CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Transportation Sciences

Complex evaluation of NM B2B Flight Data accuracy

Komplexní vyhodnocení přesnosti letových dat NM B2B

Master's Thesis

Study programme: Technology in Transportation and Telecommunications
Study field: Air Traffic Control and Management

Thesis's supervisors: doc. Ing. Vít Fábera, Ph.D.
                        Mgr. David Sitenský
                        Mathieu Cousy

Bc. Marián Pánis

Prague 2023

CZECH TECHNICAL UNIVERSITY IN PRAGUE
Faculty of Transportation Sciences
Dean's office
Konviktská 20, 110 00 Prague 1, Czech Republic

K614................. **Department of Applied Informatics in Transportation**

MASTER'S THESIS ASSIGNMENT
(PROJECT, WORK OF ART)

Student's name and surname (including degrees):

Bc. Marián Pánis

Study programme (field/specialization) of the student:

master's degree – PL – Air Traffic Control and Management

Theme title (in Czech): **Komplexní vyhodnocení přesnosti letových dat NM B2B**

Theme title (in English): **Complex evaluation of NM B2B Flight data accuracy**

Guidelines for elaboration

During the elaboration of the master's thesis follow the outline below:

- Analyze data provided by Eurocontrol via NM B2B Service
- Define the source of radar data
- Define the source of ADS-B data and the way they were obtained
- Provide radar data and ADS-B data for statistical analysis
- Identify an appropriate statistical method for data analysis
- Create and design a tool for Flight data accuracy evaluation
- Compare results in differrent stages of flight
- Create a Web-Based system for the presentation of the results

Graphical work range: according to the supervisor's recommendation
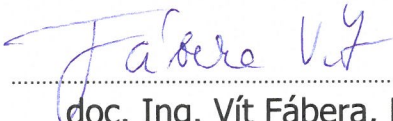

Accompanying report length: min. 55 pages (including figures, tables, graphs)

Bibliography: Ludvik Kulčák: Air Traffic Management, 2002

                             Matloff, Norman: Probability and Statistics for Data
                             Science, 2019

                             Pecinovský R.: Python - kompletní příručka jazyka pro
                             verzi 3.9, Grada, 2020

Master's thesis supervisor: **doc. Ing. Vít Fábera, Ph.D.**

                                         Mgr. David Sitenský

                                         Mathieu Cousy

Date of master's thesis assignment: **July 15, 2022**
(date of the first assignment of this work, that has be minimum of 10 months before the deadline
of the theses submission based on the standard duration of the study)

Date of master's thesis submission: **May 15, 2023**
a) date of first anticipated submission of the thesis based on the standard study duration
   and the recommended study time schedule
b) in case of postponing the submission of the thesis, next submission date results
   from the recommended time schedule

L. S.

....
doc. Ing. Vít Fábera, Ph.D.   prof. Ing. Ondřej Přibyl, Ph.D.
head of the Department   dean of the faculty
of Applied Informatics in Transportation

I confirm assumption of master's thesis assignment.

....
                             Bc. Marián Pánis
                          Student's name and signature

Prague .... July 15, 2022

Abstract

The Network Manager business-to-business services (NM B2B) allow other companies to use their data in their own systems. Before the actual implementation into the software solutions, it is crucial to recognize the nature and the precision of desired data. The emphasis is applied specifically to the NM B2B Flight data, which obtains information about the predicted flight trajectories. The thesis analyses the NM B2B Flight data accuracy in three different geographical areas of Europe while using multiple surveillance information sources for the complex evaluation. The areas with the highest accuracy were identified, as well as the disparities between individual sectors. The results of the work provide important insights into the behavior of NM B2B Flight data, which are crucial for further utilization. This study also presents the software module, which is capable of storing the data and evaluating it in real time. This feature enables the users to observe how the trajectory and the data precision evolve throughout the flight. Overall, the combination of the software solution and the accuracy analysis offers the benefits of utilizing the NM B2B data.

Keywords: accuracy, analysis, flight data, NM B2B, position, software module

Abstrakt

NM B2B služba sieťového manažéra (NM B2B) umožňuje externým spoločnostiam využívať ich dáta vo vlastných systémoch. Pred samotnou implementáciou softvérových riešení, je potrebné rozpoznať povahu a presnosť požadovaných údajov. Dôraz je kladený predovšetkým na NM B2B letové dáta, ktoré obsahujú informácie o predpokladaných letových trajektóriách. V diplomovej práci je analyzovaná presnosť letových dát NM B2B v troch rozdielnych oblastiach Európy s využitím viacerých zdrojov prehľadových informácii z dôvodu väčšej komplexnosti vyhodnotenia. Boli identifikované oblasti s najvyššou presnosťou, ako aj rozdiely medzi jednotlivými sektormi. Výsledky práce poskytujú dôležité poznatky o správaní letových dát NM B2B, ktoré sú kľúčové pre ich ďalšie využitie. V tejto štúdii je predstavený aj softvérový modul, ktorý je schopný ukladať dáta a vyhodnocovať ich v reálnom čase. Táto funkcia umožňuje používateľom sledovať, ako sa počas letu vyvíja trajektória a presnosť údajov. Celkovo kombinácia softvérového riešenia a analýzy presnosti predstavuje výhody využitia údajov NM B2B.

Kľúčové slová: analýza, letové dáta, NM B2B, poloha, presnosť, softvérový modul

Acknowledgment

Declaration

I Hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/200 Coll., the Copyright Act, as amended, in particular that the Czech Technical University in Prague has the right to conclude a license agreement on the utilization of this thesis as a school work under the provisions of Article 60 (1) of the Act.

In Prague, 14th of May 2023

…………………………………………...

Signature

Contents

List of Figures

List of Tables

List of abbreviations

ACK	Acknowledgement message
ADEP	Aerodrome of Departure
ADES	Aerodrome of Destination
ANSP	Air Navigation Services Provider
API	Arrival Planning Information
API	Application Programming Interface
ARO	Air Traffic Services Reporting Office
ATS	Air Traffic Service
BDS	Binary Data Services
CACD	Centralised Airspace and Capacity Database
CDM	Collaborative Decision Making
CPR	Correlated Position Report
CTFM	Current Tactical Flight Model
CTOT	Calculated Take-off Time
DCS	Data Collection Service
DPI	Departure Planning Information
EA	Enterprise Architect
EOBT	Estimated Off-Block Time
ETFMS	Enhanced Tactical Flow Management System
FIR	Flight Information Region
FL	Flight Level
FPL	Flight Plan
FSA	First System Activation
FTFM	Filled Tactical Flight Model
FDPA	Flight Data Processing Areas
FR24	Flight Radar 24
GAT	General Air Traffic
GNSS	Global Navigation Satellite Systems
HTTP	Hypertext Transfer Protocol
HTML	Hypertext Mark-up Language
ICAO	International Civil Aviation Organization
IFPS	Integrated initial flight plan processing system

IFPZ	IFPS Zone
IFPU	Initial Flight Plan Processing Unit
IFR	Instrument Flight Rules
INS	Inertial Navigation System
JSON	JavaScript Object Notation
MLAT	Multilateration system
NM	Network Manager
NM B2B	Network Manager Business-to-Business Service
NMOC	Network Manger Operations Centre
OS	OpenSky
PSR	Primary Surveillance Radar
REDIS	Remote Dictionary Server
RPL	Repetitive Flight Plan
RTFM	Regulated Tactical Flight Model
SSR	Secondary Surveillance Radar
TDOA	Time Difference of Arrival
TOBT	Target Off-block Time
UML	Unified Modelling Language

1 Introduction

The knowledge of the predicted aircraft position during flight is necessary not only for air traffic controllers but also for various companies and stakeholders closely working in the field of air traffic management. Monitoring the flight can be difficult, especially for companies that do not own radars or don't have access to radar data.

The European Organisation for the Safety of Air Navigation, Eurocontrol, as the leading figure in managing air traffic in European skies, has been named by the European Committee as the Network Manager. This organization has developed a service that allows other companies and stakeholders involved in air traffic management to use the air traffic data in real-time. It is being achieved through a specialized NM B2B Service, which allows the stakeholders to gain and update data in 3 main categories: Airspace Services, Flight Services, and Flow Services. In this thesis, we mainly focus on Flight Services which we use to gather the necessary information about the predicted aircraft position by Eurocontrol systems.

Data accuracy on a professional level is critical to achieving relevant results. While simulating the aircraft's position, we aim to ensure the accuracy is as good as possible. In this Master Thesis, we analyze the precision of the flight data shared via NM B2B Service by Eurocontrol. For the statistical analysis, we use several statistical methods such as descriptive, and inferential statistics combined with spatial analysis. It allows us to see the outcome in a comfortable and distinguishable way. During the study, we compare flight data from several sources. The main aim is to see how precise the NM B2B Flight data is compared to ADS-B, secondary surveillance radar, and MLAT data. The data from all the sources will be stored in a database for the following statistical analysis. Additionally, the analysis outcomes will be displayed in a Web-Based System that can compute the results and plot the graphs in real time. The whole system will be connected to the database that can be updated anytime.

Consequently, it is described how we obtained the surveillance data for this comparison. The sources used for the statistical analysis of the NM B2B Flight data accuracy are Flighradar24, OpenSky Network, and Secondary Surveillance Radar data from French Airspace. The matter of how they differ is discussed in the thesis's upcoming sections.

Also, for this work to be successful, it will be crucial to observe if there are any visible patterns of data accuracy on concrete flight routes or regions. Discovering geographical areas where the NM B2B Flight Data accuracy differs from the allowed values could also be beneficial.

1.1 Objectives of the work

The objectives of the work correspond to the Master Thesis Guidelines stated in the official Assignment of the Master Thesis. In general, we can divide all the objectives into three main subgroups. The division group reflects the actual nature of the matters within the group. The three main aspects are:

1) Air Traffic Control and Management
   o From the aviation point of view, we will discuss all the necessary details about the NM B2B Service, ADS-B, and Flight radar data. We will also examine how these data are gathered and point out their differences. Additionally, we will discuss what could be the cause of certain imprecisions and how to improve it in time.
2) Statistical Analysis
   o As we will be working with an enormous amount of data, it is inevitable to use them appropriately. We will try to discover any geographical areas, concrete route trajectories, or any other patterns that could affect the NM B2B Flight data accuracy.
3) Software engineering
   o In order to obtain and store data, it is necessary to create software tools that will be capable of doing so. The main aim is to create a Web-Based System that will compute the calculations from the data stored in the database. After the outcomes have been calculated, the tool will display them in the form of charts and graphs on a Webpage accessible from anywhere.

2 Air Traffic Control and Management

The aviation industry is the fastest-improving mode of transportation, with the number of flights steadily growing yearly. It all started in 1903 when the Wright brothers took off with an aircraft heavier than air. Since then, we could observe an incredible improvement in technology and the general support of this industry. More than 120 years past, and nowadays, we can distinguish more than tens of thousands of aircraft flying in the sky at any given moment. We can safely state that the aviation industry has opened the doors to the world for the vast majority of people. The fact that we can travel from one part of the world to a completely different part thousands of kilometres away, is just exciting, and still, it feels unreal.

As there are constantly more and more aircraft in the air, we must ensure that the airspace's capacity meets the demand. The capacity of the airspace may be influenced by many factors, such as the capacity of the airports, ATC staff, Airspace Management, ATC industrial actions, weather, and other environmental issues. It is necessary but often challenging to predict the capacity needed. Once an operator submits the flight plan, we must check if it meets the capacity of the airspace. If any problem is encountered, the flight plan must be changed. But in the case that any unexpected changes happen, it leads to delays that can vary in severity. That's why the predictions of the flight trajectories must be recalculated in case any deviation from the actual position occurs.

In this study, we are focussing on Air Traffic Control and Management above the specific but representative geographical area of Europe. The organization that has developed and maintained a tool for creating several models in relation to predicting and calculating the position of the aircraft is known as Eurocontrol.

2.1 Eurocontrol

Despite the immense advancements in the aviation industry, safety must be ensured in the best possible way. Therefore, the safety of passengers and everyone involved is the top priority with the highest level of importance. There exist several organizations on a global and regional level that try to enforce regulations and principles in order to achieve steady improvements while ensuring the safety of air traffic. One of such organizations is the European Organisation for the Safety of Air Navigation, also known as Eurocontrol, whose main aim is to support European aviation.

Eurocontrol plays a crucial role in supporting European aviation. It was founded in the year 1963, and until now, it has 43 member states (41 European states and two members from

outside Europe). As we can see in Figure 1, the whole Eurocontrol Area consists of almost all the countries from European Union as well as other additional countries, so it can eventually cover most of the geographical area of Europe. This significant fact allows European aviation to develop even further and faster. [1]
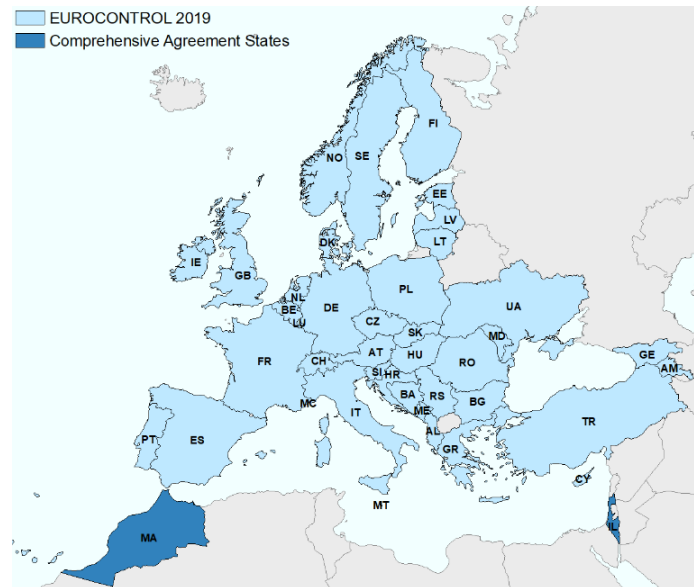
Figure 1: Eurocontrol Area [2]

To coordinate such a big and fragmented area, there must be a leading figure that could maintain steady development in an appropriate manner. Following this fact, the European Commission has stated by The Network Functions Implementing Regulation (EU) No 2019/123 all the tasks and responsibilities of the Network Manager, which is responsible for managing the whole area. Eurocontrol, an organization that supports European aviation, has been appointed as the Network Manager until the end of 2029. [3]

One of the main aims of Eurocontrol, in the role of Network Manager, is to improve the performance of European aviation. The body responsible for contributing network's operations is called Network Manager Operations Centre (NMOC). As mentioned before, to improve effectiveness and maintain the required level of safety, it is necessary that the demand, meaning the number of flights in a concrete area, meets the capacity of the requested airspace. This is exactly what NMOC is executing as a management of the traffic flows in 3 phases: strategic, pre-tactical, and tactical phase. Moreover, the scope of NMOC is more extensive as they also collect aeronautical and airspace data. Also, they improve and enable Flight planning through specialized systems and dispose of post-operational analysis, which allows further development. [4][5]

In the following chapters, we will focus on the two services provided by NMOC's network operations [4]:

- Flight Plan Filling and Management
- Network Manager Business-to-Business Web Services

2.1.1 Integrated Initial Flight Plan Processing System

An integrated initial flight plan processing system, developed by Eurocontrol, is a system that offers the service called centralized flight plan processing and distribution service. As the name implies, this system enables unified processing of all flight plans within the IFPS Zone (IFPZ). IFPS Zone is an area shown in Figure 2. In this zone, the NM is responsible for processing, validation, and distribution of flight plans to the relevant Air Traffic Control centres. In Figure 2, we can also see 2 areas that are marked yellow and named as FPM_Copy. FPM_Copy represents an area in which NM is not responsible for the validation or correction of flight plans, but all the flight plans (for aircraft flying through FPM_Copy) are being distributed to these areas as well. [6][7][8]
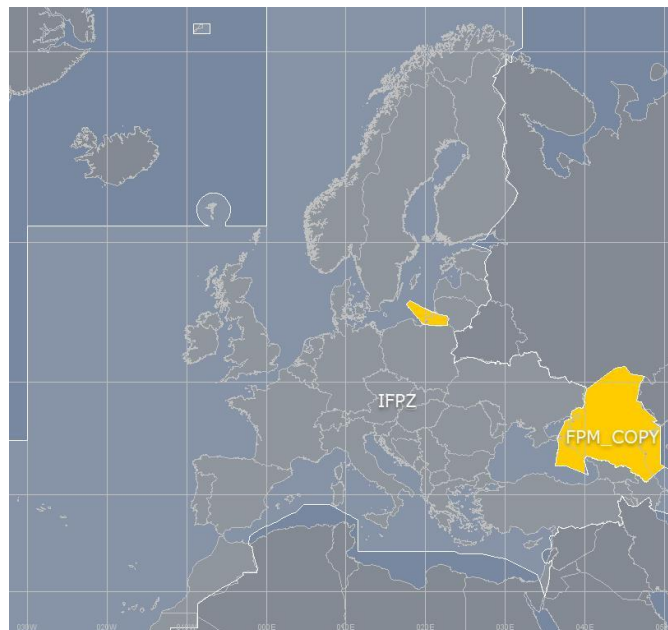
Figure 2: IFPS Zone + FPM Copy [6]

The IFPS involves two units responsible for the validation and distribution of flight plans [7]:

- IFPU1 – located in the capital city of Belgium, Brussels
- IFPU2 – located in France in the city called Brétigny

A flight plan must be created and submitted to both IFPUs for every flight flying under Instrument Flight Rules/General Air Traffic within the IFPZ. After the flight plan has been processed and acknowledged, meaning that the flight operator has received an ACK message, it is then distributed to ATC providers in affected areas. [9]

The diagram in Figure 3 depicts the flow of flight plan data between Aircraft operators, IFPS, ETFMS, and ATC.
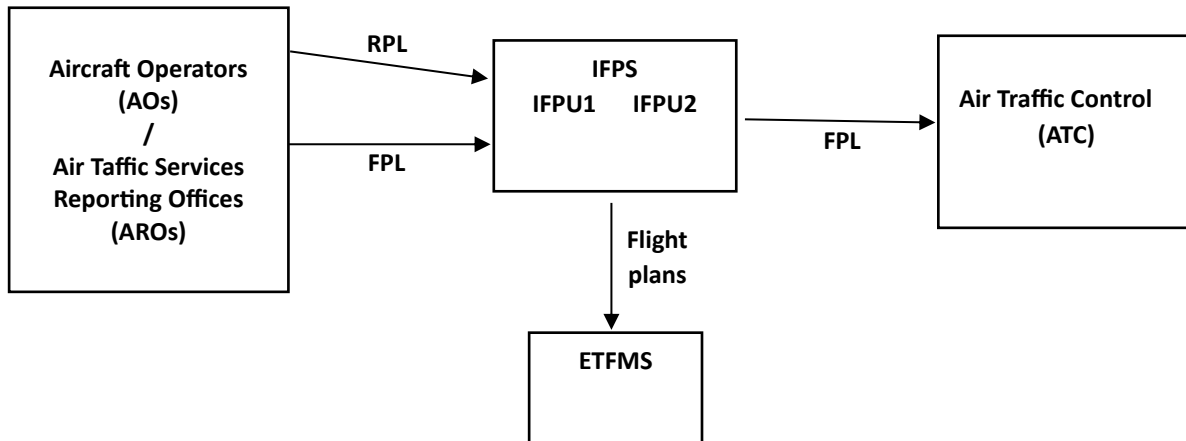
Figure 3: Flight plan data flow [10]

2.1.2 Enhanced Tactical Flow Management System

Enhanced Tactical Flow Management System, also known under the abbreviation ETFMS, is used for handling airspace, environment, and flight data from various sources to create an accurate image of the airspace situation within the IFPS Zone. As mentioned in the previous chapter about the IFPS system, the flight plan is accepted and distributed once it is assigned as acknowledged (ACK). In such cases, the IFPS sends all the flight plans to the relevant Air Navigation Service Providers (ANSPs) and other NM systems for further processing. [11]

Thus, the flight plan data are sent to the ETFMS under two conditions [12]:

- once the flight plan is acknowledged. The operator also receives the ACK message
- maximum of 20 hours before the Estimated Off-Block Time (EOBT)

If the Flight Plan has been sent to IFPS more than 20 hours before EOBT, it will not be present in the ETFMS system until 20 h prior to EOBT. If the FPL has been created less than 20 hours before EOBT, it will be sent to the ETFMS immediately after receiving the ACK message. This

way, it allows managing the whole flight from the phase of several hours pre-departure until the landing of the aircraft at the destination. [12]

Data received by ETFMS [13]:

- **FPL** - Flight plan data (AOs/IFPS)
    - o the Flight plan data sent from the IFPS system after processing the FPL data gathered from Aircraft operators (AOs)
- Aircraft position data and Flight data update (ATC/AO)
    - o **CPR** – Correlated Position Report
        - the Air Navigation Service Providers send the Correlated Position Report messages for airborne IFR/GAT flights at intervals of every 30 seconds. The ANSPs collect flight information from available surveillance sources and use them to calculate the actual 4D position of aircraft, which is then provided for further processing to the ETFMS system. [14]
    - o **API** – Arrival Planning Information
        - The airport systems send the Arrival Planning Information messages to provide more accurate information about the arrival routes, landing times, taxiing, holding, and information about flight delays, etc. This allows the ETFMS to modify the predicted trajectories and the times before the actual landing and arrival of the aircraft at the airport's gate.
    - o **APR** – Aircraft Position Report
        - For flights that do not originate from the NMOC area, the aircraft operator is responsible for sending an Aircraft Position Report message. The main aim of this message is to inform Eurocontrol systems about the actual position of the flight so that the ETFMS system can precisely predict when the aircraft will be arriving at the IFPZ at least 2 hours before the arrival. The information contained within this message includes the actual time of aircraft passing over the significant point and the estimation of the time of entering the specific area.
    - o **DPI** – Departure Planning Information
        - Information about the flight departure gathered by the airport systems is being sent via a Departure Planning Information message. Thanks to the knowledge of the taxi times, target off-block times, and take-off times, the ETFMS system can adjust the predictions of the flight trajectory models and create clearer forecasts for any IFR/GAT flights.

