Project 060 Analytical Methods for Expanding the AEDT Aircraft FLEET Database

Georgia Institute of Technology

Project Lead Investigator

Dimitri N. Mavris
Director, Aerospace Systems Design Laboratory
School of Aerospace Engineering
Georgia Institute of Technology
270 Ferst Drive, Mail Stop 0150
Atlanta, GA 30332-0150
404-894-1557
dimitri.mavris@ae.gatech.edu

University Participants

Georgia Institute of Technology (aka Georgia Tech)

- P.I.: Prof. Dimitri Mavris
- FAA Award Number: 13-C-AJFE-GIT-104
- Period of Performance: October 1, 2022 to September 30, 2024
- Tasks:
    1. Enhancement of the Aviation Environmental Design Tool (AEDT) Fleet database (Fleet dB)
    2. Analytical method development

Project Funding Level

The current FAA funding for this project is $150,000 from September 21, 2022 to September 30, 2024. The Georgia Institute of Technology has agreed to a total of $150,000 in matching funds.

Investigation Team

- Dr. Dimitri Mavris (P.I.), All Tasks
- Dr. Michelle R. Kirby, All Tasks
- Dr. Mayank Bendarkar, All Tasks
- Dr. Ameya Behere, All Tasks
- Dr. Styliani I. Kampezidou, All Tasks
- Cristian Puebla-Menne (graduate student), All Tasks

Project Overview

The AEDT relies on aircraft noise and performance (ANP) data provided by aircraft manufacturers to support the calculation of aircraft trajectories and noise at receptors by using aircraft performance information and noise–power–distance relationships for specific aircraft/engine combinations. In the ANP/Base of Aircraft Data (BADA) workflow, ANP performance data are 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 are 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 the certification noise characteristics and other performance parameters. However, a problematic issue is that the best substitute according to 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, thus resulting in overly conservative noise and emissions estimates.

The goal of this research is to increase the accuracy of AEDT noise and emissions modeling of aircraft not currently in the ANP database. Georgia Institute of Technology will identify and review aircraft not currently modeled in the AEDT, and will collect information and necessary data to better understand the characteristics of these aircraft. Various statistical analysis methods will be used to classify the aircraft into different 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 ANP and noise data for the aircraft. Validation data from certification data or airport planning documents will be gathered to validate the methods. After validation, the models will be applied to develop ANP and noise data for the aircraft. Finally, recommendations and guidelines will be developed for implementing the developed data in the AEDT, to expand the AEDT Fleet dB to include noise and performance data for aircraft currently not in the ANP database.

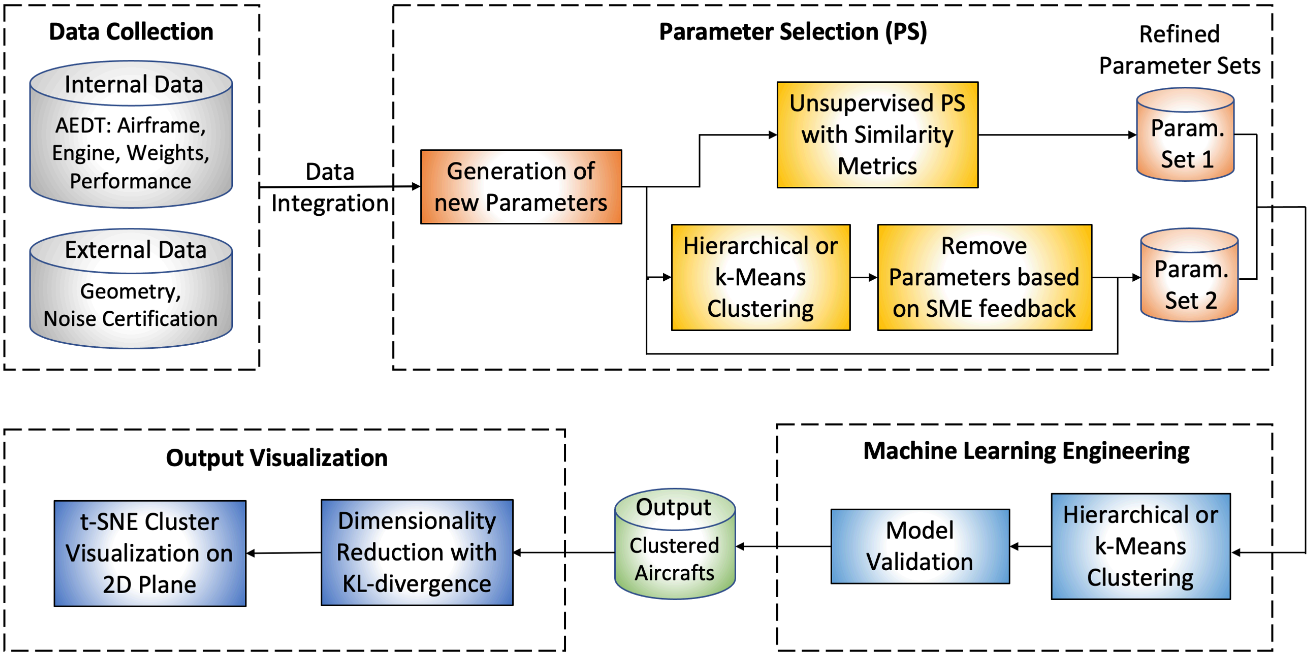


Figure 1. Overview of ASCENT Project 60 tasks and workflow.

