# Project 060 Analytical Methods for Expanding the AEDT Aircraft Fleet Database

## Georgia Institute of Technology

## Project Lead Investigator

Principal Investigator:
Professor Dimitri N. Mavris
Director, Aerospace Systems Design Laboratory
School of Aerospace Engineering
Georgia Institute of Technology
Phone: (404) 894-1557
Fax: (404) 894-6596
Email: dimitri.mavris@ae.gatech.edu

Co-Principal Investigator:
Dr. Yongchang Li
Senior Research Engineer
Aerospace Systems Design Laboratory
School of Aerospace Engineering
Georgia Institute of Technology
Phone: (404) 385-2776
Fax: (404) 894-6596
Email: yongchang.li@ae.gatech.edu

## University Participants

**Georgia Institute of Technology**

- PIs: Dr. Dimitri Mavris, Dr. Yongchang Li
- FAA Award Number: 13-C-AJFE-GIT-065
- Period of Performance: June 5, 2020 to June 4, 2021
- Tasks:
    1. Identification and Review of Aircraft not in AEDT
    2. Analytical Method Development

## Project Funding Level

The current FAA funding for this project is $150,000 from June 5, 2020 to June 4, 2021. The Georgia Institute of Technology has agreed to a total of $150,000 in matching funds.

## Investigation Team

- Dr. Dimitri Mavris (PI), Georgia Institute of Technology, oversees the entire project.
- Dr. Yongchang Li (Co-PI), Georgia Institute of Technology, oversees the entire project and leads the research team in analyzing the AEDT ANP aircraft and developing the analytical methods to expand the FLEET database.
- Dr. Michelle R. Kirby (research staff), Georgia Institute of Technology, oversees the entire project and supports all of the research activities.
- Dr. Holger Pfaender (research staff), Georgia Institute of Technology, provides consultation and support.
- Zhenyu Gao (graduate student) and Ying Chen (undergraduate student), Georgia Institute of Technology, conduct the development of analytical methods derived from statistical learning methods and techniques.
- Bogdan Dorca (graduate student), Georgia Institute of Technology, performs literature study on databases for collecting performance, emissions and noise data for aircraft not modeled with ANP data.

- Chrysoula Pastra (graduate student) and Justin Coleman (undergraduate student), Georgia Institute of Technology, work on investigating and validating the ANP aircraft substitution method.
- Fabio Chiappina (undergraduate student), Georgia Institute of Technology, conducts the analysis to match ANP aircraft substitution with the equipment in AEDT FLEET DB.

# Project Overview

The Aviation Environmental Design Tool (AEDT) relies on the aircraft noise and performance (ANP) data provided by aircraft manufacturers to support the calculation of aircraft trajectories and noise at receptors using aircraft performance information and noise-power-distance (NPD) relationships for specific aircraft/engine combinations. In the ANP/BADA (Base of Aircraft Data) workflow, the ANP performance data is also used in the calculation of emissions inventories and air quality dispersion. However, not all aircraft in the fleet are represented in the ANP database. When ANP data is not available for a specific target engine/airframe combination, AEDT uses a substitute aircraft from the ANP database to model the target aircraft by closely matching certification noise characteristics and other performance parameters. A problematic issue, however, is that the best substitute based on noise criteria does not always match the best substitute for emissions criteria. In addition, substitute aircraft do not capture the environmental benefits of newer aircraft with noise and emissions reduction technologies, resulting in overly conservative noise and emissions estimates.

The goal of this research is to improve the accuracy of AEDT noise and emissions modeling of aircraft not currently in the ANP database. Georgia Tech will identify and review the aircraft not currently modeled in AEDT and collect information and necessary data to better understand the characteristics of the aircraft. Various statistical analysis methods will be utilized to classify the aircraft as different aircraft types in terms of size, age, technologies, and other engine/airframe parameters. Quantitative and qualitative analytical methods will be identified and evaluated for each aircraft type to develop the ANP and noise data for the aircraft. Validation data from real-world flight and physics-based modeling will be gathered to validate the methods. The Environmental Design Space (EDS) will be employed to generate NPD curves for the aircraft using physics-based modeling and simulation of new and existing aircraft designs and technologies which can support the method validation analysis. After the methods are validated, they will be applied to develop ANP and noise data for the aircraft. Finally, recommendations and guidelines will be developed for how to implement the developed data in AEDT to expand the AEDT FLEET database to include noise and performance data for the aircraft currently not in the ANP database.

# Task 1 – Identification and Review of Aircraft not in AEDT

Georgia Institute of Technology

## Objective

The objective of Task 1 is to identify the aircraft that are not currently modeled with ANP data in AEDT for noise and emissions modeling. In the FLEET database (DB), specific aircraft engine/airframe combinations are defined by a series of ANP and noise coefficients that are used with the BADA and SAE-AIR-1845 algorithms to conduct performance, emissions, and noise modeling. The FLEET DB contains representative aircraft of the entire fleet, with some modeled with ANP data and others represented by a substitution aircraft. This Task involves the identification of the aircraft that do not have ANP data and are represented by a substitution aircraft.

## Research Approach

The aircraft that are not currently modeled with ANP data are identified by reviewing the AEDT FLEET DB and conducting a literature survey. The identified aircraft of interest are further investigated to understand the gaps between them and the substitution aircraft in terms of performance, noise, and emissions. This involves reviewing the existing literature on these aircraft and acquiring the information and data necessary to better under the engine/airframe characteristics of these aircraft. In addition, the ANP data in the FLEET DB are studied to summarize the key parameters for which the analytical methods will develop ANP data. The existing ANP aircraft substitution methods and the current substitution methods implemented in AEDT are also investigated to support the development of analytical methods.

**Task 1.1: Review the ANP Target-substitute Aircraft Pairs**

The objective of this task is to review and identify target-substitute aircraft pairs. First the ANP aircraft substitution table was discovered from Eurocontrol's website [Eurocontrol, 2017] in the form of a spreadsheet through a literature study. This spreadsheet was provided for jet and heavy propeller-driven aircraft to help noise modelers map the aircraft from a given traffic sample to the types available in the ANP database. In the absence of a list directly indicating which aircraft in the AEDT

database are substitution aircraft, MATLAB functions were created to facilitate searching through the AEDT database and the acquired ANP aircraft substitutions spreadsheet. Once the AEDT aircraft substitutions list is acquired, these MATLAB functions can quickly match the target aircraft to the aircraft in the AEDT database and ANP substitutions spreadsheet and record the relevant data. Example calls to these MATLAB functions are provided below.

In order to sort out the target-substitution aircraft pairs, the first step is to identify the substitution aircraft in the ANP aircraft substitution spreadsheet and their associated data from AEDT FLEET DB. This was conducted by developing a MATLAB function which takes in an ANP ID as its only parameter and outputs all entries in the ANP substitutions spreadsheet and AEDT aircraft database with an ANP ID matching the input. The outputs are in the form of a cell array that can then easily be written to a spreadsheet for further analysis. For example, calling this function with the input "DHC8" yields the following output shown in Figure 1.

