# Project 033 Alternative Fuels Test Database Library

## University of Illinois at Urbana-Champaign

## Project Lead Investigator

Tonghun Lee
Professor
Mechanical Science & Engineering
University of Illinois at Urbana-Champaign
1206 W. Green St.
Urbana, IL 61801
517-290-8005
tonghun@illinois.edu

## University Participants

**University of Illinois at Urbana-Champaign**

- PI: Tonghun Lee, Professor
- FAA Award Number: 13-C-AJFE-UI, Amendment 33
- Period of Performance: October 1, 2020, to September 30, 2021
- Tasks:
    1. Generation II online database: JETSCREEN, ALIGHT and NewJET and domestic airport connection
    2. Machine-learning-based online analysis

## Project Funding Level

FAA funding Level: $200,000
Cost sharing: Software license support from Reaction Design (ANSYS)

## Investigation Team

- Tonghun Lee (Professor, University of Illinois at Urbana-Champaign): overall research supervision
- Ji Hun Oh (graduate student, University of Illinois at Urbana-Champaign): database development and machine-learning-based analysis
- Alex Solecki (graduate student, University of Illinois at Urbana-Champaign): database development

## Project Overview

This study seeks to develop a comprehensive and foundational database of current and emerging alternative jet fuels by integrating relevant pre-existing jet fuel data into a common archive that can support scientific research, enhance operational safety, and provide guidelines for the design and certification of new jet fuels. In light of the September 2021 White House statement on advancing the future of sustainable aviation fuels (SAFs) in America, the database now has even greater potential to serve the national agenda. In previous years of this project, efforts were focused on the integration and analysis of pre-existing jet fuel data from various government agencies and individual research groups. In 2020, we converted all the compiled data to a new nonstructured query language (NoSQL) format by using a JavaScript object notation (JSON) schema, thus allowing the data to be analyzed in a flexible manner with various programming languages. To this end, we have launched the second generation of our online database, the Alternative Jet Fuels Test Database (AJFTD), which utilizes the new nonrelational database structure. This version is equipped with interactive analysis functions for users and flexible methods for plotting and downloading data. In the previous year, we extended this effort to incorporate advanced machine-learning algorithms in the analysis process. Additionally, we have integrated our database with the database assembled by the European Jet fuel screening and optimization (JETSCREEN) program. In the future, data acquisition from domestic and international airports will help further develop the database and support its use as a destination for all sustainable-aviation-related property and test data.


**Figure 1.** Generation II national Alternative Jet Fuels Test Database web interface (altjetfuels.illinois.edu).

We hope that the database will ultimately serve not only as a comprehensive centralized knowledge base utilized by the jet fuel research community but also as a resource that can enhance global operation efficiency and safety. Future efforts will include linking real-time fuel usage and certification data from domestic and international airports. Connecting our database with ongoing European projects, such as ALIGHT and NewJET, will help create avenues for future database development in this area. With the expected prolific diversification of new alternative jet fuels in the near future, the ability to track critical fuel properties and test data from both research and operation perspectives will be highly valuable for the future of commercial aviation. Furthermore, increasing the breadth of data categories available in the database - from fuel data to global usage trends—will increase its relevance to a broader audience. We hope that the ongoing website development and improvement of the user interface will also allow the general public to engage with high-level information thus increasing the awareness of SAFs and further supporting the nationwide visibility of sustainable aviation.

## Task 1- Generation II Online Database Updates: JETSCREEN, ALIGHT, NewJET, and Domestic Airport Connection

University of Illinois at Urbana-Champaign

### Objective(s)

The main objective of this task is to upgrade and debug the generation II online national AJFTD functions and link the database to the European JETSCREEN, ALIGHT, and NewJET programs. The generation II database has been designed with a new architecture allowing for flexible analysis and scaling based on a NoSQL data format. This format can accommodate various data types that can be easily accessed by any common programming language, and basic analysis functions have been directly built into the web interface. After the launch of the generation II web interface, extensive efforts in the past year have been focused on upgrading the functionality and addressing bugs according to user feedback. We have also converted much of the data to comma-separated value (CSV) format to enable future machine-learning-based analysis, as further discussed in Task 2. The specific goals in Task 1 are:

- Test and improve the functionality of the generation II online web interface and database structure
- Pilot connection with the European ALIGHT and NewJET programs
- Establish methods for acquiring real-time airport fuels data (previously delayed because of the COVID-19 pandemic; efforts restarted as of September 2020)

## Research Approach

### Generation II Database Debugging and Upgrade

A beta version of the generation II database was launched online in the summer of 2019. The web interface of the generation II database is shown in Figure 1. All functionality of the previous database has been maintained, and the security login features have been migrated from the previous version. The generation II web interface, much like that of generation I, is an HTML-oriented program built on a layer of metadata, which supports search functions for users. The tree structure applied to organize the data folders in the first database has been retained in this version, thus allowing users to access the data in a similar manner. The main difference is the addition of an inner core that houses the JSON files and is where the test data reside. Currently, the database has grown to contain more than 25,000 separate fuel records.