The flowchart in Figure 1 presents an overview of the project approach. The first step is to identify the necessary aircraft parameters that will be used to better estimate the substitution aircraft. These parameters are already included in the internal data (Fleet dB) or will be collected from external resources.

Task 1 – Enhancement of the AEDT Fleet Database

Georgia Institute of Technology

Objective

The objective of Task 1 is to identify aircraft that are not currently modeled with ANP data in the AEDT for noise and emissions modeling. In the Fleet 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 for the entire fleet; some aircraft are modeled according to ANP data, whereas others are represented by substitution aircraft. This task involves the identification of aircraft that do not have ANP data, and determining whether those substitutions can be improved or enhanced.

Research Approach

Creating the ANP extension database

Aircraft without ANP data in AEDT

The aircraft 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 identify gaps between them and the substitution aircraft, in terms of performance, noise, and emissions. This step involves reviewing the existing literature on these aircraft and acquiring the information and data necessary to better determine their engine/airframe characteristics. In addition, the ANP data in the Fleet dB are studied to summarize key parameters for which the analytical methods can be used.

The Fleet dB consists of 3,626 airframe/engine combinations; only 269 have available ANP data (native), whereas the remaining 3,357 do not (proxy). The proxy aircraft have a unique equipment ID (the primary key in the SQL database) and a default equipment ID, which is assigned as the equipment ID of the closest native aircraft, in terms of ANP similarity. The native aircraft have a matching equipment ID and default equipment ID. This substitution enables proxy aircraft to borrow ANP data from the native aircraft for the purposes of conducting environmental analyses and studies. The Fleet dB uses proxy aircraft because of the intensity of effort required to generate the required ANP definition by a manufacturer; this use of proxies was assumed to be reasonable for modeling purposes on an average basis decades ago.

Down-selecting aircraft of interest in the fleet database

To focus the efforts on aircraft types with U.S. operations in the Fleet dB, we conducted a filtering process on the proxy aircraft to create an initial ANP extension database. Filtering was applied to the original 3,626 unique equipment IDs to establish a subset of engine/airframe combinations, to be denoted as aircraft in later discussions, for which external data would be gathered. The first filter eliminated the military and cargo designation codes and small SIZE_CODE aircraft. The next filter eliminated military and general aviation, according to the AIRCRAFT_TYPE designation. This filtering reduced the number of unique equipment IDs to 2,443. With an initial focus on U.S. applications of AEDT, airframe models that are not operated in the United States or are out of production were eliminated. These filters reduced the total airframes for which external data are required to a manageable number of 107. Notably, each airframe could have multiple engine types, thus resulting in a total of 990 native and proxy aircraft remaining, as listed in Table 1. For the remaining EQUIP_IDs, the AIRFRAME_MODEL names were grouped to determine the number of unique airframes.

External aircraft database literature study

To augment the Fleet dB to establish a new ANP extension database, we collected external data for the 990 unique equipment IDs from various sources into AEDT by AIRFRAME_MODEL and the ENGINE_MODEL, as the primary definitions of what the equipment ID was intended to represent in the actual fleet. This information was helpful in determining which performance, emissions, and noise parameters were used for the substitution algorithm in the initial applications of Task 2. In particular, the following categories of data were gathered:

- **Airframe**: general aircraft information and classifications; example: maximum range
- **Engine**: important engine specifications; example: bypass ratio
- **Aircraft**: information on an airframe/engine combination; example: maximum takeoff weight (MTOW)
- **Aircraft geometry**: example: wing area
- **Emissions**: main emission indices; example: unadjusted fuel flow during takeoff
- **Noise**: certification noise; example at the three conditions: flyover, sideline, and approach

Table 1. AEDT Fleet database down-selected equipment IDs of interest.

Family	Members	Number of variants in Fleet dB
A220	2	10
A320	8	132
A330	4	70
A340	4	37
A350	2	2
A380	1	9
ATR 42	5	22
ATR 72	2	6
B737	10	78
B747	4	37
B757	2	15
B767	4	143
B777	8	89
B787	3	32
BAE 146	3	13
CRJ	11	33
Dash 8	7	38
EMB120	1	3
EMB135/145	12	113
EMB170/175	6	20
EMB190/195	8	88

As the external data gathering process began, challenges arose in identifying what the unique equipment ID in the flying fleet represents. Per Volpe's guidance, a given equipment ID entry is defined by several primary variables in the Fleet dB, which are then linked to other variables for the specific modeling of interest: fuel burn, noise, or emissions. The AIRFRAME_MODEL is intended to be the general description of an airframe, per the manufacturer. For example, the "Airbus A319-100 series" AIRFRAME_MODEL has 26 entries that should represent the various engine options on that series. The AIRFRAME_MODEL coupled with the ENGINE_MODEL is intended to define the actual engine on the specific variant of the airframe series in the fleet. Of note, ENGINE_MODEL is linked to the emissions modeling via the ICAO_UID.