ANPproxy

60x21 cell

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 'SOURCE' | 'RECORD_NUMBER' | 'ICAO_CODE' | 'AIRFRAM... | 'AIRFRAME_TYPE' | 'ENGINE_MANUFACTURER' |
| 2 | 'EASA TCDSN Heavy Prop' | 'B1169' | 'AT43' | 'ATR-GIE A... | 'ATR 42-200' | 'Pratt & Whitney Canada' |
| 3 | 'EASA TCDSN Heavy Prop' | 'B1216' | 'AT43' | 'ATR-GIE A... | 'ATR 42-200' | 'Pratt & Whitney Canada' |
| 4 | 'EASA TCDSN Heavy Prop' | 'B1170' | 'AT43' | 'ATR-GIE A... | 'ATR 42-300' | 'Pratt & Whitney Canada' |
| 5 | 'EASA TCDSN Heavy Prop' | 'B1171' | 'AT43' | 'ATR-GIE A... | 'ATR 42-300' | 'Pratt & Whitney Canada' |

AEDT

50x17 cell

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 'ANP ID' | 'Description' | 'Type' | 'Airframe Model' | 'Count' | 'Code' |
| 2 | 'DHC8' | 'DASH 8-100/PW121' | 0 | 'ATR 42-400' | 2 | 'PT6A45' |
| 3 | 'DHC8' | 'DASH 8-100/PW121' | 0 | 'ATR 42-500' | 2 | 'PT6A45' |
| 4 | 'DHC8' | 'DASH 8-100/PW121' | 0 | 'Antonov 32 Cline' | 2 | 'T56-1' |
| 5 | 'DHC8' | 'DASH 8-100/PW121' | 0 | 'ATR 42-200' | 2 | 'PW120' |

**Figure 1.** Queried data for ANP aircraft "DHC8."

The ANP proxy output shown above is a 60 x 21 cell array and contains many more columns of data than those shown, including maximum takeoff weight (MTOW), noise chapter, propeller type, and more. Similarly, the AEDT output is a 50 x 17 cell array that also contains various additional data.

Next, the target aircraft in the ANP aircraft substitution spreadsheet were studied. Each target aircraft is represented as a unique combination of airframe type and engine type. Another MATLAB function was developed which takes as its parameters an airframe type and engine type, outputting cell arrays with all data from the ANP substitutions spreadsheet with airframe type and engine type matching that of the input. It also outputs all entries in the AEDT database with an ANP ID matching that found in the ANP substitutions list. For example, calling this function with the inputs "BAe ATP" as airframe type and "PW126" as engine type yields the output shown in Figure 2.

ANP

7x10 cell

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 'ICAO... | 'AIRFRAME_MANU... | 'AIRFRAME_TYPE' | 'ENGINE_MAN... | 'ENGINE_TYPE' | 'ENGINE_NUM... | 'MTOW_KG' | 'MLW_KG' | 'NOISE... | 'ANP_PROXY' |
| 2 | 'ATP' | 'BAE Systems (Oper... | 'BAe ATP' | 'Pratt & Whitn... | 'PW126' | 2 | 22930 | 22250 | 4 | 'HS748A' |
| 3 | 'ATP' | 'BAE Systems (Oper... | 'BAe ATP' | 'Pratt & Whitn... | 'PW126' | 2 | 22930 | 22250 | 4 | 'HS748A' |
| 4 | 'ATP' | 'BAE Systems (Oper... | 'BAe ATP' | 'Pratt & Whitn... | 'PW126' | 2 | 22930 | 22250 | 5 | 'HS748A' |
| 5 | 'ATP' | 'BAE Systems (Oper... | 'BAe ATP' | 'Pratt & Whitn... | 'PW126' | 2 | 22930 | 22250 | 5 | 'HS748A' |
| 6 | 'ATP' | 'BAE Systems (Oper... | 'BAe ATP' | 'Pratt & Whitn... | 'PW126' | 2 | 23678 | 23133 | 4 | 'HS748A' |
| 7 | 'ATP' | 'BAE Systems (Oper... | 'BAe ATP' | 'Pratt & Whitn... | 'PW126' | 2 | 23678 | 23133 | 5 | 'HS748A' |

AEDT

75x17 cell

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 'ANP ID' | 'Description' | 'Type' | 'Airframe Model' | 'Count' | 'Code' | 'Model' |
| 2 | 'HS748A' | 'HS748/DART MK532-2' | 0 | 'Saab 2000' | 2 | '4AL003' | 'AE3007A' |
| 3 | 'HS748A' | 'HS748/DART MK532-2' | 0 | 'Saab 2000' | 2 | 'PW127A' | 'PW127-A ' |

**Figure 2.** Queried data for ANP aircraft with "BAe ATP" airframe type and "PW126" engine type.

The function finds the ANP ID associated with the aircraft in the ANP output cell array—"HS748A" in this example—and searches the AEDT database for this ANP ID, outputting all rows with a matching ANP ID.

With a complete list of which aircraft in AEDT are substitution aircraft, these MATLAB functions will prove quite useful in easily accessing ANP and AEDT data for each of these aircraft.

In addition to these MATLAB functions, the basic ANP data for each of the 112 unique ANP proxy aircraft IDs in the ANP aircraft substitutions spreadsheet were queried from the AEDT FLEET database. Each of the aircraft obtained from these SQL queries was compared in airframe and engine model to all entries in the ANP aircraft substitutions list; if the SQL aircraft matched an ANP aircraft in both airframe model and engine model, then the matching entries from each spreadsheet were recorded; that is, the airframe and engine model combination was recorded exactly as it appeared both in the SQL database and in the ANP substitutions spreadsheet. An example of this procedure is shown below.

First, ANP data for all 112 unique ANP proxy aircraft were queried from SQL and stored in a spreadsheet. For each of the 2466 rows in the resulting spreadsheet, the listed airframe model and engine model combination were to be searched for within the ANP substitutions spreadsheet. For example, in Table 1 we consider a particular row in the spreadsheet comprised of SQL data:

**Table 1.** Queried data from AEDT FLEET database for an aircraft with equipment ID 128

| EQUIP_ID | ANP_AIRPLANE_ID | AIRFRAME_ID | MODEL (AIRFRAME) | ENGINE_ID | MODEL (ENGINE) |
|---|---|---|---|---|---|
| 128 | 737 | 4588 | Boeing 737-200 Series | 1263 | JT8D-9 series |

The ANP aircraft substitutions spreadsheet is thus searched for a match to "Boeing 737-200 Series" and "JT8D-9" series. Though no aircraft in the ANP spreadsheet has this exact string as its airframe type, "737-200" does exist within the ANP substitutions spreadsheet; this is considered an airframe match. Next, we locate aircraft in the ANP spreadsheet with "JT8D-9" as their engine type, and an engine match was found as well. Since there are aircraft in the ANP spreadsheet with both airframe type of "737-200" and engine type of "JT8D-9", this is considered a double match—a match of both airframe and engine simultaneously.

In Table 2, a different row in the SQL data spreadsheet is considered.

**Table 2.** Queried data from AEDT FLEET database for an aircraft with equipment ID 2360.

| EQUIP_ID | ANP_AIRPLANE_ID | AIRFRAME_ID | MODEL (AIRFRAME) | ENGINE_ID | MODEL (ENGINE) |
|---|---|---|---|---|---|
| 2360 | 7478 | 5224 | Boeing 747-8F | 1357 | CF6-80C2B1F |

Though the ANP spreadsheet does have aircraft with the airframe type "747-8F" and aircraft with the engine type "CF6-80C2B1F", there are no aircraft that match both criteria simultaneously. Thus, a match is recorded for airframe type and for engine type, but a double match is not recorded.

The task of matching aircraft in the SQL database to aircraft in the ANP aircraft substitutions list remains in progress, as this task requires more than two thousand comparisons and cannot be easily automated by a computer script due to the different formats in the two databases. Upon receiving a list indicating which AEDT aircraft are substitution aircraft, this same matching process can be applied to the substitution aircraft in AEDT, and the relevant data in the ANP aircraft substitutions spreadsheet can thus be easily obtained.

**Task 1.2: Aircraft Database Literature Study**

The main objective of this task is to collect data from various publicly available databases for a wide range of aircraft. This information is helpful when determining which performance, emissions, and noise parameters will be used for the substitution algorithm.

