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

- P.I.: Professor Dimitri Mavris
- FAA Award Number: 13-C-AJFE-GIT-104
- Period of Performance: October 1, 2021 to September 30, 2022
- Tasks:
    1. Identification and review of aircraft not in the Aviation Environmental Design Tool (AEDT)
    2. Analytical method development

Project Funding Level

The current FAA funding for this project is $150,001 from October 1, 2021 to September 30, 2022. The Georgia Institute of Technology has agreed to a total of $150,001 in matching funds.

Investigation Team for All Tasks

- Dr. Dimitri Mavris (P.I.), Georgia Institute of Technology
- Dr. Michelle R. Kirby, Georgia Institute of Technology
- Dr. Mayank Bendarkar, Georgia Institute of Technology
- Dr. Dushhyanth Rajaram, Georgia Institute of Technology
- Dr. Ameya Behere, Georgia Institute of Technology
- Styliani I. Kampezidou (graduate student), Georgia Institute of Technology
- Cristian Puebla-Menne (graduate student), Georgia Institute of Technology

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 (NPD) 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 improve 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 as 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 real-world flight and physics-based modeling 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 database to include noise and performance data for aircraft currently not in the ANP database.

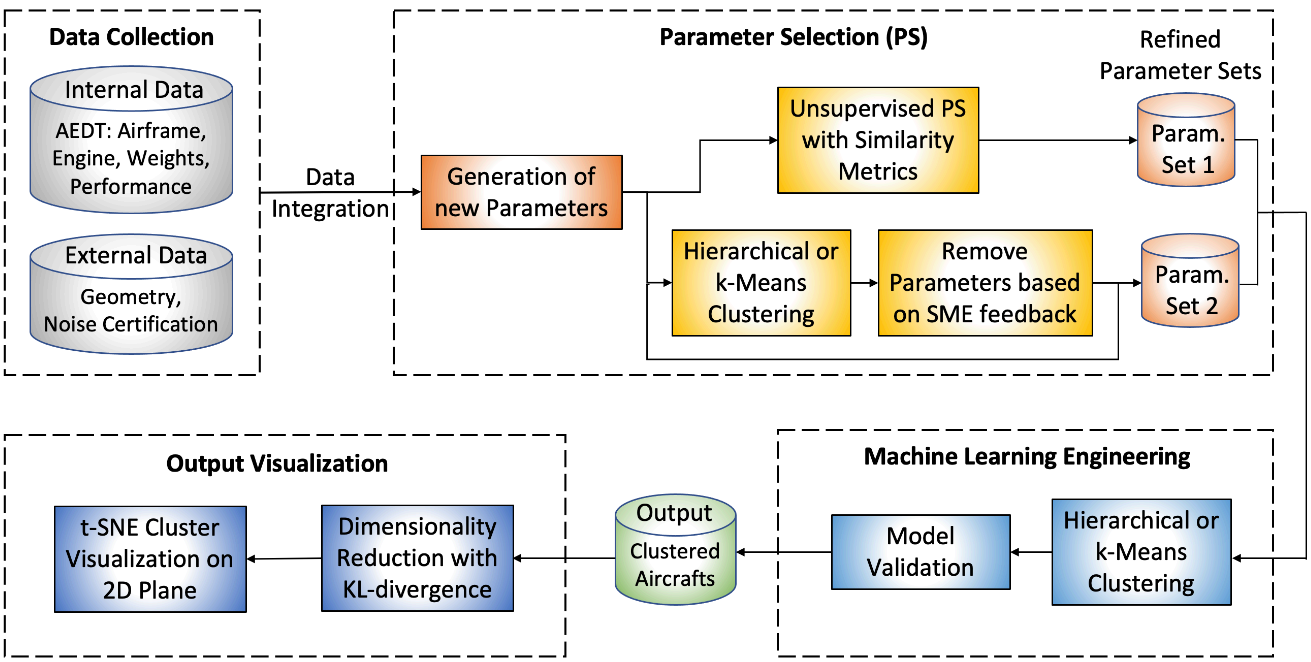


Figure 1. Overview of ASCENT Project 60 tasks and workflow.

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 database) or will be collected from external resources.