Because AIRFRAME_MODEL provides a generic description of the aircraft, a "common aircraft name" was created to relate the equipment ID entry to what is actually flying in the commercial fleet, for example, the A319-100 series was considered to represent the A319, and a new column was added to the ANP extension database. This renaming enabled the initial gathering of the external data, as listed above, and the analytical methods to be tested, as described in prior annual reports. Beyond AIRFRAME_MODEL and ENGINE_MODEL, other Fleet dB variables are used for fuel burn and noise modeling. ACFT_DESCR is the mapping to the ANP native for that proxy aircraft for noise modeling, which is then coupled to the ANP_AIRPLANE_ID. MANUF_DESC is the mapping to the BADA representation for fuel burn modeling, represented as BADA_ID or BADA4_ID. Combining these four primary descriptors/parameters of the unique equipment ID led to a questionable understanding of what the entry specifically represented.

Initial noise certification data population for the ANP extension database

To initially populate the ANP extension database with noise data, we used two sources of the European Union Aviation Safety Agency (EASA) certification noise level databases until the questionable unique entries could be rectified: one for jets and one for propellers for the three certification noise levels. The limit, margin, and cumulative noise values in EPNdB units were extracted. The methods used for matching comprised the following steps. In the ANP database, a total of 990 airframe/engine candidate combinations for noise data population were selected. The population procedure was started by selection of a specific airframe of interest (for example, the Airbus A321-200 Series AIRFRAME_MODEL and "common aircraft name"). For that airframe, a specific engine, ENGINE_MODEL, among the different options available, was selected (for example, the CFM56-5B3/2P). After the specific airframe/engine combination was defined, the exact same combination was searched and selected in EASA certification noise level database. For matching to be performed, the selected airframe/engine combination in EASA was required to be unique. To ensure this unique matching, we used a set of successive selection criteria involving the following sequence of steps:

- Use the EASA type certificate database (TCDS) to verify that the variants are actually on the airframe; use the EASA certification noise level database (e.g., MAdB Jets) to cross-reference that the engine is certified for noise.
- Use the EASA TCDS to verify that the engine emissions and thrust parameters in the ANP database are correct.
- When differences are found, they are identified and registered by matching the ANP Equipment ID and EASA Record number.
- For the certified airframe/engine combination in the EASA certification noise level database, select the MTOW.
- If no unique combination is obtained, proceed to select the maximum landing mass.
- If the combination still has more than one option, the maximum cumulative noise level can be selected.
- In cases in which multiple airframe/engine combinations have the same noise values, the first entry is selected.
- Finally, if more than one combination remains after application of the preceding criteria, the most recent modification date for the data of the remaining combinations is selected. This modification date corresponds to the most recent date when the existing values for the selected combination were entered in the database.

The rationale underlying these selection criteria was to choose the most representative noise value of the combination selected. After a unique combination is found, the corresponding noise value is transferred from the EASA database to ANP. To increase the number of combinations available for which noise values were obtained, we selected engines with similar designation codes for some airframes. In this case, the criterion for selection was a direct comparison of the main parameters (bypass ratio, overall pressure ratio, and rated thrust) of the similar engines. If the parameters were within 5% of each other, the combination was considered valid and was added to the ANP database. Unfortunately, this process yielded data for less than 50% of the 990 unique equipment IDs of interest.

Of note, the initial noise certification data gathering was an extremely labor-intensive process.

Deep-dive into the unique equipment IDs

Upon further investigation of a unique equipment ID entry for which noise data could not be established, we identified erroneous and questionable entries when the ENGINE_MODEL was cross-referenced to FAA or EASA airframe TCDS. Four variables in the Fleet dB represent the proper entry in the flying fleet: AIRFRAME_MODEL, ACFT_DESCR, MANUF_DESC, and ENGINE_MODEL. Continuing with the same example from above with the "A319-100 series," we identified seven specific variants of the A319-100 from a review of the EASA and FAA airframe TCDS, which differ according to the specific engine and engine manufacturer on the variant. For simplification, the A319 variants can be described as a A319-1XX, where "XX" is a number defining the engine type related to the engine manufacturer and the maximum thrust level of the engine. For example, the EASA TCDS states that the following engines are on A319-1XX variants. The green highlighted numbers below designate the different engine manufacturers. The yellow highlighted numbers are the thrust variant of the engine family model. Any letter or number thereafter indicates a change in the combustor or modification to the original engine type certification, as shown with blue highlighting. The numbering and naming conventions regarding the letter after the CFM 56-5 and before the thrust designation number for the CFM International manufacturers are unknown. This naming or numbering convention is not universal, and varies by airframe and engine manufacturer, but is representative.

- A319-111 two CFMI CFM 56-5B5 jet engines (MOD 24932)
- A319-112 two CFMI CFM 56-5B6 jet engines (MOD 25287), or CFM 56-5B6/2 jet engines (MOD 25530)
- A319-113 two CFMI CFM 56-5A4 jet engines (MOD 25238), or CFM 56-5A4/F jet engines (MOD 23755)
- A319-114 two CFMI CFM 56-5A5 jet engines (MOD 25286), or CFM 56-5A5/F jet engines (MOD 23755)
- A319-115 two CFMI CFM 56-5B7 jet engines (MOD 27567)
- A319-131 two IAE V2522-A5 jet engines (MOD 26152)
- A319-132 two IAE V2524-A5 jet engines (MOD 26298)

In cross-referencing of the EASA or FAA airframe TCDS, the certified engines on the interpreted AIRFRAME_MODEL (and hence the "common airframe name") to the ENGINE_MODEL in AEDT yielded several errors, including incorrect thrust variants, typographical errors in the ENGINE_MODEL entry, variants not certified for noise in any database, engines not certified for emissions, nonexistent AIRFRAME_MODEL, or incorrect engine on the airframe. In each of these cases, the issues are being thoroughly documented.