With regards to performance, the following open-source databases have been identified:

- **AEDT ANP** (FLEET + FLEET-FULL databases). The most complete performance database available, the FLEET database contains multiple performance parameters for a wide variety of airframe/engine combinations. The FLEET-FULL database, while it is not publicly available, contains information for all the registered aircraft throughout the world. Minimum and maximum values for the same airframe/engine combinations are available for certain aircraft parameters (e.g., MTOW) depending on what aircraft equipment the airline is using onboard.
- **Aircraft Performance Database** [Eurocontrol, 2020]. This database contains performance data for a wide variety of aircraft. It includes data regarding:
    - Aircraft geometry (e.g., wingspan, height, length).
    - Aircraft structure (e.g., tail configuration, wing position, engine type, landing gear configuration).
    - Aircraft performance parameters (e.g., MTOW, range).
    - Aircraft profile (e.g., takeoff, climb, approach, landing, and respective speeds, distances, rates of climb/rates of descent (ROC/ROD)).

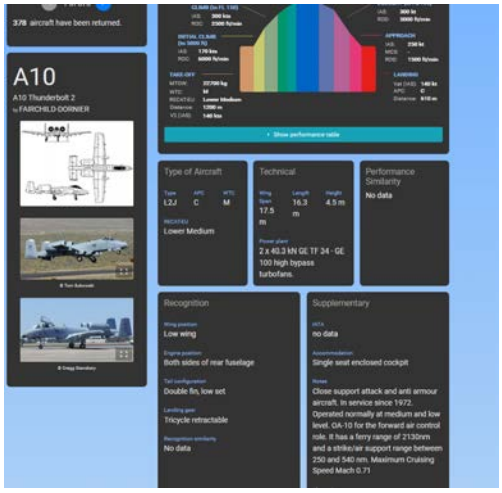


**Figure 3.** Aircraft performance database example.

- **RisingUp** [RingsUp Aviation, 2020]. This database contains performance data for a wide variety of general aviation aircraft (e.g., Cessna, Beechcraft). It includes information regarding:
    - Aircraft performance parameters (e.g., gross weight, empty weight, fuel capacity range).
    - Aircraft profile (e.g., takeoff, climb and landing and respective speeds, distances, ROCs/RODs, ceiling).

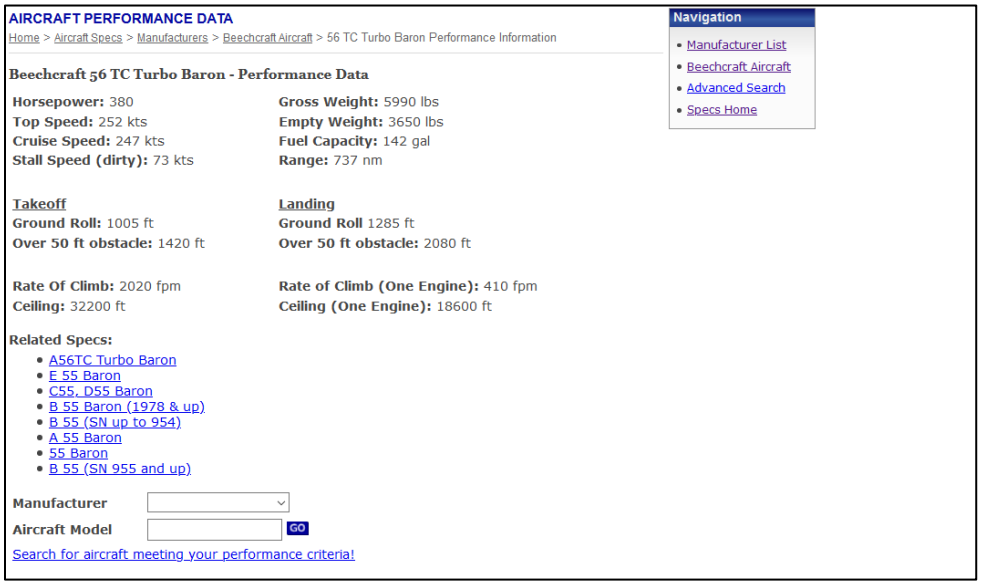


**Figure 4.** RisingUp Aviation Database example

- **OpenSky** [The OpenSky Network, 2020]. This database provides limited information about a specific existing aircraft (e.g., an A320 that is currently operating). It includes information such as:
    - Engine type.
    - Aircraft owner, airline operator, number of years in service, etc.
    - Live tracking of aircraft including altitude, velocity, track, current location.

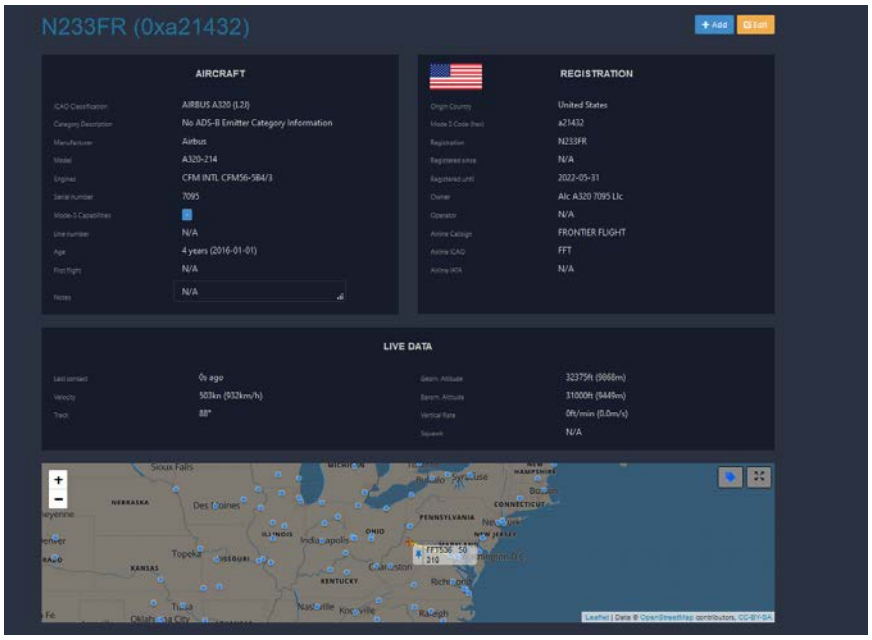


**Figure 5.** OpenSky Aircraft Database example

- **International Air Transport Association (IATA)**. This source only provides a passenger data exchange database, which does not specifically relate to performance.

With regard to emissions, the following open-source databases have been identified:

- **ICAO Aircraft Engine Emissions Databank** [International Civil Aviation Organization (ICAO), 2020]. This Excel-based database covers turbojet and turbofan engines for which the emissions are regulated (27 kN or higher static thrust). It contains engine emissions for CO, NOx, and UHC as well as a few engine performance parameters like bypass ratio and thrust rated output.

With regards to noise, the following open-source databases have been identified:

- **EASA** certification noise databases [EASA, 2020]. A very large Excel database consisting of aircraft/engine types and effective perceived noise in decibels (EPNdB) and noise levels for lateral, flyover, and approach. It is a collection of four noise databases that address heavy propeller driven aeroplanes, jet aeroplanes, light propeller driven aeroplanes, and rotorcraft.
- **ICAO Noise Database (NoiseDB)** [ICAO, 2020]. Similar to the EASA certification noise database.

In order to see if these databases contain enough information, a case study involving the aircraft with ANP ID "737800" was performed and data were gathered for the 98 combinations of:

- Airframe (e.g., Boeing 737-800 Series, Boeing 737-800 with winglets, Boeing 737-900 Series, Boeing 737-900-ER, MC-21-300, etc.)
- Engine (e.g., CFM56-7B24, CFM56-7B27/2, LEAP-1B25, PW1130G-JM, etc.)