Task 1 - Identification and Review of Aircraft not in the AEDT

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 database, 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 database contains representative aircraft for the entire fleet; some aircraft are modeled according to ANP data, whereas others are represented by a substitution aircraft. This task involves the identification of aircraft that do not have ANP data and are represented by a substitution aircraft.

Research Approach

Creating the AEDT FLEET Extension Database

Aircraft without ANP data in AEDT

The aircraft not currently modeled with ANP data are identified by reviewing the AEDT Fleet database 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 database are studied to summarize 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.

The Fleet database consists of 3,626 airframe and 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. Apparently, the native aircraft have a matching equipment ID and default equipment ID. This substitution enables proxy aircraft to borrow ANP data from native aircraft for the purposes of conducting environmental analyses and studies. Figure 2 below illustrates the Fleet database breakdown in terms of ANP data availability as well as the current efforts in extending the available parameters. The additional parameters collected from external resources are summarized in Table 1.

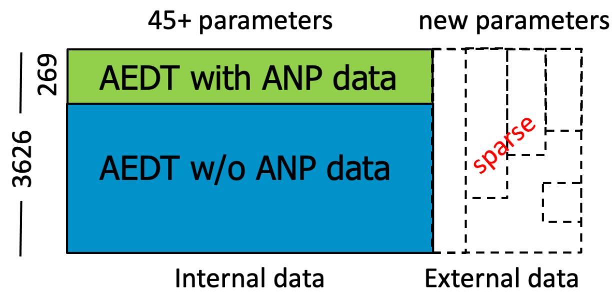


Figure 2. Fleet database breakdown with respect to ANP data availability.

Aircraft Database Literature Study

The main objective of this task is to collect data from various databases for a wide range of aircraft. This information is helpful in determining which performance, emissions, and noise parameters will be used for the substitution algorithm. In particular, we are interested in the following categories of data:

- **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**: example: flyover noise level

Overview of Tables Available in the Fleet database (Internal Data) and Associated IDs:

Some of the internal data collected from the Fleet database correspond to:

- FLT_EQUIPMENT (provides the AIRFRAME_ID and ENGINE_ID for each equipment EQUIP_ID)
- FLT_AIRFRAMES (contains airframe information that can be accessed by using the AIRFRAME_ID from the corresponding EQUIP_ID in the FLT_EQUIPMENT table)
- FLT_ANP_AIRPLANES
- FLT_FLEET
- FLT_ENGINES (contains information on engines and emissions that can be retrieved by using the ENGINE_ID from the corresponding EQUIP_ID in the FLT_EQUIPMENT table)
- FLT_ENGINES.MODEL
- FLT_CAT_DESIGNATIONS

- FLT_CAT_ICAO_TYPES
- FLT_BADA_ACFT
- FLT_ANP_AIRPLANE_NOISE_GROUPS
- FLT_NOISE_CERTIFICATION

The Fleet database contains 3,626 EQUIP_IDs, 848 unique AIRFRAME_IDs and 686 unique ENGINE_IDs. The FLT_NOISE_CERTIFICATION table has a total of 8,288 records (rows). Among the 3,626 equipment types, only 535 (15%) have noise certification records. All these records have a one-to-many match; i.e., for a certain equipment type, multiple matches exist in the FLT_NOISE_CERTIFICATION table. The number of matches ranges from 2 to more than 100. Efforts by the FAA to identify a unique path from EQUIPMENT_IDs to unique NOISEDB_IDs are ongoing; hence, our team will focus on collecting noise parameter values from up-to-date reliable external data sources. Regarding noise parameter values from the Fleet database, two potential routes for retrieving NOISEDB_IDs are proposed in Figure 3.