Among the 990 unique equipment IDs of interest, we identified 123 unique AIRFRAME_MODEL names. That list includes a subset of aircraft families, thus greatly reducing the number of TCDS that must be investigated, because a family is typically certified under one type certificate, regardless of technology generation. An aircraft family is defined by the commonality of almost all elements of the aircraft, but is differentiated by the length of the fuselage or the maximum thrust of the engine, or is within a similar type designation. A similar concept holds true for an engine family with the maximum allowable thrust as the differentiating factor, although different combustors can exist on the same engine family. For example, the "A319-100 series" is a member of the A320 family. The A320 family has four members, A318, A319, A320, and A321, in two generations of technology levels, the "ceo" and the "neo," but still fall within the same type designation as that of the original application. Similarly, the Boeing B737 has three generations of technology, denoted Classic, Next-gen, and Max, and the associated family members have different B737-X00 designations, where "X" denotes the generation of technology. As described previously, a family member can have variants, which are usually differentiated by the engine manufacturer and maximum permissible thrust.

As with the initial gathering of the noise certification data, this effort is extremely labor intensive, requiring a line-by-line investigation of a unique equipment ID entry and cross-referencing with multiple databases, and therefore is not complete at the time of this report. The databases or sources used include the EASA noise certification database, the International Civil Aviation Organization emissions databank, the EASA and FAA TCDS, and the Fleet dB. After discrepancies are finalized, a comprehensive document will be provided to the AEDT development team, so that refinements can be made to the current entries of the Fleet dB. A benefit of this investigation was the discovery of the wealth of data contained in the airframe and engine TCDS—including certified information on geometry, performance, capability, operating limits, weights, and more—which could be used in lieu of the initial external data gathered for the ANP extension database that did not rely on certified data.

This effort will be ongoing in the coming year.

Collection of airframe and engine TCDS data

As a result of the deep dive, a wealth of data was determined to exist in the EASA and FAA TCDS, for both airframes and engines, that could augment the ANP extension database. For each aircraft family listed in Table 1, the airframe TCDS was downloaded from the respective certification authority website. For EASA, the website is https://www.easa.europa.eu/en/document-library/type-certificates. For the FAA, the website is https://drs.faa.gov/browse/TCDSMODEL/doctypeDetails. The airframe was used as the main designation to determine the associated TCDS; for each airframe, the engine variants could be determined and used to download the engine TCDS. EASA and the FAA use different formats and levels of content for not only the airframe but also the engine parameters provided. A further complication is that each manufacturer provides varying levels of detail of those parameters.

To simplify the data gathering process and structure, we established an overarching list of parameters by cross-referencing the EASA and FAA content, as listed in Table 2, noting that the units between sources could vary between English and metric. Resources were directed to gather the data manually, in parallel to the noise certification data gathering; this effort was also extremely labor intensive, entailing reading of every page of each airframe and engine TCDS to extract the appropriate and desired information, and compiling it into a new certification database that would ultimately be merged into the ANP extension database. If a particular parameter was not contained in a TCDS, it was noted as "can't find" and would be resolved later.

The TCDS data gathering was completed after an exhaustive effort. A summary of the amount of data collected by manufacturer category is as follows: Airbus provided the most comprehensive information, with approximately 70% of the desired parameters obtained; the percentage of parameters collected for Boeing aircraft varied between 20% and 50%; and all remaining airframe manufacturers had approximately 50% of the parameters populated. In the review of the percentages of parameters collected, engineering judgement of the potential key drivers of noise, emissions, and fuel burn—such as dimensions, capability, and performance limits—was used to identify key gaps. The only further option identified to potentially fill these gaps was to investigate the information provided by the manufacturers in the airport planning manuals, which are usually available online. At present, the airport planning manuals for each airframe under consideration are being compiled.

Table 2. Airframe and engine TCDS parameters gathered.

Airframe TCDS parameters	Engine TCDS parameters
Airframe	Airframe
Certification date	Engine
Engine	Engine TCDS #
Engine TCDS #	Certification date
Sea Level Static thrust (lbs)	Engine description
Maximum continuous thrust (lbs)	Overall length (in)
Maximum engine speed, N1 rpm (%)	Overall width (in)
Maximum engine speed, N2 rpm (%)	Overall height (in)
Wing area (ft^2)	Dry weight (lbs)
Wingspan (ft)	Takeoff thrust (lbs)
Fuselage height (ft)	Maximum continuous thrust (lbs)
Fuselage length (ft)	Maximum take-off shaft (HP)
Fuselage diameter (ft)	Maximum continuous shaft (HP)
Aircraft height (ft)	Flat rating ambient temperature: takeoff (˚C/[˚F])
Maximum operating altitude (ft)	Flat rating ambient temperature: maximum continuous (˚C/[˚F])
MTOM (kg)	Maximum engine speed, N1 rpm (%)
Maximum Zero Fuel Weight (kg)	Maximum Engine speed, N2 rpm (%)
Maximum Landing Weight (kg)	Maximum engine speed, N3 rpm (%)
Mean aerodynamic chord (m)	
Maximum seat capacity (basic)	
Maximum seat capacity (option)	
Total maximum baggage/cargo loads (kg)	Engine Exhaust Gas Temperature (red line takeoff) (˚C/[˚F])
Fuel capacity (kg)	Engine EGT (red line maximum continuous) (˚C/[˚F])
Wheels	Engine EGT (maximum indicated takeoff) (˚C/[˚F])
Tires	Engine EGT (indicated maximum continuous) (˚C/[˚F])

Milestone

Developed a framework for new external data to be used in Task 2.

