Project 033 Alternative Fuels Test Database Library (Year IV)

University of Illinois Urbana-Champaign, University of Dayton Research Institute

*this report covers portion of University of Illinois

Project Lead Investigator

Tonghun Lee
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

- P.I.(s): Tonghun Lee, Associate Professor
- FAA Award Number: 13-C-AJFE-UI-022
- Period of Performance: 8/15/2017 to 8/14/2018
- Task(s):
    1. Conversion of Database into Non-Relational Format
    2. Inclusion of Additional Data: GC x GC, Metron, PQIS

Project Funding Level

Funding Level: $140K ($165K total, $25K contracted to UDRI)
Cost Share: Software license support from Reaction Design (ANSYS)

Investigation Team

- Anna Oldani (Graduate Student, University of Illinois at Urbana-Champaign): Compilation of fuel test data and development of database.
- Siddhartha Jahorie (Graduate Student, University of Illinois at Urbana-Champaign): Compilation of fuel test data.

Project Overview

This study seeks to create a comprehensive, foundational database of current and emerging alternative jet fuels by integrating relevant pre-existing jet fuel data into a common archive that can provide guidelines for design and certification of new jet fuels in our future, as well as aid federal work including fuel certification. The effort has focused on the integration and analysis of pre-existing jet fuel data from various government agencies and individual research groups with oversight from the Federal Aviation Administration (FAA). In this year, we are making a significant transition of data into a non-relational format, allowing for real time analysis of all the data and also enabling our database to link with a similar effort in Europe (JETSCREEN). We hope that the database will one day serve as 'the comprehensive and centralized knowledgebase' shared by the academic, government, and industrial communities in fuels research and policy, possibly facilitated on a cyber-based infrastructure. With ongoing prolific diversification of new jet fuels, this effort to integrate dispersed information is critical in providing the FAA with an overview of the latest developments and to support many other tangential fields of research in government, industry, and academia impacted by integration of new alternative jet fuels.

Task 1 – Conversion of Database into Non-Relational Format

University of Illinois at Urbana-Champaign

Objective(s)

The main objective of this project is to establish a foundational database current and emerging alternative jet fuels (AJFs) and fuel blends by compiling relevant data and storing this data in a centralized location for easy retrieval and analysis. The format for storing data should be consistent for all fuels and easily accessible by some common programming languages. The focus this year was coming up with a digital non-relational database structure what could be linked with other databases and readily analyzed. The intention of creating this database is to facilitate the large-scale analysis of AJF data and ease the creation of statistical correlations and surrogate models for AJFs. The goals of this project are:

- Select a suitable database for optimal storage and retrieval of jet fuel data
- Design a centralized data structure for compiling data
- Conversion of AFRL data to JSON format
- Preliminary analysis of AFRL fuel data using database resources.

Research Approach

Development Strategy of a Successful Fuels Test Database (Long Term Plan)

Database selection

The selection of a database service was a key aspect of this project. The first step involves a choice of relational (SQL) or non-relational (NoSQL) database. The major differences between these databases are summarized in Figure 1. A NoSQL database was considered more appropriate for two major reasons. First, documents and reports gathered for fuels had different formats and varying amounts of data. For example, data for one fuel may include viscosity, density and heat of combustion, while data for another fuel may only include density. Furthermore, multiple tests may be performed for a single fuel property. Data for one fuel may include five (5) measurements of viscosity, while data for another may only include one. Due to these considerations, a flexible schema was considered to be of primary importance when selecting a database service. SQL databases require that data fit into predefined tables, which could be problematic. Second, one of the goals of this initiative included collaboration with the European Agency, JETSCREEN (JET fuel SCREENing and optimization), which also studies AJFs. JETSCREEN's AJF data is already hosted by MongoDB, a NoSQL database service. Since MongoDB uses an unstructured schema, interfacing with a SQL database using a fixed schema would impose greater complexity compared to interfacing with a NoSQL database. The acid and base properties mentioned in Figure 1 describe consistency models used by the different databases. The BASE model used by NoSQL databases is a much weaker consistency model that the ACID model. However, a flexible schema cannot be used within the confines of the ACID consistency model. Therefore, NoSQL databases sacrifice consistency for flexibility. JOIN operations accommodate the merging of SQL tables to compare data across different tables. By exploiting the flexible schema of NoSQL databases, one can eliminate the need for JOIN operations.