The catalog of data currently available in the database is assembled primarily from four separate sources. The fuels with POSF (Air Force Research Laboratory [AFRL] fuel database code) number designations were added from the internal database maintained by the AFRL at the Wright Patterson Air Force Base. The second dataset was obtained from the Petroleum Quality Information System (PQIS) reports of the Naval Air Systems Command (NAVAIR) and corresponds to a compilation of fuel data geared primarily toward government use. The third set was provided by Metron Aviation, which compiled fuel properties from samples collected at airports through a previous ASCENT project. The dataset resulting from this study has proven valuable by providing a landscape of fuels currently used in commercial aviation and will guide our future efforts focused on capturing this type of data in real time. The final dataset was obtained from the National Jet Fuel Combustion Program (NJFCP) within ASCENT.

After the launch of the generation II database, significant changes have been made to fix bugs and upgrade various aspects of the database. Several key changes to the database are summarized below.

- During the integration of the database with JETSCREEN, we modified the labeling structure to ensure that files were coded as JETSCREEN data and were separately searchable. Similar to how the Metron data are labeled according to the airport from which the data were retrieved, JETSCREEN has been added as a separate search label. This search filter can be integrated with additional search filters to allow users to view tests for a specific fuel type from the JETSCREEN group if desired.

- The Export and Compare features also required updates. Although we had worked with JETSCREEN extensively to create a standard JSON format, the files generated from the AJFTD and JETSCREEN had minor differences, which caused the current code on the database to fail occasionally. Slight differences among the JETSCREEN files themselves were also causing errors in attempts to compare the data with FAA files as well as with other JETSCREEN files. The Export and Compare features were updated to work around these issues and to support the comparison of all test files in the database.

- The display of data in the database was also changed to allow for more privacy and security. Authors of the files (which included student names) were removed from the JETSCREEN display. The sharing function, which was added to share selected FAA data with JETSCREEN via Amazon Web Services (AWS), was set to display for only administrator accounts. These shared FAA files sync with the AWS S3 bucket every hour. The JETSCREEN bucket on AWS is checked each day for new files, which are then downloaded to the website.

- Efforts were made to convert the JSON format in the database to CSV format for select files to enable machine-learning-based analysis, which will be addressed further in Task 2. The actual files being stored will use the NoSQL JSON format, which is more conducive to maintaining a flexible database. However, certain parts of the data to be analyzed with machine learning will need to be converted to CSV format, for which multiple Python-based machine-learning scripts are available. In the future, a process to automate this conversion in real time may be necessary.

- Online viewing of 2D gas chromatography (GCxGC) data is being made available to users (beta testing in progress). The scripts that process data uploads have been revised to properly process the section containing GCxGC data and now present this information on the web portal. This development is important because it is the first of many upcoming improvements to the user interface that will allow users to engage with data on the website itself in a comprehensive and high-level fashion before downloading and processing raw files. The data are available in both tabulated format and bar chart format by carbon number. Figure 2 shows an example of the tabulated GCxGC data for a sample fuel. The table organizes compositional data by chemical family and further by subgroups within that family, and the final values are the species within the subgroup, ordered by carbon number. The Expand feature allows for more focused viewing of compositional data by chemical family and subgroup. Users can choose to view composition data by weight or volume percent with a toggle switch at the top of the table. Figure 3 depicts a graphical representation of the fuel composition by carbon number and chemical family, which further elucidates the distribution of the fuel species by carbon number and also gives users a general idea of the most abundant species in the fuel sample. The scroll-over feature allows users to preview numerical data for a given bar section. Users may also zoom by using a mouse scroll wheel. For ease of viewing of less abundant species, the chemical families represented in the bar chart can be toggled on and off. The *y* axis automatically adjusts accordingly for optimal viewing. As with the table, the bar chart can also be viewed by weight percent or volume percent.