- **FSA** – First System Activation
  - ETFMS system receives the First System Activation message from ANSPs. As the name implies, they are being sent after the aircraft starts to be active in Flight Data Processing Areas (FDPA). For example, once the aircraft depart from the airport, the FSA message must be sent to the ETFMS system. Also, if the aircraft enters another FDPA, another FSA must be sent. It contains information about the actual take-off time, estimated time of arrival to the specific area as well as the concrete point of entry and the flight level that the aircraft is expected to have while entering.

- ETFMS system also receives other types of data that cover the information about the state of the airspace comprised with the Environment data that are being sent through the ADR systems.

The way how the data are gathered and then shared with the appropriate systems and service providers is explained in the following chapters.

2.1.3 Data Collection Service

To maintain a clear view of the airspace, it is necessary to collect real-time information about every flight flying through IFPZ. This is all done through the centralized Data Collection Service (DCS), which collects the flight plan data and flight update messages from IFPS and ETFMS. In this way, we can divide the collected data into two groups [15]:

1. Flight Plan Data
2. ATC and Airport Data

As the flight plan is created by the aircraft operator and the IFPS has assigned this flight with an ACK message, this data must be stored and shared with other systems, such as ANSPs and airports.

All the real-time data update messages, such as CPR, FSA, DPI, etc., must be stored as well to achieve the best possible situational awareness of the actual state of the airspace. It is a key factor for evaluating the demanded volume of the flights and regulating this amount to the allowed capacity of the sectors. [15]

Once all the data are successfully gathered, we can then distribute them to supply other ATC-related systems and companies with the actual picture of the airspace. In conclusion, we can say that the DCS system allows ETFMS to create and update the predictions of the flight. [15]

The DCS data are also being distributed to various stakeholders of Eurocontrol via NM B2B Service (see Chapter 2.1.5).

2.1.4 Flight Profile Models

The Flight Profile can be described as a model that contains all the information about the intended or actual flight trajectory. It contains all the information about the significant points on the route, flight levels, and the expected times when the aircraft should be flying over the significant points. Thanks to the vast number of specialized systems cooperating with each other, the Network Manager can create the Flight Profile Models and then distribute them to other systems.

Flight Trajectory Models used within NM B2B Service can be divided into three categories [12]:

1. Filled Tactical Flight Model
2. Regulated Tactical Flight Model
3. Current Tactical Flight Model

The Filled Tactical Flight Model (FTFM) represents the flight trajectory that is derived from the IFPS. It means that, in the beginning, it contains the information from the flight plan created by the AOs. But FTFM is also being updated by the CDM, Ready messages, and NM regulations before the departure. It can be inferred that the FTFM represents the most actual flight plan trajectory for all IFR/ GAT flights before the aircraft is airborne. [12]

The Regulated Tactical Flight Model (RTFM) contains information about the flight trajectory from IFPS but is not updated by any flight update messages or regulations. It is mainly used for statistical purposes as it allows us to compare the first filled flight plan to the flight plan that has been shifted. [12]

The Current Tactical Flight Model (CTFM) is a model that consists of flight profile data derived from the ETFMS. Thanks to this fact, the flight trajectory in the CTFM is being gradually updated by the flight update messages (FSA, CPR, DPI, etc.). Another condition for the recalculation of the CTFM point profile is the reception of changed flight plan, CTOT or ETA. As the name also implies, it allows the users to retrieve the actual flight data during the whole flight. [12]

The thresholds for the updating of the CTFM point profile upon receiving the CPR message [12]:

- Deviation of the vertical profile from the actual one by more than 400 ft
- Deviation in the lateral direction by more than 10 NM
- Time deviation of more than 1 minute

As the objective of this work is to analyze the accuracy of the NM B2B data, therefore we focus on the CTFM model. The main reason is that this model enables us to gather the most up-to-date data about every IFR/GAT flight passing through IFPZ in real-time.

2.1.5 NM B2B Service

Eurocontrol has developed a service to share air traffic related data with their stakeholders. It is all being done through system-to-system access, where other companies can use the data in their own platforms and applications. The service that makes it possible is called Network Manager Business-to-Business Service (NM B2B). It operates over a secure network infrastructure, and it is constantly available for its users. [16]

We can divide all the data being shared via NM B2B Service into three main categories:

- Airspace Services
- Flow Services
- Flight Services

The Airspace services allow users to retrieve information related to airspaces, routes, and aerodromes. All those data are present in the Centralised Airspace and Capacity Database (CACD), which is a primary source of data within the Airspace Service. This work required the use of this service as it was necessary to know the exact locations of all the designated points and the locations of all the aerodromes within IFPZ. [12] A designated point is a point used in aviation for describing an ATS route, flight trajectory of the aircraft, or other navigation purposes. [17] The way of obtaining such data is described in the Bachelor's Thesis [18].

The Flow services provide data related to air traffic flow and capacity management. Thanks to this service, we can gather current information about air traffic control sector capacity, restrictions, regulations, and airport capacity information. As the purpose of this work is to analyze the accuracy of the flight data rather than focusing on flow and capacity management, the Flow Services have not been used for the fulfillment of this Master´s thesis. [12]

The Flight services give us access to the flight plan data and flight data. The flight data is defined by the NM as the information about the flight in an execution phase. It is all achieved thanks to the ETFMS, which uses flight update messages in real-time to modify the initial flight

plan according to the current conditions and create accurate flight profile models. As we are mainly focused on the Flight Services area for the vast amount of the most necessary information needed to fulfill the work assignment, detailed information about this service can be found in a separate chapter. [12]

The way how to access the NM B2B Service has been described in the Bachelor's Thesis named "*The usage of flight data for the application showing the actual position and their trajectory*". [18]

2.1.5.1 Flight Services

The Flight Services provided by the NM B2B allow us to obtain the flight data about all the IFR/GAT flights passing through IFPZ. In figure 4, we can observe the way how to request flight data from NM B2B Service. It allows us to retrieve flight data in so-called Flight Lists, which are basically lists of all the data that suit the stated requirements.

We can obtain the Flight Lists in several ways [12]:

- Flight List by Aircraft Operator
    - This function lets us obtain the Flight Data of all the aircraft of the specific Aircraft Operator. To describe the AO, the NM B2B Service must receive a request with the ICAO code of the Aircraft Operator (e.g., Air France, ICAO code: AFR).
- Flight List by Aerodrome
    - The Flight List contains information about all the flights departing from or arriving at the specified aerodrome. To specify the desired Aerodrome, we will have to use the ICAO code of the aerodrome (e.g., the Václav Havel Airport Prague, ICAO code: LKPR).
- Flight List by Airspace
    - The Flight List holds the Flight Data about all the flights that are passing through desired Airspace. If at least one point on the Flight route stated in the Flight plan in IFPS/ETFMS is located within the required Airspace, the flight will be included in the Flight List. To define the whole airspace of a specific country, it is necessary to state the first two characters of the ICAO code of the country (e.g., Airspace of Slovakia, ICAO code: LZ).
- Flight List by Aircraft Registration

- o The Flight List will provide us with the Flight Data that belongs to the specified Aircraft Registration mark (e.g., GEUYC). Therefore, for the specified time window, we will be able to retrieve all the flights of the concrete aircraft.
- Flight List by Point
  - o The Flight List contains Flight Data about all the flights that on their route pass through the specified Point. The point must be described as a Designated Point with its exact ICAO code (e.g., BODAL).

For the purpose of this work, the Flight List that suits us the best is the Flight List by Airspace. It allows us to observe a substantial area using only one request. To define the request, we need to fill the get_flight_data function (Attachment 1) with four required attributes:

- Airspace
  - To define the whole airspace of France, we will use the first two characters of the ICAO code of the French airspace, which is LF.
- Start Time
  - Start Time attribute defines the starting value of the time window from which we wish to access the data. In our case, in Figure 4, we have defined the start time value to be NOW – 0.8 hours, which refers to 48 minutes.
- End Time
  - End Time attribute is used as a second parameter defining the time window. In Figure 4, we can see that it is represented by NOW + 0.2 hours minutes, which refers to 12 minutes.
- Flight Fields
  - Flight Fields are used for describing requested information about the flight. Through the NM B2B Service and especially the Flight Service, we distinguish several types of Flight Fields such as: calculated Off Block Time, actual Off Block Time, aircraft Type, estimated Take Off Time, calculated Take Off Time, actual Take Off Time, taxi Time, estimated Time Of Arrival, calculated Time Of Arrival, Actual Time of Arrival, requested Flight Level, aircraft Operator, delay, runway Visual Range, ICAO Route, route Length, FTFM Point Profile, RTFM Point Profile, CTFM Point Profile, SSR Code, filed Registration Mark, target Time Over Fix, fuel Consumption Indicator, alternate Aerodromes, etc.

- In Figure 4, we can see that in our case, we have chosen 4 attributes within the Flight Fields.

```
#NM B2B request
flight_data = b2b_request_response.get_flight_data(
    'LF',
    now - timedelta(hours=0.8),
    now + timedelta(hours=0.2),
    ['aircraftType', 'estimatedTimeOfArrival', 'delay','ctfmPointProfile'])
```

Figure 4: NM B2B Flight Data request

By running the request for Flight Data as shown in Figure 4, we receive information about the aircraft type, Estimated Time of Arrival (ETA), delay, and CTFM Point Profile for all the flights that are on their route present inside the French airspace within the specified time window.

The Aircraft Type is in the NM B2B Service, represented by the official ICAO code of the concrete type of the aircraft (e.g., B737, A320).

The Estimated Time of Arrival, also known as ETA, is the predicted time when the aircraft should be arriving at the aerodrome of destination. It is derived from the FTFM Point Profile. It means that the ETA is being updated until the departure of the aircraft. Once the aircraft is airborne, the ETA in the FTFM is not going to be shifted due to any changes encountered during the flight.

The Delay of the aircraft represents the time difference between the Calculated Take off Time and the Requested Take off Time.

The CTFM Point Profile contains all the information describing the whole trajectory of the aircraft. As the name implies, the flight data gathered from CTFM are being gradually updated due to the Flight Update Messages. As we are analysing the Flight Data accuracy provided via NM B2B Service, we can also say that we will observe the precision of the data from the CTFM profile distributed by the Flight Services. [12][13]

2.2 Surveillance information sources

The surveillance system is a system whose main role is to detect actual position together with other information about the aircraft and present them to the air traffic controllers. Therefore, it is an irreplaceable part of the aviation industry. In this chapter, we will explore some of the surveillance information sources used for the NM B2B flight data accuracy analysis. [19]

2.2.1 Secondary Surveillance Radar

The Secondary Surveillance Radar (SSR) belongs to the group of independent cooperative surveillance systems. They can calculate the aircraft's position but require direct cooperation with the aircraft's onboard systems. As shown in Figure 5, the system relies on two main components: The radar antenna and the onboard transponder. As the radar antenna rotates, it sends a signal to the aircraft at a frequency of 1030 MHz. Once the aircraft's onboard system processes this signal, the transponder sends the reply on a frequency of 1090 MHz. [20]
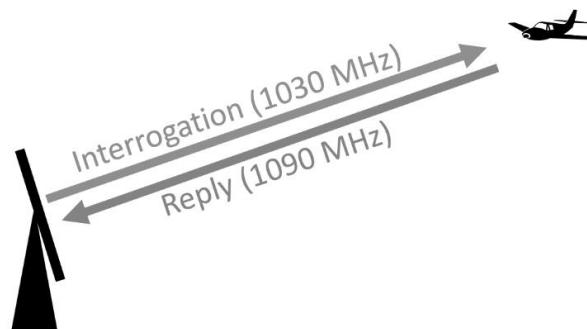
Figure 5: SSR principle [21]

When comparing SSR to the Primary Surveillance Radar (PSR), they brought a big advantage with the fact that we can obtain not only the aircraft's position but also other additional information. The reply from the SSR transponder can contain different types of information depending on the SSR mode type. [22]

- **SSR Mode A**
  - o When the transponder detects the SSR Mode A request, it will send back the information about the flight's identification consisting of four octal digits given by the ATC.
- **SSR Mode C**
  - o The transponder includes in the reply message the information about the current flight level. Flight level information is, in this case, calculated with respect to the 1013,25 hPa, which is known as the standard pressure.
- **SSR Mode S**
  - o The SSR Mode S is the most advanced one of the three mentioned.
  - o The main reason why it was created, was to enable the selection of specific aircraft and enable the request for bigger sets of information.
  - o It operates in two periods: The all-call period, and the roll-call period.

○ In the All-Call period, the goal of the SSR is to recognize all the aircraft within reach. To do so, it must send two requests. The first one is meant for aircraft that possess SSR mode S transponders. In this case, the reply message from the aircraft contains the 24-bit ICAO identification, which is unique for every aircraft. The second request in the All-Call period is intended to receive the message from the aircraft with a Mode A transponder. [22]

○ In the Roll-Call period, the SSR uses so-called Selective requesting. In the beginning part of the request message, the SSR Mode S includes the specific ICAO identification of the aircraft. The following part of the request states the group category of the required information. After the Mode S transponder recognizes its own ICAO identification included in the request, it sends back the reply with all the information declared in the initial request. [23]

The SSR Mode S transponders are directly working with the Binary Data Services (BDS) Registers. BDS Registers contain sets of information that are collected from systems onboard the aircraft. Every register is assigned its own code. The SSR declares the code of the specific BDS Register in the request message, and consequently, the transponder sends all the information within the register back to SSR. [24]

2.2.2 MLAT

Multilateration systems (MLAT) belong to the group of independent surveillance technologies as they do not require any message carrying the aircraft's position information from the onboard systems.

MLAT can compute the aircraft's position by itself using the method called Time Difference of Arrival (TDOA). On the ground, we have several stations that receive the signals emitted from the aircraft on the frequency 1090 MHz. [25]

To calculate the position of a given aircraft in three-dimensional navigation, we need at least 4 receivers (ground stations) and a Central Processing Station. As all receiving stations are based in different locations, the signal from the aircraft will arrive at the different receivers at different times. This is the crucial part of the computation. The receivers focus on the time when the signal arrived to them. These times are measured and sent directly to the Central Processing station for further calculations. If we take one pair of receiving stations with their TOAs, we can calculate the TDOA. Once we multiply the TDOA by the speed of propagation, we get the difference in the distance between Aircraft – 1st receiving station and the Aircraft – 2nd receiving station. Thanks to this fact, we can create a hyperboloid in a space. "A

hyperboloid is the set of all points in a space, the difference of whose distances from two fixed points (the foci) in the space is constant." [26] To be able to calculate the aircraft position, it is necessary to create at least 3 hyperboloids in space whilst the place of intersection indicates the position of the aircraft. The graphical representation of the MLAT system determining the position of the aircraft is shown in Figure 6. [25][26]
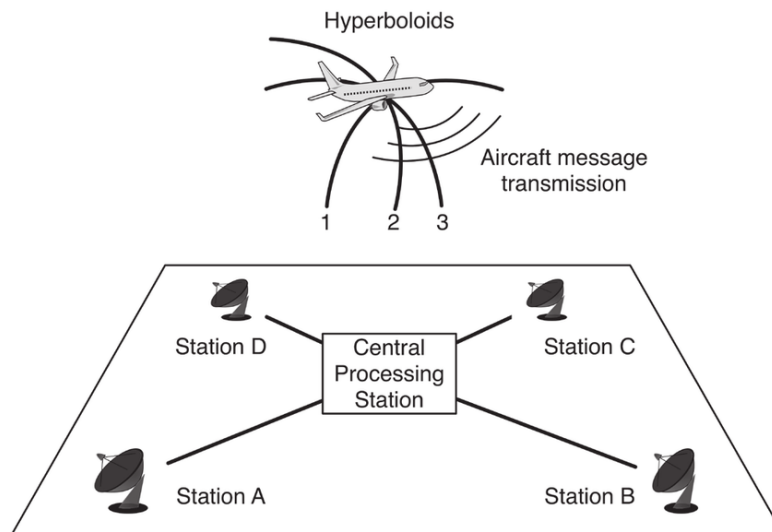
Figure 6: MLAT principle [26]

2.2.3 ADS-B

The Automatic Dependent Surveillance-Broadcast, also known by the abbreviation ADS-B, is a technology that depends on the onboard calculation of the aircraft's current position. Thanks to the systems on board the aircraft, that are used for computation of the position, such as the Global Navigation Satellite System (GNSS) and the Inertial Navigation System (INS), we can broadcast this position via ADS-B. As the name implies, it is broadcasted automatically in a specified time interval. Together with the position, we can derive from the ADS-B message other various types of information like the callsign. altitude, velocity, heading, etc. As it solely relies on the onboard systems, it belongs to the group of Dependent surveillance technologies. [28]

After the required information is collected and calculated by the systems in the aircraft, it is ready to be broadcasted. The ADS-B transmitter sends all the information to the whole area, meaning that anyone who receives the message might be able to decode it and work with the ADS-B data. This is the principle of how FlightRadar24 and OpenSky Network operate. For both associations, they have deployed thousands of receivers capable of reading the ADS-B messages, which finally enables them to provide it to their web servers and, from there, to the

application displaying the actual position and information about the aircraft throughout the world. [29][32]
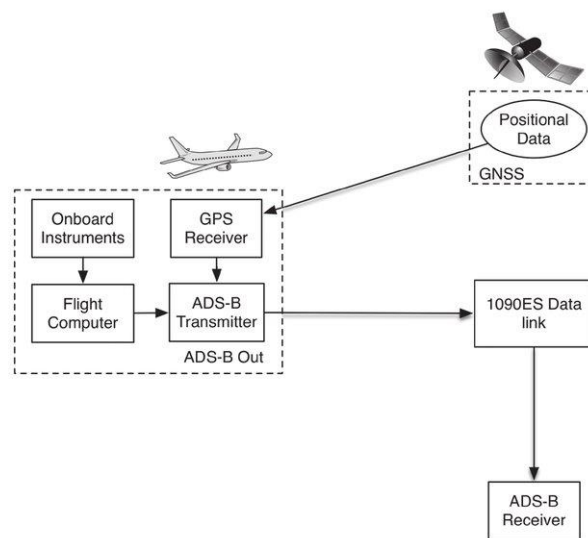
Figure 7: ADS-B technology principle [28]

The whole principle of how the ADS-B works is described in Figure 7. It can be transmitted via the 1090 MHz Extended Squitter, which is a 112 bits message that carries all the information. [29]

2.2.4 OpenSky Network

The OpenSky Network is a Swiss non-profit association that collects flight data primarily from ADS-B and Mode-S receivers and provides them to their users for research and non-commercial purposes. The data from the aircraft onboard transponders are collected by the ADS-B receivers. Afterwards, they are sent voluntarily to the accessible Database on OS servers which makes it possible to retrieve live or historical data from anywhere. [30]

The coverage of these receivers is smaller in comparison to FR24, but it is still very sufficient for our study as it provides a significant amount of data. The OpenSky Network plays an important role in this Master Thesis as it is used as another reference source of flight data for the final evaluation. The way how we access the real-time data from OpenSky Network is explained in Chapter 3. Eventually, we might be able to observe any significant differences when also compared to the FlightRadar24 data.

Both OpenSky and Flightradar24 data with used technologies and their accuracy depend on several factors such as the quality of data, coverage of the receivers in a region. Also, the

quality and the age of the receivers can affect the precision of the position information we obtain using these Services. As stated in the previous research about the accuracy of OpenSky positional data [30][31] the mean value of horizontal deviation is 295,55 m. The precision of the OpenSky data will be better than 900 m with the probability level of 95%. As all of the systems bring certain errors, we have to state, that it is indeed accurate enough for recognising the behaviour and accuracy of the NM B2B Flight Data. [31]

2.2.5 Flight Radar 24

One of the sources of Flight Data used for the analysis of the NM B2B flight data accuracy is Flight Radar 24 (FR24). This Swedish company offers a high-quality service that allows users to track most flights throughout the whole world. To enable this feature, it uses the combination of various data sources such as previously mentioned ADS-B, MLAT, and Radar data. [33]

The primary source of information is the ADS-B system, as the whole network coverage of ADS-B receivers reaches the amount of more than 35 thousand. In the areas outside the coverage of these receivers or in the case of aircraft that are not equipped with the required ADS-B on-board system, the FR24 uses the MLAT and radar data. ADS-B coverage within the area of Europe exceeds 80%, which is sufficient for our analysis, as we tend to analyze the area not outside the IFPZ. [32]

Thanks to the FR24, we can proceed with the analysis as their service allows us to retrieve real-time information about the flights, such as the callsign, position, or flight level. All the mentioned parameters are consequently used for the NM B2B Flight Data accuracy analysis. As the data in FlightRadar24 are not based on Flight Plans but on actual surveillance data, it becomes very interesting to observe the difference in positional information with the NM B2B Service. The FlightRadar24 data also bring the errors up to hundred metres, similarly like the OpenSky data. [33]

3 Data analysis

Data analysis is a process that involves comprehensive techniques that help us understand the data we are working with. Flight data derived from NM B2B Service have great potential in the air traffic management industry. As for now, there is no known document that would depict the precision of these data. The accuracy of the data we use within our systems is crucial for the determination of the concrete application, which will be working with the specific data. To point out the overall accuracy, we will use various techniques to get the most relevant results. [34]

In general, the data analysis used in this work comprises the following steps [35]:

- Defining the evaluation criteria
- Collecting, exploring, and pre-processing data
- Statistical analysis
- Validating results

3.1 Evaluation criteria

The accuracy of the NM B2B Flight data is going to be evaluated in several ways. We will use different approaches that will help us deeply understand the data. To achieve a relevant and effective analysis, we must first define what evaluation criteria are being considered. It includes the statement of what attributes are going to be used in the analysis as well as the way they are going to be classified and eventually evaluated.