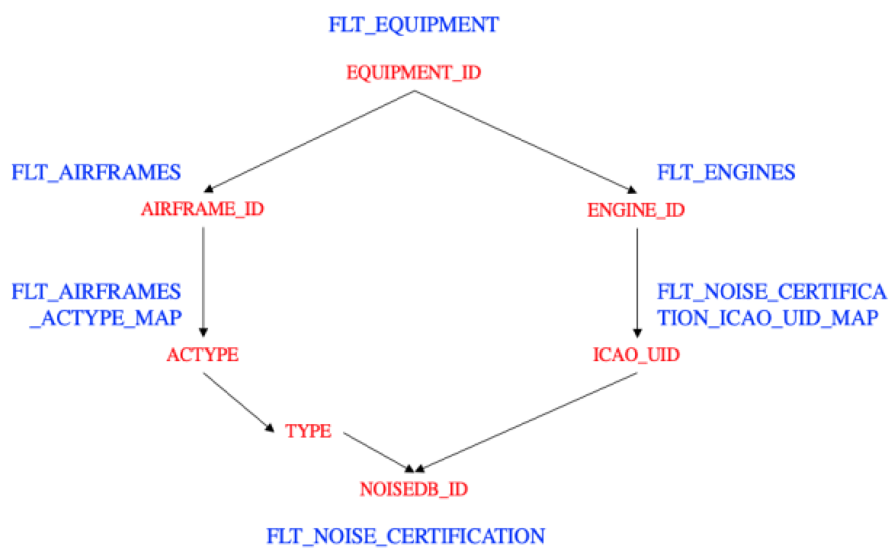


Figure 3. Two alternative Fleet database routes for identifying the one-to-many choices from EQUIPMENT_ID to NOISEDB_ID.

To create an initial database to use for the analytical methods in Task 2, we applied filtering to the original number of unique equipment IDs to establish a subset of engine/airframe combinations 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 unique equipment IDs to 2,443. For the remaining EQUIP_IDs, the AIRFRAME_MODEL names were grouped to determine the number of unique airframes. With an initial focus on U.S. applications of AEDT, airframe models not operated in the United States and the production status of the airframe models were eliminated. For future efforts of this research, a broader set of aircraft types can be included to extend beyond U.S. operations. These filters reduced the total airframes for which external data are required to a manageable number of 138. Notably, each airframe could have multiple engine types.

Data Sources Used

The external databases used to extend AEDT's available parameters are summarized in Table 1 below. To augment the existing ANP database for the unique engine/airframe combinations that do not have ANP data, we identified external data sources. The databases considered to populate the extended Fleet database table (Figure 2) are provided below. First, information from the internal Fleet database was used, and then, various other databases were retrieved to populate the external parameters:

- AEDT ANP (Fleet + FLEET-FULL databases): The Fleet database, the most comprehensive performance database available, contains multiple performance parameters for a wide variety of airframe/engine combinations. Although it is not publicly available, the FLEET-FULL database contains information for all registered aircraft worldwide.

Minimum and maximum values for the same airframe/engine combinations are available for certain aircraft parameters (e.g., MTOW), depending on the aircraft equipment used by an airline on board.

- Janes: This database contains aircraft information such as certification date, weight, range, production date, and status, as well as corresponding engine information (i.e., thrust, number of engines, power, and speed). This database also contains wing and fuselage dimensions, maximum payload, and number of passengers (Janes, n.d.)
- EASA Emissions Databank v28 (EASA, 2023) This Microsoft Excel-based database covers turbojet and turbofan engines for which emissions are regulated (static thrust of 27 kN or higher). It contains engine emissions for CO, nitrogen oxides (NO_x), and unburned hydrocarbons as well as several engine performance parameters, including bypass ratio and thrust-rated output
- Jet Engines: This database contains engine information only, such as thrust type, SFC type, airflow, overall pressure ratio, fan pressure ratio, bypass ratio, speed, and engine dimensions (Meier, 2021).
- Piano v2.2: This database contains aircraft and engine combination data, including wing and fuselage dimension information, weight, payload, and specific air range evaluations (Lissys Ltd., n.d.)
- European Union Aviation Safety Agency (EASA) Type Certificate Data Sheet (TCDS): This database contains aircraft-specific variants and designation of the engines certified on the aircraft, in addition to geometric and performance information. (EASA 2023).
- Bluebook: This database contains aircraft/engine combination information, such as thrust, maximum speed, recommended speed, stall speed dirty, fuel, gross weight, empty weight, range, length, height, and wingspan
- Elsevier: This database contains aircraft and engine information, including thrust, number of passengers, weight, payload, fuselage dimensions, and wing dimensions (Jenkinson, et al., 2001a; Jenkinson, et al., 2001b).
- EASA certification noise databases (EASA, 2023) https://www.easa.europa.eu/en/domains/environment/easa-certification-noise-levels. This very large Excel database consists of aircraft/engine types, effective perceived noise in decibels (EPNdB), and noise levels for lateral, flyover, and approach. EASA is a collection of four noise databases that address heavy propeller-driven airplanes, jet airplanes, light propeller-driven airplanes, and rotorcraft
- DGACv2.30: This noise certification database offered by French authorities will be used along with EASA noise certification values to collect up-to-date noise certification levels (NoisedB, 2023).
- Online photographic material: For the purposes of identifying wingtip presence, wing location, and engine location, pictures available online have been considered.