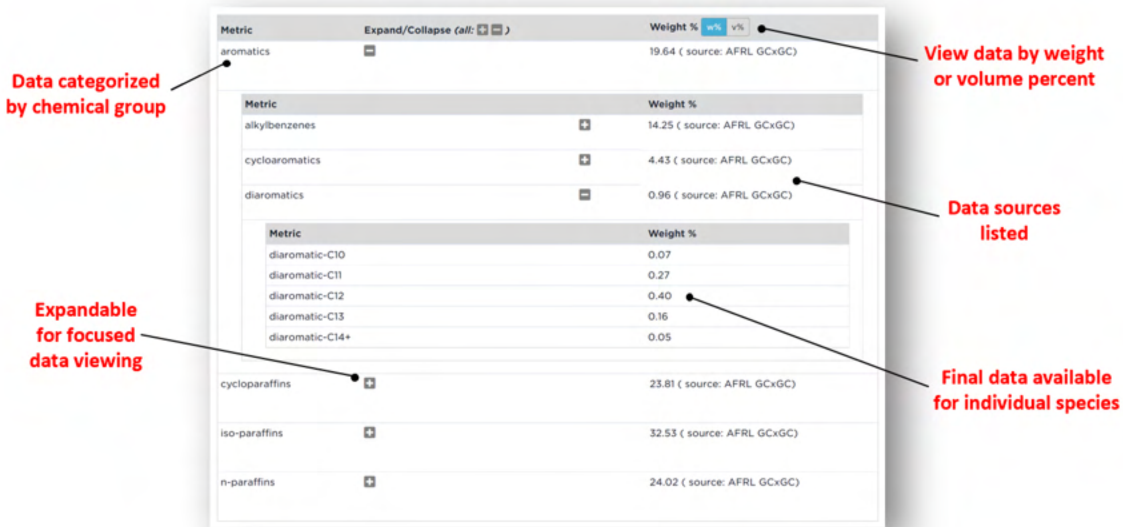


**Figure 2.** Updates to data viewing on the AJFTD website: view of GCxGC data in tabulated format.

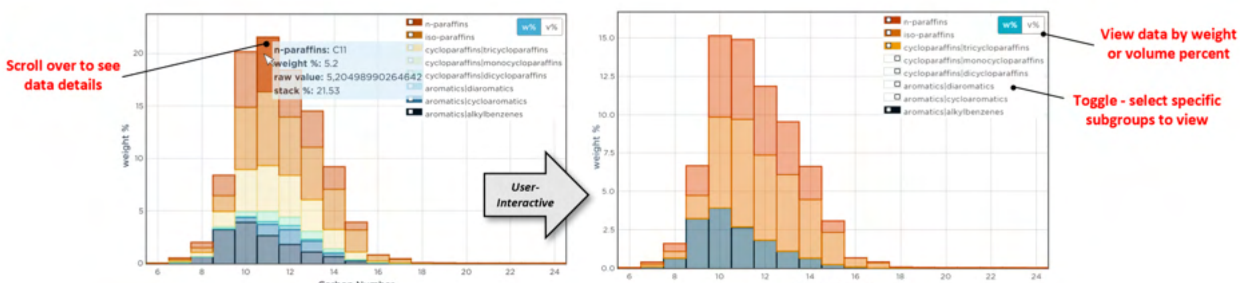


**Figure 3.** Updates to data viewing on AJFTD website: view of GCxGC data in carbon number bar chart format.

**Integration of the Database with JETSCREEN, ALIGHT, and NewJET**

In 2020, we completed the integration of our database with the European JETSCREEN program. The JETSCREEN program was initiated to provide fuel producers, air framers, and aero-engine and fuel system original equipment manufacturers (OEMs) with knowledge-based screening tools for fuels and have a similar database that could be linked with ours. We first started discussing a potential merger with the JETSCREEN database in 2018, after which we started methodically synchronizing the data structure so that a merger could be possible. After extensive beta testing, the two databases were first linked in the spring of 2020. Data sharing between AJFTD and JETSCREEN ended in 2020, when the JETSCREEN project was completed and archived.

After the completion of JETSCREEN, AJFTD will continue acquiring new data through connections with new European programs ALIGHT and NewJET. ALIGHT is a program aiming to assess and improve the supply chain, integration, and use of SAFs and smart energy solutions through examining and optimizing operations at Copenhagen Airport. NewJET is a research program headed by the University of Birmingham aiming to improve pathways for production of new SAFs. We anticipate using a similar data-sharing structure to that used by JETSCREEN to connect these programs with AJFTD (process outlined in Figure 4).