The evaluation criteria used in this work can be described as follows:

- Positional accuracy
    - The positional accuracy as an evaluation criterion will highlight the precision the NM B2B Flight data when compared to positional data from multiple surveillance sources.
- Altitude accuracy
    - The altitude accuracy will depict the differences in terms of altitude information from NM B2B Service and surveillance sources.

The data fields derived from NM B2B Service used in the accuracy evaluation:

- The data fields derived from NM B2B Service used in the accuracy evaluation:
    - Aircraft position
    - Flight Level
- The accuracy will be assessed in different ways. We determine and focus on the following:
    - Flight Information Region (FIR)
    - Geographical area
    - Flight Route (defined by the Aerodrome of Departure and the Aerodrome of Destination)
    - Time period (defined by the 24-hour intervals)
    - Flight (defined by the concrete callsign)

3.2 Collection and exploration of data

This chapter is focused on the way of gathering flight data from various sources for the upcoming evaluation. NM B2B Flight data contain so much information that we must decide and clearly determine whether to store only those data that are relevant for further analysis. Besides NM B2B Flight data, it is also required to collect and pre-process data from surveillance information sources:

- FlightRadar24
- OpenSky Network
- Secondary Surveillance Radar data from French airspace

The surveillance information sources will be used as the reference for the NM B2B Flight Data evaluation. As for the FlightRadar24 and OpenSky Network, they both contain information that are primarily based in the calculations from the onboard system and received through the ADS-B system. Comparing NM B2B Flight data also to the data from the secondary surveillance radars will give us a better overview of the whole system's accuracy. We can think of it in the way that more sources of data using different approaches could lead to more complex analysis. The collection of mentioned information is being achieved by using the Python 3 programming language, which allows us to store the data in a database system. Specifically, in this study, we use the Python programming language for accessing the SQL and Couchbase databases.

In Table 1, we can see which surveillance information sources were used for the evaluation of the NM B2B flight data accuracy in specified geographical areas.

Table 1: The utilisation of surveillance information sources in specified regions

Geographical region	Surveillance information source		
	OS	FR24	SSR
West IFPZ	✓	✓	✓
Central IFPZ	✓	✓	✗
South-East IFPZ	✓	✓	✗

Consequently, the areas on which we performed the analysis are displayed on a map in Figure 8. The specification of three focus areas was made on behalf of achieving an appropriate overview of how the data accuracy differs in different regions.
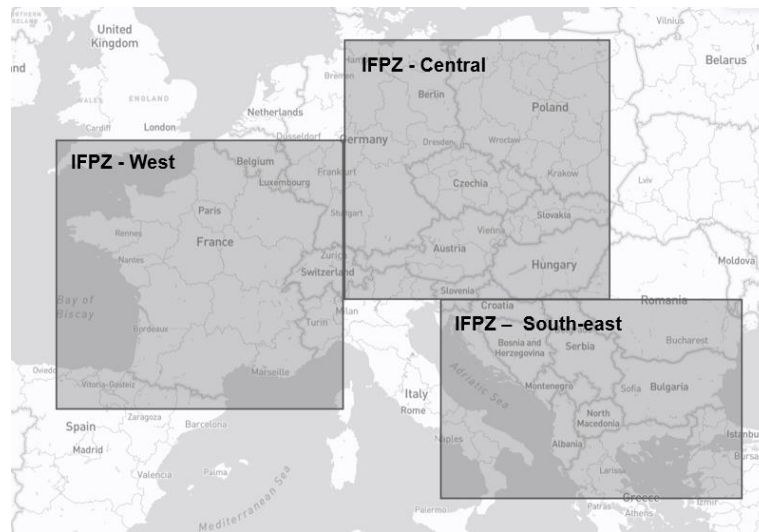
Figure 8: Focus areas

3.2.1 NM B2B

As mentioned in Chapter 2, we will be gathering data using the command function that returns the Flight List by Airspace. From all the possible ways how to retrieve data from NM B2B Service, we have chosen to use the Flight List by Airspace because of one particular reason. In this way, we can cover a much wider area instead of just information gathered for the specific but smaller area, for example, from the Flight List by Aerodrome. The most important information used for the accuracy evaluation is involved within the CTFM Point Profile, which contains all the necessary information used for calculating the position of the aircraft. We are talking about the trajectory of the aircraft described in the form of all the points on the flight route and times at which the aircraft is supposed to fly over the specific set of points accompanied by the flight level maintained on the route segments.

The flight data retrieved from the NM B2B Service are being received in JSON format. An example of this data format is shown in Figure 9. As this response from NM B2B also contains information that is not subject to our analysis, we will need to pre-process it. This process involves the filtration of the data in a way that will allow us to store only that information relevant to our analysis.

```
{"flight":{
        "flightId": {
            "id": "AT01522847",
            "keys": {
                "aircraftId": "EJU3642",
                "aerodromeOfDeparture": "LKPR",
                "aerodromeOfDestination": "LEBL",
                "estimatedOffBlockTime": "2021-07-19 09: 15"}},
        "aircraftType": "A319",
        "estimatedTimeOfArrival": "2021-07-19 11: 20",
        "ctfmPointProfile": [{{
                "timeOver": "2021-07-19 09: 34: 15",
                "flightLevel": {
                    "unit": "F",
                    "level": 240,
                    "point": {
                        "pointId": "DOBEN"},
                    "flightPlanPoint": "True"},
                "timeOver": "2021-07-19 10: 03: 21",
                "flightLevel": {
                    "unit": "F",
                    "level": 390}},
            "point": {
                "nonPublishedPoint-GeoPoint": {
                    "position": {
                        "latitude": {
                            "angle": "463527",
                            "side": "NORTH"
                        },
                        "longitude": {
                            "angle": "0110337",
                            "side": "EAST"}}}}}}}
```

Figure 9: Sample of NM B2B data in JSON format

From Figure 9, it is clearly visible that we won't need all the gathered data. The concrete data fields that are being stored for further analysis are:

- Flight ID
  - {"flight": {"flightId": {"id": "AT01522847"}}}
  - A unique identifier of a flight plan assigned by the IFPS system (e.g., AT01522847).
- Aircraft ID
  - {"flight": {"flightId": {"keys": {"aircraftId": "EJU3642"}}}}
  - The callsign of the aircraft, which can be represented by the registration mark assigned by the state of registration (OKMAA) or by the designator of the aircraft operator accompanied by the number of the flight (EJU3642).
- Aerodrome of Departure
  - {"flight": {"flightId": {"keys": {"aerodromeOfDeparture": "LKPR"}}}}
  - ICAO designator of the aerodrome of departure (LKPR)
- Aerodrome of Destination
  - {"flight": {"flightId": {"keys": {"aerodromeOfDestination": "LEBL"}}}}

- o ICAO designator of the aerodrome of destination (LEBL)
- Time Over
    - o {"flight": {"ctfmPointProfile": {"timeOver": "2021-07-19 09: 34: 15"}}}
    - o The predicted time at which the aircraft is supposed to be flying over the significant point.
    - o The attribute of Time Over can be found within the CTFM Point Profile, which represents the most actual trajectory of the flight.
- Flight Level
    - o {"flight": {"ctfmPointProfile": {" flightLevel ": {"unit": "F","level": 240}}}}
    - o The expected Flight Level that the aircraft is supposed to maintain while passing through the significant point.
- Position
    - o Point ID
        - ▪ {"flight": {"ctfmPointProfile": {" pointId ": " DOBEN "}}}
        - ▪ The significant point on the flight route that is represented by the ICAO designator. While the position is described by such a point ID, it is not accompanied by any other position information, meaning the geographical coordinates. Because of that, we have created a Points.JSON file (Attachment 1) which contains information about the geographical coordinates of all the designated points as well as the aerodromes.
    - o nonPublishedPoint – GeoPoint
        - ▪ {"flight": {"ctfmPointProfile": {"point": {"nonPublishedPoint-GeoPoint": {"position": {"latitude": {"angle": "463527","side": "NORTH"},"longitude": {"angle": "0110337","side": "EAST"}}}}}}
        - ▪ In case the point on the flight route is not represented by the ICAO designator, it is represented by the latitude and longitude coordinates. Both coordinates are described by the angle and the side.

3.2.2 FR24

To evaluate the accuracy of NM B2B Flight data, we need to have information from other sources that will be used as a reference. Flight Radar 24 collects an enormous amount of data about the aircraft throughout the whole world. This information is being shared for non-commercial purposes through the Application Programming Interface (API). This tool enables us to access the actual data of FR24 using the API as a Python library. After the installation,

we can obtain numerous information such as the list of the airports, airlines, zones and most importantly, the list of the actual flights. [32][36]

Consequently, in Figure 10, we filter only those data that are relevant for our analysis. We collect almost the same data as previously from NM B2B Service. For the upcoming evaluation, it is necessary to obtain information about the callsign and position of the aircraft in the form of latitude and longitude coordinates, as well as the actual flight level. After requesting the data in this way, we get the information about the flights obtained by the API at the actual time of the request. Thus, the time attribute will be close to the actual time.

```python
for f in flights:
    callsign_fr.append(str(f.callsign))
    lat_fr.append(f.latitude)
    lon_fr.append(f.longitude)
    time_fr.append(datetime.fromtimestamp(f.time))
    fl_level_fr.append(f.altitude)
```

Figure 10: Filtering FR24 data

Afterward, the specific data fields are loaded into separate lists. It allows us to keep track of the concrete flight using the indexes of the elements within the lists.

As there are certain limitations for the amount of data gathered within specified time window, we decided to use the source of Flight Radar 24 data for the specified parts of IFPZ. Another reason for doing so was to observe any difference in the accuracy evaluation of NM B2B Flight data using the other sources like OpenSky and Secondary Surveillance Sadar in this concrete area.

3.2.3 OpenSky Network

Another source of flight data used in the NM B2B analysis is the OpenSky Network which uses similar ways of obtaining information about the aircraft's position. The way how to get the OpenSky Network data is shown in Figure 11. [37]

```
    lat_min = 35
    lon_min = -9
    lat_max = 55.28450035530949
    lon_max = 14
    username =
    password =
    url_page = "https://"+username+":"+password+"@opensky-network.org/api/states/all?
lamin="+str(lat_min)+"&lomin="+str(lon_min)+"&lamax="+str(lat_max)+"&lomax="+str(lon_max)
```

Figure 11: OpenSky Data request [37]

Access to this data is done through the REST API. It is the form of an API that uses HTTP requests to obtain the data. In Figure 11, we can see that we try to access the data of the flights that are present in the specified polygon, which is represented by the maximum and minimum latitude and longitude. By a single request, we receive the actual flight data corresponding to the airspace situation at the moment of making the request. [37]

The parameters of the flights that can be obtained through the OpenSky Network are numerous. For the purposes of the work, we retrieve only the data regarding the Callsign, Latitude, Longitude, and Altitude.

The OpenSky flight data were collected for the analysis of the NM B2B Flight data precision in the areas of the western, central and south-eastern part of IFPZ.

3.2.4 SSR

The approach for accessing the Flight data gathered from the secondary surveillance radars differs from obtaining data from the previous sources. The usage of SSR data on behalf of this study was granted by the ACHIL laboratory at the ENAC University in Toulouse. The SSR data are stored in an internal database that utilizes the Couchbase platform. Couchbase offers a NoSQL database which has many advantages. The format of the data tables does not have to be necessarily fixed. It is possible to adjust the format of stored data with intention to adapting to changing requirements.

Thanks to the ACHIL laboratory we are able to access the SSR data in a JSON based format through the Couchbase platform. The format in which we obtain the data is shown in Figure 12. It is apparent that the data within are already filtered only for those required by this study. From this database we obtain the information such as the callsign, day, the time of receiving the position data followed by the actual aircraft position described by the geographical coordinates of latitude and longitude.

```
{
    "rstna": {
        "callsign": "RYR68AB",
        "day": "2023-03-25",
        "trajectory": {
            "15: 04: 45": {
                "FL": 240,
                "lat": 46.04555,
                "lon": 2.38766
            }
        }
    }
}
```

Figure 12: SSR Couchbase database structure

This way we can obtain the SSR data from all the Secondary Surveillance Radars within the French Airspace. This implies that for the accuracy evaluation of NM B2B Flight Data in the western part of IFPZ and specifically in the region of the French Airspace, we used all of the available source such as FR24, OS and SSR.

3.2.5 Database

After the collection of the relevant data has been performed, we must store them in a suitable way. That's why several database tables have been created and filled with the necessary information needed for the actual evaluation. The language of the database structure has been determined to be a Structured Query Language (SQL). SQL databases have many advantages as it enables us to store data in structured entities, which makes them easy to read and analyse. With the intention of comparing the NM B2B Flight Data with multiple surveillance information sources, we will create separate SQL tables for every reference source, as well as for the observing geographical. In the upcoming examples, we will refer to OS and FR24, but it is important to state that the structure of tables will be the same for all the database tables.

For every surveillance information source and geographical area, we will have 2 database tables:

- Position
- Flight Level

As the names of the database tables imply, the first will contain information that is associated with the position analysis. This reflects that all the information regarding the mutual position between the same aircraft obtained from two different sources will be stored within this

concrete database table. The second database table contains the information regarding the flight level information from both sources (NM B2B and the chosen reference source).

3.2.5.1 Position – database table

The database table that carries the name Position consists of several attributes, which in a way, refer to the relationship between information gathered from two different sources. Figure 13 represents the structure of the database table Position_os_c, its individual attributes also with their corresponding data type. In this example, also shown in Figure 13, we can observe that most of the attribute names contain "os" in their name. This is supposed to represent the fact that the database stores and compares data from OpenSky and NM B2B. In such way, we may explore the "fr" or "ssr" element in the other database tables.

Position_OS_C	
<u>ID</u>	
Callsign	(text)
nm_time	(real)
os_time	(real)
nm_lat	(real)
nm_lon	(real)
os_lat	(real)
os_lon	(real)
Infront	(text)
position_os	(text)
distance_os	(real)
perpendicular_d	(real)
offset_d	(real)
Adep	(text)
Ades	(text)
Category	(text)
Fir	(text)

Figure 13: Position databse table

In this work, we have created Position database tables separating different reference sources and geographical areas in this way:

- Position_FR_W – Flightradar24 used as a reference in western part of IFPZ.
- Position_OS_C – OpenSky Network used as a reference in central part of IFPZ.

- Position_OS_SE – OpenSky Network used as a reference in south-eastern part of IFPZ.

- Position_SSR – Secondary Surveillance Radar data used as a reference.

At this point, we will proceed to the definition of all the attributes within the Position database table.

ID

The ID within the database table is used as a Primary Key. Every time we load a single row of data to the database, we don't need to refer to the same aircraft. That is the reason why we have chosen to assign a random ID. It clearly states that every row entity has the same severity, and we differ between all of the data from different or the same aircraft in the same manner.

Callsign

The callsign of the aircraft must be present in the database for one particular reason. With the intention of analysing the precision assigned to concrete aircraft, we will use to filter by the same callsign. This allows us to see the differences in data accuracy between concrete aircraft.

NM_time

NM_time attribute represents the timestamp used for the position calculation from the NM B2B Flight data. As the CTFM Point profile obtained from NM B2B Service contains the whole trajectory with the Time over attributes, we can use it for the position prediction at a particular time. To compare two pieces of information on the position from two different sources, we need to have the timestamps of both pieces of information aligned. Therefore, the NM_time attribute must be the same as the OS_time attribute.

OS_time

After the retrieval of the data from the reference source, in this case, meaning OpenSky Network; we obtain the information about the timestamp. This information represents the concrete time, when the reference source received the information about the specific aircraft. For this reason, we must use the same time in the position calculation from NM B2B data.

NM_lat

The latitude is calculated from NM B2B Flight data at a particular time. As the CTFM Point Profile describes the whole trajectory of the flight, we can interpolate and find any position of the aircraft at a given time. In this work, we use the type of linear interpolation which has proven to be sufficient for desired analysis. Once we have the information about two data points (Point 1 and Point 2) with the corresponding time_over attributes (T1 and T2), we can estimate the position of the aircraft at time T.

The linear interpolation in this master thesis is used in the following way:

- At first, we need to state the relative position of time T between two time intervals T1 and T2 [18]:

$$t = \frac{T - T_1}{T_2 - T_1} \tag{1}$$

  Where the value of t is from the interval (0,1). If the aircraft Is located at Point 1, the value of t will be 0 and consequently 1 if it is situated at Point 2. Time T1 and T2 are obtained from NM B2B Service, while time T is the information from the reference source (OS).

- Consequently, we can use the information about the relative position of time T to calculate the position of the aircraft at this time. To calculate the latitude coordinate, we use the following formula [18]:

$$X = (1 - t) X_1 + t X_2 \tag{2}$$

  Based on the coordinates of Point 1 (X_1, Y_1) and Point 2 (X_2, Y_2) we calculate the X coordinate at the specific time T. The X coordinate represents the latitude value of the aircraft that is comprised within the attribute NM_lat in the database table.

NM_lon

The longitude attribute based on the NM B2B Flight data is calculated in the same manner as previously shown with the calculation of latitude coordinate.

The formula used for the calculation of the longitude from CTFM Point Profile at a given time T (OpenSky timestamp) [18]:

$$Y = (1 - t) Y_1 + t Y_2 \tag{3}$$

OS_lat

The latitude coordinate is obtained from OpenSky API in a decimal form. This information is directly used for the calculation of the distance between NM B2B and OS position information.

OS_lon

The longitude information representing the position of the aircraft gathered from OpenSky API is expressed in decimal form. In this way, we do not need to pre-process it further, and we may use it in calculation straightaway.

Distance_os

The distance_os attribute presents the actual distance in nautical miles between two points on Earth:

- Point 1 [nm_lat, nm_lon] – the position of this point has been calculated through a linear interpolation of trajectory from NM B2B Flight data at a given time.
- Point 2 [os_lat, os_lon] – The position of this point has been obtained from the surveillance information source (in this case OS).

The great-circle distance between two points on the Earth is calculated using the Haversine formula [38]:

$$a = \sin\left(\frac{\Delta lat}{2}\right)^2 + \cos(nm\_lat) \cos(fr\_lat) \sin\left(\frac{\Delta lon}{2}\right)^2 \tag{4}$$

$$c = 2 \; atan2\left(\sqrt{a}, \sqrt{1 - a}\right) \tag{5}$$

$$distance_{fr} = \frac{R \, c}{1.852} \tag{6}$$

The Haversine formula assumes that the Earth is a perfect sphere, so we must admit some error in the final results. In this concrete equation, we use the variable R as an Earth's radius, which is 6 371 km. With the intention of getting the result in nautical miles rather than kilometres, we must divide the result by 1,852. [38]

Perpendicular_d

The distance_os contains information about the overall distance between two points. But in the further analysis we must see if the aircraft position obtained from surveillance information

source is also misplaced from the actual predicted route by NM B2B Service. It is important to recognize whether the value of distance_os has been caused by the perpendicular component or there has been some delay in a position along the line.

The attribute perpendicular_d contains the information about the perpendicular distance between the point and the line. To calculate the perpendicular distance of the point (OS) to the line formed by two points from NM B2B we use this formula [39]:

$$perpendicular\_d = \frac{|(os_1 - nm_1) \times (fr_1 - nm_2)|}{|nm_2 - nm_1|} \tag{7}$$

Where os_1 represents the point obtained from OpenSky and nm_1 and nm_2 are the points obtained from NM B2B. [39]

Offset_d

Another element that we must take into account is the information about the offset distance. By the offset distance we mean the distance between the points os_1 and nm_1 along the line which is formed by points nm_1 and nm_2.

As we previously calculated the overall distance between the points os_1 and nm_1, as well as the perpendicular distance between them, we can use this information for the calculation of the offset distance. As the distance values don't necessarily exceed 20 NM, it is sufficient for our needs to use the Pythagorean theorem:

$$offset_d = \sqrt{distance\_os^2 - perpendicular\_d^2} \tag{8}$$

Even though the usage of the Cartesian coordinates and the Pythagorean theorem doesn't consider the curvature of the Earth's surface, the difference of the error in comparison with the spherical equations is negligible. As the Earth is not the perfect sphere and the distance values are relatively small in the magnitude, it is reasonable to state that these equations are appropriate for the purpose of this study. [40]

Infront

The information stored within the attribute Infront states which of the data source, whether NM B2B or OS are in front or delayed on their path. The knowledge of such fact can help us understand and analyse if the NM B2B data are delayed or the predictions are further in time when comparing to the real position stated by the Surveillance Information sources.

To figure out which of the positions are more in front on the flight path, we will use the haversine formula. We have to calculate two distance values:

- Distance between points nm_1 and nm_2
  - The Point nm_1 is in this case represented by the calculated position based on NM B2B Service. The calculation is done for the timestamp of the Surveillance information source.
  - The Point nm_2 is the following point on the flight trajectory defined by the NM B2B Service
- Distance between points os_1 and nm_2
  - The point os_1 is defined as the actual position of the aircraft from the Surveillance information source (OS)
  - The point nm_2 is the following point on the flight trajectory.

After we have found out which of the mentioned distances is shorter, we can assign to the attribute Infront within the Database table that either NM B2B or OS are closer to the consequent point of the flight trajectory.