Non-Relational (NoSQL)	*Relational (SQL)*
Highly scalable	**Less scalable**
Flexible schema - data can be inserted/altered anytime without issue	**Structured schema** - data has to fit into predefined tables
Does not support JOIN operations	Supports JOIN operations
BASE properties	ACID properties
Does not use SQL as query language	Mainly uses SQL as query language

Figure 1. Comparison of relational and non-relational databases

Once it was decided to use a NoSQL database, the next choice is to decide on a NoSQL service. While using MongoDB would be the most convenient choice to interface with JETSCREEN, DynomoDB was selected for several logistical reasons. Using university resources, DynamoDB was the easiest NoSQL database service to use. The University of Illinois has an agreement with Amazon AWS, which provides the DynamoDB service. As such, Amazon AWS representatives are present on campus to assist in the setup of the database and handle any potential problems. Because the IT services at the university do not have any staff experienced with NoSQL databases, implementation of any other NoSQL database would have to be done in-house or arranged with an external organization. MongoDB is notoriously difficult to setup and therefore it was decided to avoid doing this in-house. Using an external organization would also involve additional challenges as such an organization would need the right clearances to work with government data. This was not an issue with DynamoDB due to their partnership with the University of Illinois. Since DynamoDB and MongoDB both store data in the JSON (JavaScript Object Notation) format, interfacing between the two services is not expected to be difficult.

Data Schema

Both MongoDB and DynamoDB stores data using the JSON (JavaScript Object Notation) format. Therefore, the next step involved converting data to the JSON format. The JSON format stores data using key, value pairs. Values can also be keys pointing to other values resulting is a complex schema with nested keys. This is often the most appropriate way to store data. In order to access data, these specific keys are required. Since JETSREEN already had a schema in place, our schema was created by modifying JETSCREEN's schema to account for additional data that we collected. Since the same keys are used, integrating and sharing data would be fairly straightforward.

Conversion of AFRL data to JSON format

Data received from AFRL mostly included fuel lab reports from the Air Force Petroleum Office (AFPET). These reports were in pdf form and had to be converted to a format that is easily accessible by code. The Able2Extract software was used to convert AFPET lab reports to suitable excel tables. Since many reports, particularly old reports, were scanned pdfs, these tables were also manually checked for errors and the ensure consistency in formatting. After converting AFPET reports to excel tables, scripts were written in python 3 to extract the data from these excel tables and create JSON files. The scripts mapped key phrases in the names of fuel tests to specific keys defined in the data schema. The process of converting AFRL data to a JSON format is illustrated in Figure 2. It should be noted that a single JSON file was created for each POSF number. Because there were multiple AFPET lab reports for a single POSF number, these were combined into a single JSON file.