An Excel document was made containing various parameters of interest for all of the possible combinations. For this case study, the following ANP parameters which are necessary for substitution were identified:

- Engine related parameters (ENGINE_CODE, ENGINE_MOD_CODE, ENGINE_MOD_DES, Number of Engines, Location of Engines, Type of Engine, Location of Engines, Engine Bypass ratio)
- Performance related parameters (INTRO_YEAR, SIZE_CODE, Maximum range, Max takeoff gross weight, Max landing gross weight, Static rated thrust or 100% thrust, in or out of production)
- Noise related parameters (EPNdB lateral/flyover/approach/cumulative levels, limits and margins)

**Table 3.** B737-800 airframe/engine combinations

| EQUIP_ID | ANP ID | BADA ID | AIRFRAME | ENGINE | ENGINE | ENGINE | ENGINE | FLT_AIRFRAMES | FLT_AIRFRAMES | ENGINE |
|---|---|---|---|---|---|---|---|---|---|---|
| EQUIP_ID | ANP ID | BADA3 ID | AIRFRAME | ENGINE | ENGIN_CODE | ENGINE_MOD | ENGINE_MOD | ENGINE_COUNT | ENGINE_LOCATION | BPR |
| EQUIP_ID | ANP ID | BADA3 ID | AIRFRAME MODEL | ENGINE MODEL | UID | ENGINE_MOD_CODE | ENGINE_MOD_DES | Number of Engines | Location of Engines | Engine Bypass ratio |
| 210 | 737800 | B738 | Boeing 737-800 Series | CFM56-7B24 | 140 | NONE | No engine modification. | 2 | W | 5.2 |
| 211 | 737800 | B739 | Boeing 737-900 Series | CFM56-7B24 | 140 | NONE | No engine modification. | 2 | W | 5.2 |
| | 737800 | B738 | Boeing 737-800 with winglets | CFM56-7B27/2 | 140 | NONE | No engine modification. | 2 | W | 5 |
| | 737800 | B739 | Boeing 737-900-ER | CFM56-7B27E/B1 | 140 | NONE | No engine modification. | 2 | W | 5.1 |
| | 737800 | B738 | MC-21-300 | CFM56-5B2/3 | 140 | NONE | No engine modification. | 2 | W | 5.5 |
| | 737800 | B738 | Boeing 737-800 with winglets | CFM56-7B24 | 140 | NONE | No engine modification. | 2 | W | 5.2 |
| | 737800 | B738 | Boeing 737-800 Series | CFM56-7B26/3 | 140 | NONE | No engine modification. | 2 | W | 5.1 |
| | 737800 | B738 | Boeing 737-800 Series | CFM56-7B26 | 140 | NONE | No engine modification. | 2 | W | 5.1 |
| | 737800 | B739 | Boeing 737-900 Series | CFM56-7B26/3 | 140 | NONE | No engine modification. | 2 | W | 5.1 |
| | 737800 | B739 | Boeing 737-900 Series | CFM56-7B26 | 140 | NONE | No engine modification. | 2 | W | 5.1 |
| | 737800 | B739 | Boeing 737-900-ER | CFM56-7B26/3 | 140 | NONE | No engine modification. | 2 | W | 5.1 |

**Task 1.3: Summary of ANP Data in AEDT FLEET DB**

To understand the ANP data that need to be developed for the target aircraft, the ANP data in AEDT FLEET database were studied. A summary of ANP data was made to identify the locations and data structures of all ANP parameters. Going forward, this summary is expected to facilitate the identification of variables that play important roles in developing ANP data for the target aircraft as well as provide fast searches of relevant variables. A full list of summarized tables is given below:

- FLT_AIRFRAMES
- FLT_ANP_AIRPLANES
- FLT_ANP_AIRPLANE_NOISE_GROUPS
- FLT_ANP_AIRPLANE_NPD_CURVES
- FLT_ANP_AIRPLANE_PROCEDURES
- FLT_ANP_AIRPLANE_PROCEDURES_EXT
- FLT_ANP_AIRPLANE_PROCEDURES_MAP
- FLT_ANP_AIRPLANE_PROFILE_POINTS
- FLT_ANP_AIRPLANE_PROFILE_POINTS_EXT
- FLT_ANP_AIRPLANE_PROFILES
- FLT_ANP_AIRPLANE_THRUST_GENERAL
- FLT_ANP_AIRPLANE_THRUST_JET
- FLT_ANP_AIRPLANE_THRUST_PROP
- FLT_ANP_AIRPLANE_THRUST_TSFC_COEFFICIENTS

The summary is made using an Excel sheet in a way that is easy to sort and filter information. Table 4 below shows a fraction of the complete ANP summary table, which only includes FLT_AIRFRAMES. As a small part of the complete table, Table 4 gives an example of how data is sorted in the ANP summary. It contains six main columns plus a "Note" column. The function of each column is described below:

- Table Name: This column provides information regarding which table in the FLEET database a variable belongs to. This column can be used to look at a specific table in FLEET database.
- Variable Name: This column shows the variable name in the FLEET database, as well as links between tables.
- Variable Description: This column includes the descriptions of the variables. It can be used to find and locate similar variables across the entire FLEET database. For example, by searching the term "engine", you will be able to find the location and information for variables such as "Number of engines", "Location of engines", "Type of engines", "Engine bypass ratio", etc. In addition, this column contains all the necessary information to locate crucial parameters that are mentioned in the aircraft grouping and substitution rules in Doc 29 (described in the next Task).
- Discrete (Y/N): Whether the variable is discrete or not. If the variable is discrete, it typically only contains a limited number of options. For example, for variable "Location of engines", it has three discrete levels: Fuselage/Tail, Internal, and Wing. In contrast, a continuous variable usually spans across a numeric range. The variable "Maximum range" is an instance of a continuous variable.
- Variable Possible Values: Possible values of a variable. For each variable, it is in the form of either a range or a set that includes all discrete values.
- Code Description (if applicable): If the variable is discrete and its possible values are non-numeric, this column contains explanations of the abbreviation codes.

**Table 4.** Example of the ANP summary table

| Table Name | Variable Name | Variable Description | Discrete (Y/N) | Variable Possible Values | Code Description (if applicable) |
|---|---|---|---|---|---|
| FLT_AIRFRAMES | MODEL | Detailed aircraft models | N | 1136 models | |
| FLT_AIRFRAMES | ENGINE_COUNT | Number of Engines | Y | 1, 2, 3, 4, 6, 8 | |
| FLT_AIRFRAMES | ENGINE_LOCATION | Location of Engines | Y | I, W, F | F Fuselage/Tail<br>I Internal<br>W Wing |
| FLT_AIRFRAMES | DESIGNATION_CODE | Link to FLT_CAT_DESIGNATIONS table | Y | G, C, M | C Civil<br>G General Aviation<br>M Military |
| FLT_AIRFRAMES | EURO_GROUP_CODE | Link to FLT_CAT_EUROS_GROUPS table | Y | SS, PP, JB, H2, TP, LB, JR, JL, JS, H1, JM | H1 Helicopter Light<br>H2 Helicopter Heavy<br>JB Jet Business<br>JL Jet Large<br>JM Jet Medium<br>JR Jet Regional<br>JS Jet Small<br>PP Propeller<br>SS Supersonic<br>TP Turboprop |
| FLT_AIRFRAMES | MAX_RANGE | Maximum range | N | 513 possibilities within [11, 11000] | |
| FLT_AIRFRAMES | INTRO_YEAR | First year airframe was certified for flight operations | N | 59 different years | |
| FLT_AIRFRAMES | USAGE_CODE | Link to FLT_CAT_USAGE table | Y | O, H, A, P, C, B | A Attack/Combat<br>B Business<br>C Cargo/Transport<br>H Helicopter<br>O Other<br>P Passenger |
| FLT_AIRFRAMES | SIZE_CODE | Link to FLT_CAT_SIZES table | Y | T, H, S, L, M, V | H Heavy<br>L Large<br>M Medium<br>S Small<br>T Light<br>V Very Light |
| FLT_AIRFRAMES | ENGINE_TYPE | Type of Engine | Y | J, T, P | E Electric<br>J Jet<br>P Piston<br>R Rocket<br>T Turboprop/Turboshaft |

# Task 2 – Analytical Method Development

Georgia Institute of Technology

## Objective

The objective of this task is to formulate analytical methods that can be used to develop ANP data for the substituted aircraft in the AEDT FLEET database (DB). In this task, various statistical analysis techniques are investigated to analyze the distribution of the aircraft characteristics to distinguish different aircraft types. After the aircraft are classified into different aircraft types, the most appropriate analytical methods will be identified for each aircraft type to develop the ANP performance and NPD data by generating correction factors that can be applied to the substitution aircraft or new data for the aircraft. Multiple methods will be developed for different aircraft types that will span the current fleet depending on the data availability to develop the ANP performance data. This also requires an understanding of how to create correction factors or data suitable for integration into AEDT from higher fidelity noise and performance models. In addition, the substitution method implemented in AEDT is also to be examined and will be kept for use on some aircraft types if it is the most appropriate method. Georgia Tech will evaluate each method and identify the one most appropriate for the corresponding aircraft type.