Position_os

In the attribute Position_os we present whether the position obtained by the Surveillance information source (OS) lies on the left or on the right side of the flight trajectory defined by the NM B2B Service. To determine so, we will use the cross product of the vectors. The first vector A is formed by the points nm_1 and os_1 and the second vector B by the points nm_1 and nm_2. The cross product of two vectors will result in the third vector C which is perpendicular to the plane containing first two vectors. The mathematical method used in the calculation is as follows [41]:

$$A \times B = C \tag{9}$$

$$(os_1 - nm_1) \times (nm_2 - nm_1) = C \tag{10}$$

Where C is the resulting vector which consists of three components (x,y,z). But to determine the position of the point os_1 relative to the line (nm_1 and nm_2) it is sufficient to calculate only the x component of the vector C.

$$c_x = a_y * b_z - b_y * a_z \tag{11}$$

Where a_y, b_z, b_y, a_z are the components of the vectors A and B. Eventually the position of the point from OS is determined by the sign of the x component of the vector C. If the x component

of the cross product is positive, then the OS position is located to the left of the line created by NM B2B data. If it is zero, it means they are both on the same line. And if the x component of the vector C is negative, it states that the OS position is to the right of the line. Consequently, we can use this information in analysing whether the position of NM B2B in relation to the Surveillance information source deviates to which side in some geographical areas or routes. [41]

Adep

We store the information about the Aerodrome of Departure because it allows us to proceed with further analysis of a given route defined by the aerodrome of departure and destination.

Ades

The information about the Aerodrome of Destination is again stored as an ICAO identification of the airport. For the additional analysis of the concrete route, it is necessary to include this information within the database as well.

Category

The information about the category of the flight differs in two ways:

- D – Domestic flight
  - The flight which on a whole flight trajectory doesn't leave an airspace of a specific country is considered to be a domestic flight. Such flight will be assigned a letter D in the Category attribute within the database.
- I – International flight
  - The flight which, on its flight route, leaves one airspace and enters the airspace of another country is considered to be an international flight. In this case, we will assign the category attribute with the letter I.

This valuable information gives us the option to analyze whether there are any accuracy differences in NM B2B Flight data between the Domestic and International flights. As domestic flights are not leaving one airspace on the entire flight trajectory, it could be less affected by the regulations, which could lead to easier predictions of the flight position as well as better accuracy.

Fir

Flight information regions, also known by the abbreviation FIR, are the regions of the airspace strictly defined by their dimensions. We have chosen to include this information within the database as we can then compare the precision of NM B2B Flight data in different FIRs respectively.

3.2.5.2 Flight Level – database table

The database table that carries the name Flight Level is meant to store data for the analysis of the vertical profile of the flight. It is important to load such data within the database to identify any patterns in areas, routes or specific time intervals where the accuracy differs from the normal condition.

The attributes used in this database table are similar to those used in the Position database table as we tend to maintain the same approach of the upcoming analysis. The structure of the database table is shown in Figure 14.

FlightLevel_FR_W	
<u>ID</u>	
Callsign	(text)
Time	(text)
FL_nm	(real)
FL_fr	(real)
FL_dif	(real)
nm_lat	(real)
nm_lon	(real)
Adep	(text)
Ades	(text)
Category	(text)
Fir	(text)

Figure 14: Flight level database table

As we can observe in Figure 14, the only entities within the database table that are different from the previous one are:

- **FL_nm**
  - The flight level of the specific aircraft derived with the same timestamp as we obtained from the Surveillance information source.

47

- **FL_fr**
  - The information about the flight level gathered from the Surveillance information source (in this case we are talking about the FR24)
- **FL_dif**
  - The difference between two previously mentioned FL_nm and FL_fr. The information displayed under the FL_dif attribute will be used for the further analysis of the vertical profile precision of NM B2B Flight data.

3.3 Statistical analysis

Statistics is a science that combines mathematics and logic with the intention of understanding the relationship and patterns from the collected data. It is widely used in any sphere of our lives. The aviation industry depends on the quality of the used data within their systems. In the same way, we can conclude that all the companies that use the NM B2B Service in their applications must know the accuracy of the data they receive. Before implementing such data and before choosing the purpose for the utilisation, we need to know the basic information about their precision.

For the accuracy evaluation of NM B2B Flight data we will be analysing several samples of data that will differ from each other by the area and the source of the surveillance information used as a reference. At the end, it will allow us to observe and analyse any differences between them. From the statistical point of view, we distinguish two widely used statistical techniques:

- Descriptive statistics
- Inferential statistics

The inferential statistics method works with the utilisation of hypothesis for making the conclusions about the samples of data. In order to satisfy the objectives of this study we will use the first mentioned method called the Descriptive statistics. In the following chapters we will analyse the NM B2B data accuracy using the tools of descriptive statistics. [42]

3.3.1 Descriptive statistics

The descriptive statistics is a method of describing, summarizing, and understanding the data we collect. It is a useful way of visualising data in the form of graphs and charts for better understanding. The main aim is to focus on historic and real-time data and not to focus on making the predictions about the future. The results can point out to important facts about the concrete data samples and can be furthermore used as a foundation for another research. [42]

After we have successfully obtained data, filtered, and stored them in a proper way, we can proceed with the analysis. In general, we distinguish 3 most commonly used types of descriptive statistics [42]:

- Measures of Frequency
- Measures of Central Tendency
- Measures of Dispersion

The closer explanation of every type with the application is comprised in the following chapters.

3.3.1.1 Measures of frequency

As the name implies the measures of frequency focus on the density of occurrence of specific values within the data sample. Therefore, we can begin with the evaluation and analysis of the accuracy of the NM B2B Flight Data.

At first, it is important to state the attribute we are observing. The most significant entity which defines the accuracy of the NM B2B Flight data focused on in this study is the information about the position information of the aircraft. We have been gathering data from three surveillance information sources (FlightRadar24, OpenSky Network and Secondary Surveillance Radars). In the next steps of this chapter, we will focus on the following attributes from the databases:

- Distance_os
- Perpendicular_d
- Offset_d
- FL_dif

As the FlightRadar24 and OpenSky Network platforms collect data about the aircraft's positions in similar manner, we have decided to use in the next examples each source for different area of IFPZ zone. The FR24 was used for analysis of the western part of the IFPZ while OpenSky Network was used for the analysis of the NM B2B Flight Data accuracy in the South-Eastern part.

In the first histogram in Figure 15 we can see the distribution of the distance_fr values. This value represents the difference between the position information gathered from NM B2B and FR24, measured in nautical miles. The amount of data within this concrete sample is 200 000 and was measured in regular time intervals over 10 days in the Western part of IFPZ.
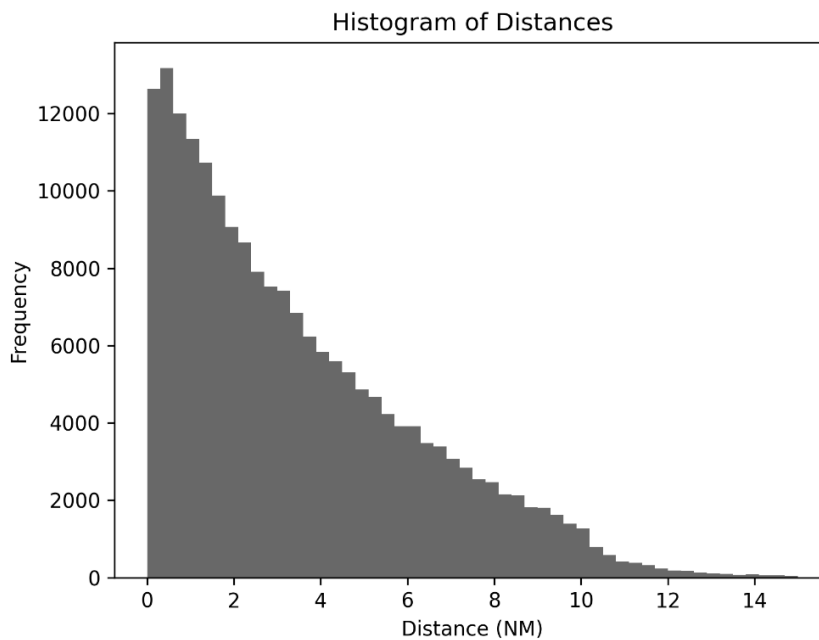
Figure 15: Histogram - FlightRadar24 – West of IFPZ

The information represented by the histogram in Figure 15 distinctly points out that most of the distance values differences falls into the interval of 0-2 nautical miles. The apex is situated within the vicinity of 0.5 nautical miles and exhibits an exponential decay pattern until reaching a magnitude of 15 nautical miles. At the value of 10 NM mark, a notable decrease is observable. This fact could potentially suggest and support the fact that positional data obtained by NM B2B are being recalculated upon reaching the 10 nautical mile threshold as stated in ATFCM Manual [14].

As a comparison of another surveillance source and explored IFPZ area we will advance to the OpenSky Network data from the South - Eastern part of IFPZ. The amount of data depicted in the Figure 16 is 40 000. We can observe the distance_os values which represent the difference between the position announced from NM B2B and OpenSky Network measured in nautical miles. A very similar pattern is depicted as previously shown in Figure 15. The peak is again located at the mark of 0,5 nautical miles and afterwards we see the decrease in an exponential pattern down the mark of 14 nautical miles.
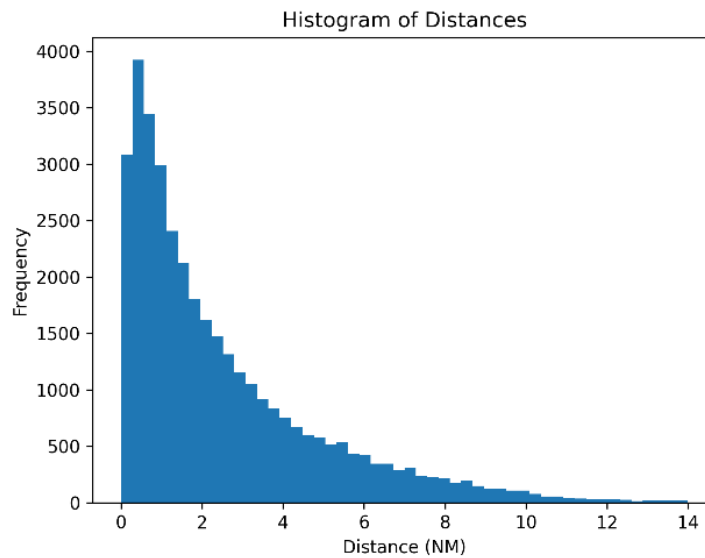
Figure 16: Histogram - OpenSky Network SouthEast

Following this, it is appropriate to figure out which of the two components, meaning the perpendicular and offset distance, make the most significant influence on the overall distance value. The relationship between the distance components is displayed in Figure 17. From Figure 17, we can see the blue trajectory representing the predicted trajectory by the CTFM from the NM B2B Service. We aim to point out the values of the overall, perpendicular, and offset distance between the position information from NM B2B and Surveillance sources.

Figure 17: Distance components

For depicting the relation between the perpendicular and offset distance, we will at first use the FlightRadar24 data in Western part of IFPZ. In the following Figure 18 we can explore two histograms. Each of them is focused on one of the distance components. Histogram a) in Figure 18, depicts the distribution of the offset distance values. While the distribution of

perpendicular distance values is represented by the histogram b). We can now conclude several important facts. The perpendicular distance seems to have much less significant influence on the overall distance in comparison with the offset distance. The perpendicular distance values exhibit a notable peak situated in the interval of 0-0,5 nautical miles. Subsequently, we can observe a substantial decrease in the histogram. In comparison with the offset distance values, the peak is located more around 0,5 nautical miles followed by more consisted decrease in exponential pattern. But on the other hand, we can also observe that perpendicular distance values in histogram b) reach more values to the negative extremes as well. The influence on the overall distance will be examined in the following sections.
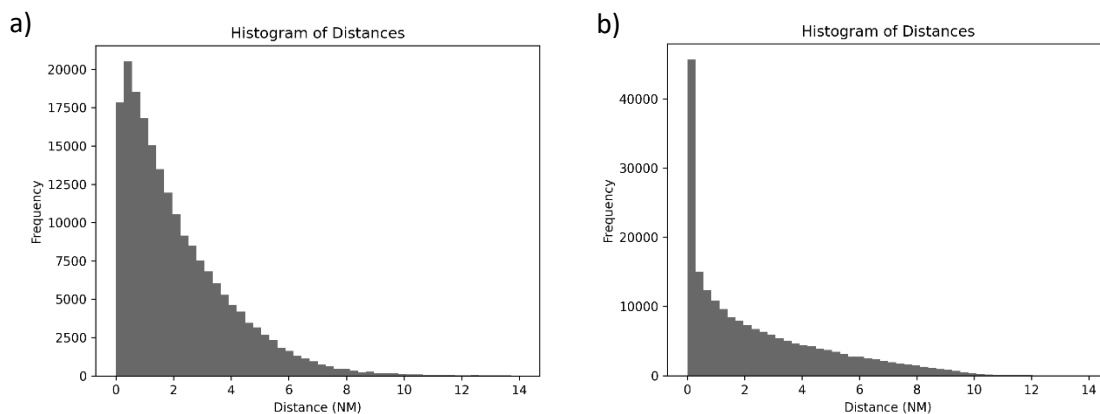
Figure 18: Histograms of Offset a) and Perpendicular distance b) – FR24

In a similar manner we will progress with the comparative analysis of the perpendicular and offset distance related to OpenSky Network data in the south-eastern part of IFPZ. In Figure 19 we can explore two histograms. The histogram a) shows the distribution of the offset distance values while the histogram b) of the perpendicular distance values.
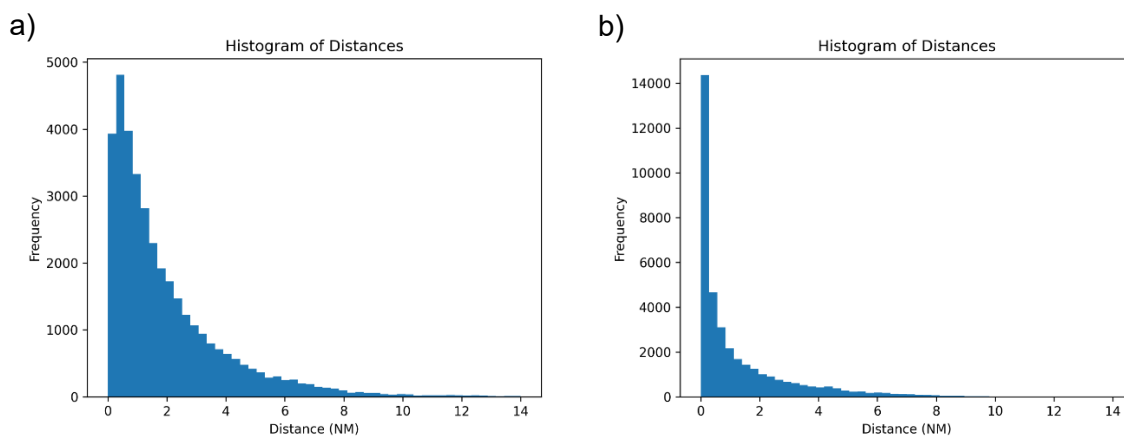

Figure 19: Histograms of Offset a) and Perpendicular distance b) – OS

When comparing figure 18 to figure 19 we can see that the relationship between the perpendicular and offset distance values are considerably similar. Following this fact, we can suppose that the NM B2B Flight data accuracy doesn't significantly change in different areas of IFPZ. Also, the amount of the lowest values of the position information is apparent in the histogram b). This points out to the idea that in a bigger way the NM B2B position could potentially vary from the actual one mostly in time difference. As time difference is directly proportional to the value of the offset distance, proving the fact that the offset distance represents the difference in position along the actual flight trajectory created by CTFM Point Profile.

At present, we can proceed to the examination of the accuracy of the vertical profile information gathered from NM B2B Service. In this case we will analyse the relationship between the FlightRadar24 data. We are comparing the flight level information at a given points described by the CTFM Point Profile. The distribution of the attribute FL_dif is represented by the histogram in Figure 20. The amount of data used in this concrete analysis is 45 000. From the Figure 20 we can see that over 44 000 of all the data are around 0 ft. This proves that the accuracy of the vertical profile information is very precise.
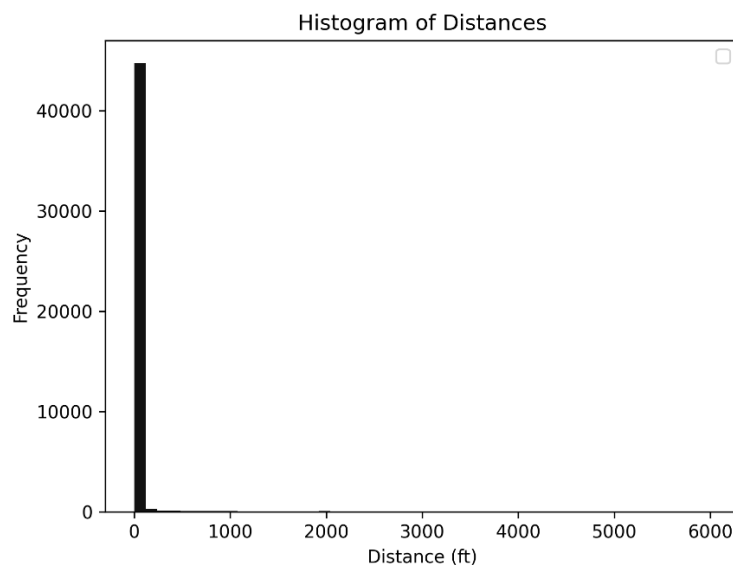
Figure 20: Flight Level analysis - FR24

As there are very little amount of data that are not fitting the desired quality, in the figure 21 we can see the extreme values of the Flight level information accuracy derived from NM B2B Service. In Figure 21 we can explore the histogram of data of which the difference was bigger than 1000 feet. It is visible that only few hundreds of data fell out of the margin less than 1000 feet. The peak is located in the area of 2000 ft that could potentially indicate that an Air Traffic

Controller suggested the change of the Flight Level, which wasn't updated at the right time in the CTFM Point Profile.
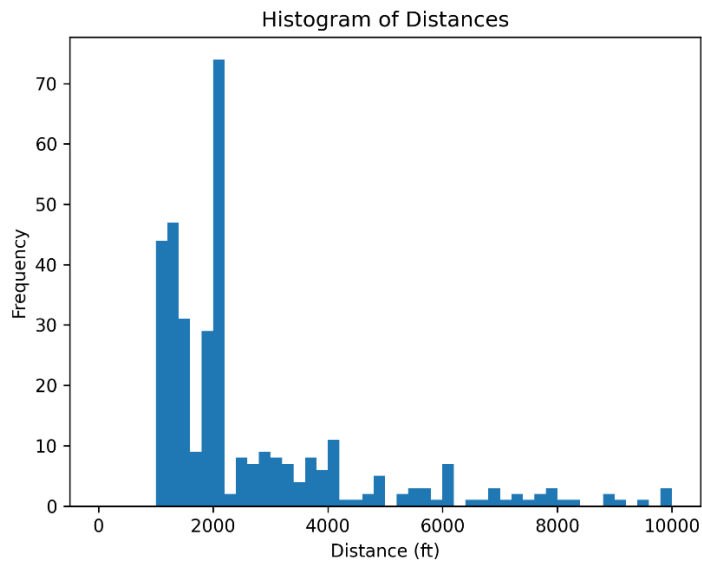
Figure 21: Histogram of Flight level extreme values

In this study, we have experienced a high precision of the vertical profile data described by the flight level differences in several data samples. The majority of the data samples of more than 99% showed that the flight level differences between NM B2B Flight Data and either OpenSky Network or FlightRadar24 were less than 150 ft. Therefore, we concluded that the flight level differences between the sources are highly accurate and decided to focus our analysis on position data.

3.3.1.2 Measures of Central Tendency

The Measures of Central Tendency is one of the most common used methods of the Descriptive Statistics. The central tendency indicates that using these measures we will focus on the typical value around which the other values tend to cluster. We distinguish three most commonly used measures of central tendency [42][44]:

- Mean
  - Mean is also known as the average value of the data sample.
  - The formula used for the calculation of the mean:

$$\bar{x} = \frac{\sum x_i}{n} \tag{12}$$

- ▪ Where x represents the sample of the data while n is the number of data within the specific data sample
- Median
  - o Median value represents the middle score of the dataset (A), which is arranged in a particular order.
  - o The formula for calculating of the median is:
    - ▪ If the number of data is even:

$$Mdn \; = \; \frac{A\left[\frac{n}{2}\right]+A\left[\frac{n}{2}+1\right]}{2} \tag{13}$$

    - ▪ If the number of data is odd:

$$Mdn \; = \; A\left[\frac{n+1}{2}\right] \tag{14}$$

  - o After the calculation we get the middle value of the dataset.
- Mode
  - o Mode represents the value from the dataset with the highest frequency. In our case we use continuous data which could make it complicated for calculating the concrete value with the highest occurrence. With the intention of finding the mode within our data we will create periodic intervals and determine the mode using those intervals.
  - o Therefore, there is no direct formula for calculating the mode of the dataset. To find the mode we will group the values into intervals and the concrete interval with the biggest amount of data can be assigned as our mode.

Table 2: Measures of Central Tendency- Distance

Surveillance information source	Mean [NM]	Median [NM]	Mode [interval in NM]
FR24 - West IFPZ	3,5149	2,7585	0 – 1
OS – West IFPZ	3,4656	2,7360	0 – 1
SSR – West IFPZ	3,8865	2,1148	0 – 1
FR24 – Central IFPZ	3,3429	1,9350	0 – 1
OS – Central IFPZ	3,3893	1,9624	0 – 1
FR24 – Southeast IFPZ	3,3695	1,8710	0 – 1
OS – Southeast IFPZ	3,6517	1,7497	0 – 1

In Table 2, we can explore how the measures of central tendency differ in comparison with the utilization of different surveillance information sources as well as in different areas of IFPZ. The measures of central tendency were calculated for the dataset of 40 000 values obtained from the database tables referring to Position. It is apparent that the values of Mean are nearly the same, as the mean of the overall distance differences between NM B2B and FR24 is 3,5149 nautical miles obtained in the western part of IFPZ. In comparison with the OpenSky Network data from the Southeast part of the IFPZ, the mean reached the value of 3,6517 nautical miles. When we look at the values of modes, all of them suggest that most of the measured distance differences belong to the interval of 0 to 1 nautical mile. The only significant difference is observable in the measure of central tendency called the median. The median of measured distance differences in the western part of IFPZ is significantly higher. As the median value represents the middle value of the whole dataset, it could suggest that the values obtained in the western part are more evenly distributed. In comparison with the data calculated in the central and south-eastern part of IFPZ, the median value suggests that in the dataset, there are more extreme values present.