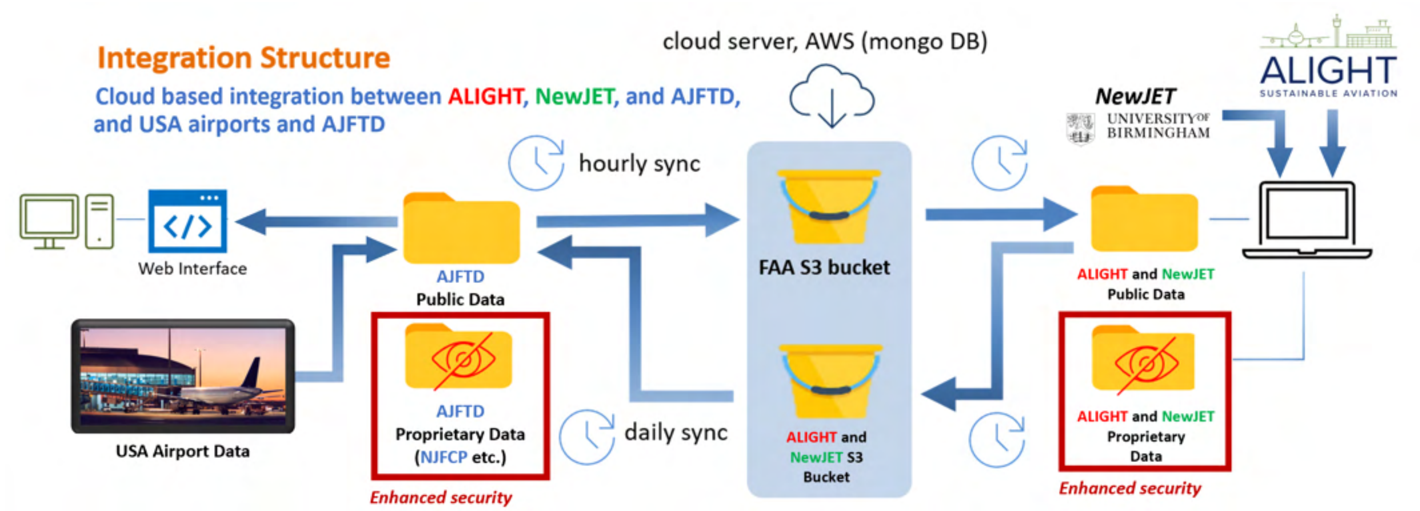


**Figure 4.** Plans for future data sharing with the European sustainable aviation programs ALIGHT and NewJET.

The JETSCREEN integration process has helped streamline a data-sharing method that ensures frequently updated and well-secured data flow between two projects; thus, this method of data sharing is expected to be used in future integration with the aforementioned ALIGHT and NewJET programs. Previously, the JETSCREEN and FAA databases were joined by a common cloud storage. AWS was selected as the server to store the shared data, mainly because of its affiliation with the University of Illinois. S3 buckets (Amazon database structure) were created for both FAA and JETSCREEN to share their JSON files. Each could pull files from the other's folder, but read and write access was granted to only the owners of the bucket. The FAA data are shared with its S3 bucket via altjetfuels.illinois.edu. All public FAA data on the website will have an option to be shared with JETSCREEN, which can be toggled by administrators. The website syncs hourly with the bucket to upload newly shared data. No proprietary data are shared with the FAA S3 bucket. Any files uploaded to the FAA bucket can be viewed and downloaded by JETSCREEN. For downloading new JETSFCREEN data to the website, a script is run daily to check JETSCREEN's S3 bucket for newly shared data. Any new files are then downloaded to our local database and can be accessed by users.

We note that the actual interface of the database is decided by each entity. We have adopted an open web interface; ALIGHT has an operational website; and NewJET is currently in the process of developing their web interface. Additionally, we note that near the end of the JETSCREEN data integration period, new security measures were implemented for protecting proprietary data at both JETSCREEN's and AJFTD's ends. As database integration with new programs continues, the process of handling proprietary data from different sources may need to be streamlined to account for variations in preferences between different program managers.

The completion of the JETSCREEN database integration process was a major first step in linking many other fuel databases across the globe. From this joint effort between FAA and JETSCREEN, we established a foundation for data sharing that can be used again with other programs, such as ALIGHT and NewJET. The ultimate goal of international database integration is to help monitor and evaluate fuels used in international airspace and provide an accurate picture of how fuel composition and usage trends are changing with time. As new fuels are integrated into the global supply chain, having a means to keep track of their properties will become critical. Such an interconnected database will ensure our ability to provide the most representative information needed for research on, and certification of, new SAFs. The impacts of database integration are outlined below in Figure 5.