Major Accomplishments

Populated new extension database, and created additional certification database.

Publications

Bendarkar, Mayank V., Michelle Kirby, Styliani I. Kampezidou, Cristian Puebla-Menne, and Dimitri N. Mavris, "Exploring Analytical Methods for Expanding the AEDT Aircraft Fleet Database for Environmental Modeling", AIAA Aviation 2023 Forum, doi.org/10.2514/6.2023-4216.

Outreach Efforts

Bi-annual ASCENT meetings.

Awards

None.

Student Involvement

Styliani I. Kampezidou and Cristian Puebla-Menne (graduate students). Collected the TCDS information

Plans for Next Period

Finalize the ANP extension database, and document issues uncovered with the existing Fleet dB for the AEDT development team.

References

Cirium. (n.d.) *Aviation Analytics.* https://www.cirium.com/
Janes | Login for world leading open source defense intelligence. (n.d.). https://customer.janes.com/janes/home
European Union Aviation Safety Agency. (2023). *ICAO Aircraft Engine Emissions Databank.* https://www.easa.europa.eu/domains/environment/icao-aircraft-engine-emissions-databank
Meier, N. (2021). *Jet Engine Specification Database.* http://www.jet-engine.net/
Lissys Ltd. (n.d.) *Piano's Aircraft Database.* https://www.lissys.uk/dbase.html
Aircraft Bluebook. (2023). *Aircraft Bluebook – Spring 2023 Vol. 23-01.* https://aircraftbluebook.com/Tools/ABB/ShowSpecifications.do
Jenkinson, L., Simpkin, P., Rhodes, D. (2001). *Civil Jet Aircraft Design: Aircraft Data A File.* https://booksite.elsevier.com/9780340741528/appendices/data-a/default.htm
Jenkinson, L., Simpkin, P., Rhodes, D. (2001). *Civil Jet Aircraft Design: Aircraft Data B File.* https://booksite.elsevier.com/9780340741528/appendices/data-b/default.htm
European Union Aviation Safety Agency. (2023). *EASA Certification Noise Levels.* https://www.easa.europa.eu/en/domains/environment/easa-certification-noise-levels
NoisedB. (2023). *Noise Certification Database Version 2.34.* https://noisedb.stac.aviation-civile.gouv.fr/bdd
Aircraft Bluebook. (2023). *Aircraft Bluebook – Spring 2023 Vol. 23-01.* https://aircraftbluebook.com/Tools/ABB/ShowSpecifications.do
European Union Aviation Safety Agency. (2023). *EASA Certification Noise Levels.* https://www.easa.europa.eu/en/domains/environment/easa-certification-noise-levels

Task 2 – Analytical Method Development

Georgia Institute of Technology

Objective

The objective of Task 2 is to develop analytical methods and solutions that can improve the modeling of aircraft types (airframe/engine combinations) that are not included in the ANP database. In this process, machine learning (ML) and data-mining (DM) approaches are used to analyze aircraft features (both internally and externally collected), ANP data, and environmental output data, as well as to gain insights and evidence of better model substitution and approximation. The following research questions can be answered while developing these more advanced analytical methods:

- How can substitutions be better assigned for aircraft types not included in the ANP database?
- How can representative aircraft models be better chosen to develop more ANP data, with the aim of more sufficiently covering the entire population?
- Which aircraft features should be used in the identification of aircraft substitution?
- How can the current ANP data be better used to approximate the remaining aircraft with greater flexibility?

Research Approach

The data-driven analytical methods used in this task are based primarily on ML and DM techniques. The solution for each research question consists of multiple ML/DM algorithms. In general, the analytical techniques that are useful in this project can be classified into five categories: clustering, dimensionality reduction, regression, feature selection, and data visualization. Table 2 presents examples and objectives for all five categories.

The method is outlined in Figure 1. The process begins with a data fusion step, wherein various data sources are queried and merged with the AEDT Fleet dB to create the ANP extension database, as explained in Task 1. The resulting database contains 3,626 airframe/engine combinations with 112 columns. The total number of airplanes with nitrogen oxide (NO_x) emissions data is 2,361, which decreases to 520 when noise data are also included. Of these, 269 aircraft have data from the ANP database. Most of the efforts of the past year focused on finalizing the full dataset. Consequently, the present report summarizes the cumulative progress made on analytical method development performed over the past several years.

Past efforts explored three broad areas to synthesize ANP data for aircraft lacking these data. Of these, the first step involved exploring unsupervised clustering to group similar aircraft by using the enriched dataset from Task 1. Native aircraft (with ANP data) within each cluster can be considered potential substitutes for other aircraft without ANP data within each cluster.