Table 3 represents the measures of central tendency calculated from the perpendicular distances. It is apparent that we measured the lowest values of perpendicular distance on average in the area of Central IFPZ.

Table 3: Measures of central tendency - Perpendicular distance

Surveillance information source	Mean [NM]	Median [NM]	Mode [interval in NM]
FR24 - West IFPZ	2,4319	1,5998	0 – 1
OS – West IFPZ	2,4931	1,6442	0 – 1
SSR – West IFPZ	2,5374	1,3814	0 – 1
FR24 – Central IFPZ	1,8516	1,4924	0 – 1
OS – Central IFPZ	1,9334	1,3838	0 – 1
FR24 – Southeast IFPZ	2,0373	1,4471	0 – 1
OS – Southeast IFPZ	2,1679	1,6039	0 – 1

In contrary, Table 4 shows the measures of central tendency that come from the measured values of the offset distances. We can observe that the values of median are close to each other. But we can see apparent difference in the mean values. We will explore the possible reasons in the upcoming chapters in further analysis.

Table 4: Measures of Central Tendency - Offset Distance

Surveillance information source	Mean [NM]	Median [NM]	Mode [interval in NM]
FR24 - West IFPZ	2,5256	1,3126	0 – 1
OS – West IFPZ	2,0730	1,5735	0 – 1
SSR – West IFPZ	2,7356	1,4291	0 – 1
FR24 – Central IFPZ	2,3586	1,4043	0 – 1
OS – Central IFPZ	2,6072	1,4555	0 – 1
FR24 – Southeast IFPZ	2,8791	1,5798	0 – 1
OS – Southeast IFPZ	3,1142	1,3557	0 – 1

The big disadvantage of the measures of central tendency is the fact that it doesn't really say much about our datasets. Even when two of our sets have the same mean or median values it doesn't necessarily mean that they are the same.

The measures of central tendency showed us another overview of the collected data samples. As we could see, the majority of measured values of either overall, perpendicular or offset distance were located in the interval of 0 – 1 nautical miles. Another interesting thing to point out is that on average, the results measured using OpenSky data were sufficiently close to the results when gathering the data from FlightRadar24.

3.3.1.3 Measures of Dispersion

To understand how the data are distributed in the dataset sample we will use the measures of dispersion. We distinguish two main types of measures of dispersion [45]:

- The absolute measures of dispersion
- The relative measures of dispersion

As we are comparing datasets with the values assigned by the same units, it is sufficient to use the absolute measures of dispersion, which are composed of [45]:

- Range
  - The range value is being calculated as a difference between the largest and the smallest value of the dataset. It doesn't reveal much information about the datasets besides that it allows us to see the maximum and the minimum values.
- Interquartile range

- o The interquartile range represents the middle 50% of the whole dataset. It is calculated as a difference of the first quartile (first 25% from the dataset) and the third quartile (75% of the dataset).
  - o The main advantage of the interquartile range is that it is not affected by the extreme values and focuses on the area where most of the obtained data are clustered.
- Variance
  - o Variance is the measure that tells us how the data are spread across the dataset from the mean of the dataset. The higher the value of variance, the further the individual data points are located from the mean value. The variance s^2 of the dataset X, where n represents the number of data can be calculated by the formula:

$$s^2 = \Sigma_1^n \frac{(x_i - \bar{x})^2}{n-1} \tag{15}$$

- Standard deviation
  - o The standard deviation, also considered as the most commonly used measure of dispersion, represents the average amount of dispersion of individual data points from the mean value of the dataset. It is less sensitive to the extreme values and provide more accurate information about the behaviour of the observing data. The value of standard deviation can be calculated by following formula:

$$SD = \sqrt{\Sigma_1^n \frac{(x_i - \bar{x})^2}{n-1}} \tag{16}$$

Consequently, we can apply the measures of dispersion on gathered data of the database table Position. The amount of data within the used data sample is 40 000. The Table 5 exploits the range, interquartile range, variance and standard deviation of the three attributes (overall distance, perpendicular distance, and offset distance measured between the position information gathered from NM B2B and FR24). The range values suggest that the highest influence in the overall distance value was made by the offset distance. It could potentially say that some of the aircraft position data were delayed for quite some time as the extreme values of the overall distance reached over 129 nautical miles. The interquartile range shows that most of the data within the given dataset fell into the interval of around 3 nautical miles. We can also observe that the variance value is more affected by the extreme values in the dataset.

The value of the standard deviation on the other hand shows that the individual data points are on average dispersed from the mean value only by the 2 to 3 nautical miles.

Table 5: Measures of dispersion - FR24 – West IFPZ

FR24 and NM B2B	Overall distance	Perpendicular distance	Offset distance
Range	129,0365	49,0789	129,0375
Interquartile range	4,0008	3,4197	2,3521
Variance	9,3675	6,6223	4,4636
Standard deviation	3,0606	2,5734	2,1127

Table 6 shows the measures of dispersion calculated for the data gathered when comparing the distance components between the NM B2B and SSR sources. We can observe similar pattern than in Table 5, as the most of the extreme values were caused predominantly by the offset distance. The distance values within the data sample range up to 151,74 NM with a standard deviation of 7,25 NM.

Table 6: Measures of dispersion - SSR

SSR and NM B2B	Overall distance	Perpendicular distance	Offset distance
Range	151,7408	49,9002	150,0824
Interquartile range	3,9817	3,3348	2,0362
Variance	54,5923	17,6101	37,5862
Standard deviation	7,2521	4,1964	6,1307

In the following Table 7, we depict the measures of dispersion observed from data after using the OpenSky Network data in the South-eastern part of IFPZ. We can see that the range of the same amount of 40 000 data in another part of IFPZ is significantly different. The extremes are more than 3 times bigger, which has a high influence on the variance and standard deviation values. The interquartile range proves the fact that most of the data fell close to the mean value in the interval from 1,6 to 3 nautical miles. [46] Further analysis of the concrete locations where we obtained the unusual extreme values will be performed in the Point Pattern (Spatial analysis).

Table 7: Measures of dispersion - OpenSky Network – Southeast IFPZ

OS and NM B2B	Overall distance	Perpendicular distance	Offset distance
Range	247,5648	31,6288	247,3440
Interquartile range	3,0358	1,6470	2,2598
Variance	82,2083	3,0800	81,1055
Standard deviation	9,0669	1,7551	9,0058

3.3.2 Spatial statistics

In the previous chapter we discussed the need of descriptive statistic for our study. Thanks to the descriptive statistics we could obtain the first insights about the NM B2B Flight data accuracy. The problem is that while using these methods we get the physical information about various factors of precision, but we do not know if the mentioned accuracy varies in different areas. To help us visualise the data on a map and identify any patterns in the accuracy of our data we will use the spatial statistics. [47]

With the combination of descriptive and spatial statistics, we will accomplish not only the numerical and graphical results but also the spatial distribution of data across the whole dataset. Using various methods of spatial statistics will help us determine how the accuracy changes in different Flight information regions or specific areas of the IFPZ zone.

Two methods of spatial statistics that will be used within this study are [47]:

- Spatial analysis
- Point Pattern analysis

Both methods are widely used for analysing and identifying any spatial relationships in observed data. In the upcoming steps we will proceed with the spatial analysis of three different areas of IFPZ:

- Western part of IFPZ
- Central part of IFPZ
- South-eastern part of IFPZ

These concrete three groups of the IFPZ zone could point out to any differences in data accuracy in between them. We have chosen these specific areas in order to observe any visible behaviour of different datasets. We had to deal with several limitations including the

amount of data retrieving from the NM B2B Service as well as Surveillance information sources. We gather data in small periodic intervals to cover the whole area. It is preferred to focus more on the chosen areas and obtain more data within them to achieve more accurate results, rather than covering the whole IFPZ zone using less data. Based on this fact, we can conclude that it will eventually provide a sufficient overview of the overall NM B2B Flight data accuracy.

As the NM B2B Service allows us to obtain information based on the particular airspace, we have decided to choose the French airspace for the western part, the Czech airspace for the central part, and the Greek airspace for the south-eastern part of IFPZ.

Western part of IFPZ

The data obtained from the French airspace were accepted for the analysis of the western part of IFPZ. We have obtained flight data information from NM B2B Service for every flight which on his route entered a French airspace. With the intention of their accuracy analysis, we have also collected flight data information from the surveillance information sources such as the FliightRadar24, OpenSky Network, and Secondary surveillance radars. In fixed time intervals of 4 minutes the data from NM B2B and both surveillance sources were pre-processed and stored in the database with all the necessary information.
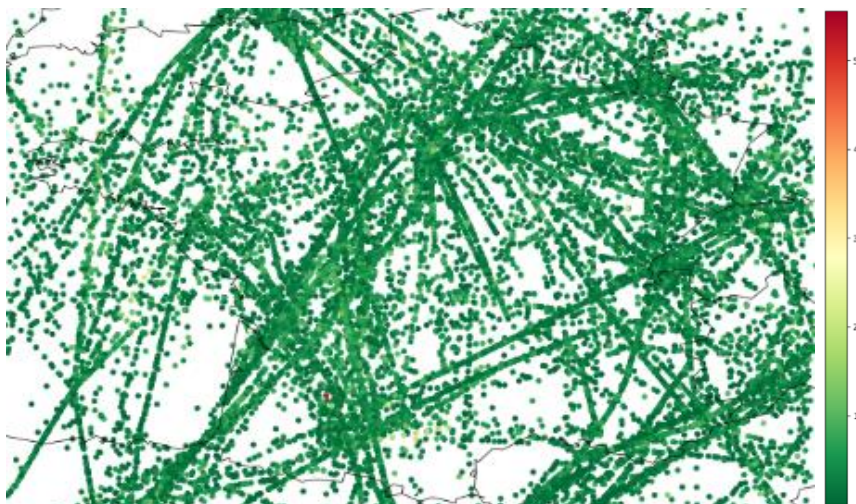
Figure 22: Data accuracy in the western part of IFPZ

Figure 22 represents the distribution of gathered data from both surveillance and NM B2B sources in the western part of the IFPZ zone. We can observe visible point patterns that allow us to distinguish the concrete routes and flight trajectories over the French airspace. Each point on the map in Figure 22 represents a single value of the overall distance measured between the position information of NM B2B and surveillance information source. To make the

results appropriate, we calculated the distance difference of the same timestamps and stored them within the database.

Besides the data distribution, we can observe another phenomenon in Figure 22. The colouring of the individual data points symbolizes the value of the distance difference measured at each point. The dark green shade represents that the difference in position between NM B2B and surveillance information is close to 0 nautical miles, while the dark red colour of the point refers to the extreme values which exceed more than 50 nautical miles difference.

To see how the data are clustered within the concrete areas, we will use the point pattern analysis method, which comprises the utilization of the Kernel density. [48] It is a capable tool that enables us to visualise the data over the map in a recognisable manner. The kernel density, as shown in Figure 23, is achieved by applying the kernel function and calculating the density estimation. The kernel function weights every point in the dataset while creating a smooth distribution density of the data over the map. The more data we observed in neighbouring areas the darker the colour of the area will be presented. [48]
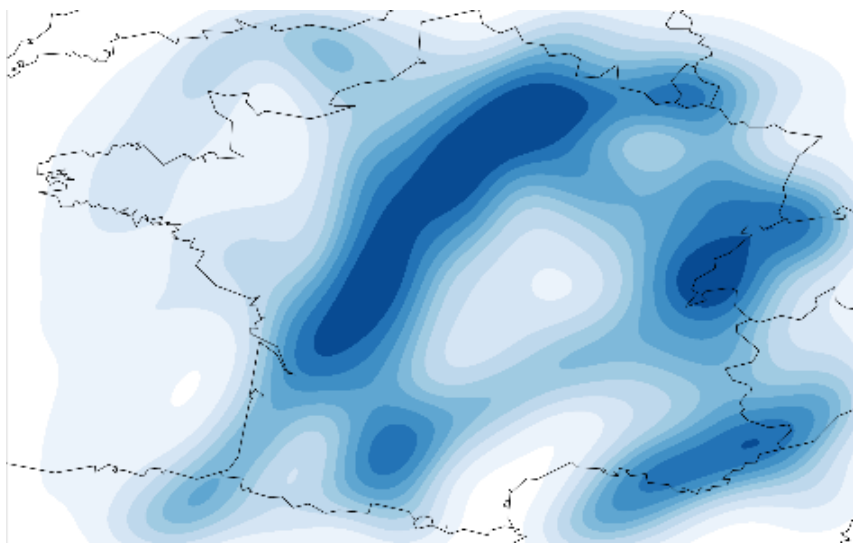
Figure 23: Kernel density - western part of IFPZ

In Figure 23 we can see the most populated areas within the whole dataset. Most of the gathered position information points corresponds to the biggest airports in observed area. We can clearly distinguish the airports in the cities like Paris, Bordeaux, Toulouse, Nice, Lyon, and Geneve which depicts the busiest airports in the area. Also, we can observe that the most frequent route is the route between Paris and Bordeaux which from there continues down to the direction of Spain.

We could suppose that the more crowded area by the number of the flights, the bigger the regulations and the bigger the differences could be observed. But the opposite appears to be the true. In Figure 24 we applied the kernel density estimation for only those data from the database in which the difference in distance was less than one nautical mile. We can explore in which areas we obtained the most accurate position information from the NM B2B Service. As the aircraft approach the airports in Paris, Nice, Marseille or Geneve, the precision of the position information substantially rises. It could suggest that the airports with the most frequent flights require a higher refresh rate of data as the situation of big demand depends on the quality of the data.
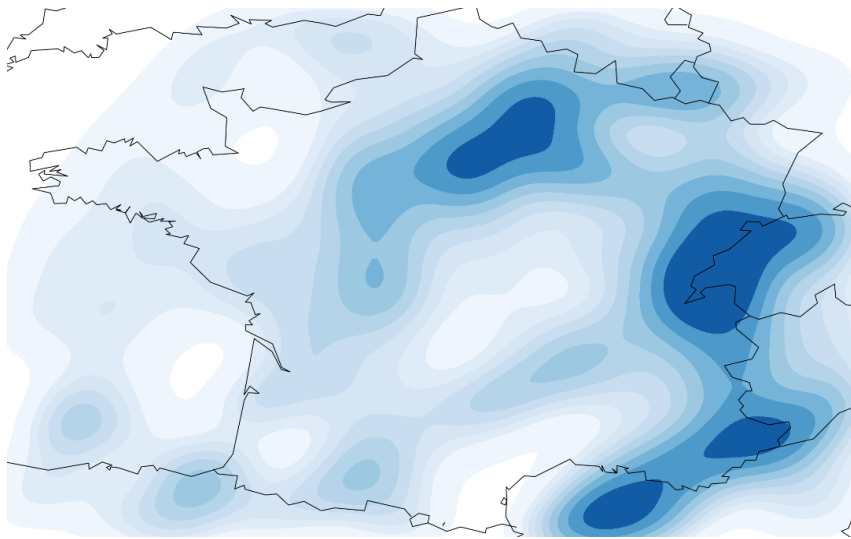

Figure 24: Kernel density - most accurate areas of west IFPZ

The visible proof is depicted in the following Figure 25. The line graph shows the distance difference between NM B2B and surveillance position information on the flight route from Heathrow Airport in London to the destination of Milano Malpensa Airport. On this route, the aircraft pass through the two most accurate areas in this region, which are the area between the cities Paris and Rennes and the second is the western border of Switzerland with France. In Figure 25, we can see two substantial decreases. Each of them corresponds to the area depicted in Figure 24.

Another example that supports this statement is shown in Figure 26. In this line graph, we can see the behaviour of all the distance differences between NM B2B and FR24/OS for all the flights on the route between Toulouse Blagnac Airport and Paris Orly Airport. During most part of the flight trajectory from Toulouse, the position accuracy obtained by NM B2B reflects around the value of 4 nautical miles. When reaching the area with higher precision of the flight data, an evident decrease in position difference occurs.
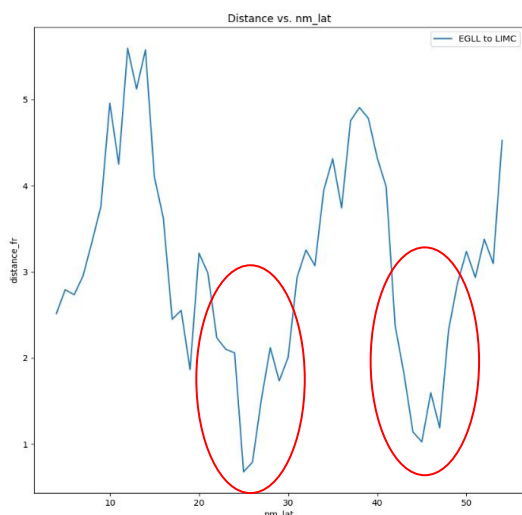
Figure 25: Position accuracy of NM B2B - EGLL to LIMC
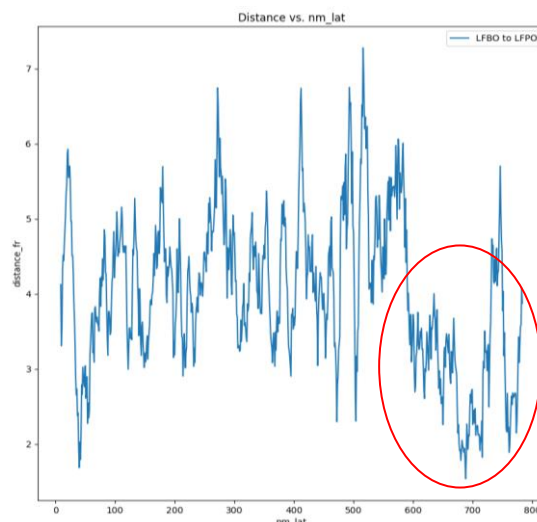
Figure 26: Position accuracy of NM B2B - LFBO to LFPO

Central part of IFPZ

For the spatial analysis of the NM B2B flight data accuracy in the central part of IFPZ we gather data of all the flights that on their trajectory pass through the airspace of the Czech Republic. The surveillance information sources used for this analysis are the OpenSky Network and the FlightRadar24. In Figure 27, we can observe the distribution of all the individual points within the database.

Figure 27: Data distribution - central part of IFPZ

Consequently, the colouring of the data points refers to the actual precision of the position information obtained from the NM B2B Service. One concrete area with the evident red shade is observable. Besides this area, the rest of the data remains relatively precise and evenly spaced in the whole region. The pattern of the data distribution also displays the most used flight routes. By applying the kernel density function on the whole dataset, we are able to

distinguish which places in the specific region are more occupied by the position points than others. It is evident, that while we use the Czech airspace as a reference source of NM B2B data, most of the data points will take place in the region of Czech Republic. The Figure 28 demonstrates the way of distribution density of observed data.
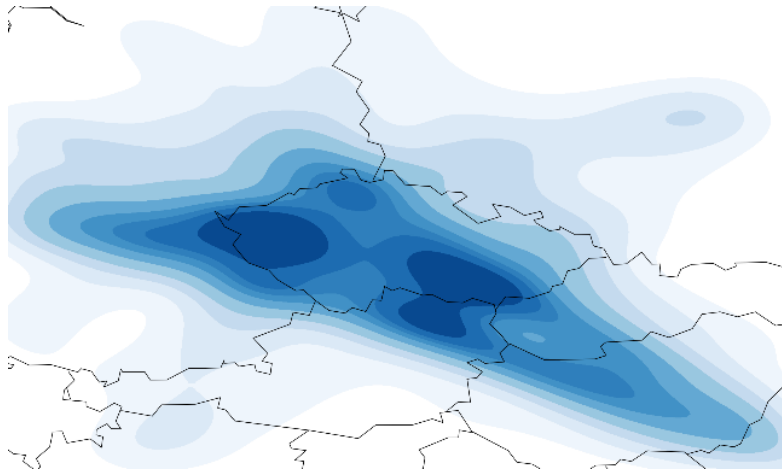
Figure 28: Kernel density distribution - Central IFPZ

If we want to display the most accurate areas, we will use the kernel density for all the NM B2B position data, of which the difference with the surveillance information doesn't exceed the amount of one nautical mile. Such a situation is presented in Figure 29. We can observe a similar pattern that we demonstrated in the west IFPZ region. The most accurate areas are located near the biggest airports, which in this case is the Prague airport together with the Schwechat airport in Vienna.
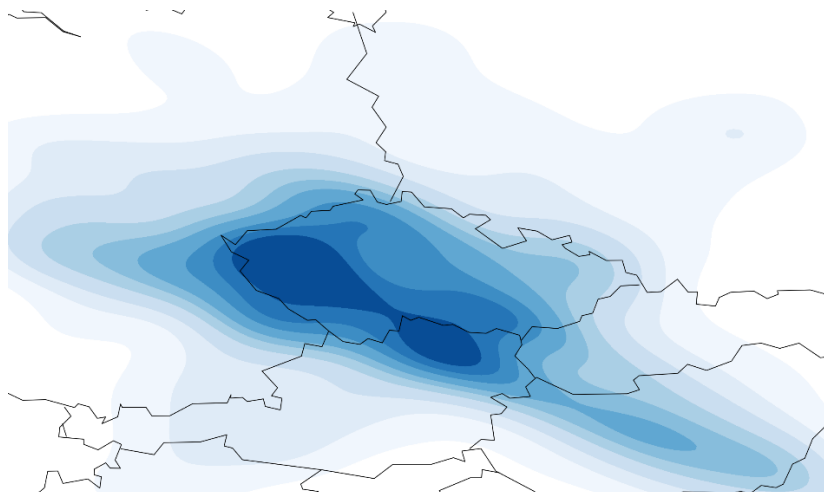
Figure 29: Kernel density of most accurate areas in Central IFPZ

The most accurate areas are not located directly above the biggest airports because of one simple reason. When the aircraft approach the airport, its position can quickly change from the

predicted one because of the actual situation at the airport. Also, when the aircraft is departing from the specific airport, it is not possible to predict the exact second when it will be in the air. Because of that, the prediction within the CTFM Point profile updates right after the departure, and therefore the accuracy improves substantially. These factors contribute to the fact that the accuracy of NM B2B Flight data appears to be higher in areas close to the most significant airports in a given region.