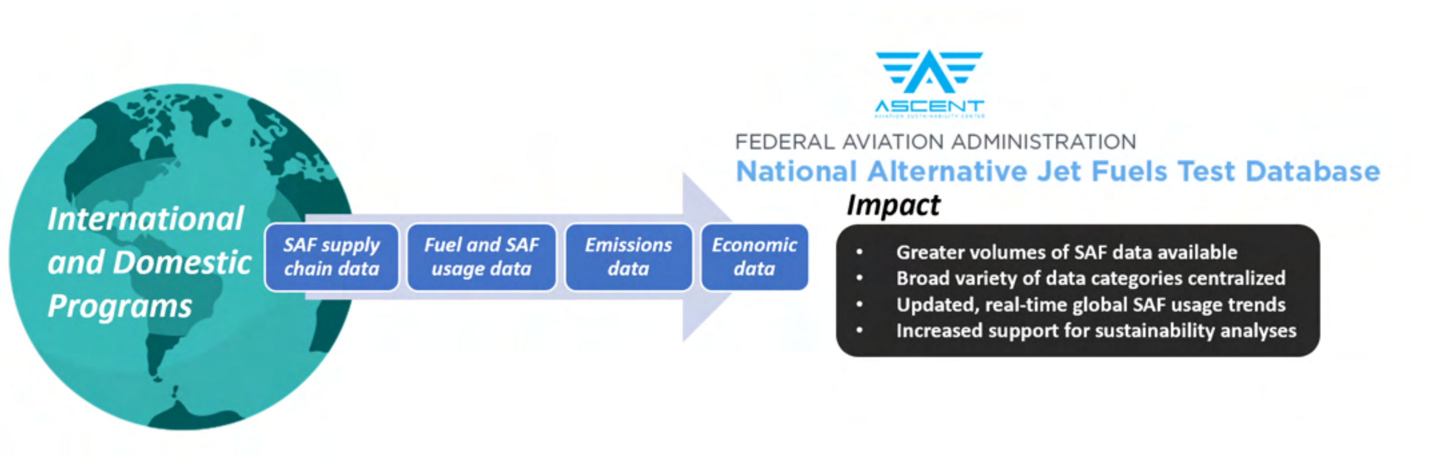


**Figure 5.** Plan to expand AJFTD's available data categories through data sharing with domestic airports and the current international SAF projects ALIGHT and NewJET.

This year, in addition to opening pathways for international database integration, we have resumed our efforts to intercept fuel samples and usage data from a selection of US airports (activities previously delayed because of the COVID-19 pandemic). National motivation to achieve a more sustainable and emission-free aviation sector grew in the Fall of 2021 with the launch of the White House Sustainable Aviation Fuel Grand Challenge. This initiative outlines various public and private grants, as well as policy changes, supporting the research, development, and implementation of SAFs. As this topic gains national visibility, the platform for further development of the AJFTD will also grow.

The goal of connecting AJFTD with domestic airports is to be able to assess the potential for zero-carbon operation via integration of SAFs. This aim can be accomplished by collecting actual fuel supply data from domestic airports and applying advanced analysis techniques to determine both the current status and future prospects for optimal integration of SAF. Real-time fuel property data can be used to analyze the current fuel supply infrastructure, thus enabling the determination of optimal SAF integration strategies for the future. This effort is expected to lay a foundation for achieving the sustainability targets set by both the United States (SAF Grand Challenge) and the global community (Renewable Energy Directive [RED II] of the European Union and the Paris agreement) by providing the data and analytic tools for sustainable pathways toward zero-emission airport operation through SAF integration.

## Milestone(s)

**3 months**

- Debugging and optimization of the data structure in the generation II database
- Completion of JETSCREEN database integration security features

**6 months**

- Communication initiated with Airlines for America (A4A) contacts for the airport data integration plan
- Completion of most of the debugging of the generation II database and further improvements to online analysis tools

**9 months**

- File upload script supporting the identification of fuel compositional data along with property data
- Discussions with domestic airport contacts to establish the next steps for real-time airport data interception.

**12 months**

- Modification of View, Compare, and other functions to include the presentation of fuel compositional data
- Planning and communication with ALIGHT and NewJET contacts

## Major Accomplishments

**Initiation of Domestic Airport, ALIGHT, and NewJET Connections**

Correspondence with the relevant program contacts for the aforementioned programs has been initiated and plans to proceed with international program integration are in place. These three connections will serve as new sources for acquiring greater amounts of fuel data as well as the expansion of the data categories available in the database to include categories such as usage and emissions data. Continued international collaboration will increase the long-term potential for support and data sharing with other international programs as they are formed. Connections with domestic airports will also support the long-term reliability of in-country data acquisition if consistent avenues for data sharing are built and maintained, thus providing the database with the most up-to-date and relevant information available.

**Modifications to the Generation II Online Database and Inclusion of Fuel Compositional Data on the Website**

Continuing improvements to the online database will include the availability of fuel compositional data viewing, representing a significant step in streamlining the use of the database as an analytic tool as well as a source of raw data files. Such modifications will not only improve the user experience, by enabling more meaningful inspection and interaction with fuel data, but also support further development of the machine-learning interactive capabilities we hope to include in future versions of the database.

