX B: Case Study on Unexpected Arrival Delays at Frankfurt Airport

### Case 1: Arrival demand over capacity on EDDP airport

The arrival capacity is determined as **5 per 15 minutes** by considering calculated 99% confidence nominal capacity. However, the arrival and departure demand is infinity according to DDR2 data (see Table 6.1).

**Table A.1 :** Sample from DDR2 data.

| |
|---|
| 10/02/2011;EDDP;00:00;23:59;**999**;_;AD;**A**;B |
| 10/02/2011;EDDP;00:00;23:59;**999**;_;AD;**D**;B |
| 10/02/2011;EDDP;00:00;23:59;999;_;AD;G;B |
| 10/02/2011;EDDP;00:00;23:59;30;_;TV;G;B |
| 999: Infinity A: Arrival D: Departure |

Since there arrival demand is below the capacity and there is no propagating demand from the previous time windows, there is no delay between 22:00 and 22:15 as seen on Table 6.2 However, between 22:15 and 22:30 the arrival demand is over the capacity by one (gets propagated). That over capacity causes one flight to shift to next time window. Since, the next time window is at full capacity another flight is shifted to the next time window, too, which is 378. Since 3+1 (3 existing + 1 propagated) demand is less then 5, there is no propagating delay effect left at 22:45-23:00 time window.

**Table A.2 :** Arrival Demand in EDDP Airport.

| Time Window | Time Range | Arrival Demand | Propagated |
|---|---|---|---|
| 375 | 22:00-22:15 | 3 | 0 |
| 376 | 22:15-22:30 | 6 | 0 |
| 377 | 22:30-22:45 | 5 | +1 |
| 378 | 22:45-23:00 | 3 | +1 |

### Case 2 and 3: Arrival demand over capacity on LGIR airport

The arrival capacity of LGIR airport is 2 with respect to both DDR2 data (see Table 6.3) and calculated 95% confidence nominal capacity.

**Table A.3 :** Sample from DDR2 data.

| |
|---|
| 22/02/2011;LGIR;00:00;23:59;**8**;_;AD;**A**;B |
| 22/02/2011;LGIR;00:00;23:59;7;_;AD;D;B |
| 22/02/2011;LGIR;00:00;23:59;20;_;AD;G;B |
| 22/02/2011;LGIR;00:00;23:59;20;_;TV;G;B |

There is an extreme arrival delay at time window 1367 and 1369 as seen on Table 6.4, which is more than 1 hour. So, arrival capacity should be more than 2.

**Table A.4 :** Arrival Demand in LGIR Airport.

| Time Window | Time Range | Arrival Demand | Propagated |
|---|---|---|---|
| 1361 | 4:45-5:00 | 1 | |
| 1363 | 5:15-5:30 | 2 | |
| 1365 | 5:45-6:00 | 3 | |
| 1366 | 6:00-6:15 | 3 | +1 |
| 1367 | 6:15-6:30 | 4 | +2 |
| 1368 | 6:30-6:45 | 2 | +4 |
| 1369 | 6:45-7:00 | 5 | +4 |
| 1370 | 7:00-7:15 | 3 | +7 |
| 1371 | 7:15-7:30 | 4 | +8 |
| 1373 | 7:45-8:00 | 1 | +8 |

There is a major problem at determining capacity. One reason for that is the capacity is calculated from scheduled flights, whose initial off block times are not normalized. Although there is a slight difference there could be special cases, especially when the capacity is small.

**Table A.5 :** Capacity Calculation with Different Methods.

| Capacity Calculation Method | Capacity | Confidence |
|---|---|---|
| Not Normalized (Scheduled Demand) | 2 | 95.26% |
| Not Normalized (Scheduled Demand) | 3 | 98.65% |
| Normalized (Departure Queue) | 2 | 94.32% |
| Normalized (Departure Queue) | 3 | 98.15% |

Considering Table 6.4 the capacity value is calculated as 2 if scheduled demand is used for demand data set (current method). However, the capacity value is calculated as 3 if departure queue is used as data set. Using departure queue might make sense because it is the data set being used for actual schedule simulation.