In order to prove the statement that the accuracy levels differ based on the specific area, we will, in this case, use inferential statistics. Since, we observed from the histogram in Figure 15, that the distribution of our distance values deviates from the normal distribution, we will use the non-parametric Mann-Whitney U test. It is a test that will be used to analyse the difference in NM B2B data accuracy in two different areas. The condition for the utilisation of Mann-Whitney test is satisfied as the distribution of our data sample is not normal and we use two independent groups of data [49]. The first step of the current analysis is to define the null hypothesis:

- H0: There is no significant difference in the distribution of the accuracy of the NM B2B Flight data in the area 1 and outside the area 1.
    - Area 1 is represented by the dark blue shade in Figure 29
- H1: There is a significant difference in the distribution of the accuracy of the NM B2B Flight data in area 1 and outside area 1

At first, we will separate data from our database into 2 groups. The first one will contain all the data points of the flights measured between NM B2B and OS/FR24 that are situated in area 1, and the magnitude of their distance value is less than 1 nautical mile. The second group contains data points that are located outside area 1, and also the value of distance difference is less than 1 nautical mile. For each group, we calculate the median of the distance values and compare them by the non-parametric test called the Mann-Whitney U test. The significance level also referred to as alpha, is set to 0,05. [50] After applying the Mann-Whitney U test on the observed areas, we obtained the following p-value:

- P-value = 0.0050957

As the p-value is smaller than the significance level (alpha), we can reject the null hypothesis. This fact indicates that there is a significant difference in the distribution of the accuracy of NM B2B Flight data in area 1 and outside the area. This evidence supports the statement which we obtained in the western part of IFPZ, that the accuracy levels of position information from NM B2B flight data are significantly higher in areas with busier airports.

The south-eastern part of IFPZ

For the analysis of the position information of NM B2B Flight data in the south-eastern part of IFPZ, we will use the data from the surveillance information sources called such as the OpenSky Network and FlightRadar24. The data from NM B2B Service used in this analysis consist of all the flights which on their route flew over the airspace of Greece.

First, we will have a look at the distribution of the data sample from the dataset. In Figure 30, we can explore an observable pattern of the most frequent flight routes in the region. The colouring of the individual data points represents the actual distance difference between the position information between the NM B2B and OS/FR24 sources. We can notice that most of the data points are evenly coloured by the dark green shade, with only a few points reaching the extreme values of more than 30 nautical miles.
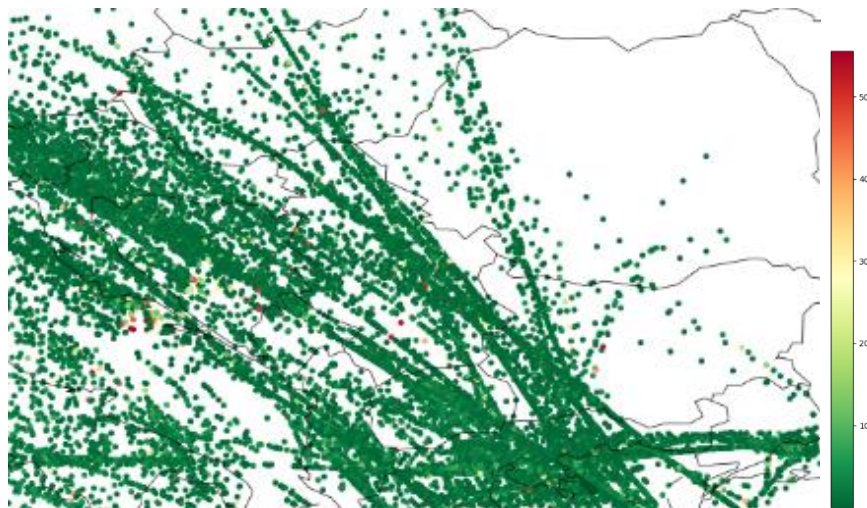
Figure 30: Distribution of data points - southeast IFPZ

In the following steps, we will examine the most accurate areas within the south-eastern part of IFPZ. It is important to see whether we can observe a similar pattern in accuracy as in the western as in the central part of IFPZ. The kernel density function, as a useful tool of Point pattern analysis, has been applied to discover and visualise the most accurate areas. In this case, by the most accurate areas, we exploited all the data points at which the distance difference between NM B2B and the surveillance information source is less than one nautical mile. In Figure 31, we can straightaway detect the same point pattern as in the previous examples. The most accurate areas in the observed area are located at the biggest and busiest airports, such as Skopje, Sofia, Sarajevo, Thessaloniki, and Zadar. The accuracy worsens above the area of Kosovo and Montenegro.
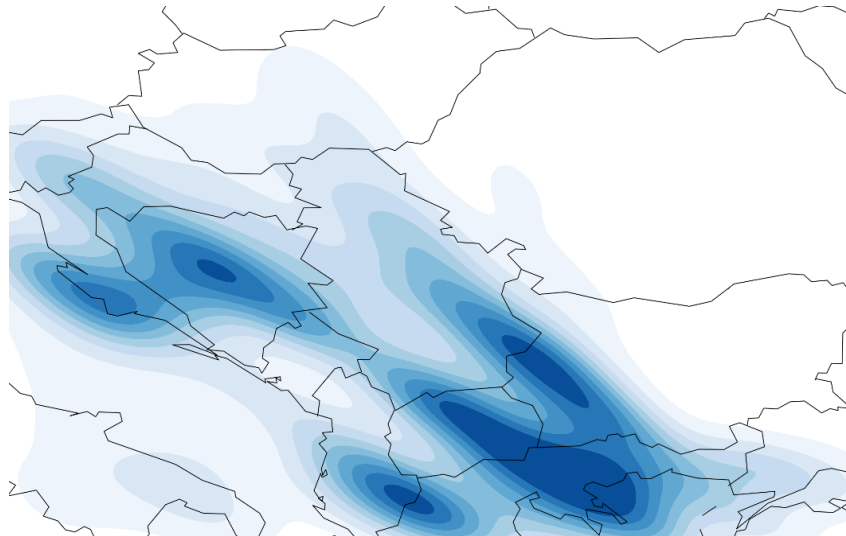
Figure 31: Kernel density - accurate areas - South-east IFPZ

3.4 Results validation

<u>The comparison of NM B2B Flight data in IFPZ parts</u>

In this chapter we will explore the differences in NM B2B Flight data accuracy in 3 previously observed parts of IFPZ. With the intention of discovering any patterns in concrete areas, we have divided each part into 5 sectors. The dimension of the sectors has been chosen according to already discovered information about the accuracy of the whole area comprised in the previous chapter. The priority is put to cluster the data with similar point patterns into the same sector. The arrangement of the sectors is displayed in Figure 32.
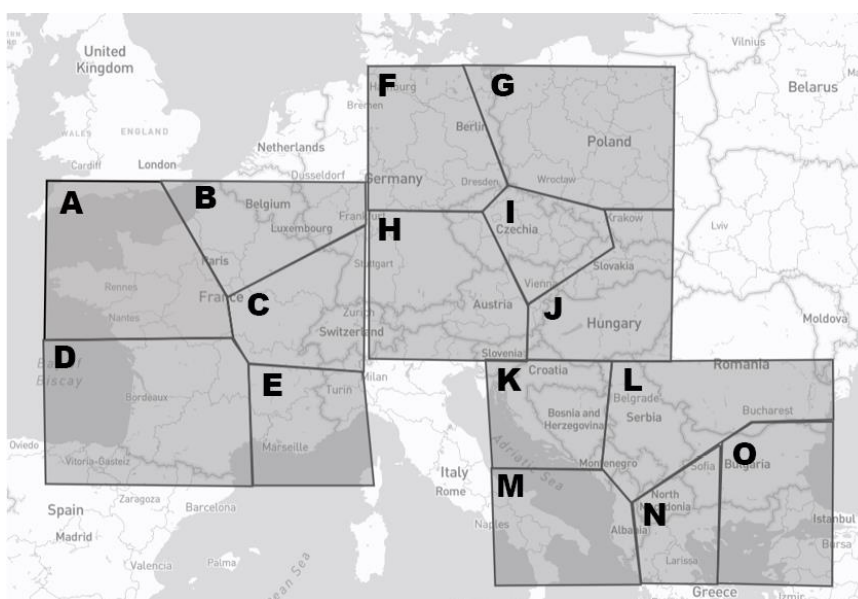
Figure 32: Sectorisation for comparative analysis

From figure 32, it is apparent that we cover representative parts of three IFPZ zones. For each of the sectors, we have calculated the mean value of the overall distance, perpendicular distance, and offset distance between NM B2B and the surveillance information source (OS). The OpenSky Network data were chosen to evaluate the differences in the individual sectors in order to compare data gathered from the same source. As proved below, the Mann-Whitney statistical test showed that the results of the accuracy measured from FR24 are not significantly different than using the OS as the reference source.

- H0: There is no significant difference in the accuracy of NM B2B data compared to OpenSky or FlightRadar24
- H1: There is a significant difference in the accuracy of NM B2B data compared to OpenSky or FlightRadar24

The calculated p-value using the Mann-Whitney test in which we compared the accuracy of the NM B2B data gathered when compared to OpenSky and FR24 was 0,63. As the p-value is greater than 0,05, we cannot reject the null hypothesis. Therefore, we can suppose that there is no significant difference when using either OS or FR24 data for the NM B2B accuracy evaluation. Thus, it is reasonable to use the OpenSky data for the following evaluation of the accuracy of NM B2B data in individual sectors.

The concrete values of the distance attributes are displayed in Table 8. These values are also accompanied by information about the concrete position of NM B2B in relation to the reference surveillance position. The particular values in the table depict the percentage of occurrence within the whole area of the specific IFPZ zone. By this calculation, we tried to identify any sector where the NM B2B position deviates from the actual OpenSky position information in a specific manner.

As shown in Table 8, we have gathered all the information calculated for each sector. We have divided all of the results into 3 main groups that we can analyse separately. At this point we will proceed with the comparison of the calculated values in different sectors.

Table 8: Comparison of data accuracy in different sectors

Sector	Distance [NM]			NM B2B offset position		NM B2B perpendicular position	
	Overall	Perpendicular	Offset	Delayed	In-front	Left	Right
A	3,7955	2,8101	2,1003	21,60%	24,93%	16,41%	20,85%
B	3,3369	2,2139	2,1764	21,30%	19,20%	23,01%	28,99%
C	3,4270	2,2866	2,2009	18,72%	19,29%	19,89%	26,47%
D	3,8641	2,7562	2,3266	19,71%	19,30%	24,64%	11,36%
E	3,2340	2,0508	2,2335	18,66%	17,29%	16,05%	12,33%
F	2,8565	1,7056	2,0663	18,84%	27,79%	26,92%	17,64%
G	4,8148	3,2331	3,2677	11,22%	23,60%	19,45%	1,4%
H	2,7142	1,6423	1,8632	29,17%	0,79%	9,91%	18,81%
I	2,5686	1,3842	1,9077	10,53%	19,96%	11,39%	32,42%
J	2,7450	1,6367	1,9738	30,24%	27,87%	32,33%	29,72%
K	3,0584	0,9681	2,6696	32,49%	30,21%	11,46%	26,81%
L	3,0945	0,9852	2,6311	9,13%	16,43%	21,71%	18,51%
M	5,3027	1,4056	4,7664	14,72%	21,59%	31,28%	25,60%
N	3,7354	1,3687	3,1479	17,22%	22,88%	14,13%	9,52%
O	3,8674	1,5509	3,2091	26,40%	8,89%	21,42%	19,55%

Distance values

For analysing the disparities in particular sectors, we have calculated the mean values of overall, perpendicular, and offset distances between NM B2B and Surveillance position information. We can observe how the distance values differ in sectors within the same area of IFPZ as well as the other IFPZ areas.

As for the western part of the IFPZ region, meaning the sectors A – E, the mean of overall distance values reveal that the highest accuracy has been achieved in sectors B, C, and E. This fact supports the statement mentioned in the previous chapter, where we have displayed the kernel density of the areas with the most of the 0 to 1 nautical mile distances. A similar pattern can be interpreted from the perpendicular and offset distance values. Exceptionally we can see interesting observations in sector A. In this specific sector, we have measured data that result in the highest average overall and perpendicular distance value in comparison with other sectors from the same IFPZ area. On the other hand, we have observed the lowest value of the mean offset distance. This comparison of those three values shows which of the perpendicular and offset distances have the biggest influence on the inaccuracy of the overall distance.

In the area of the central part of the IFPZ region, meaning the sectors F – J, the most accurate area according to the overall distance mean value is sector I. The sector I was also previously recognised as the one with the highest frequency of the most accurate NM B2B positional data. The perpendicular distance values measured in the whole central IFPZ region represent better accuracy when compared to the offset distance values. The highest value of the overall distance was measured in sector G. This could indicate that more extreme values have been collected from this area.

In the south-eastern part of IFPZ, represented by the sectors K – O, we detected that the K and L sectors contain the most accurate areas in relation to the overall distance values. On the other hand, the least accurate sector is sector M. When we look at the values of the perpendicular and offset mean values, we can interpret it in a way that impact on the overall distance as the offset distance. This could indicate that the aircraft's position after the departure from airports in this area does not update soon enough, so it results in a higher value of the overall distance measured.

In Table 8, it is evident that the average values of distance attributes vary in between each of the observed IFPZ areas. The highest values of the overall distance appear to be in the western and south-eastern parts. Additionally, the lowest perpendicular values were measured in the south-eastern part, and the lowest offset distances were in the central part of IFPZ. This could be caused by many factors, including the density of the airspace, weather conditions, flight route patterns, airspace design, and used equipment.

NM B2B offset position

By the offset position, we accomplished the information which states whether the NM B2B position is either delayed or in front of the actual position retrieved from the surveillance information sources. The main aim was to recognize if there is a specific sector of the IFPZ area in which we could discover a higher presence of concrete attributes of the offset position. The values represented in Table 8 depict the percentual share of each attribute within the whole IFPZ part composed of 5 sectors.

The percentage of the flight positions that were delayed in the western part of IFPZ has the range of 18,66% measured in sector E and 21,60% in sector A. It suggests that delayed flights are evenly distributed in the whole part of IFPZ. A higher range can be discovered from the NM B2B data that were in front of the actual position. The biggest difference between delayed and in-front positions is discovered in sector A. It indicates that in this sector, we obtained more

information of in front position, which means that the actual position of the aircraft is delayed. The other sectors contain similar information in both delayed and in-front position information.

As for the central part if IFPZ, the sector with the lowest percentage of delayed NM B2B data is sector I, which also contains the biggest amount of data within the region. An interesting fact is that the highest percentage share is in sector J which is right next to sector I. The values assigned to sector H indicate that almost all the data measured in this sector were NM B2B delayed, which refers to the fact that the actual position was in front.

The data measured in the south eastern part indicate that the most accurate sector L contains most of the NM B2B positional information that is in front rather than delayed. While the opposite phenomenon can be observed in sector O.

NM B2B perpendicular position

The information about the position of the NM B2B flight data compared to the position information from surveillance sources is interpreted in the percentual share of the observed IFPZ part. This value is calculated in the same manner as previously the offset position.

The highest extremes are discovered in the central part of IFPZ, specifically in the sectors G and H. Consequently, also in sector N, we can observe that the vast majority of NM B2B position data are located on the left side to the actual position.

The overall accuracy of NM B2B data in focus areas

The evaluation of the NM B2B Flight data accuracy in three focus areas (West, Central, and Southeast IFPZ) is depicted in Table 9. We obtained information from multiple surveillance sources in each area. With the intention of obtaining a better overview, we used all available information sources to point out the mean value of calculated distances between the NM B2B and the surveillance source. In addition to the mean values, we calculated the confidence interval, which represents the range where we can expect that the true population mean lies. For combining all the sources and calculating the values, we used the bootstrap statistical method. [51] It is a statistical method that uses the resampling technique for the estimation of the overall mean distance, as well as the confidence interval. By generating multiple samples, we were able to calculate the mean of all the sample means.

The observed values in Table 9 show that the overall distance values vary on average between 3 to 4 nautical miles in each of the focus areas. This approach and its results also support the statements from previous chapters.

Table 9: Overall distance evaluation in focus areas

Focus area – Surveillance source	Bootstrap mean [NM]	95% confidence interval
West IFPZ – OS/FR24/SSR	3,60	3,59 – 3,62
Central IFPZ – OS/FR24	3,32	3,21 – 3,39
Southeast IFPZ – OS/FR24	3,67	3,58 – 3,78

4 Software engineering

Following the objectives of this work, we will create a software module that will be able to calculate and evaluate the accuracy of NM B2B Flight data in real-time. The software module comprises several Python files that will be responsible for accessing NM B2B Service and Surveillance information sources, then loading and storing data within the database, calculating the accuracy of data, and presenting the results in three web-based applications. These applications should allow the user to access the database from anywhere and plot the graphs and charts which depict the measured accuracy based on various conditions the user stated. [52]

In the sphere of software engineering, we will use the Unified Model Language (UML) for the visualization of particular steps. This language comprises many various graphs, and while not all of them suit, we will be using only those which are necessary for the representation of our software module. As for the representation of concrete graphs, the application called Enterprise Architect is used in this study. [53]

4.1 Requirements

Before the implementation of any software product, we must declare all the necessary requirements we are trying to achieve. For the graphical visualisation we use the Use case diagrams described by the UML language. We can categorize the requirements of the software systems into four main groups: business, user, functional and non-functional requirements. [53]

4.1.1 Business requirements

The first group of requirements that we must acknowledge is composed of Business requirements. It accompanies all the conditions we demand the application to perform in the final state. The reasons of creating the software system within this work can be described by the following business requirements shown in Figure 33. [53]
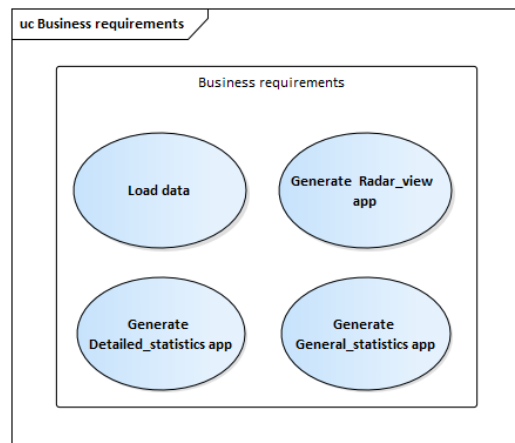
Figure 33: Business requirements

1. Load data from NM B2B Service and Surveillance information sources
2. Generate web-application Radar_view
3. Generate web-application General_statistics
4. Generate web-application Detailed_statistics

As we want to allow the user to see various types of statistics in a web-based system, we have to load the data from external sources and visualise them in an appropriate way in 3 web applications.

4.1.2 User requirements

The user requirements contain the demands on the application that describe what the actual user is going to observe. They are based on the business requirements, while in our system, we can describe them by the use case diagram in Figure 34. [53]
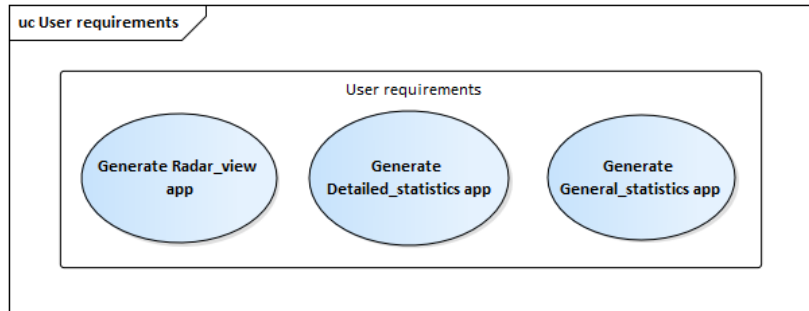
Figure 34: User requirements

1. Generate web-application Radar_view
2. Generate web-application General_statistics
3. Generate web-application Detailed_statistics

According to the user requirements, the output of our system will be accessed by users through three web applications.

4.1.3 Functional requirements

The functional requirements describe what the system has to offer to the users. It accompanies the loading of the data, all the calculations, and the final visualisation of the outputs. Figure 35 depicts the functional requirements of our software system.