According to engineering judgement and prior research on key drivers of noise, emissions, and fuel burn, a set of parameters to define a unique engine/airframe combination were established, which include internal AEDT data and external data. The purpose of the additional parameters is to enhance the information for a particular combination, so that a better substitute aircraft can be identified to represent the environmental impact of that combination (performance, noise, and emissions). In Table 1, the first column indicates the broader parameter group; the second column provides the parameter details, which could be an existing Fleet database parameter or a newly added one; and the third column shows internal or external resources from where the parameters were collected. This ANP extension database will serve as the basis for the analytical method approach in Task 2.

Table 1. Summary of external data sources used to extend the AEDT FLEET database.

Parameter group	Collected parameters	Resources
Performance	Typical cruise speed, typical range	Jet Engines, Piano, Janes, Bluebook, public resources
Weights	Maximum payload	Piano, Janes, Elsevier
Geometry	Wingspan, wing area, fuselage height/ width/length, typical number of passengers	Piano, Janes, Bluebook, public resources
Engine	B/P ratio, pressure ratio, thrust, emissions	Jet Engines, Piano, Janes, Elsevier, International Civil Aviation Organization Engine Emissions Databank, Purdue Engineering, Campbell Hill
Noise	Flyover, lateral, approach	EASA noise database

To populate ANP database with noise data, we used two sources from EASA certification noise level database (EASA, 2023).For jet airframe/engine combinations, MAdB Jets (210408) and its updated version, MAdB Jets (20220331), were used; for propeller-driven aircraft, MAdB Heavy Prop (21325) was used. The data extracted from these sources consisted of the following information:

- Lateral noise level
- Flyover noise level
- Approach noise level

In each case, along with the noise level, the limit, the margin as well as cumulative noise values in EPNdB units were extracted. The methods used for matching comprised the following steps. In the ANP database, a total of 996 airframe/engine candidate combinations for noise data population was selected. The population procedure was started by selecting a specific airframe of interest (for example, the Airbus A321-200 Series). For that airframe, a specific engine 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 TCDS to verify that the variants existing in the EASA certification noise level database (e.g., MAdB Jets) for the airframe selected are certified.
- 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 more than one airframe/engine combination have the same noise values, the first entry is selected.
- Finally, if more than one combination remains after the 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 is to choose the most representative noise value of the combination selected. After a unique combination is found, the corresponding noise values are transferred from the EASA database to ANP. To increase the number of combinations available for which noise values were obtained, engines with similar designation codes were selected 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. The application of these criteria enabled the generation of the noise values for the airframe/engine combinations used in Task 2.

Challenges in Data Integration

Multiple challenges exist in collecting data from external resources and integrating them into the extended AEDT table:

- The data quality from websites other than those of the FAA, manufacturers, or certification organizations may be questionable.
- The external data are sparse, thus generating challenges for machine learning (ML) model training.
- The integration of multiple databases can be labor-intensive, unless automation is introduced to bypass it.

After incorporation of noise parameter values for flyover, lateral and approach noise from the EASA certification noise level database, the number of available airframe and engine combinations was only 438.

Milestone(s)

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

None

Outreach Efforts
Biweekly calls
Bi-annual ASCENT meetings

Awards
None

Student Involvement
Styliani I. Kampezidou (graduate student)
Cristian Puebla-Menne (graduate student

Plans for Next Period
Continue gathering certification data

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, 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 utilized to approximate the remaining aircraft with more 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.

In this project, the data collection and integration (Task 1) and analytical method development (Task 2) have been conducted in parallel. The proposed analytical approaches have been applied to selected problems. However, notably, because the master data set has not yet been finalized, as described in previous sections, the concepts of the analytical methods are demonstrated primarily through notional or incomplete data sets. The present report highlights the progress made over the past year and is not cumulative.

The method is outlined in Figure 1. It begins with a data fusion step, wherein different data sources are queried and merged with the AEDT Fleet database 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 NO_x emission data is 2,361 which decreases to 520 when noise data are also included. Of these, 269 aircraft have data from the ANP database.