## APPENDIX C: ALL_FT+ Detailed Fields

**Table A.6 :** ALL_FT+ Detailed Fields

| # | Field | Type | Size | Comment |
|---|---|---|---|---|
| 1 | Origin Airport ICAO ID | char | 4 | e.g. LTBA |
| 2 | Destination Airport ICAO ID | char | 4 | e.g. EDDF |
| 3 | Aircraft ID | char | 7 | Found less than 7, e.g. THY141 |
| 4 | Aircraft Operator ICAO ID | char | 3 | e.g. THY |
| 5 | Aircraft Type ICAO ID | char | 2-4 | e.g. A320 |
| 6 | AOBT | datetime | 14 | Actual Off Block Time, format: YYYYMMDDHHMMSS, e.g. 20130322191000 |
| 7 | IFPS ID | char | 10 | The IFPS unique identification key for the flight, format: XXnnnnnnnn (two letters AA or BB); e.g. AA83161603 |
| 8 | IOBT | datetime | 14 | Initial Off Block Time, format: YYYYMMDDHHMMSS, e.g. 20110703171500 |
| 9 | Original Data Quality | char | 3 | FPL<br>RPL |
| 10 | Flight Data Quality | char | 3 | FPL<br>RPL |
| 11 | Data Source | char | 3 | FPL:<br>RPL: Repetetive Flight Plan<br><br>FNM:<br>MFS:<br><br>AFP:<br>AFI: |
| 12 | Exemption Reason Type | char | 4 | NEXE:<br>EMER:<br>SERE(?):<br>HEAD:<br>AEAP: |
| 13 | Exemption Reason Distance | char | 4 | NEXE:<br>LONG: |
| 14 | Late Filer | boolean | 1 | Y or N |
| 15 | Late Updater | boolean | 1 | Y or N |
| 16 | North Atlantic Flight | boolean | 1 | Y or N |
| 17 | COBT | datetime | 14 | Calculated Off Block Time, format: YYYYMMDDHHMMSS, e.g. 20130322191000 |
| 18 | EOBT | datetime | 14 | Estimated Off Block Time, format: YYYYMMDDHHMMSS, e.g. 20130322191000 |
| 19 | LOBT | datetime | 14 | Last Off Block Time, format: YYYYMMDDHHMMSS, e.g. 20130322191000 |
| 20 | Flight State | char | 2 | TE: Terminated – Expected Landed (most frequent)<br>CA: Canceled<br>AA: ATC Activated – Reported Airborne<br>TA: TACT Activated – Expected Airborne<br>FI: Filed |

| | | | | |
|---|---|---|---|---|
| 21 | Previous | char | 2 | FI: (most frequent)<br>SI: Slot Issued<br>NE: Not Exempted (Flight State: CA or FI)<br>FS: Slot Calculated (Flight State: mostly TE) |
| 22 | Suspension Status | char | 2 | Suspended with the following status:<br>NS: Not Suspended<br>NR: Not Reported as Airborne<br>RC: FCM Required<br>RV: FP Revalidation<br>SM: Slot Missed<br>TV: Traffic Volume Condition |
| 23 | TACT ID | num | 6 | The ETFMS unique identification key for the flight |
| 24 | SAM CTOT | datetime | 14 | Slot Allocation Message<br>Calculated Take Off Time format:<br>YYYYMMDDHHMMSS, e.g. 20130322101014<br>(only about 13% is available) |
| 25 | SAM Sent | boolean | 1 | Slot Allocation Message Sent |
| 26 | SIP CTOT | datetime | 14 | Slot Improvement Proposal Message<br>Calculated Take Off Time format:<br>YYYYMMDDHHMMSS<br>(Not present : SIP Sent only N) |
| 27 | SIP Sent | boolean | 1 | Slot Improvement Proposal Message Sent (Only N) |
| 28 | Slot Forced | boolean | 1 | F: Forced<br>N: Not Forced |
| 29 | Most Penalizing Regulation ID | char | | e.g. LGRDK02 |
| 30 | Regulations Affected by Number of Instances | num | 1 | 0-6 |
| 31 | Excluded from Number of Instances | num | 1 | 0-3 |
| 32 | Last Received ATFM Message Title | char | 3 | REA: Ready to Depart<br>SWM: SIP Wanted Message<br>RFI: Ready for Direct Improvement<br>FCM: Flight Confirmation Message<br>SPA: Slot Proposal Acceptance<br>SMM: Slot Missed Message<br>SRJ:<br>RJT: |
| 33 | Last Received Message Title | char | 3 | FPL: Flight Plan Message<br>DEP: ICAO Defined Departure Message<br>CHG: ICAO Defined Change Message<br>DLA: Delay Message<br>ACH: ATC Flight Plan Change Message<br>APL: ATC Flight Plan Message |

| | | | | |
|---|---|---|---|---|
| 34 | Last Sent ATFM Message Title | char | 3 | SRM: Slot Revision Message<br>SAM: Slot Allocation Message<br>SLC: Slot Requirement Cancellation Message<br>FLS: Flight Suspension Message<br>DES:<br>RRP:<br>SIP: Slot Improvement Message |
| 35 | Manual Exemption Reason | char | 1 | N: nil<br>S: set<br>R: reset<br>(only N) |
| 36 | Sensitive Flight | boolean | 1 | Y or N |
| 37 | Ready for Improvement | booelan | 1 | Y or N |
| 38 | Ready to Depart | booelan | 1 | Y or N |
| 39 | Revised Taxi Time | num | 1 | In seconds? (min: 0 max: 2700)<br>Only 1.5% present |
| 40 | TIS | num | 1 | Time to Insert in Sequence<br>In seconds (most frequent:600) |
| 41 | TRS | num | 1 | Time to Remove from Sequence<br>In seconds (most frequent:300) |
| 42 | TBS Slot Message | char | 3 | To be sent slot message<br>SLC: Slot Requirement Calculation Message<br>Not present (only 1 SLC) |
| 43 | TBS Proposal Message | char | 3 | To be sent proposal message<br>Not present |
| 44 | Last Sent Slot Message | char | 3 | SRM: Slot Revision Message<br>SAM: Slot Allocation Message<br>SLC: Slot Requirement Cancellation Message<br>ERR:<br>FLS: Flight Suspension Message<br>DES:<br>16% present |
| 45 | Last Sent Proposal Message | char | 3 | RRP:<br>SIP: Slot Improvement Message<br>RRN:<br>0.1% present |
| 46 | Last Sent Slot Message Time | datetime | 14 | Format: YYYYMMDDHHMMSS<br>Same presence rate with field #44 |
| 47 | Last Sent Proposal Message Time | datetime | 14 | Format: YYYYMMDDHHMMSS<br>Same presence rate with field #45 |
| 48 | Flight Count Option | char | 1 | N: normal<br>P: proposal<br>(only N present) |