Figure 35: Functional requirements

1. Load data from NM B2B and Surveillance information sources
2. Pre-process data
3. Calculate the position of NM B2B based on a same timestamp as obtained by Surveillance source
4. Calculate the overall distance between NM B2B and surveillance source
5. Calculate the perpendicular distance
6. Calculate the offset distance
7. Calculate the side position attribute (left or right) of the NM B2B to the surveillance source
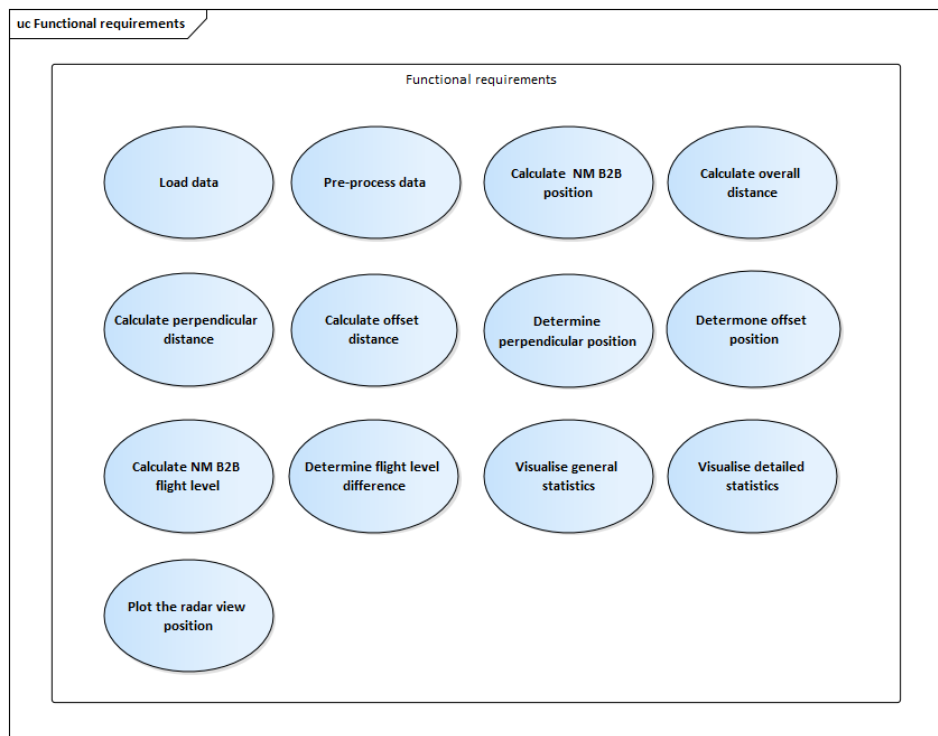8. Calculate the time position attribute (infront or delayed) of the NM B2B to the surveillance source
9. Calculate the flight level based on the timestamp from the surveillance source
10. Determine flight level difference between NM B2B and surveillance source
11. Allow the user to see the radar view
12. Allow the user to see the general statistics
13. Allow the user to see the detailed statistics

4.1.4 Non-functional requirements

The non-functional requirements represent the attributes such as the operational features that include the chosen programming language, usage of the external database, and the way of updating the database. [53] We can describe the non-functional requirements of our system in the following way:

1. Programming language Python 3
2. Programming language JavaScript
3. Programming language HTML
4. Django framework
5. Software module will keep the historical data in the database
6. The usage of external REDIS database NM B2BD

4.2 System diagram

Our software system and its functionality are described in Figure 36. The type of diagram which best fits the purpose of the system description is known as the Data Flow Diagram. The visual notation is based on the UML language while using the concrete symbols of the Enterprise Architect software. [53]
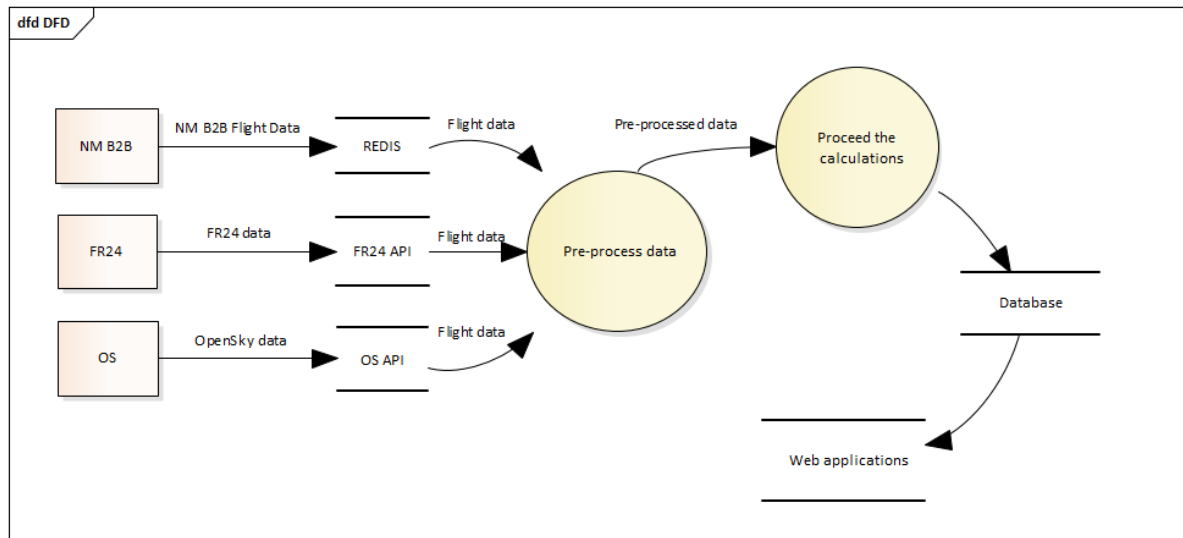
Figure 36: Data flow diagram

In Figure 36, it is apparent that the data originate from three different external sources such as the NM B2B, Flight Radar 24, and OpenSky. The NM B2B flight data are obtained through the REDIS database, which is being run by the company CS SOFT, a.s.. On the other hand, the flight data from both surveillance sources are gathered through their concrete APIs. Collected data are consequently filtered and pre-processed for upcoming calculations. The results and necessary data information are afterwards stored within our database, which is used as the final source of data for the visualisation and evaluation of NM B2B flight data accuracy through the web applications.

4.3 Design and Implementation

In this chapter, we will focus on the design and the following implementation of the web-based system which will provide the evaluation and overview of the NM B2B Flight Data accuracy. The whole system has been programmed in the framework called Django. It allows us to implement any calculation, store the data within the database and eventually provide the results for users in the form of web applications.

4.3.1 Database

The database within our system comprises all the required information needed for the accuracy evaluation. The structure of the database and its elements have already been described in the previous Chapter 3. The objective of this section is to specify the programming method by which the data are stored within the system. All the codes are responsible for the calculations, and the collection of data are programmed in Python programming language within the

views.py file. This Python file is located in Attachment 1 of this work. Within the views.py file, we can observe several functions from which, three of them are responsible for the actual pre-processing, calculations, and storing of the data to our database. These functions are:

- Database_fr_w
- Database_fr_c
- Database_os_w
- Database_os_c
- Database_os_se

These functions perform the same task, but each of them is focused on a different part of the IFPZ area. The W, C, and SE attributes suggest the usage of the Flight Radar 24 and the OpenSky Network in the western, central, and south-eastern parts of IFPZ. The structure of the code is the same in every function. They differ only in the way of defining the target areas and the utilization of the surveillance source.

The calculations within the functions reflect the same pattern described in the chapter Data Analysis. After collecting all the necessary information and calculating the remaining values, we must store them in separate database tables. We have stored all that information within the list called Result. These database tables are created in SQL manner through the python programming language by the sqlite3 library. Figure 37 depicts that we are at first creating a table with the required elements. When the database table is created, we may proceed with the storing process. The storing of data is done by the INSERT function, which loads the previously prepared result list.

```
con = sqlite3.connect('Database_W.db')
cur = con.cursor()
cur.execute('''CREATE TABLE IF NOT EXISTS fldata
                (ID INTEGER PRIMARY KEY AUTOINCREMENT,callsign text, nm_time text, fr_time text, nm_lat real, nm_lon real, fr_lat real,
fr_lon real, infront text, position_fr, distance_fr real, perpendicular_d real, offset_d real, adep text, ades text, category text, fir text)''')


    cur.executemany("INSERT OR IGNORE INTO fldata (callsign, nm_time, fr_time, nm_lat, nm_lon, fr_lat, fr_lon, infront, position_fr, distance_fr,
perpendicular_d, offset_d, adep, ades, category, fir) VALUES (?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?)",result)
    con.commit()
```

Figure 37: Creating the database

In the same way, we have created the Flight Level database tables that load the data from the list of needed values called fl_result.

4.3.2 Radar View application

The first web application within our web-based system is the application called Radar View. The purpose of this app is to show the user, how the NM B2B data differ from the surveillance data in real-time. This fact incorporates several limitations. We had to choose one airspace which is going to be used for the retrieval of the NM B2B Flight data. In this case, the chosen airspace is the French airspace. It means that the user will be able to see all the flights, which on their route pass through the airspace of France. Another limitation is the amount of data we can retrieve from the NM B2B Service per hour. In order not to exceed the maximum limit, we have decided to use the Database which was implemented in the Bachelor's Thesis [18]. More detailed information about the database and the data within the tables can be found in this document.

Functions of the Radar View app:

- Display the aircraft position from the NM B2B Service
- Display the aircraft position from the OpenSky Network
- On-hover show the basic information about the flight
- On-click connect the points by the line and display the distance between them

There are 3 files within the Attachment 1 that are responsible for this application: views.py, urls.py, and radar_view.html. The views.py file contains 2 concrete functions (OS_data_url, B2B_data_url) that gather the data from external sources and render them in JSON format as an HTTP. The file urls.py defines which function is being called by the specific URL query. The final and most important file that visualizes the points on the map with the necessary information is the radar_view.html. The map plotting was done using the source called OpenLayer. [54]

In Figure 38, we can observe the final version of the Radar View application. The blue markers represent the position from the NM B2B Service, while the orange markers depict the aircraft's position from OpenSky Network.

Figure 38: Radar view app

The Radar view application will allow the user to observe the real-time difference of the aircraft position. When analysing the data accuracy, it can be useful in comparative analysis while visually inspecting the area.

4.3.3 General statistics application

The web application which is used for the overall description of the NM B2B Flight data accuracy within the region of French airspace is called The General Statistics. The code is programmed for this specific airspace but can be adjusted for any airspace within the IFPZ area. This concrete chapter and the application can be used as practical example of the accuracy evaluation.

The functions of the application:

- Allow the users to distinguish the most and the least accurate areas
- Display the bar graph of the average position accuracy within different FIRs in French airspace
- Plot the bar graph of the average position accuracy of the Domestic and International flights
- Plot the pie chart where the user will be able to observe the percentual share of data whose accuracy is better than 10 NM within the data sample

The files used for the programming of the General Statistics application from Attachment 1 are: views.py, urls.py, and general_statistics.html. The Python file views.py contains a function called General_statistics, which is responsible for rendering the map, making the necessary calculations, and calling the general_statistics.html file. The HTML file, where all the data are used and visualised with the intention of achieving the final product which, is shown in Figure 39.
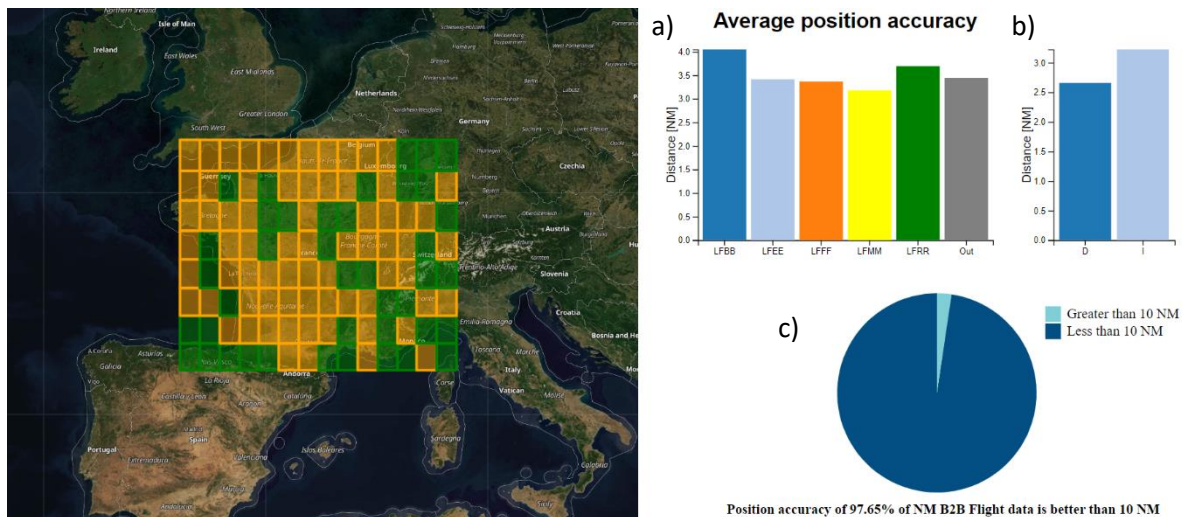
Figure 39: General statistics application

The source of information for this application is the Position database table. For the rendering of the map to the background of the HTML element, we have used the folium map. The source code within the General_statistics function loads the data from the database, then creates a rectangle shape that is composed of even squares with a width and height of 1 degree. Afterwards, all the data are grouped into their squares, and the average accuracy of the overall distance is calculated for each of the squares. The coloring is based on the condition shown in Figure 40.

```python
if avg_accuracy > 3:
    color = 'orange'
elif avg_accuracy >10:
    color = 'red'
else:
    color = 'green'
```

Figure 40: Condition of the colour shades

In Figure 39, we can also see 2 bar charts. Bar chart a) shows the average accuracy measured in concrete FIRs, while bar chart b) depicts the average accuracy of the domestic and international flights. The pie chart c) then shows the percentual share of the data points where we the measured overall distance was better than 10 NM. In this case, we have used the data sample with 10 000 data points, which we can interpret in the way that the accuracy of 97,68% from the chosen data sample is better than 10 NM.

4.3.4 Detailed statistics application

The web application that carries the name Detailed Statistics can be potentially used for further real-time analysis of the NM B2B Flight data accuracy. The objective was to allow the user to retrieve only specific data that fulfill user's needs. This way, we were able to create the application with the following functions:

- Allow the user to select a specific flight
- Allow the user to select a specific date
- Allow the user to select a specific route
- Display all the data points (which fulfill user's stated conditions) on the map
- Plot the line graph of the overall distance
- Plot the line graph of the perpendicular distance
- Plot the line graph of the offset distance

The files within the Attachment 1 which are responsible for rendering this application are views.py, urls.py, and detailed_statistics.html. The urls.py is again used for declaring the concrete URL query. The Python file called views.py contains the Detailed_statistics function, which is responsible for getting the user's input from the HTML and retrieving only those data from the Position database that fulfill the user's input conditions. Once the required data are filtered, we may plot the data on a folium map. Folium is the Python library that we have used for plotting individual data points on a map. The colors of the markers represent the source of position information. The red color is assigned to Flight Radar 24, and the blue coloring represents the points from NM B2B Service. In this function, we also divide the necessary information into separate Python lists: nm_list, distance_list, perp_distance_list, and offset_distance_list. The elements within those lists are used by the detailed_statistics.html file for the creation of the line graphs. The average values of three kinds of distances we calculated within the function of the views.py file. All the rendered information is then filled into the library called context, which renders all the data into the HTML. In the HTML file, we created several HTML elements which hold the concrete information. For the graphical visualization of the line graphs, we use the D3 JavaScript library [55][55]. Thanks to the D3 library, we are able to create an SVG element which we use as our graph for the graphical representation of data.

The final design of the Detailed Statistics application is shown in Figure 41. It consists of 5 main components:

- Filter options
- Average distance values
- Map
- Line graph of the overall distance
- Line graph of the perpendicular distance
- Line graph of the offset distance

In this way, the user can the accuracy on a chosen route or during a specific day, or he may also combine several conditions to make the filtering even more precise.
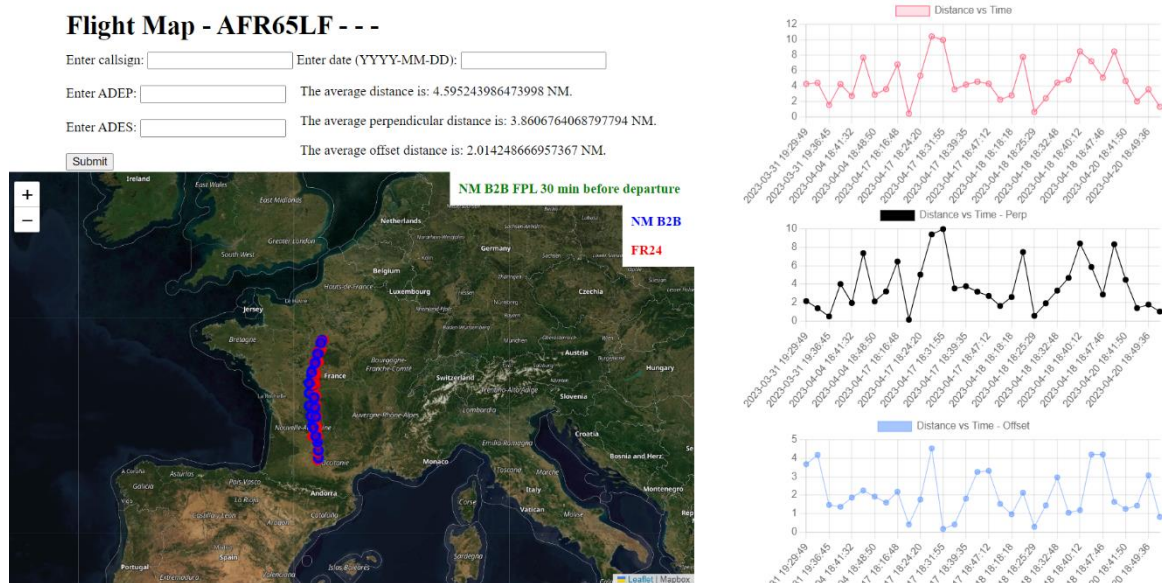
Figure 41: Detailed statistics application

5 Discussion

Upon analysing the NM B2B Flight data that we focused on in this study, we have revealed several important findings. As for now, there are no known publicly available documents that specifically discuss the accuracy of NM B2B data. As stated in the introduction of the work, it is necessary to point out the behaviour of data that we desire to use within our systems.

The concrete model that was evaluated in this study is the CTFM Point Profile created by the system ETFMS. It provides the most actual information about flight trajectory prediction. The access to the NM B2B Service, which is the source of the CTFM point profile information, is

provided by the company CS SOFT a.s.. The main aim is to study whether appropriate accuracy is maintained throughout the whole flight, as well as to determine any changes in different geographical areas. Before the actual evaluation, it was important to state how these data are updated by the Network Manager. As is apparent from the first chapters of this work, several update messages, such as FSA, CPR, DPI, etc., influence the data within the CTFM model with the intention of making the highest possible precision of the prediction. It is important to state that the CTFM point profile was created with the intention of providing the real-time air traffic situation within the IFPS zone. It is not designed to be used as a primary source for the radar view as it may deviate from the actual position in a significant way. On the other hand, the main purpose of the CTFM point profile is to support air traffic flow and capacity management. It plays an important role in predicting and monitoring the sector counts, the workload of air traffic controllers, and optimizing the traffic flow. Even by taking these facts into account, we tried to evaluate the accuracy of the NM B2B Flight Data by comparing them to several surveillance information sources. Within this study, we used data from sources like Flight Radar 24, OpenSky Network, and Secondary Surveillance Radars.

One of the limitations of this study is the fact that we compare the flight data on the whole flight trajectory, excluding the standard instrument departure and standard arrival route. As the CTFM profile doesn't include detailed information about the concrete points on SIDs and STARs, it would not be appropriate to make the position approximation at this stage of flight. The main reason is that it wouldn't be appropriate to include these calculations that might mislead the results of the NM B2B data accuracy.

We focused on position deviations as well as flight level differences. The deviations of flight level information were analysed for an en-route phase. This approach was determined in order to detect any pattern which would suggest that the change in the vertical profile of the flight might not be caught by the NM B2B Flight data. But after observations of several data samples in different areas, we discovered an appropriately high precision, which is also presented by the histogram in Chapter 3.3.1.1. The precision of the flight level differences is less than 200 ft with a probability level of 99%. Accuracy up to 200 ft can be primarily assigned as the error caused by the equipment used and the sudden meteorological change. After revealing these results, we have concluded to focus more closely in this study on the position data deviations.

For the final evaluation, we used several statistical methods that helped us achieve a complex overview of the precision of NM B2B data. We have combined descriptive and inferential statistics together with spatial and point pattern analysis methods. Each of them played an important role in this study. Descriptive statistics helped us understand the basic information about the data, such as the distribution of the measured values, mean median, range, and

much more. The histograms in Chapter 3.3.1.1 showed that most of the distance values are located near 0 NM, and with higher values, we observed an exponential decrease in frequency. One of the main findings of this thesis was the evaluation of the overall distance between the position information from NM B2B and the Surveillance information sources. We have focused on three different parts of the IFPZ. When we look at the western part, the mean value of the overall distance between NM B2B and Surveillance sources is 3,51 NM. If we take into consideration the CTFM recalculation threshold, we can conclude that the overall distance values are less than 9,6 NM with a probability level of 95%. These facts imply that 95% of the distance values will be less than 9,6 NM, which satisfies the lateral deviation threshold of 10 NM. The same approach can be used for the central and south-eastern parts of IFPZ. For the south-eastern part, the mean value of the overall distance attribute was 3,65 NM. It shows that with a probability level of 95%, the values of distance will also be less than 10 NM. In the case of central IFPZ, the mean value is 3,13 NM which states that with the probability level of 95%, the measured values will be in magnitude smaller than 8,5 NM. The 10 NM threshold may seem large when we compare it to the minimum separation distance in the en-route phase, which is 5 NM. This fact suggests that it should not be used for real-time tracking. On the other hand, it depicts that the accuracy is sufficient for short-term decision-making and managing the air traffic flow. But the precision of NM B2B Flight data could be improved by adjusting the 10 NM threshold. This threshold seems to be taking three important factors into account: systems performance, the requirement of the accurate position, and the potential workload of ATCs.