## Research Approach

The existing aircraft substitution methods documented in Doc 29, a European Civil Aviation Conference (ECAC) report from 2016, are studied and validated. This will help with the development of analytical methods for the aircraft not modeled with ANP data in AEDT. Various statistical learning methods are investigated to understand the advantages and disadvantages of each method in order to select the most appropriate method for each aircraft type.

**Task 2.1: ANP Aircraft Substitution Method**

The objective of this task is to study and validate the ANP substitution method from literary research. Doc 29 provided a comprehensive description of the substitution process which will be reviewed and studied in this section.

The ANP database contains noise and performance characteristics for various different airframe and engine combinations. In Doc 29, it is stated that when performing these combinations, it is often required that certain groupings of aircraft types with similar noise and performance characteristics be created and they are all represented by one aircraft category. There are three main reasons why this type of grouping would be necessary: 1) insufficient information, especially for forecast scenarios, 2) lack of separate data from different aircraft models or variants, and 3) a need to decrease modelling time and cost. Out of these three reasons, decreasing the time and cost is the most important due to the tedious nature and the magnitude of calculations that involves multiple aircraft variants operating at an airport. Furthermore, since the differences in noise performance are often relatively small, creating these groupings saves both time and money. Another reasoning behind the groupings is that very often the noise contours in airports are dominated by a few aircraft types, and therefore it is considered to be more efficient to emphasize the most significant noise-generating aircraft types rather than all.

The aircraft are initially grouped based on certain characteristics that are directly related to sound emission and the performance of the aircraft. These characteristics include the maximum takeoff mass (MTOM) of the aircraft, the type of engine, the number of engines, the bypass ratio, the installation of the engines, the type of operation, and lastly the ICAO noise certificate. The MTOM is a parameter that is widely used and is fairly simple, dividing existing aircraft into the three categories of light, medium, and heavy. The type of engine refers to the common engines that are usually paired with turbojet, turbofan, and turboprop aircraft and the number of engines in each configuration. When evaluating the bypass ratio, there is a distinction between the turbojet and turbofan aircraft due to the relation of the parameter-to-sound emissions. The distinction leads to a differentiation between turbojet and turbofan aircraft with low, medium, and high bypass ratios. The "installation of the engines" refers to fuselage-mounted engines or wing-mounted engines. This distinction is made due to the fact that studies have shown that lateral sound emission does depend on the installation of the engine. The "type of operation" is also a very important parameter used for aircraft grouping which might differ between departures and arrivals. The type of operation may also be extended with respect to takeoff procedures when considering more modern aircraft such as wide-bodied twins where reduced takeoff thrust is widely used. The final parameter that is considered, when the grouping is performed, is the ICAO certificate. This parameter is used for grouping if no other information is available, since for landing operations, the certified approach noise level can be a trusted way of recognizing the operational noise. When it comes to departures, unfortunately, the deep cut back method that is used is not representative of real-life operations. For the groupings to make sense, these parameters have to be used in a combinatorial manner, but the difficult part is selecting the correct combination for the grouping. A rule of thumb is that all possible combinations should be used so that a large group can be generated with many similar noise characteristics. Not compromising accuracy acoustic equivalency and noise significance, however, should also be considered. Acoustic equivalency is defined as the comparable noise produced by two aircraft and is expressed in terms of event level $L_{\mathrm{Max}}$ or $L_E$ at multiple points on the ground or noise footprints. Acoustic equivalency depends on a variety of factors, such as operating procedures and aircraft mass, which means that all aircraft that are acoustically equivalent will not necessarily be grouped together, which is why an acoustic equivalency criterion is demanding in resource terms. Noise significance comes from the notion that the total noise at an airport is mostly driven by a small number of aircraft types. This is used so that the less noisy aircraft types can be grouped together in a simple way that can increase efficiency in potential noise studies.

Having defined the parameters used for grouping and the considerations for ensuring that accuracy and efficiency is not lost, a grouping approach can be set up by following three main steps. The first step consists of the introduction of a fundamental aircraft category scheme based on all of the combinations of parameters that can be used for grouping as mentioned above, such as MTOM, type of engine, number of engines, etc. The second step consists of identifying the different aircraft categories of low significance and grouping together based on a simple grouping scheme, such as engine type and takeoff mass. The third step combines the remaining groups depending on their acoustic equivalencies.

As mentioned earlier, one of the common reasons why a substitution is necessary is that the ANP database might be missing certain information about some possible airframe engine combinations. Therefore, since the information or model provided for this combination is not enough to model the required full set of operations at a specific airport, a substitution aircraft will be used instead in order to provide a noise model. This substitution aircraft will be similar to the ANP aircraft and is often referred to as a proxy aircraft. There are two options when using a proxy aircraft. One of the options is using the proxy aircraft as a one-to-one substitution, which means that it will be used, as is, without making any adjustments. The second option involves making adjustments to the NPD data or the number of movements of the proxy aircraft in the input operations. It is highly recommended that adjustments be made when a proxy is used so that the accuracy of the noise models is maintained. Two different types of adjustments are Method A and Method B. In Method A, a new entry is created in the ANP database and is defined as a duplicate of the proxy aircraft with the adjusted NPD data. The NPD data of the proxy are corrected by adding decibel adjustments, which are calculated using Equation (1) and (2). These equations are valid only for aircraft that have been certified under the ICAO Annex 16 Volume I Chapters 1, 3, 4, and 14, and different adjustments are made for arrival and departure as seen below.

$$\Delta_{dep} = \frac{FO_{LEVEL_{miss}} + LAT_{LEVEL_{miss}} - FO_{LEVEL_{proxy}} - LAT_{LEVEL_{proxy}}}{2} \tag{1}$$

$$\Delta_{arr} = APP_{LEVEL_{miss}} - APP_{LEVEL_{proxy}} \tag{2}$$

Equation (1) calculated the noise adjustment for departure and Equation (2) calculated for arrival. The subscript “miss” refers to the aircraft with the missing information, while the subscript “proxy” refers to the proxy aircraft. $FO_{LEVEL}$ is defined as the flyover noise level, $LAT_{LEVEL}$ is defined as the lateral noise level, and $APP_{LEVEL}$ is the certified approach noise level. For the aircraft that are certified under Chapters 6 and 10, a different adjustment calculation is performed using the Equation (3).

$$\Delta_{certif} = CERTIF_{LEVEL_{miss}} - CERTIF_{LEVEL_{proxy}} \tag{3}$$

$CERTIF$ is defined as the overflight and takeoff levels in decibels for Chapter 6 and 10 aircraft, respectively. The adjustments that are made in Method A are reflected in all noise metrics, including the maximum sound level metrics such as $L_{Amax}$.