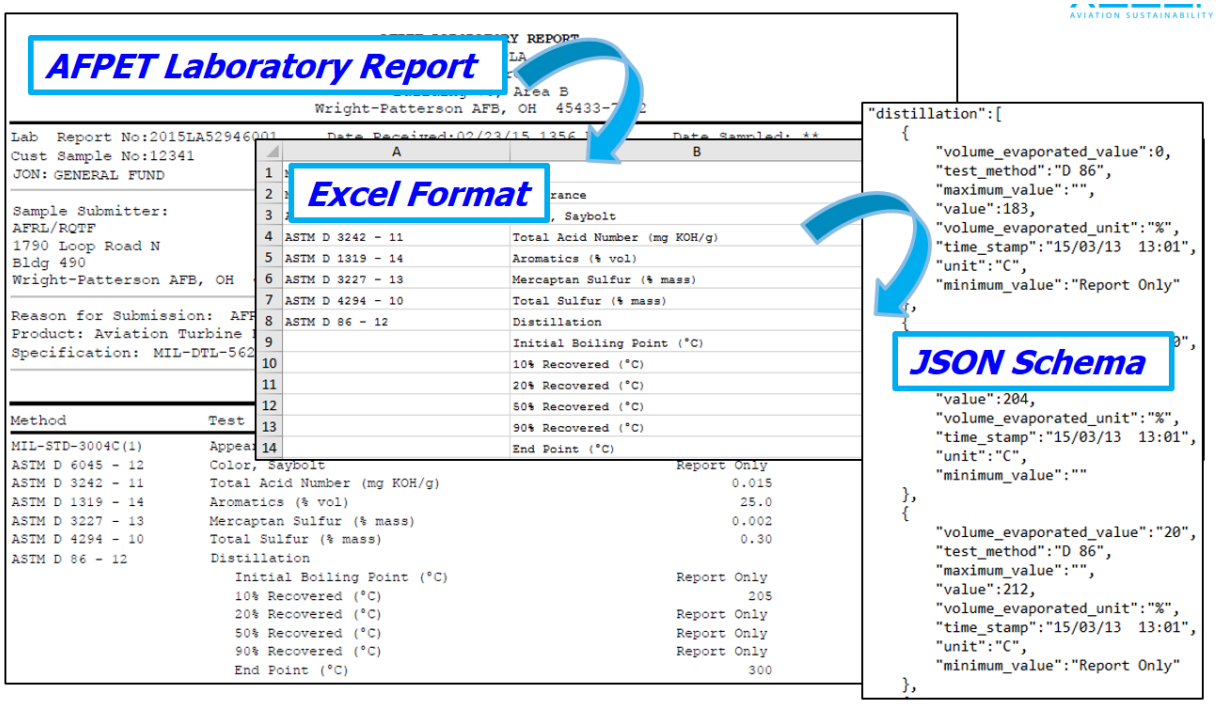

Figure 2. Conversion of AFRL Laboratory Reports from pdf files to excel tables to serialized JSON files

Preliminary analysis of data

Once data for AFRL fuels were stored in JSON files, scripts could be easily created in most programming languages to access this data and perform numerical analysis. Uploading the data to DynamoDB allows the data to be shared by other collaborators. In the future, we hope to use the capabilities of DynamoDB to interface with the AJFTD site to allow users to easily sort, filter and perform simple analysis on fuels.

Figures 3 and 4 demonstrate some of the analysis that can be performed using the database. Figure 3 shows the distillation curves for Category A and C fuels and the average distillation profile of all AFRL fuels. The unusual boiling curve of C-1, the asymmetric boiling curve of C-2 and the flat boiling curve of C-5 can be easily observed from this plot. The effect of farnesane on the boiling range of a fuel can be observed by comparing the distillation curve for C-3, which is composed of 64% A-3 and 36% farnesane, and A-3. The distillation temperatures for C-3 is higher than A-3 up to about 90% recovered, and from 90%-100% recovered the distillation temperatures are quite similar. This occurs since farnesane is a C15 hydrocarbon and lower carbon number components generally evaporate first. The average distillation distribution is very close to that of A-2, further demonstrating the suitability of A-2 as a nominal fuel.