From the view of the spatial and point pattern analysis, the utilization of the Kernel density was proven to be highly essential in this study. By plotting the results of the kernel density function on a map, we could clearly see the differences in NM B2B Flight Data accuracy in different geographical areas. We have discovered that most of the NM B2B data points, with the accuracy better than 1 NM, are situated near the busiest airports in the observed area. To prove this statement depicted from the usage of the kernel density function, we used both descriptive and inferential statistical methods. When we plotted all the measured distance values on several trajectories for the flights passing through the most accurate areas on a line graph, we could observe an apparent increase in data precision. An interesting finding was the fact, that even if the aircraft just flies over the most accurate area, the precision of the NM B2B Flight data improves. The usage of inferential statistics also supported this statement. By the calculation of the p-value, which resulted as much smaller than 0,05, we were able to reject the null hypothesis that implied no significant difference in values of distance between the area with the most accurate data and outside the area. There are several possible reasons, why the accuracy improves near the busiest airports. Better ground-based infrastructure with a high

density of radars could be one of the reasons for getting the updates on aircraft position more often. Another fact that could be pointed out is that the areas in western and central IFPZ recognized as the most accurate are all located nearby the A-CDM airports. As the airport collaborative decision-making improves the predictability of air traffic, we can assume that it also plays an important role in the predictions of the NM B2B Flight data. [56] Another factor affecting the predictions of the CTFM point profile could be the density of the air traffic nearby the major airports. We experienced higher accuracy for aircraft flying over the mentioned airports. The higher demand in such areas could result in more frequent position updates, as we need to achieve the appropriate level of safety by keeping the distances between the aircraft. This might require that the systems send the CPR messages more often, which also leads to the improvement of the NM B2B Flight Data accuracy.

Consequently, we divided each area into 5 sectors and compared the distance values between each other. An interesting finding is shown in Table 8 where the most accurate data are by the average situated in central Europe, followed by the western and south-eastern Europe. These differences could be caused by different weather conditions or the density of the airspaces as well as the quality of the input data. In the south-eastern part of Europe, there is no A-CDM airport. This fact could influence the accuracy of the NM B2B data, and therefore we observed worse overall distance values. Also, coastal areas that we can observe in France or Greece are affected by more distinct weather patterns that might worsen the NM B2B data accuracy.

Besides the overall distance, we also focused on the perpendicular and the offset distance between the NM B2B and the Surveillance sources. We observed an interesting pattern when examining the differences between the western, central, and south-eastern parts of IFPZ. We discovered that the perpendicular distance values were the most precise in the south-eastern part, followed by the central and western parts. Considering this fact, it might suggest that the less complex airspace with less density of flights can result in better accuracy of the perpendicular distance. The offset distance values resemble the same pattern as the overall distance values, which might suggest the same reason of the disparities between the sectors.

To make the evaluation more complex, we attempted to review the position of data points from NM B2B Service in relation to the surveillance position information. We calculated and determined whether the NM B2B position is situated on the left or the right side, as well as whether it is in front or behind the actual position. By comparing individual sectors with each other within the same IFPZ area, we recognized that most of the data points in sector H are situated behind. It means that most of the time, the actual position of flights in this sector is slightly in front of the prediction included in the CTFM point profile. For sector G, we also

discovered that the majority of the NM B2B data points are situated slightly to the left side in relation to the actual position.

Another way to analyze the NM B2B flight data accuracy is by dividing the flights into two categories: domestic and international flights. We have proved that overall, for the domestic flights, we observed better accuracy when compared to the international flights. The reason could be that domestic flights usually fly through fewer airspace sectors as they tend to operate in smaller geographical area. On the other hand, international flights could encounter more complex flight trajectories and therefore the accuracy could be affected by the coordination of multiple air traffic control centres in different countries and regions.

With the intention of making the most of the obtained flight data, a web-based system was created. It composts of three web applications, which are directly connected to the database tables filled with NM B2B and surveillance data. The creation and maintenance of the database make it possible to retrieve also historical data. Consequently, the users of the applications will be able to create statistics for any flights from the database. It also depicts the fact, that the database can be used for further analysis of NM B2B Flight data.

The first application within the web-based system is called Radar View. The main purpose of this application was to plot actual aircraft position from two sources at once. The sources used for the flight tracking are NM B2B Service and OpenSky Network. It clearly displays the reason why the NM B2B Flight data are not specifically meant for radar view in ATC systems. The differences between the actual positions are clearly visible in real-time. This kind of application could be, for example, used by other companies for performing simulations in concrete areas of IFPZ.

The second application carries the name General Statistics. It allows the user to see the differences in data accuracy of the French airspace. Also, it points out the percentage share of the data, which satisfies the CTFM recalculation threshold of 10 NM.

The third and most complex application is called Detailed Statistics. It allows the users to filter solely according to their needs. It is also capable of plotting the whole trajectory on the map while recognizing different data sources. Further, it calculates the average overall, perpendicular, and offset distance on the selected flight routes. And another interesting feature is that the user will be able to observe how the distance values changed in time. This web application could be used for comparing the NM B2B Flight data accuracy in different days, on different flight routes as well as for different aircraft.

6 Conclusion

The Network Manager business-to-business service brings significant opportunities for external companies as they may use its data in their own systems. With this intention in mind, it is essential to carefully analyze and recognize its behavior, so then we can adequately implement them for appropriate purposes. Access to this service was made possible by the company CS SOFT a.s., which allowed us to perform the analysis of the NM B2B data for the purposes of the study.

The research, made on behalf of the thesis, focused on the accuracy evaluation of the NM B2B Flight data and provided valuable results for enhancing aviation technologies. The analysis performed was characterized by its complexity, which was achieved by several factors. The utilization of multiple statistical methods, such as descriptive and inferential statistics, point pattern, and spatial analysis techniques, helped us to achieve a clear overview of the observed data behavior. To make it possible, multiple surveillance information sources were used in three robust but different geographical areas. The position information used for the analysis and gathered from NM B2B Service is generated in the form of the CTFM point profile. It incorporates the most up-to-date information on the aircraft's position supplied by the ETFMS system. Furthermore, we aimed our attention at evaluating multiple measures, such as the overall, perpendicular, and offset distance between the NM B2B and the Surveillance position of the aircraft. As both sources use different approaches for determining the aircraft's position, we could observe considerable differences. In each of the observed geographical areas, the accuracy of NM B2B Flight data was proven to be better than 10 nautical miles for 95% of the data.

One of the most significant findings was the fact that the accuracy of the NM B2B Flight data substantially increases in the areas of major airports in cities like Paris, Geneve, Nice, Prague, and Vienna. Several other factors, such as the weather conditions and the density of the airspace, play an important role in data accuracy as well.

At last, we aimed to create a software solution consisting of several web applications that would point out the pros and cons of the NM B2B data and allow the users to specifically analyse the accuracy on the chosen route or on a given day. The database that is used as the primary source for the web-based system holds and updates the data in real time, making it possible to carry on with the analysis for further research. The web applications were essential in identifying the real behavior of the NM B2B data.

The findings of the thesis can be beneficial for the aviation industry as they provide valuable insights into the accuracy and behavior of the NM B2B Flight data. Additionally, the outcomes of the analysis, together with the software solution, may prove to be useful for other external companies to better understand the nature of NM B2B data which broadens the potential of this service. In conclusion, the evaluation of NM B2B Flight data accuracy presents a solid foundation for further research and emphasizes the development of the service for achieving greater accuracy and reliability which could consequently ensure better efficiency of the flight operations.

Bibliography

[1]   MENDES VIDEIRA, Idalina. NM Interoperability strategy, Version 1. 2016. Available at: https://www.eurocontrol.int/sites/default/files/publication/files/nm-interoperabilitystrategy-may-2016.pdf

[2]   EUROCONTROL Area. Aviation Intelligence Unit Portal. Aviation Intelligence Unit Portal [online]. Copyright © 2023 [cit. 01.11.2022]. Available at: https://ansperformance.eu/definition/eurocontrol-area

[3]   EUROPEAN UNION. Commission Implementing Regulation (EU) 2019/123 of 24 January 2019 laying down detailed provisions for the implementation of the common basic standards on aviation security. Official Journal of the European Union, L 25, 1-119.

[4]   Network operations. EUROCONTROL. [online] Copyright © EUROCONTROL [cit. 10.02.2023]. Available at: https://www.eurocontrol.int/network-operations

[5]   Network Manager Operations Centre (NMOC). SKYbrary Aviation Safety. SKYbrary Aviation Safety [online]. Copyright © SKYbrary Aviation Safety, 2021 [cit. 10.01.2023]. Available at: https://skybrary.aero/articles/network-manager-operations-centre-nmoc

[6]   NM Main Areas of Operations. EUROCONTROL.[online]. Copyright © EUROCONTROL 2017 [cit. 15.12.2022]. Available at: https://www.nm.eurocontrol.int/STATIC/NM_AREA/

[7]   WOOLIN Andy, IFPS Users Manual, Network Manager, Version 27.0, 2023. IFPS manual 2023. Available at: https://www.eurocontrol.int/sites/default/files/2023-03/eurocontrol-ifps-user-manual-27-0.pdf

[8]   Flight plan filing and management. EUROCONTROL[online]. Copyright © EUROCONTROL [cit. 10.12.2023]. Available at: https://www.eurocontrol.int/service/flight-plan-filing-and-management

[9]   KULČÁK, Ludvík. Air Traffic Management. Brno: Akademické nakladatelství CERM. 2002. ISBN 80-7204-229-7. Language: Czech

[10]   KOOLEN Hans, Enhanced Tactical Flow Management System (ETFMS). Presentation. 2013. Available at: https://slideplayer.com/slide/753420/

[11]   Enhanced tactical flow management system (ETFMS).EUROCONTROL[online]. Copyright © EUROCONTROL [cit. 10.12.2023]. Available at: https://www.eurocontrol.int/system/enhanced-tactical-flow-management-system

[12]   NM NOP/B2B TEAM. NM 23.0.0 - NOP/B2B Reference Manuals – FlightServices, Verzia 23.0.0.5.26. 2019. Sensitive document file.

[13]   KOOLEN, Hans a COLIBAN, Ioana. Flight Progress Messages. version 2.800. 2022. Available at: https://www.eurocontrol.int/publication/flight-progress-messages-fpmdocument

[14]   SFYROERAS, M. and VEGA, V. ATFCM Operations manual. Network Manager. EUROCONTROL. Version: 27.0. 2023. Available at: https://www.eurocontrol.int/sites/default/files/2023-03/eurocontrol-atfcm-operations-manual-27-0-wef-20230425.pdf

[15]   Data collection service (DCS).EUROCONTROL[online]. Copyright © EUROCONTROL [cit. 11.12.2023]. Available at: https://www.eurocontrol.int/service/data-collection-service

[16]   Network Manager business-to-business web services (NM B2B). EUROCONTROL[online]. Copyright © EUROCONTROL [cit. 03.01.2023]. Available at: https://www.eurocontrol.int/service/network-manager-business-business-b2b-web-services

[17]   Coding guidelines per subject - (ICAO) AIP Data Set - AIXM Confluence [online]. [cit. 03.01.2023]. Available at: https://ext.eurocontrol.int/aixm_confluence/display/ACGAIP/Coding+guidelines+per+subject

[18]   PÁNIS, Marián. The use of flight data for Application showing actual position and flight route. Czech Technical University. 2021.

[19]   ICAO. Air Traffic Management. Doc 4444. 16th edition. 2016. Available at: https://ops.group/blog/wp-content/uploads/2017/03/ICAO-Doc4444-Pans-Atm16thEdition-2016-OPSGROUP.pdf

[20]   SVYD, Iryna. Model and method for request signals processing of secondary surveillance radar. *IEEE 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM)*. IEEE, 2019. p. 1-4. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8779347

[21]   Secondary Surveillance Radar (SSR). SKYbrary Aviation Safety[online]. Copyright © SKYbrary Aviation Safety, 2021 [cit. 01.03.2023]. Available at: https://skybrary.aero/articles/secondary-surveillance-radar-ssr

[22]   GALATI, G., LEONARDI, M., MAGARO, P. and PACIUCCI, V. Wide Area Surveillance using SSR Mode S Multilateration: advantages and limitations. Tor Vergata University, DISP and Vito Volterra Centre. 2013. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1605606

[23]   SCHAFER, M., STROHMEIER, M., STROHMEIER, SMITH, M. PINHEIRO, R., LENDERS, V., MARTINOVIC, I. OpenSky Report 2016: Facts and Figures on SSR

Mode S and ADS-B Usage. 2017. Available at:

https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7778030

[24]   TOPKOVÁ, Tereza and PLENINGER, Stanislav. Czech Technical University in
       Prague. Department of Air Transport. MATEC Web Conf. 2018. Available at:
       https://www.matec-
       conferences.org/articles/matecconf/pdf/2018/93/matecconf_bultrans2018_01006.pdf

[25]   MANTILLA – GAVIRIA, I., LEONARDI, M., GALATI, G., BALBASTRE-TEJEDOR, J.
       Time-difference-of-arrival regularised location estimator for multilateration systems.
       The Institution of Engineering and Technology. IET Radar, Sonar and Navigation.
       2014. Available at:
       https://ietresearch.onlinelibrary.wiley.com/doi/pdfdirect/10.1049/iet-rsn.2013.0151

[26]   CARTER, M. Hyperbola. Presentation. 2016. Available at:
       https://slideplayer.com/slide/7853484

[27]   Automatic dependent surveillance – broadcast (ADS-B). EUROCONTROL.[online].
       Copyright © AIRBUS [cit. 03.01.2023]. Available at:
       https://www.eurocontrol.int/service/automatic-dependent-surveillance-broadcast

[28]   STROHMEIER, Martin, SMITH, Matt, SCHAFER, Matthias, LENDERS, Vincent and
       MARTINOVIC, Ivan. Assessing the Impact of Aviation Security on Cyber Power.
       2016. DOI: 10.1109/CYCON.2016.7529437. Available at:
       https://www.researchgate.net/publication/305719081_Assessing_the_Impact_of_Avi
       ation_Security_on_Cyber_Power

[29]   About Us. The OpenSky Network - Free ADS-B and Mode S data for Research
       [online]. Copyright © 2013 [cit. 01.01.2023]. Available at: https://opensky-
       network.org/about/about-us

[30]   STROHMEIER, M. and OLIVE, X. and LUBBE, J. and SCHAFER, M. and
       LENDERS, V. Crowdsourced air traffic data from the OpenSky Network 2019—
       2020. Earth System Science Data. Vol. 13. 2021. DOI: 10.5194/essd-13-357-2021.
       Available at: https://essd.copernicus.org/articles/13/357/2021/

[31]   SCHAFER, M., STROHMEIER, M., LENDERS, V., MARTINOVIC, I. WILHELM, M.
       Bringing up OpenSky: A Large-scale ADS-B Sensor Network for Research. IPSN-14
       Proceedings of the 13th International Symposium on Information Processing in
       Sensor Networks. Germany. 2014. doi: 10.1109/IPSN.2014.6846743

[32]   Live Flight Tracker - Real-Time Flight Tracker Map. Flightradar24. [online].
       Copyright © 2023 Flightradar24 AB [cit. 01.01.2023]. Available at:
       https://www.flightradar24.com/how-it-works

[33]   KALAGIREVA, Karina and RADKOV, Veselin. Displaying the air situation through
       the collection and processing of Flight information on Flightradar24 project. Scientific

research and education in the Air Force – AFASES. 2016. Department of Aeronautics, Technical University of Sofia, DOI: 10.19062/2247-3173.2016.18.1.36.

[34]   BRUCE, P., BRUCE A. and GEDECK, P., Practical Statistics for Data Scientists. Second Edition. O Reilly media. 2017. ISBN: 978-1-492-07294-2

[35]   JOSHUA, F., WILEY, A. Descriptive Statistics and Exploratory Data Analysis. Indiana USA. 2015. Available at: https://link.springer.com/chapter/10.1007/978-1-4842-0373-6_8

[36]   FlightRadarAPI · PyPI The Python Package Index [online]. Copyright © 2023 [cit. 01.03.2023]. Available at: https://pypi.org/project/FlightRadarAPI/

[37]   The OpenSky Network API documentation — The OpenSky Network API 1.4.0 documentation. GitHub Pages [online]. (cit. 10.02.2023) Available at: https://openskynetwork.github.io/opensky-api/

[38]   Calculate distance and bearing between two Latitude/Longitude points using haversine formula in JavaScript. Movable Type — Information Design & Management [online]. Copyright © 2002 [cit. 09.10.2022]. Available at: https://www.movabletype.co.uk/scripts/latlong.html?

[39]   Point-Line Distance--3-Dimensional -- from Wolfram MathWorld. Wolfram MathWorld: The Web's Most Extensive Mathematics Resource [online]. Copyright ©1999 [cit. 04.01.2023]. Available at: https://mathworld.wolfram.com/Point-LineDistance3-Dimensional.html

[40]   BURROUGH, P., MCDONNELL, P. and LLOYD, Ch. Principles of Geographical Information Systems. Third Edition. 2015. ISBN: 9780198742845

[41]   LARSON, Ron and EDWARDS, Bruce. Multivariable Calculus. Ninth Edition. Cengage Learning. 2020. ISBN-13: 978-0-547-20997-5

[42]   FISHER, M., MARSHALL, A. Understanding descriptive statistics, Australian Critical Care, Volume 22, Issue 2, 2009, Pages 93-97, ISSN 1036-7314, https://doi.org/10.1016/j.aucc.2008.11.003.

[43]   MANIKANDAN, S. Measures of central tendency: Median and mode. Postgraduate Corner. 2016. Available at: https://gyansanchay.csjmu.ac.in/wp-content/uploads/2022/10/Measures_of_central_tendency_Median_and_mode.pdf

[44]   CHAKRABATY, D. Measuremental Data: Seven Measures of Central Tendency. International Journal of Electronics and Applied Research (IJEAR). Vol:8, Issue:1, 2021. Available at: https://www.researchgate.net/profile/Dhritikesh-Chakrabarty/project/HUMAN-RESOURCES-MANAGEMENT-AND-DEVELOPMENT-FROM-GOOD-PRACTICES-TO-SUSTAINABLE-DEVELOPMENT-AND-EFFICIENCY-

2/attachment/61c5b6ccf5675b211b19209a/AS:1104510952976384@16403473405
32/download/Paper-150.pdf?context=ProjectUpdatesLog

[45]   MANIKANDAN, S. Measures of dispersion. *Journal of Pharmacology and Pharmacotherapeutics,* 2011.. doi:https://doi.org/10.4103/0976-500X.85931

[46]   KAUR, P., STOLTZFUS, J. and YELLAPU, V. Descriptive statistics. Department of Research and Innovation, The Research Institute. 2018. Available at: https://www.ijam-web.org/article.asp?issn=2455-5568;year=2018;volume=4;issue=1;spage=60;epage=63;aulast=Kaur

[47]   SCHABENBERGER, Oliver, GOTWAY, Carol A. *Statistical methods for spatial data analysis.* CRC press, 2017. Available at: https://books.google.fr/books?hl=sk&lr=&id=I20NDgAAQBAJ&oi=fnd&pg=PP1&dq=statistical+analysis&ots=HudGxBlvq9&sig=EA77L1HJC9xafzQJyeCuuywfo2A&redir_esc=y#v=onepage&q=statistical%20analysis&f=false

[48]   Chapter 11 Point Pattern Analysis. Intro to GIS and Spatial Analysis. Main repos. mgimond.github.io [online]. Available at: https://mgimond.github.io/Spatial/chp11_0.html

[49]   ROSNER, B. and GROVE, D. The use of the MannWhitney test for clustered data. Statistics in medicine. Statist. Med. 18, 1387}1400. 1999. Available at: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291097-0258%2819990615%2918%3A11%3C1387%3A%3AAID-SIM126%3E3.0.CO%3B2-V

[50]   RICE, John A. *Mathematical statistics and data analysis.* Cengage Learning, 2006. Available at: https://books.google.fr/books?hl=sk&lr=&id=7SI8AAAAQBAJ&oi=fnd&pg=PR4&dq=Mathematical+Statistics+and+Data+Analysis%22+by+John+A.+Rice&ots=-DVOlmDY7N&sig=FuooRGz9tY1Jvq4cB1xboLLB8nc&redir_esc=y#v=onepage&q=Mathematical%20Statistics%20and%20Data%20Analysis%22%20by%20John%20A.%20Rice&f=false

[51]   KESAR, S. and MINGE, X. Bootstrap: A Statistical Method. Rutgers University. 2008. Available at: https://statweb.rutgers.edu/mxie/stat586/handout/Bootstrap1.pdf

[52]   RICHTA Karel a SOCHOR, Jiří. Softwarové inženýrství 1. Vydavatelství ČVUT. Fakulta elektrotechnická. 1996.

[53]   ARLOW, Jim, NEUSTADT, Ila a KISZKA Bogdan. UML 2 a unifikovaný proces vývoje aplikací: objektově orientovaná analýza a návrh prakticky. Brno: Computer Press, 2007. ISBN 978-80-251-1503-9. Language: Czech

[54]   OpenLayers - Welcome. OpenLayers - Welcome [online]. . [cit. 01.02.2023]. Dostupné z: https://openlayers.org/

[55]   Home · d3/d3 Wiki · GitHub. GitHub: Let's build from here · GitHub [online].
       Copyright © 2023 GitHub, Inc. [cit. 01.02.2023]. Available at:
       https://github.com/d3/d3/wiki

[56]   EUROCONTROL Airport CDM Team. Airport CDM Manual- Version 5.0. 2017.
       Available at: https://www.eurocontrol.int/concept/airport-collaborative-decision-
       making

Attachments

Attachment 1   Web – based system

The data within the database that were used for the NM B2B Flight data accuracy evaluation are protected and, therefore, cannot be included in the attachment of the master's thesis.