Three broad areas will be explored to synthesize ANP data for aircraft lacking these data. The first step is to explore unsupervised clustering to group similar aircraft by using the enriched data set 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. Clustering also aids in identifying outliers in the data and correcting the data entries for any potential errors.

A second approach potentially customizes 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. The third approach in Figure 1 will explore hybrid models, wherein a composite model of multiple closest ANP aircraft is used to synthesize ANP data for non-native aircraft in AEDT. This approach is currently being developed and will be included in future work.

Using Clustering to Identify Representative Aircraft Model Portfolios

Groups of similar aircraft are placed in the same cluster, and dissimilar aircraft are placed in other clusters. Clustering is a typical task in unsupervised ML, and extensive methods have been reported in the literature. The choice of a specific clustering algorithm depends on the objectives of the problem. In this project, clustering can be used to achieve at least two aims. The first aim is to group similar aircraft and compare the results with the current ANP aircraft substitutions to improve the current substitution mapping and to identify gaps. Many algorithms in the literature, such as *k*-means (KM), hierarchical clustering, and DBSCAN, can all achieve this objective. The second aim is to select representative aircraft types from the population through clustering. In a basic process, 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 are also available for conducting clustering and representative aircraft selection simultaneously.

To categorize the airframe/engine combinations on the basis of the different parameters within the enriched data set, we use two clustering techniques: KM and agglomerative hierarchical clustering (AHC). In the present work, aircraft with similar performance, geometry, engine characteristics, noise, and emissions are grouped. The dimensionality of this task is equal to the number of parameters selected by using subject matter experts (SME) inputs, which are shown in Table 2. These parameters have been selected after multiple rounds of clustering experiments involving SME feedback. The effects of the parameters on the physics of noise propagation and their correlation with other parameters in Table 1 were considered.

Table 2. 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

The number of clusters is determined with the elbow method for KM clustering, wherein a suitable trade-off between error and the number of clusters is determined. Figure 4 shows the inertia (elbow) plot for selecting the number of clusters for the KM algorithm. The same number of clusters is used for AHC.

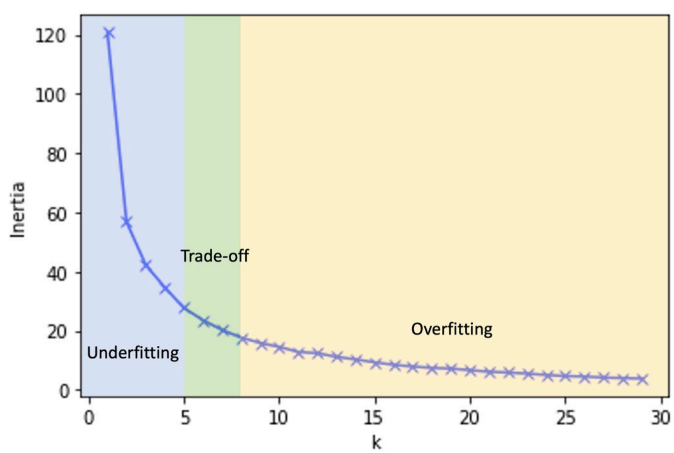
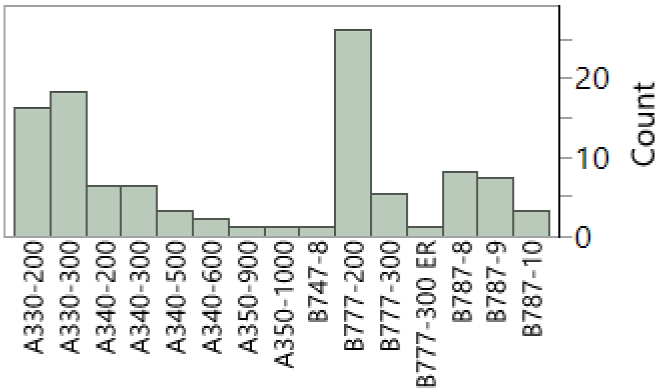


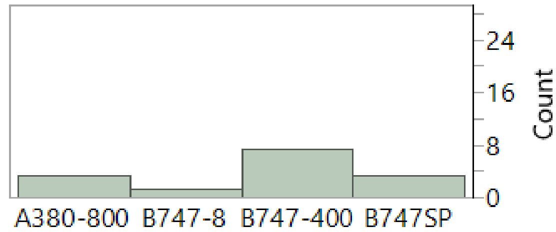
Figure 4. Inertia (elbow) plot for KM clustering.