Method B takes a different approach. It adjusts the number of movements of the proxy aircraft in the input operation, which translates to one movement of the missing aircraft to *N* movements done by the proxy aircrafts. It may be easier to implement Method B than Method A, seeing that only the input operations need to be adjusted, but applied only to equivalent sound levels such as $L_{Aeq}$.

The variable *N* that is being calculated in this method is referred to as the movement adjustment factor and is derived by comparing the certified noise level of the aircraft with the missing information to the proxy aircraft. Just as with Method A, there are different equations that are being used to calculated the movement adjustment factor for aircraft that were certified in the ICAO ANNEX 16 Volume 1 Chapters 2, 3, 4, and 14 as well as Chapters 6 and 10. For the first set of chapters, the movement adjustment factor is calculated for both the departure and the arrival using Equations (4) and Equation (5).

$$N_{dep} = 10^{\frac{\Delta_{DEP}}{10}} \tag{4}$$

$$N_{ARR} = 10^{\frac{\Delta_{ARR}}{10}} \tag{5}$$

For the second set of Chapter -certified aircraft, the movement adjustment factor can be calculated from Equation (6).

$$N_{CERTIF} = 10^{\frac{\Delta_{CERTIF}}{10}} \tag{6}$$

It is important to note that the certified noise levels for the proxy aircraft can be located in the ANP database under the FLT_NOISE_CERTIFICATION Table. When there are no certified noise levels for the missing aircraft, then the users can decide

to apply a one-to-one substitution, as was mentioned earlier, where $\Delta=0$ and N =1. Lastly, if the $\Delta$ adjustments are large or if *N* deviates greatly from 1, then this is a good indication that the proxy selection is not appropriate.

In order to select the most appropriate proxy aircraft, certain criteria have to be compared and matched to the missing aircraft. Such criteria include engine category, number of engines, engine installation, MTOW, thrust-to-weight ratio (which is defined as the static thrust divided by MTOW), certified noise levels, airframe manufacturer, and engine manufacturer. Therefore, when initiating the proxy selection process, finding and comparing these characteristics is essential. If the MTOW or engine type of the missing aircraft is unknown, then the rule of thumb is that the variant with the largest MTOW should be used with the corresponding engine type and static thrust. The most ideal substitution would consist of the proxy and the missing aircraft having identical engine categories, number of engines, and engine installation, with MTOW, thrust-to-weight, and certified levels as close as possible. Unfortunately, it is not always trivial to find a proxy that satisfies all of the criteria, therefore a degree of relaxation of certain criteria might be necessary. In conclusion, when looking for the most appropriate proxy aircraft, the engine category and installation should be identical, while different variants such as engine and MTOW of the same aircraft type should be assigned to the same proxy unless the variants are present in the ANP database. Lastly, when the missing aircraft type is present in the ANP database but does not have the same engine, MTOW or another variant should be used as a proxy.

In order to validate this substitution process, the calculations for Method A were performed for four aircraft: the 737-700, 737-800, A320-211, and A330-343. For each aircraft, the process began by looking in the ANP substitutions spreadsheet for the first instance of the substituted aircraft. From there, the aircraft's FOLEVEL, LATLEVEL, and APPLEVEL noise levels were recorded, along with the $\Delta$dep and $\Delta$arr values used in the ANP database, plus the ANP ID of the aircraft's proxy. Next, the proxy's ANP ID was used to find the specifications of the proxy aircraft in the FLT_ANP_AIRPLANES table. The three noise levels for this proxy were then found in the FLT_NOISE_CERTIFICATION table by finding the entry of the proxy aircraft type with a maximum takeoff mass matching that from the previous table. Lastly, these noise levels were used with those of the substituted aircraft to compute $\Delta$dep and $\Delta$arr which were compared to the values in the ANP database. In the case of the 737-800 and A320-211, both values were found to match, validating the substitution method. However, for the other two aircraft, only one of the values matched, with the other differing by a small amount. Repeating the calculations for other substituted aircraft with the same proxy found similar differences, suggesting that the problem is not with the substitution method but that one of the databases may be simply out of date. One additional complication was that there were several entries in the FLT_NOISE_CERTIFICATION table which matched both the proxy aircraft type and MTOW in each case. This suggests that more parameters from the proxy aircraft need to be considered when making this match to ensure the proper noise data is being used.

**Task 2.2: Review of Statistical Learning Methods**

The objective of this task is to review the statistical learning methods that can be utilized to conduct analytical analysis for the tasks of this project. This section includes: (1) Review of the data analytics process; (2) Measuring data similarity and dissimilarity; (3) Data preprocessing; (4) Summary of common classification/regression and clustering algorithms.

**2.2.1 The knowledge discovery from data (KDD) process**

In the data mining literature, the KDD process contains seven main steps. The steps' names and roles are listed below [HPK, 2011]:

1. *Data cleaning*: To remove noise and inconsistent data.
2. *Data integration*: Where multiple data sources may be combined.
3. *Data selection*: Where data relevant to the analysis task are identified and chosen.
4. *Data transformation*: Where data are transformed and consolidated into forms appropriate for mining by performing summary or aggregation operations.
5. *Data mining*: An essential process where intelligent methods are applied to extract data patterns.
6. *Pattern evaluation*: To identify the truly interesting patterns representing knowledge based on interestingness measures.
7. *Knowledge presentation*: Where visualization and knowledge representation techniques are used to present mined knowledge to users.

Steps 1 through 4 are different forms of data preprocessing, where data are prepared for mining. *Data reduction* may also be performed to obtain a smaller representation of the original data without sacrificing its integrity. This multi-step process

is also representative of a typical machine learning project, where multiple steps—from collecting data to model delivery—are needed for the entire project cycle.

### 2.2.2 Measuring Data Similarity and Dissimilarity

Dissimilarities and similarities are assessed based on the attribute values describing the objects and often involve distance measures. Similarity and dissimilarity measures are related and referred to as measures of proximity. In this project, such measures can take important roles because of the need to quantitatively determine the closeness between, for example, two aircraft types or two engines. The challenging part is that an aircraft or an engine normally has multiple attributes that are different in type. Table 5 provides a summary of the four main attribute types: nominal attributes, binary attributes, numeric attributes, and ordinal attributes. The two right columns of Table 5 includes examples of such attributes in the FLEET database.

**Table 5.** Summary of the four main attribute types

| Attribute Type | Description | ANP Example | Values |
|---|---|---|---|
| Nominal Attributes | A nominal attribute can take on two or more discrete states | ENGINE_LOCATION | **F** (Fuselage/Tail), **I** (Internal), **W** (Wing) |
| Binary Attributes | A binary attribute has only one of two states: 0 and 1 | THR_RESTOR | **N** (No), **Y** (Yes) |
| Numeric Attributes | Has continuous state within certain range | MX_GW_TKO | Within [2200, 1254430] |
| Ordinal Attributes | The values of an ordinal attribute have a meaningful order or ranking about them | SIZE_CODE | **H** (Heavy), **L** (Large), **M** (Medium), **S** (Small), **T** (Light), **V** (Very Light) |

Meanwhile, the crucial point here is that for different types of attributes, different similarity and dissimilarity measures must be used. Table 6 includes a summary of the corresponding metrics used in measure data dissimilarity for the four attributes in Table 5. The ratio of mismatches is used to measure nominal attributes, which are attributes with discrete and unordered states. Binary attributes can be either symmetric or asymmetric. Numerical attributes are the most common type in distance measuring. The Euclidean and Manhattan distances are commonly used in the literature, which are all special cases of the Minkowski distance (with $p$ equal to 2 and 1, respectively). Chebyshev distance is another useful option that is also referred to as the supremum distance, and is the maximum difference in values between the two objects. Lastly, prior to calculating distance for ordinal attributes, data normalization is required to map the range of each attribute onto $[0.0, 1.0]$.