The other two analytical techniques will be explored after updating of the full dataset from Task 1 is completed. These techniques are (a) potentially customizing ANP data by using statistical techniques and regressions to enable more flexible synthesis for ANP data rather than the currently used one-to-one substitution for aircraft without ANP data and (b) exploring hybrid models, wherein a composite model of multiple closest ANP aircraft is used to synthesize ANP data for non-native aircraft.

Using clustering to identify representative aircraft model portfolios

Results of using unsupervised clustering methods on the extension database were presented in the previous annual report and are summarized here for completeness. Two algorithms, *k*-means (KM) and hierarchical clustering, were implemented on the preliminary database, which included 520 airframe/engine combinations with available noise data. In these studies, all *n* aircraft are first partitioned into *k* clusters; one aircraft from each cluster is then selected to represent all aircraft in that cluster. Methods for conducting clustering and representative aircraft selection simultaneously will be explored after the full dataset is ready. The implemented clustering techniques also aided in identifying outliers in the data and correcting the data entries for any potential errors.

The dimensionality of clustering is influenced by the number of parameters selected for the exercise. For our preliminary explorations, inputs from subject-matter experts (SMEs) were used to determine the important parameters for emissions and noise modeling, as shown in Table 2. These parameters were selected after multiple rounds of clustering experiments involving SME feedback and focus on aircraft performance, geometry, engine characteristics, noise, and emissions.

Table 3. Selected SME parameters for clustering.

Group	Parameter	Units
Geometry	Wing area	ft^2
	Wing aspect ratio	
	Fuselage volume	ft^3
Performance	Gross weight	lbs
	Cruise Mach	
	Typical range	nm
	Number of passengers	
	Cruise altitude	ft
Engine	Pressure ratio	
	Total thrust	kN
	Bypass ratio	
Emissions	NO_x	gm/kg
Noise	Flyover noise	EPNdB
	Approach noise	EPNdB
	Lateral noise	EPNdB

For KM clustering, the elbow method is widely used to determine the number of clusters. This method provides a suitable tradeoff between error and the number of clusters. Figure 4 shows the inertia (elbow) plot for selecting the number of clusters for the KM algorithm. Approximately five to seven clusters appear to be ideal to divide the data. The same number of clusters was used for Analytical Hierarchical Clustering (AHC) to enable comparison between the outputs of the two methods. These clusters were also visualized with *t*-distributed stochastic neighbor embedding (Melit Devassy, 2020) visualization, which enables the depiction of higher-dimensional clusters in two or three dimensions.

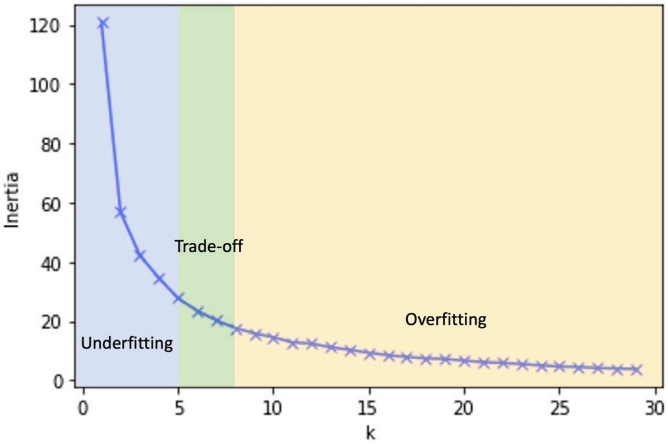


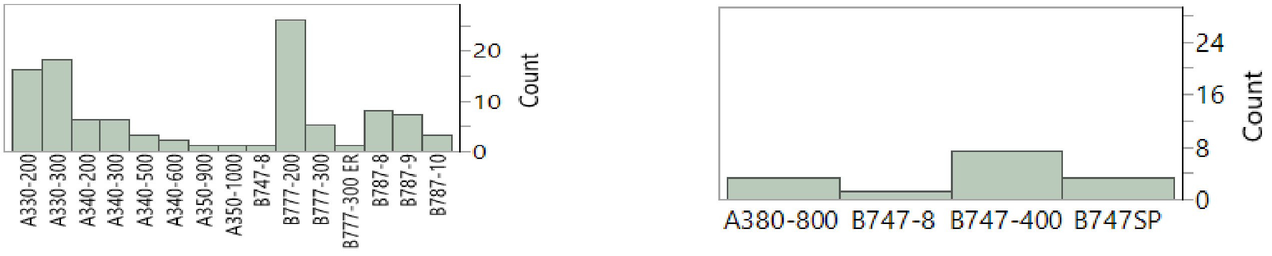
Figure 2. Inertia (elbow) plot for KM clustering.

Preliminary clustering results
The approximately 520 aircraft for which the AEDT Fleet extension database contains complete parameter data were included in the preliminary results. The results from using the AHC clustering method from last year are shown in Figure 5 for completeness. Overall, the clusters showed good agreement with real-world distinctions: larger wide-body aircraft formed cluster 0; so-called "jumbo" jets formed cluster 1; regional jets were found primarily in clusters 2 and 6; smaller wide-body aircraft were grouped in cluster 3; newer-generation small single-aisle aircraft were grouped in cluster 4; and traditional small single-aisle aircraft were grouped in cluster 5. Goodness of fit for clustering can be a difficult metric to quantify for unsupervised methods. Because we address real world airframe/engine combinations and their impacts in terms of emissions and noise, we used SME inputs and feedback to evaluate the goodness of fit.