## Publications

Blakey, S., Rauch, B., Oldani, A., & Lee, T. (2019). Advanced fuel property data platform: Overview and potential applications [Presentation]. 16th International Conference on Stability, Handling, and Use of Liquid Fuels Dresden, Germany.

## Outreach Efforts

The database has been made accessible through https://altjetfuels.illinois.edu/

## Awards

None

## Student Involvement

This project was primarily conducted by two graduate students, Ji Hun Oh and Alex Solecki.

## Plans for Next Period

In the future, detailed planning regarding the integration of the database with ALGIHT and NewJET will be undertaken. A more formal outline of the intents and goals of collaboration with these programs will guide the anticipated progress for the coming years. Interception of domestic airport data will be an increasingly important process to streamline as the national focus on SAFs increases in accordance with the 2021 Grand Challenge. Detailed planning in close coordination with our contacts at selected target airports will be the next step.

In addition, efforts to improve the functional and aesthetic features of the database are ongoing, as summarized below.

- Data presentation: Similar to the updates to data viewing and interaction on the online portal, we plan to implement a series of additional visual aids to support more meaningful user interaction with the data before downloading and analysis of raw files. One way to achieve this goal is by optimizing the Compare feature. We are currently working on presenting GCxGC data comparisons, with which users can quickly judge significant compositional differences between selected fuels. We are also redesigning the comparison graphs that are generated automatically during the use of the fuel comparison feature to make it more logical and account for situations in which the fuels being compared have different information types.

- Navigation: A site navigation guide, accessible on the main page, will outline the information available on the database website and indicate where it is located, to help new users locate relevant information more quickly. Because the prospective plans for the database include the addition of new data categories, which may reflect airport and emissions data, this feature is expected to become more important to supporting effective and efficient use of the database resources. To make finding specific fuels easier and more intuitive, we will also improve the fuel search function, by revising the tags for each fuel to include a broader range of search terms that could be used to locate a fuel of interest.

- User feedback: To implement more relevant improvements in the web portal's main interface and functionality, efforts are underway to gather more data on how the database is being used. We will be including a user feedback window on every page, through which users can submit feedback or questions directly to the website managers, thus improving the response time and the frequency of contact between our team and users. Surveys may be introduced to frequent users to allow us to better understand how the database is being used, and further improve the website's organization and functionality.

# Task 2 - Machine-Learning-Based Analysis

University of Illinois at Urbana-Champaign

## Objective(s)

The main objective of this task was to develop advanced methods based on machine-learning algorithms for analysis of the data in the alternative jet fuel database. The effort is inspired by the notion that the intricate relationships between the properties of fuels and their chemical signatures are critical but potentially too complex to be addressed with routine, classical, regression-based analysis. The effort has become increasingly important with the advent of new analytic techniques, such as GCxGC, that yield large amounts of data that are difficult to process with simple analytic algorithms. Machine learning can enable the most advanced analysis to be applied to our current data and is expected to become even more powerful as the size and diversity of data increase in the future. Previous studies have verified the efficacy of artificial neural networks in modeling the complex and obscure correlations between jet fuels' chemical compositions or structures and their physicochemical properties. The next step in this task is to exploit neural networks and deep-learning methods to address realistic challenges in SAF databases. The specific major goals are as follows:

- Identify real-world challenges in the analyses of jet fuel, specifically SAF, datasets
- Devise deep-learning-based strategies for addressing these challenges
- Perform composition–property modeling by using GCxGC and advanced machine-learning techniques

## Research Approach

### Data-driven Streamlining Based on Novelty Fuel Composition

A major challenge in SAF evaluation and qualification is the diversity and sparsity of SAF chemical compositions. If a SAF candidate exhibits compositions not explicitly characterized by D1665 conventional aviation fuels, its properties cannot be accurately correlated via existing data-driven approaches, such as quantitative structure–property relationships. This conclusion was corroborated by our preliminary studies attempting to map the comprehensive features of GCxGC to key properties in a diverse dataset: fuels with compositions not represented during the training (referred to as "novelty fuels" herein) showed high predictive errors. Therefore, there is a need to establish a means for preemptive detection and handling of such novelty fuels within the analyses pipeline. Unfortunately, most studies and literature have focused on conventional fuels, and only a few have addressed this topic.

Motivated by this challenge, we began to develop a deep-learning-based strategy to detect and quantify the degree of GCxGC novelty with respect to an existing pool of known (i.e., labeled) fuels. Using the labeled pool, our novelty-detection scheme utilizes an autoencoding neural network trained to encode more than 80 hydrocarbon groups to a few latent representations with the highest variance (or "energy") and optimal clustering characteristics, then decodes the output back to the original input. This network is implemented in Python by using PyTorch and is trained via stochastic gradient descent to update network parameters. After training, the novelty of unknown target fuels can be computed via decoded reconstruction error and outlying distances to the latent embeddings. Our preliminary results have indicated that our hybrid approach is more robust than traditional anomaly detection techniques such as *K*-nearest neighbors because it (a) accounts for all GCxGC information, (b) uses artificial neural networks that permit intricate nonlinear mappings, and (c) extracts the final score from more than one criterion.