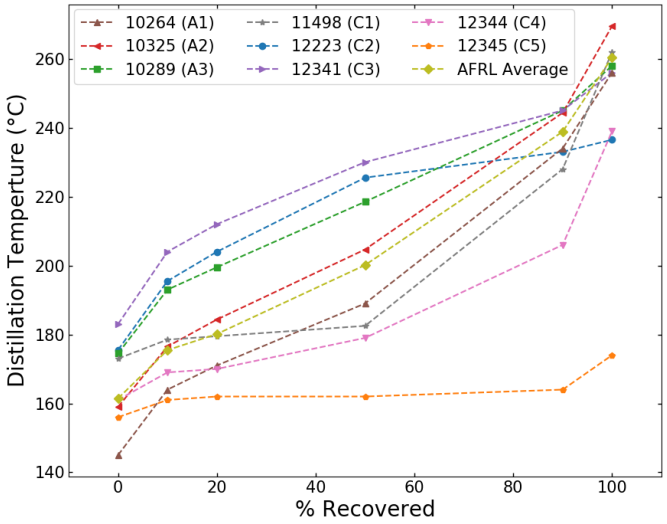

Figure 3. Distillation temperature for AFRL fuels

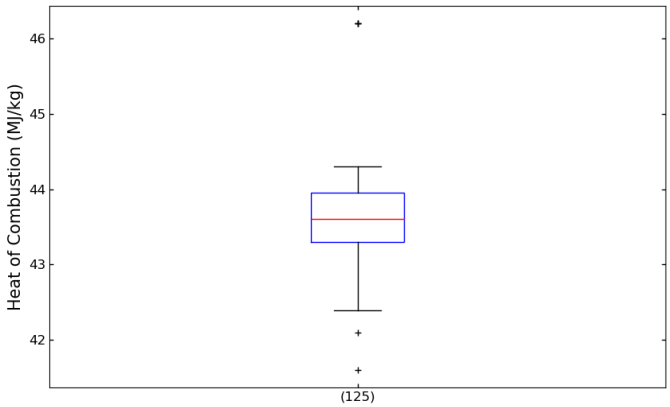

Figure 4. Boxplot showing range of heat of combustion values for AFRL fuels

Figure 4 shows a boxplot of the heat of combustion values for AFRL fuels. The range of values can be easily observed in such a plot. Outliers and erroneous data can also be easily identified. In the future, by integrating DynamoDB with the AJFTD site, it may be possible to plot data for new fuels against a range of existing data to determine if the data is within the expected range, or if it may be a possible outlier.

The analysis shown here is just a demonstration of the flexibility offered by turning the entire database into a non-relational format. The analysis enabled by this database is now only limited on the programming and can be easily extended to new data which is added to the database. The current version was carried out with customized PYTHON codes based on the output for which we were aiming. As stated above, it is anticipated that in year V, we will be integrating the functionality to the website so that other users can log in and conduct basic functions using the data. More importantly, we anticipate that this type of structure will allow the current database to be linked with other databases including the JETSCREEN database in Europe. In order to accomplish this, the format in the JSON schema were based on the existing architecture of the JETSCREEN database with minor modifications. These minor modifications were mainly required to address US specific data terms and add other testing information such as the GC x GC (will be addressed below).

Task 2 – Inclusion of Addition Data: GC x GC, Metron, PQIS

University of Illinois at Urbana-Champaign

Objective(s)

The main objective of this project is to augment the existing data by adding GCxGC data and collecting data from additional sources, with a focus on fuels in use. As more data is added to the database, more meaningful analysis can be performed. The major goals of this project are as follows:

- Collect GCxGC data and integrate into JSON schema
- Convert Metron Aviation data to JSON format
- Convert PQIS data to JSON format
- Preliminary analysis of GCxGC, Metron Aviation and PQIS data

Research Approach

Development Strategy of a Successful Fuels Test Database (Long Term Plan)

Integration of GCxGC data to JSON Schema

Gas chromatography is a technique for separating complex mixtures based on differences in boiling point/vapor pressure and polarity. Two-dimensional gas chromatography (GCxGC) subjects a sample to two dimensions of separation. The first separation is based on volatility as seen in GCMS, and the second separation is a very fast separation based on polarity. This allows identification of different hydrocarbon types and carbon numbers.