To visualize 15-dimensional clusters, we used scatterplot matrices. Figure 6 shows an example scatterplot matrix of NO_x and noise emissions for aircraft, with cluster 1 highlighted. As expected, the largest aircraft and highest thrust engines that pair with them have the highest emissions and noise signatures, and thus are located at the top right of almost every plot. Clear distinctions between clusters are not expected in this figure, which shows only 4 of the 15 dimensions used for clustering.

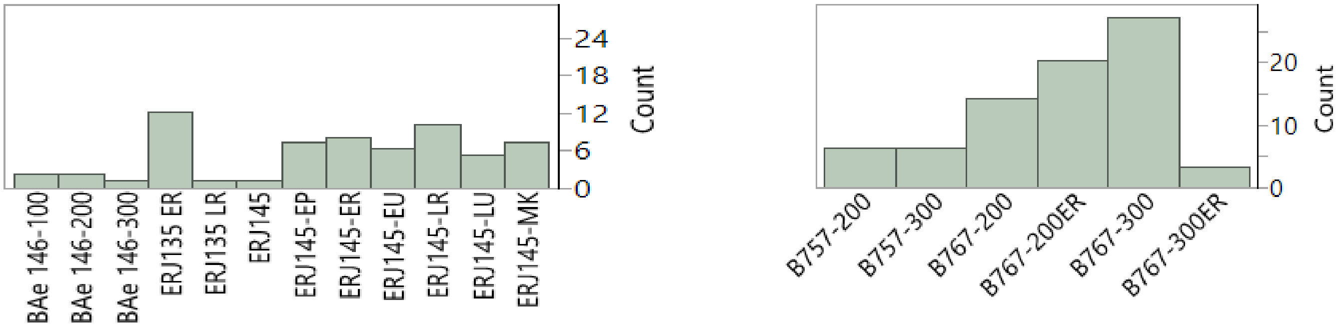
Parameter importance is difficult to gauge for unsupervised learning clustering algorithms. Therefore, to determine the importance of the parameters with the greatest effects on the clusters, we fit a supervised random forest algorithm with 100 trees to the cluster numbers while using the same 15 parameters for clustering the aircraft. A parameter importance function of this random forest was evaluated to indicate the parameter importance of the AHC clusters (Figure 7).

The idea underlying segregating the aircraft within the AEDT Fleet extension database into clusters is to observe whether aircraft with ANP data (native) are present in certain clusters with non-native aircraft. This process can help identify more suitable substitute ANP aircraft for airframe/engine combinations that do not have ANP data. Because of limitations of the dataset, the results summarized herein focus on the unsupervised clustering approach. Implementations of other analytical methods on the full dataset will be described in future reports.



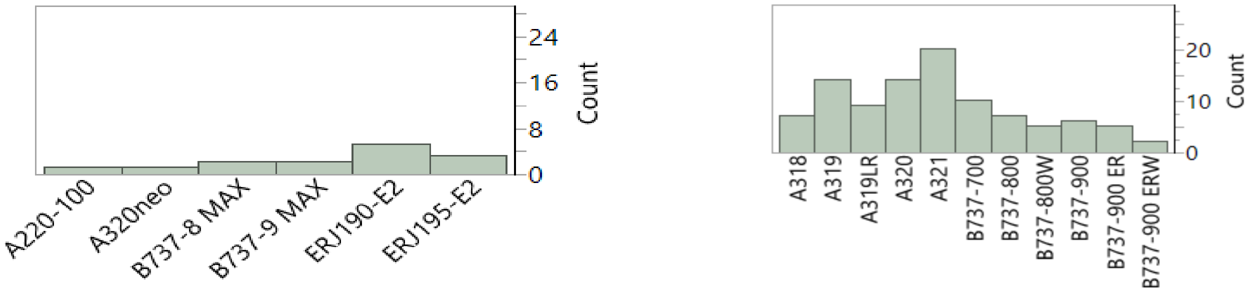
(a) Cluster 0

(b) Cluster 1



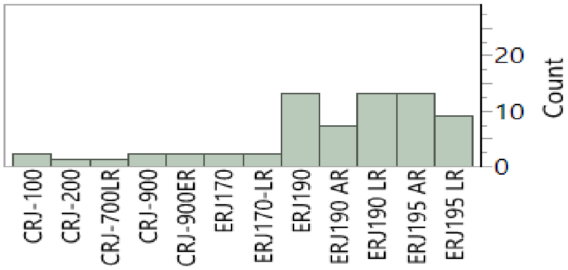
(c) Cluster 2

(d) Cluster 3



(e) Cluster 4

(f) Cluster 5



(g) Cluster 6

Figure 3. Preliminary hierarchical clustering results.