The clusters can be visualized by using scatterplot matrices and also using *t*-distributed stochastic neighbor embedding (*t*-SNE) (Melit Devassy, 2020). *t*-SNE is very useful for visualizing data with more than three dimensions, by creating a low-dimensional embedding of the original data in two-dimensional or three-dimensional space (Maaten, 2008). The embedding is generated by minimizing the Kullback–Leibler divergence over all high-dimensional data points with a gradient descent method. *t*-SNE is an updated version of the originally proposed SNE that mitigates the crowding problem and optimization problems, by using a symmetric SNE objective function and simpler gradients as well as Student's *t* distribution instead of a Gaussian distribution to evaluate the similarity of the data points in the low-dimensional space (Hinton, 2002).

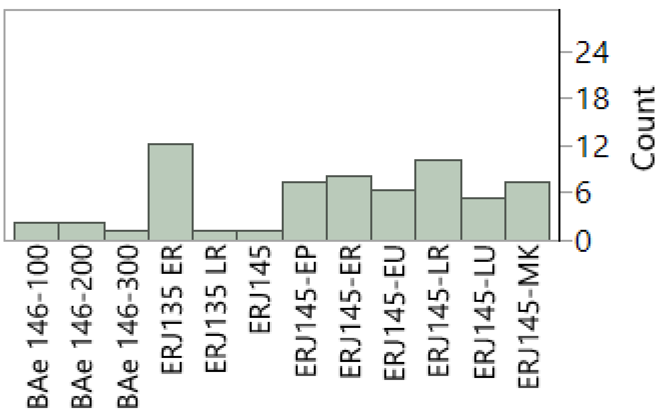
Preliminary Clustering Results:
The approximately 520 aircraft for which the ADET Fleet extension database has complete parameter data are included in the preliminary results herein. An elbow plot denoting the inertia (within cluster sum of squares) versus the number of clusters is shown in Figure 5. Approximately five to seven clusters appear to be ideal to divide the data. Although this is helpful for KM clustering, it is also used for AHC in the present work.