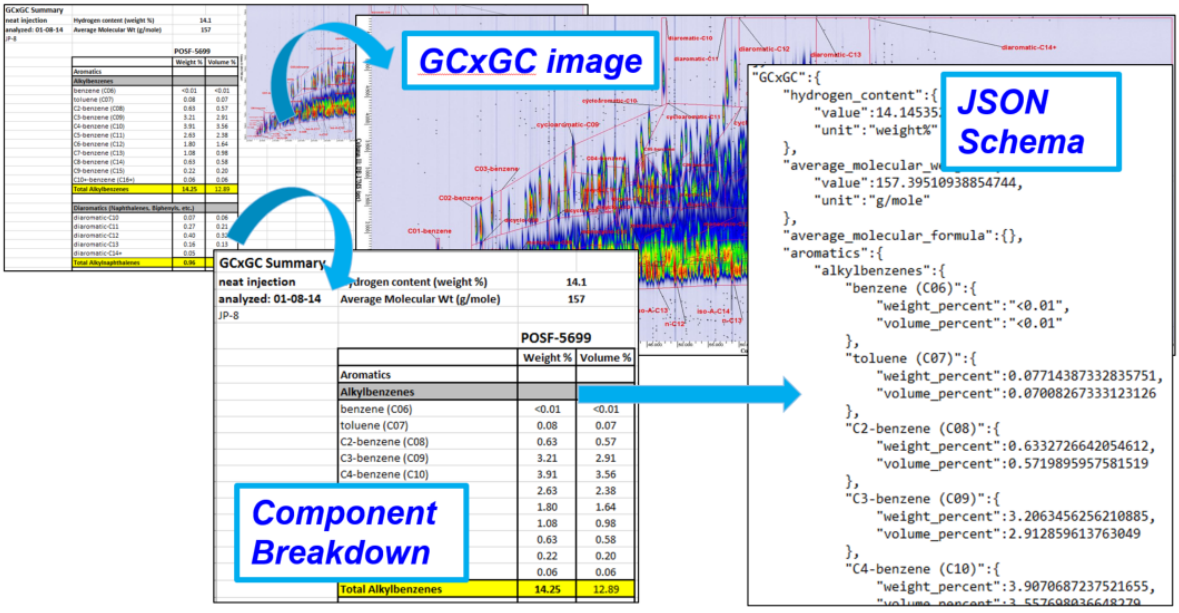

Figure 5. Conversion of GCxGC data to JSON format

The GCxGC technique provides important information for understanding fuel properties. The carbon number provides information on fuel volatility, which also affects flash point and freeze point. For example, lighter fuel fractions (with lower carbon numbers) are more reactive and lower the flash point of the fuel. Heavier fuel fractions (with higher carbon numbers) raise the freezing point of the fuel. Due to flash point and freezing point constraints, jet fuels typically do not contain significant amounts of hydrocarbons below C8 and above C17. Hydrocarbon type also imparts particular properties to fuels. For example, normal paraffins are the most reactive hydrocarbon group. Isoparaffins have the same chemical formula as

normal paraffins but are less reactive and have much lower freezing points. Cycloparaffins have lower freezing points and higher densities than both normal and isoparaffins of the same carbon number. Single-ring aromatics (alkylbenzene) are important for elastomer (O-ring) swell and have high energy density per unit volume, while multi-ring aromatics contribute to soot formation. It is worth noting that GCxGC is weak in differentiating the level of branching in iso-paraffins and side chains in aromatics and cycloparaffins. The level of branching in these molecules can impact the cetane number and hence ignition delay of fuels

GCxGC data was collected for several AFRL fuels and converted to the JSON format as shown in Figure 5. The GCxGC schema includes data on average molecular weight, hydrogen content and four major hydrocarbon types (n-alkanes, iso-alkanes, cycloalkanes and aromatics), including a carbon number breakdown.

Conversion of Metron Aviation and PQIS Data to JSON format

Additional fuel data was received from Metron Aviation and Petroleum Quality Information Systems (PQIS). Metron Aviation collected 14,823 fuel samples from 11 airports across the United States. PQIS, which is managed by the Defense Logistics Agency (DLA) has provided information on 9,201 conventional and alternative fuels used by DLA customers. These data are particularly interesting to study since they contain data of fuels in use. One downside of the Metron Aviation data is that very little identifying information if available for fuels. Therefore, we do not know what type of fuel is used or if the fuel is blended with an alternative fuel. On the other hand, PQIS fuel data can be categorized into alternative fuels, Air Force kerosene, Air Force gas and Air Force diesel. Furthermore, the alternative fuels can be categorized as hydroprocessed reformed jet (HRJ), alcohol-to-jet (ATJ), direct sugar to hydrocarbon (DSHC) or hydroprocessed depolymerized cellulosic diesel (HDCD) fuels. Data from both of these sources were received in formats that are easily accessed by code. There scripts were written to directly convert the data into JSON files.

Preliminary Analysis of GCxGC, Metron and PQIS data

This section again demonstrates the utility of the database by showing some of the analyses which can be performed with relative ease. Distillation data collected from Metron Aviation, averaged over airport locations, is shown in Figure 6. While the reporting units for distillation temperatures are supposed to be in degrees Fahrenheit, it appears that the data collected at Portland airport (PDX) were reported in degrees Celsius. The distillation temperatures for PDX, seen in Figure 6(a) appears to be too low for a kerosene range fuel. Figure 6(b) was produced by assuming that these values were in Celsius and converting them to Fahrenheit. The graphs with the converted values fit much better with the traditional boiling ranges for kerosene range fuels.