The calculated novelty score dictates the next course of action for the target fuels. Low-scoring "normal" fuels can be evaluated with high fidelity with data-driven models built from the pool of labeled fuels. We have successfully demonstrated that key properties, such as distillation, flashpoint, viscosity, and net heat of combustion, can be accurately predicted. In contrast, the remaining novelty fuels require other means, e.g., labeling by conducting fuel tests and compiling the results. We further propose a labeling strategy based on prioritizing the fuels with the highest novelty scores. This corresponds to active learning, a subfield of machine learning in which the goal is to query the unlabeled data that are richer in information to promote optimal model enhancements with minimal query iterations. This process is especially relevant to the SAF context, in which data labeling is expensive or restricted, and thus streamlining is crucial. Our novelty-based labeling strategy has been found to outperform existing query methods in terms of accuracy and robustness.

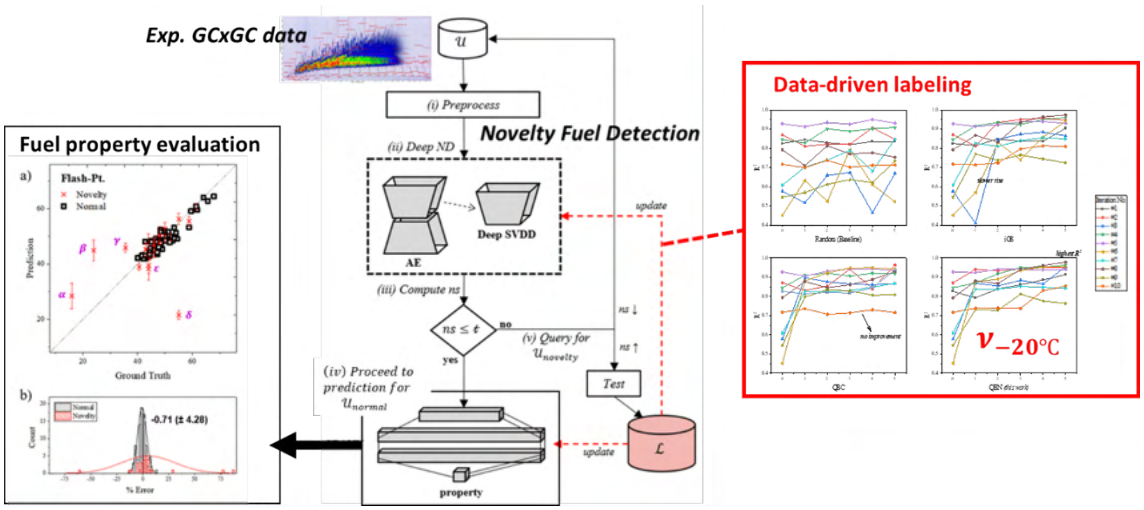


**Figure 6.** Framework for streamlining based on novelty fuel composition.

This framework is schematically shown in Figure 6. From top to bottom, the target fuel's GCxGC is screened on the pre-trained novelty fuel detection agent before being approved for downstream predictive analyses (left) or deferred/rejected to manual labeling (right). This framework was validated on a diverse and sparse dataset of 100 total fuels comprised of petroleum-based aviation fuels or SAF derived from various sources, methods, and blending ratios. The results indicated that our framework is highly effective.

**Missing Data Imputation Algorithms**

Another major challenge of real SAF databases is missing or noisy data. Such occurrences compromise the implementation of machine-learning algorithms. In these scenarios, the most frequently used approach is consolidating the dataset by discarding fuels containing missing elements (listwise deletion) or substituting missing data with averages or means computed from the available data. However, these approaches lead to either (a) reduced dataset sizes, which are known to affect the performance of machine-learning models, or (b) induce biases, because missing and non-missing data possess systemic differences; for example, achievable temperature measurements of certain properties (e.g., flashpoint or freezing point) are restricted by the apparatus, and thus the missingness mechanisms are not strictly random (Figure 7, top left diagram).