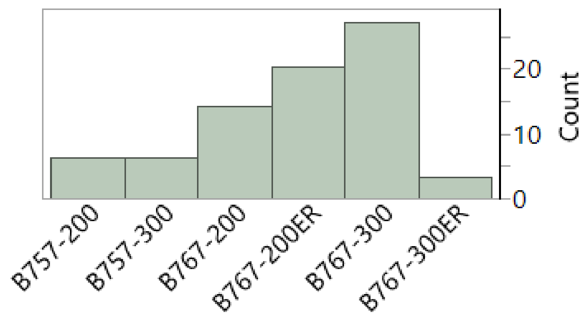
(a) Cluster 0



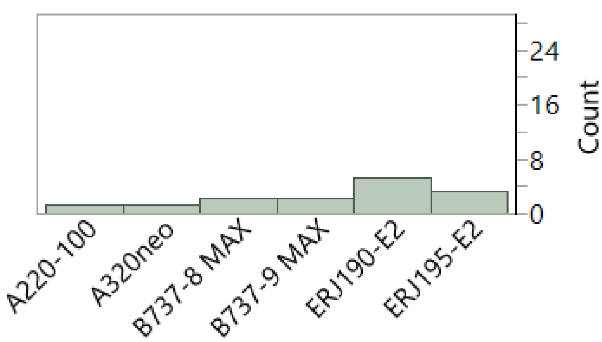
(b) Cluster 1



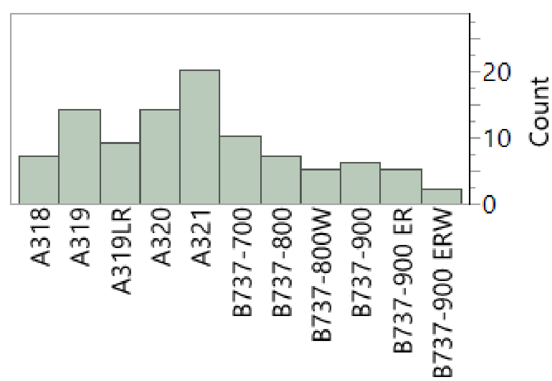
(c) Cluster 2



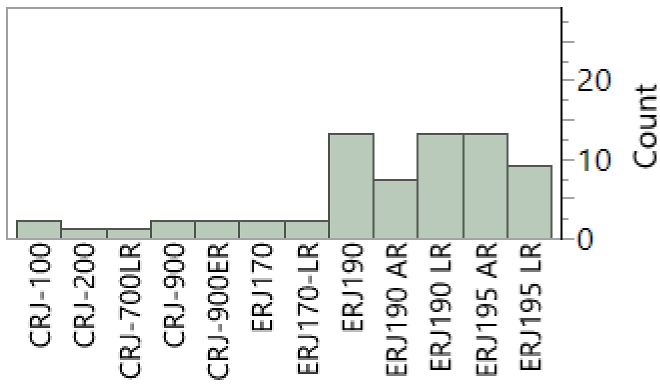
(d) Cluster 3



(e) Cluster 4



(f) Cluster 5



(g) Cluster 6

Figure 5. Preliminary hierarchical clustering results.

Results with AHC with seven clusters are shown in Figure 5. The efficacy of clustering is determined on the basis of SME feedback. Overall, the clusters show good agreement with real-world distinctions: larger wide-body aircraft form cluster 0, so-called "jumbo" jets form cluster 1, regional jets form clusters 2 and 6, smaller wide-body aircraft are in cluster 3, newer-generation small single-aisle aircraft are in cluster 4, and traditional small single-aisle aircraft are in cluster 5.

Visualizing the results of clustering poses a challenge, because the algorithm operates in 15 dimensions, but the results can be presented in only two. Figure 6 shows an example scatterplot matrix of NO_x and noise emissions for aircraft within cluster 1. As expected, the largest aircraft and highest thrust engines that pair with them have the highest emissions and noise signatures, and thus are 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.

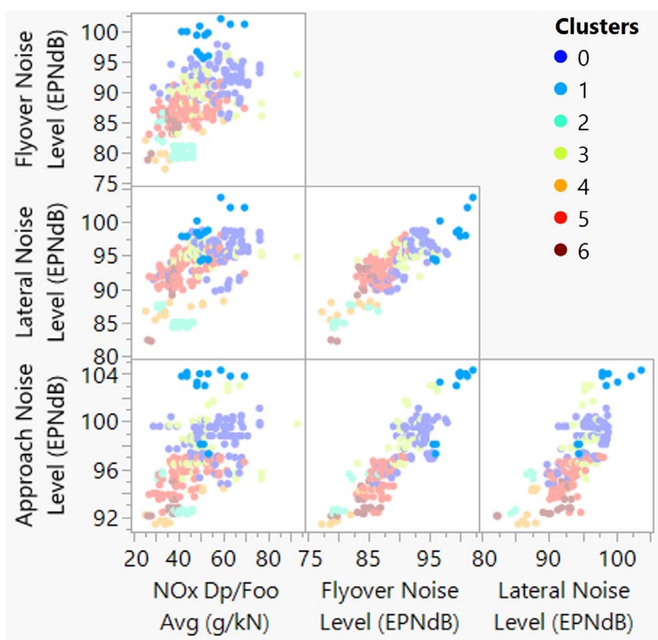


Figure 6. Scatterplot matrix of emissions and noise, with cluster 1 highlighted.

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 used to cluster the aircraft. A parameter importance function of this random forest was evaluated to indicate the parameter importance of the AHC clusters (Figure 7).

	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 7. Parameter importance for overall clustering.

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.

The present work makes two primary contributions. The first is the generation of the Fleet extension database, which enriches the AEDT Fleet database with performance, weight, emissions, and noise parameter values from openly available external data sources. The second is the exploration of various ML techniques to identify commonalities and patterns in the airframe/engine combinations. The changes to the Fleet database will be contrasted against 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 was explored on the available Fleet extension database 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*-SNE, and parameter importance calculations to help better understand the trends.

Publications

None.

Outreach Efforts

Biweekly calls

Bi-annual ASCENT meetings

Awards
None.

Student Involvement
Styliani I. Kampezidou (graduate student)
Cristian Puebla-Menne (graduate student

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