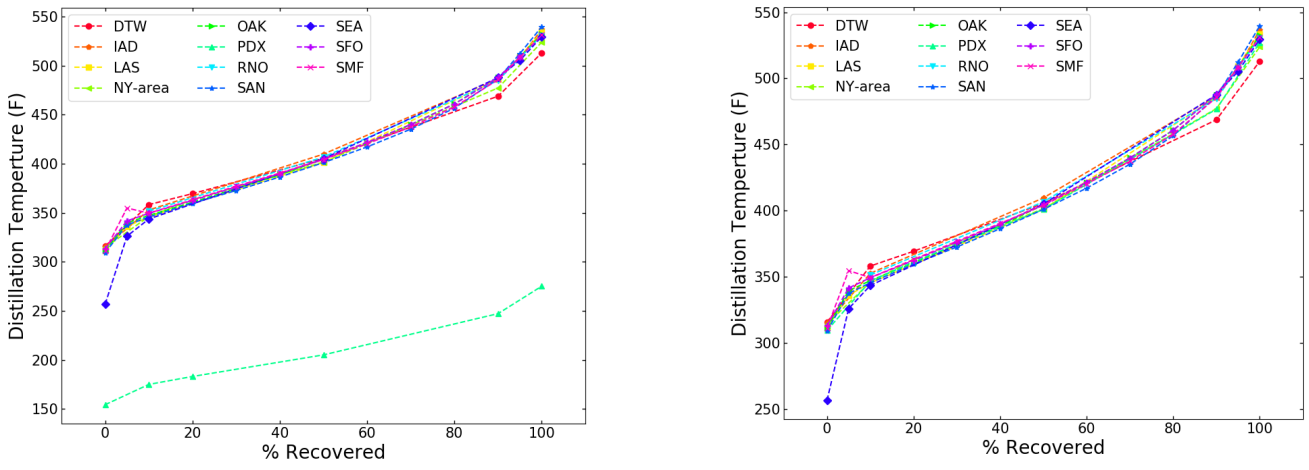

Figure 6. Metron Aviation Distillation Curves

Figure 7(a) shows the composition of category A and category C fuels that were reconstructed from the GCxGC data. The red bar represent category A fuels and the green bars represent category C fuels. Only the four major hydrocarbon groups, n-paraffins, iso-paraffins, cycloparaffins and aromatics, are shown, because the amounts of other hydrocarbon types are negligible. While an even spread of hydrocarbons can be observed for category A (conventional) fuels, the predominance of iso-paraffins is observed in category C fuels. Figure 7(b) shows the average GCxGC composition of six different types of neat alternative fuels and compares this with conventional fuels and fuel blends. Again, the predominance of iso-paraffins is observed in most alternative fuels with the exception of HDO-SK (Hydro DeOxygenated Synthetic Kerosene)/HDO-SAK (Hydro DeOxygenated Synthetic Aromatic Kerosene) and HDCJ (Hydroprocessed Depolymerized Cellulosic Jet) fuels. HDO-SAK fuels are intentionally produced to have a high aromatic content, while the cellulosic feedstock of HDCJ fuel usually produce fuels that predominantly composed of cycloparaffins and aromatics. It is interesting to note that the alternative fuels which have low iso-paraffins content are the ones which have not been approved for use as drop-in fuels when blended with conventional fuels, while the alternative fuels with high iso-paraffin content have all been approved.