Other than the individual measuring methods for the four types of attributes, an aircraft or engine typically has attributes of mixed types. Therefore, an approach is needed to combine the dissimilarity calculations from all four types into a single dissimilarity measure. The last row of Table 6 contains a brief introduction of how this can be done. When put into practice, the assignment of $d_{ij}^{(f)}$ needs to be discussed on its type. Overall, the materials here can be utilized to quantitatively measure objects with attributes of mixed types.

**Table 6.** Metrics used to measure data similarity and dissimilarity

| Type | Method of calculating dissimilarity |
|---|---|
| Nominal attributes | The ratio of mismatches: $d(i,j) = (p-m)/p$<br>($m$ is the number of matches and $p$ is the total number of attributes describing the objects). |
| Binary attributes | For symmetric binary attributes:<br>$d(i,j)$ = summation of off-diagonal numbers / total number of attributes.<br>The asymmetric binary dissimilarity therefore ignores the number of negative matches in the denominator of $d(i,j)$. |
| Numeric attributes | 1. Euclidean distance:<br>$d(i,j) = \sqrt{(x_{i1}-x_{i1})^2 + (x_{i2}-x_{i2})^2 + \cdots + (x_{ip}-x_{ip})^2}$<br>2. Manhattan distance:<br>$d(i,j) = \lvert x_{i1}-x_{i1}\rvert + \lvert x_{i2}-x_{i2}\rvert + \cdots + \lvert x_{ip}-x_{ip}\rvert$<br>3. Minkowski distance:<br>$d(i,j) = \left(\lvert x_{i1}-x_{i1}\rvert^p + \lvert x_{i2}-x_{i2}\rvert^p + \cdots + \lvert x_{ip}-x_{ip}\rvert^p\right)^{1/p}$<br>4. Chebyshev distance:<br>$d(i,j) = \lim_{h\to\infty}\left(\sum_{f=1}^{p}\lvert x_{if}-x_{jf}\rvert^h\right)^{1/h}$ |
| Ordinal attributes | Perform data normalization to map the range of each attribute onto $[0.0,\ 1.0]$:<br>$z_{if} = (r_{if}-1)/(M_f-1)$, and then use any dissimilarity method. |
| For attributes of mixed types | Process all attribute types together and combine the different attributes into a single dissimilarity matrix. Suppose that the data set contains $p$ attributes of mixed type, the dissimilarity $d(i,j)$ between objects $i$ and $j$ is defined as:<br>$d(i,j) = \sum_{f=1}^{p}\delta_{ij}^{(f)}d_{ij}^{(f)} / \sum_{f=1}^{p}\delta_{ij}^{(f)}$ |

### 2.2.3 Data Preprocessing

Real-world data are highly susceptible to noise, incompleteness, and inconsistency due to their typically huge set sizes and their likely origin from multiple, heterogeneous sources. Data preprocessing is an important step in the knowledge discovery process because quality decisions must be based on quality data. Although numerous methods of data preprocessing have been developed, data preprocessing remains an active area of research. Major data preprocessing steps are listed below (and they are not mutually exclusive):

1. *Data cleaning*: "Clean" the data by filling in missing values, smoothing noisy data, identifying or removing outliers, and resolving inconsistencies.
2. *Data integration*: Merges data from multiple sources into a coherent data store. Additional data cleaning can be performed to detect and remove redundancies that may have resulted from data integration. An attribute may be redundant if it can be "derived" from another attribute or set of attributes.
3. *Data reduction*: Obtains a reduced representation of the dataset that is much smaller in volume, yet closely maintains the integrity of the original data and produces the same (or almost the same) analytical results. Data reduction strategies include the following two types:

   a. Dimensionality reduction: The process of reducing the number of random variables or attributes under consideration. Examples: include principal components analysis, attribute subset selection, attribute construction, and wavelet transform.
   b. Numerosity reduction: The data are replaced by alternative, smaller representations. Examples include parametric models (regression, log-linear models, and nonparametric models), histograms, clusters, and sampling.
4. *Data transformation*: Can improve the accuracy and efficiency of algorithms involving distance measurements. An example is normalizing the data attempts to give all attributes an equal weight.
5. *Data discretization*: Concept hierarchy generation, where raw data values for attributes are replaced by ranges or higher conceptual levels.

**2.2.4 Algorithms**

This section provides a review of the common machine learning/data mining algorithms. The objective is to provide a list of candidate algorithms that can be chosen in this project and compare their characteristics.

Table 7 includes a summary of the common regression and classification algorithms [KJ, 2013]. We summarize these two types of algorithm in the same table for two reasons: (1) they are similar in that both belong to supervised learning methods, and (2) many algorithms can handle both regression and classification problems. Overall, the difference between regression and classification is that regression predicts continuous and numeric outputs, while classification predicts discrete outputs. Table 7 first classifies the algorithms into three categories: linear methods, nonlinear methods, and trees/rules-based methods. In the second column of Table 7, the collected candidate algorithms are listed. In the rest of the table, candidate algorithms are compared in eight different aspects:

1. Regression/Classification: What type of problem can the algorithm solve. “R”, “C”, and “R/C” stand for regression, classification, and both regression and classification, respectively.
2. Allows n<p: Whether or not the algorithm can be used when the number of observations *n* is less than the dimension of the predictors *p*.
3. Preprocessing: Which preprocessing methods must be used before applying the algorithm. “CS” stands for centering and scaling, “NZP” stands for the removal of near-zero predictors, and “HCP” stands for the removal of highly correlated predictors.
4. Interpretability: To what extent is the model interpretable. “H” stands for high, “M” stands for medium, and “L” stands for low.
5. Feature Selection: Whether or not the algorithm can perform feature selection (also referred to as model selection). “H” stands for high, “M” stands for medium, and “L” stands for low.
6. No. of Tuning Parameters: How many tuning parameters the algorithm has.
7. Robustness to Predictor Noise: Whether or not the algorithm is robust to noise. “H” stands for high, “M” stands for medium, and “L” stands for low.
8. Comp. Time: The overall computational time required when running the algorithm. “H” stands for high, “M” stands for medium, and “L” stands for low.

**Table 7.** Summary of common regression/classification algorithms

| Category | Method | Regression/ Classification | Allows $n < p$ | Pre-processing | Inter-pretability | Feature Selection | No. of Tuning Parameters | Robustness to Predictor Noise | Comp. Time |
|---|---|---|---|---|---|---|---|---|---|
| Linear | Linear Regression | R | N | CS, NZP, HCP | H | L | 0 | L | L |
| | Partial Least Squares | R/C | Y | CS | H | M | 1 | L | L |
| | Ridge Regression | R/C | N | CS, NZP | H | L | 1 | L | L |
| | LASSO/ Elastic Net | R/C | Y | CS, NZP | H | H | 1 to 2 | L | L |
| | Logistic Regression | C | N | CS, NZP, HCP | H | L | 0 | L | L |
| | Linear Discriminant Analysis | C | N | NZP | M | L | 0 to 2 | L | L |
| | Nearest Shrunken Centroids | C | Y | NZP | M | H | 1 | L | L |
| Nonlinear | Neural Networks | R/C | Y | CS, NZP, HCP | L | L | 2 | L | H |
| | Support Vector Machines | R/C | Y | CS | L | L | 1 to 3 | L | H |
| | MARS/FDA | R/C | Y | - | M | H | 1 to 2 | M | M |
| | K-nearest Neighbors | R/C | Y | CS, NZP | L | L | 1 | M | L |
| | Nonlinear Discriminant Analysis | C | N | NZP | M | L | 0 to 2 | L | L |
| | Naïve Bayes | C | Y | NZP | L | L | 0 to 1 | M | M |
| Trees/Rules | Single Decision Trees | R/C | Y | - | M | H | 1 | H | L |
| | Rule-Based Models | R/C | Y | - | M | H | 1 to 2 | H | L |
| | Bagged Trees | R/C | Y | - | L | H | 0 | H | M |
| | Random Forest | R/C | Y | - | L | M | 0 to 1 | H | H |
| | Boosting | R/C | Y | - | L | H | 3 | H | H |
| | Cubist | R/C | Y | - | L | M | 2 | H | H |
| | C5.0 | C | Y | - | M | H | 0 to 3 | H | H |

Next, we provide a summary of the common clustering methods [HPK, 2011]. Clustering belongs to a category called unsupervised learning that is different from the regression and classification algorithms contained in Table 7. In unsupervised learning, the data are unlabeled, and there is no ground truth to evaluate the prediction accuracy. The difference between supervised and unsupervised learning leads to a different summary between them. Generally speaking, the discussion and comparison of clustering algorithms have more aspects to consider. Table 8 below provides a collection of clustering methods. In this table, candidate algorithms (listed in the second column) are classified into seven categories (listed in the first column). The first four categories—partition methods, hierarchical methods, density-based methods, and grid-based methods—indicate four different angles of conducting the clustering task. The last three categories—probabilistic modeled-based clustering, high-dimensional clustering, and clustering with constraints— belong to more advanced topics generated by higher level needs.