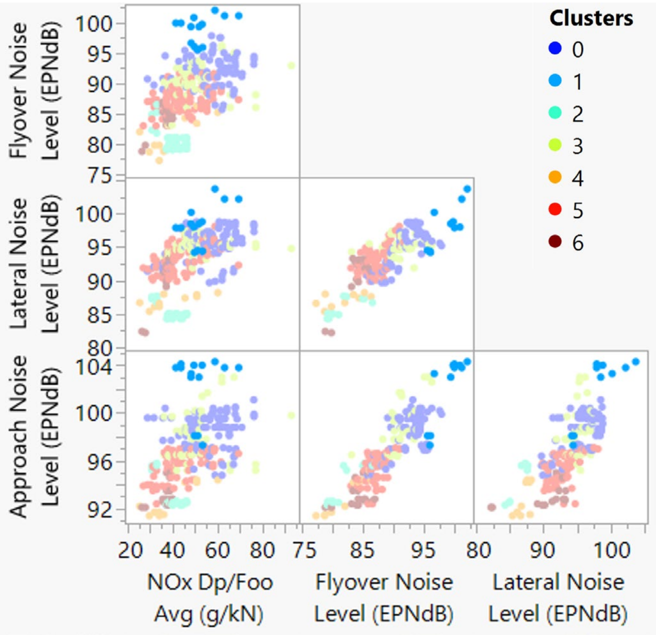
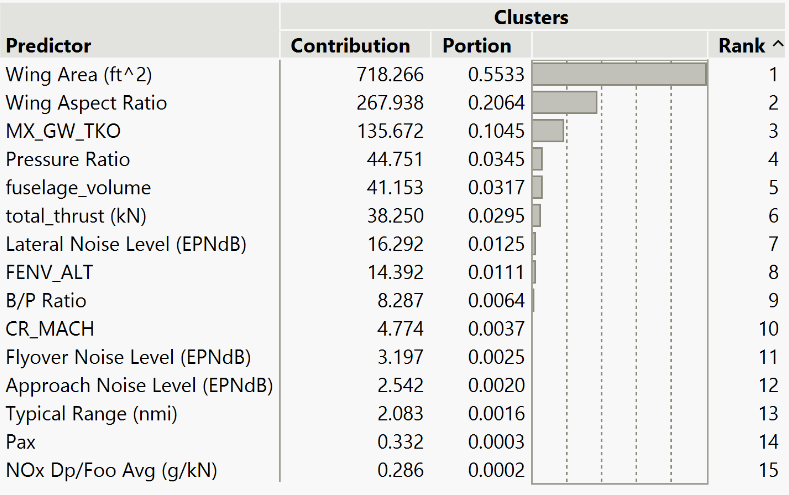


Figure 4. Scatterplot matrix of emissions and noise, with cluster 1 highlighted.



	Clusters			
Predictor	**Contribution**	**Portion**		**Rank ^**
Wing Area (ft^2)	718.266	0.5533		1
Wing Aspect Ratio	267.938	0.2064		2
MX_GW_TKO	135.672	0.1045		3
Pressure Ratio	44.751	0.0345		4
fuselage_volume	41.153	0.0317		5
total_thrust (kN)	38.250	0.0295		6
Lateral Noise Level (EPNdB)	16.292	0.0125		7
FENV_ALT	14.392	0.0111		8
B/P Ratio	8.287	0.0064		9
CR_MACH	4.774	0.0037		10
Flyover Noise Level (EPNdB)	3.197	0.0025		11
Approach Noise Level (EPNdB)	2.542	0.0020		12
Typical Range (nmi)	2.083	0.0016		13
Pax	0.332	0.0003		14
NOx Dp/Foo Avg (g/kN)	0.286	0.0002		15

Figure 5. Parameter importance for overall clustering.

The present work makes two primary contributions. The first contribution is the generation and continuous development of the Fleet extension database, which enriches the AEDT Fleet dB with performance, weight, emissions, and noise parameter values from openly available external data sources. The second contribution is the exploration of various ML techniques to identify commonalities and patterns in the airframe/engine combinations. The changes to the Fleet dB will be contrasted with the default AEDT mapping of different airframe/engine combinations to ANP native aircraft, thereby enabling the exploration of areas for improvement in fleet modeling of noise and emissions within AEDT, to improve its accuracy.

Major Accomplishments

The major accomplishments for this period performance include the following:

- A literature study was conducted on databases to collect performance, emission, and noise data for target aircraft.
- A new template was created for the Fleet extension database, and external data were gathered.
- External databases were gathered to augment the extension database with completion of 520 aircraft engine combinations.
- A literature survey was conducted on analytical methods in clustering, dimensionality reduction, feature selection, and data visualization.
- Unsupervised clustering on the available Fleet extension database was explored, to better group similar aircraft and provide insights on the parameters driving the grouping.
- The results were postprocessed by using bar charts, scatterplot matrices, *t*-distributed stochastic neighbor embedding, and parameter importance calculations, to help better understand the trends.

Publications

None.

Outreach Efforts

Bi-annual ASCENT meetings.

Awards

None.

Student Involvement

Styliani I. Kampezidou and Cristian Puebla-Menne (graduate students). Conducted research on potential analytical techniques to use for the clustering.

Plans for Next Period

- Finalize the ANP extension database to include noise certification data, to serve as the basis for Task 2
- Continue to refine analytical methods on the new database, identify gaps in the approach, and implement them on the remaining engine/airframe combinations within the FLEET database
- Validate the methods in Task 2

References

Hinton, G. E., & Roweis, S. (2002). Stochastic neighbor embedding. *Advances in neural information processing systems*, *15*.
Maaten, L. V., & Hinton, G. E. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research, 9*, 2579-2605.
Melit Devassy, B., & George, S. (2020). Dimensionality reduction and visualisation of hyperspectral ink data using t-SNE. *Forensic Science International*, *311*, 110194. https://doi.org/10.1016/j.forsciint.2020.110194