| | | | | |
|---|---|---|---|---|
| 49 | Normal Flight TACT ID | num | 1 | Only 0 present |
| 50 | Proposal Flight TACT ID | num | 1 | Only 0 present |
| 51 | Operating Aircraft Operator ICAO ID | char | 3 | e.g. THY<br>Difference with field #4<br>15% not identical if present<br>e.g. f #4: DLH f #51: CLH |
| 52 | Rerouting Reason | char | 1 | N: not rerouted<br>C: CFMU rerouting<br>O: Aircraft Operator rerouting<br>M: ATFCM rerouting (not present) |
| 53 | Rerouted Flight State | char | 1 | If rerouting reason is not N:<br>N: No Match (rerouting has been invalidated)<br>T: Time out<br>E: Executed (rerouting has been done)<br>R: Rejected<br>V: Revoked<br>P: Produced (not present) |
| 54 | Runway Visual Range | num | 1 | e.g. 200 (most frequent)<br>In what distance unit? |
| 55 | Number Ignore Errors | num | 2 | (?) mostly 0 |
| 56 | ARC ADDR Source | char | 1 | ARC ADDR (?)<br>N:<br>C:<br>M: |
| 57 | ARC ADDR | char | 6 | If f #56 is not N: |
| 58 | IFPS Registiration Mark | char | | e.g. SXBIW<br>(22% missing) |
| 59 | Flight Type | char | 1 | S: Scheduled Air Service<br>N: Non-scheduled Air Transport Operation<br>G: General Aviation<br>X: Others<br>M: Military or Police Aircraft |
| 60 | Aircraft Equipment Description | char | | e.g. DGHIJRSWXYZ<br>means:<br>source URL: |

**CURRICULUM VITAE**

**Name Surname:** **Yiğit Bekir Kaya**

**Place and Date of Birth:** **Istanbul, May 1990**

**E-Mail:** **yigit353@gmail.com**

**EDUCATION:**

**B.Sc.:** Istanbul Technical University, Computer and Informatic Faculty, *Computer Engineering, 2013*

**B.Sc.:** Istanbul Technical University, Faculty of Aeronautics and Astronautics, *Aeronautics Engineering, 2014*

**M.Sc.:** Istanbul Technical University, Graduate School of Science Engineering and Technology, *Aeronautics and Astronautics Engineering Graduate Program, 2015*

**PROFESSIONAL EXPERIENCE AND REWARDS:**

- **Data Science Researcher,** Istanbul Techincal University Control and Avionics Laboratory, March 2013-May 2015
- **Co-op Student,** Turkish Airlines Technic, March 2013-November 2013
- **Database Manager,** Eduse Interactive, July 2012-March 2013
- **3D Game Programming Intern,** Sobee, August 2011-September 2011
- **Research Intern,** Istanbul Techincal University Control and Avionics Laboratory, June 2010-August 2010

- **Young Researcher of the Year,** TUBITAK, May 2008
- **Young Researcher 1st Place in Computer Science**, TUBITAK, May 2008
- **Boeing Gradute Student Grant**, Boeing, June 2014
- **111M167 Research Grant**, TUBITAK, May 2014
- **Computer Science Scholarship**, OBSS, January 2013
- **1st Place in German Reading Contest**, Istanbul Erkek Lisesi, May 2005

**PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:**

▪ **Kaya Y. B.**, Inalhan G., 2014, A Dynamic Bayesian Belief Network Approach for Modelling the ATM Network Delays. *International Conference on Research in Air Transportation*, May 26-30, 2014 Istanbul, Turkey.

**OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS :**

▪ Ure, N. K., **Kaya, Y. B.,** Inalhan G., 2011, The development of a Software and Hardware-in-the-Loop Test System for ITU-PSAT II nano satellite ADCS, *Aerospace Conference, IEEE*, 2011 Montana, USA.

▪ E. Koyuncu, E. Baskaya, M. Cihan, S. Isiksal, M. Fidanoglu, C. Akay, A. Cetin, B. Karadag, **Kaya Y.B.,** B. Karyot, C. Haciyev, I. Ozkol, M.O. Kaya, G. Inalhan., 2014, ITU-pSAT II: High-precision nanosatellite ADCS development project, *Recent Advances in Space Technologies (RAST), 2011 5th International Conference*