The third column of Table 8 summarizes the general characteristics of the seven categories listed in the first column. Through comparing the information in the third column to the contextual knowledge of the problem, one can better select candidate algorithms. (A comparison of other characteristics, such as the computational complexity of each algorithm is outside the scope of this table.) The clustering methods provided can be used together with other unsupervised learning steps, such as dimensional reduction.

**Table 8.** Summary of common clustering algorithms

| Category | Method | General Characteristics |
|---|---|---|
| Partition Methods | K-means | Find mutually exclusive clusters of spherical shape.<br>Distance-based.<br>May use mean or medoid (etc.) to represent cluster center.<br>Effective for small- to medium-size datasets.<br>Heuristic methods: Global optimality is often computationally prohibitive. |
| | K-Medoids | |
| | CLARA | |
| | CLARANS | |
| Hierarchical Methods | DIANA | Clustering is a hierarchical decomposition (i.e., multiple levels).<br>Cannot correct erroneous merges or splits.<br>May incorporate other techniques like micro-clustering.<br>Can be distance-based or density- and continuity-based. |
| | AGNES | |
| | Chameleon | |
| | BIRCH | |
| Density-based Methods | DBSCAN | Can find arbitrarily shaped clusters.<br>Clusters are dense regions of objects in space that are surrounded by low-density regions.<br>Each point must have a minimum number of points within its neighborhood.<br>May filter out outliers. |
| | OPTICS | |
| | DENCLUE | |
| Grid-based Methods | STING | Use a multiresolution grid data structure.<br>Fast processing typically independent of *n* and dependent on grid size. |
| | CLIQUE | |
| Probabilistic Model-based Clustering | Fuzzy Clusters | Fuzzy or flexible cluster assignment.<br>Each object is assigned a probability of belonging to a cluster.<br>Each data point can belong to more than one cluster. |
| High-dimensional Clustering | CLIQUE | Conventional distance measures can be dominated by noise.<br>Defined using a small set of attributes instead of the full data space.<br>Methods include subspace clustering and dimensionality reduction. |
| | PROCLUS | |
| | PCA-based | |
| | Biclustering | |
| | MaPle | |
| Clustering with Constraints | COP-k-means | Can integrate background knowledge into cluster analysis.<br>Constraints on instances, clusters, and similarity measurement. |
| | CVQE | |

## Milestones

| Milestone | Due Date | Estimated Date of Completion | Actual Completion Date | Status | Comments (Problems & Brief Resolution Plan) |
|---|---|---|---|---|---|
| Quarterly Report (Jun) | 7/31/2020 | 7/31/2020 | 7/31/2020 | Completed | |
| A60 Kickoff Meeting | 8/13/2020 | 8/13/2020 | 8/13/2020 | Completed | |
| ASCENT Meeting | 9/29-30/2020 | 9/29-30/2020 | 9/29-30/2020 | Completed | |
| Quarterly Report (Sep) | 10/31/2020 | 10/31/2020 | 10/31/2020 | Completed | |
| Annual Report | 11/30/2020 | 11/30/2020 | 11/30/2020 | In Progress | |

## Major Accomplishments

The major accomplishments for this period performance include:

- Created MATLAB functions for searching through the AEDT aircraft database and ANP aircraft substitution list to match the target aircraft with the equipment in FLEET database.
- Matched aircraft from the AEDT FLEET database to aircraft in the ANP aircraft substitution list according to airframe model and engine model.
- Conducted literature study on databases for collecting performance, emissions, and noise data for target aircraft.
- Investigated the substitution method and how the groupings of the aircraft are performed.
- Validated the substitution method and identified gaps in the documentation and the data, since some substitution results could not be reproduced from the procedure described in Doc 29.
- Created a summary of the ANP database that assorts the data structure and location of all ANP parameters in a manner that is friendly to search and compare.
- Reviewed the knowledge discovery from data (KDD) process and common algorithms to conduct clustering and classification/regression tasks.

## Publications

**Written reports**

ASCENT quarterly reports (Jun. 2020)
ASCENT quarterly report (Sep. 2020)

## Outreach Efforts

N/A

## Awards

None

## Student Involvement

Zhenyu Gao is a fourth year PhD student. Mr. Gao has conducted a literature review on the analytical methods to develop the analytical methods derived from statistical learning methods and techniques. Mr. Gao is being trained on related tools such as INM, AEDT Tester, AEDT 3b, and AEDT 3d.

Bogdan Dorca is a third year PhD student. Mr. Dorca has performed a literature study on databases for collecting performance, emissions and noise data for aircraft not modeled with ANP data. Mr. Dorca is being trained on related tools such as INM, AEDT Tester, AEDT 3b, and AEDT 3d.

Chrysoula Pastra is a first year PhD student. Ms. Pastra has investigated the ANP aircraft substitution method and conducted a validation exercise. Ms. Pastra is being trained on related tools such as INM, AEDT Tester, AEDT 3b, and AEDT 3d.

## Plans for Next Period

The ANP aircraft substitution dataset will be further investigated to continue matching the target aircraft in the spreadsheet with the equipment in AEDT FLEET database. Additional literature review will be conducted to collect data for the target aircraft in order to obtain performance, emissions, and noise data that can be used to develop ANP data for these aircraft.

The substitution methods implemented in AEDT will be studied to identify the capability gaps for potential improvement. Georgia Tech will coordinate with the AEDT development team to get a list of target-substitution aircraft to work on.

The problems associated with ANP data development will be defined. Based on the characteristics of the problem, the most appropriate statistical learning methods will be selected to formulate the analytical methods to develop ANP data for target aircraft.

## References


- EASA, https://www.easa.europa.eu/domains/environment/easa-certification-noise-levels, 2020 (accessed: Sept. 2020)
- ECAC.CEAC Doc 29, Report on Standard Method of Computing Noise Contours around Civil Airports, Volume 1: Applications Guide, As endorsed by DGCA/147 on 7 December 2016
- Eurocontrol, https://contentzone.eurocontrol.int/aircraftperformance/default.aspx?, 2020 (accessed: Sept. 2020)
- Eurocontrol, https://www.aircraftnoisemodel.org/aircraft/substitutions, 2017 (accessed: Sept. 2020)
- ICAO, https://www.icao.int/environmental-protection/pages/modelling-and-databases.aspx, (accessed: Sept. 2020)
- Jiawei Han, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.
- Max Kuhn, Kjell Johnson, et al. Applied predictive modeling, volume 26. Springer, 2013.
- NoisedB, http://noisedb.stac.aviation-civile.gouv.fr/, 2020 (accessed: Sept. 2020)
- RisingUp Aviation, https://www.risingup.com/planespecs/ (accessed: Sept. 2020)
- The OpenSky Network, https://opensky-network.org/aircraft-database, 2020 (accessed: Sept. 2020)