To combat these issues, we have devised a machine-learning strategy to impute missing data more accurately and reliably. We use a two-fold algorithm in which missing data are fist approximated via multiple imputations via chained equations (MICE), and this is followed by a refinement stage using a denoising autoencoder, as shown in the top right schematic in Figure 7. We have implemented this process on a hydroprocessed esters and fatty acids (HEFA) property dataset with data instances randomly omitted according to a defined missingness mechanism. The results have indicated that missing data can be accurately imputed for all properties under moderate missingness scenarios. For certain properties, imputations are feasible (although with limited accuracy), even for extreme circumstances where up to 60%–70% of the data are missing, and only 5% of the fuel instances are completely labeled. In future studies, we will investigate the impact of using these imputations for machine-learning applications, to demonstrate the advantages over using a reduced dataset.

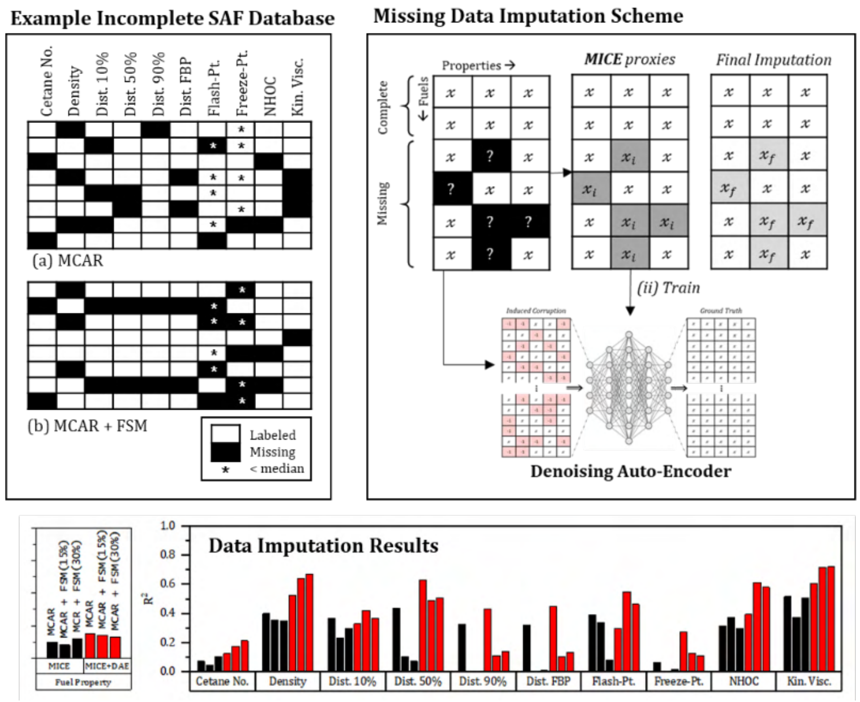


**Figure 7.** Missing fuel property data imputations.

## Milestone(s)

**3 months**

- Formalization of a machine-learning implementation plan

**6 months**

- Setting up scripts and algorithms for implementation of machine learning
- Organization of target data from the database for implementation of machine learning

**9 months**

- Implementation and optimization of machine-learning algorithms

**12 months**

- Finalization of all machine-learning tasks for publication

## Major Accomplishments

We have continued to implement advanced machine-learning algorithms in the analysis of data in our database. Specifically, we have devised deep-learning-based strategies to address relevant challenges in existing SAF databases, such as how to accommodate fuels exhibiting low predictive confidence due to novelty fuel chemical compositions and missing data. Although this work is in nascent stages, our preliminary results showed promise in tackling these challenges, especially when compared to traditional machine-learning methods. Current efforts are devoted to finalizing and refining the methods, and we plan to publish our findings, integrate our algorithms into the website database as an online analysis tool, and test the feasibility of using these methods to aid in the rapid assessment of SAF in real-world scenarios.

## Publications

The following two publications are in progress, with anticipated publication in AIAA SciTech:

Oh, J., Oldani, A., Lee, T., & Shafer, L. (2021).*Deep learning algorithms for assessing sustainable jet fuels from two-dimensional gas chromatography* [Manuscript in preparation].

Oh, J., Oldani, A., Shafer, L., & Lee, T. *Data-driven streamlining of sustainable aviation fuels via deep novelty GCxGC detection and query* [Manuscript in preparation].

## Outreach Efforts
The database has been made accessible through https://altjetfuels.illinois.edu/

## Awards
None

## Student Involvement
This project was conducted primarily by two graduate students, Ji Hun Oh and Alex Solecki.

## Plans for Next Period
We will expand our machine-learning capabilities and provide tangible performance metrics for various datasets in the database. Furthermore, we acknowledge the importance of obtaining high-fidelity assessments of chemically novel SAFs, and thus we plan to extend or rebuild our current novelty fuel detection framework. In particular, we plan to investigate more advanced deep learning techniques, such as propagation or synthesis of data via generative AI or physical structure-property relationship models, to improve model robustness to outlying cases.