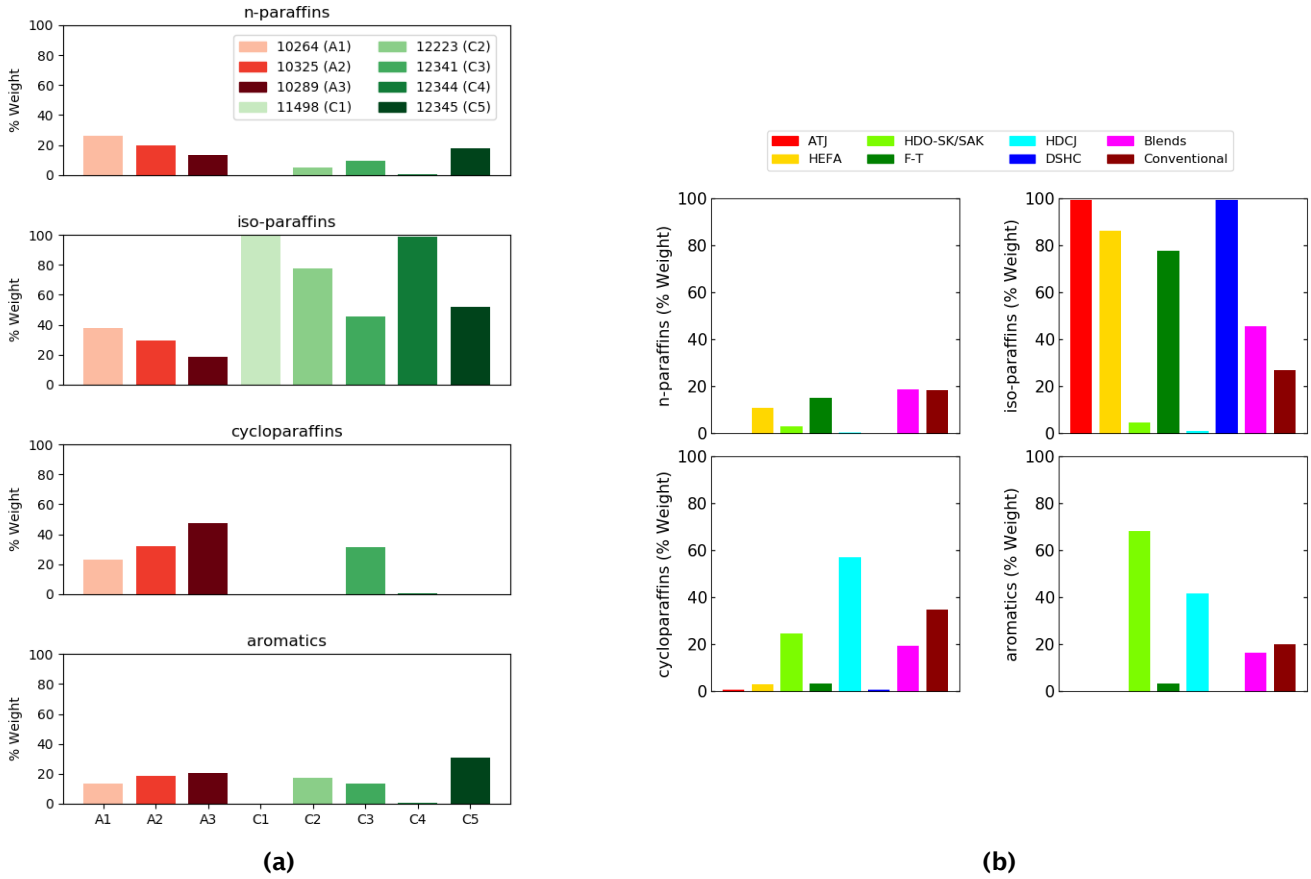

Figure 7. GCxGC graphs showing the hydrocarbon content of various fuels

In an attempt to develop a procedure for shortlisting viable candidates for drop-in fuels, Won et al. proposed the formulation of a surrogate model for predicting global combustion properties. The model would be based on a multivariate linear regression (MLR) analysis of certain combustion property targets (CPTs). Some of the suggested CPTs were derived cetane number (DCN), molecular weight (MW), hydrogen to carbon ratio (H/C ratio) and smoke point (SP). In order to perform an MLR analysis, it must be verified that the selected CPTs are independent. Figure 8 shows a correlational analysis of these CPTs using both the data provided by this paper and data from the database. The number in parenthesis are the Pearson correlation coefficients of the Won et al. data, while the number above (without parenthesis) are the Pearson correlation

coefficients for data from the database. The high correlation coefficient values for smoke point and H/C ratio suggests that these CPTs cannot be considered independent. Therefore, any meaningful MLR analysis will have to be performed with only one of these.

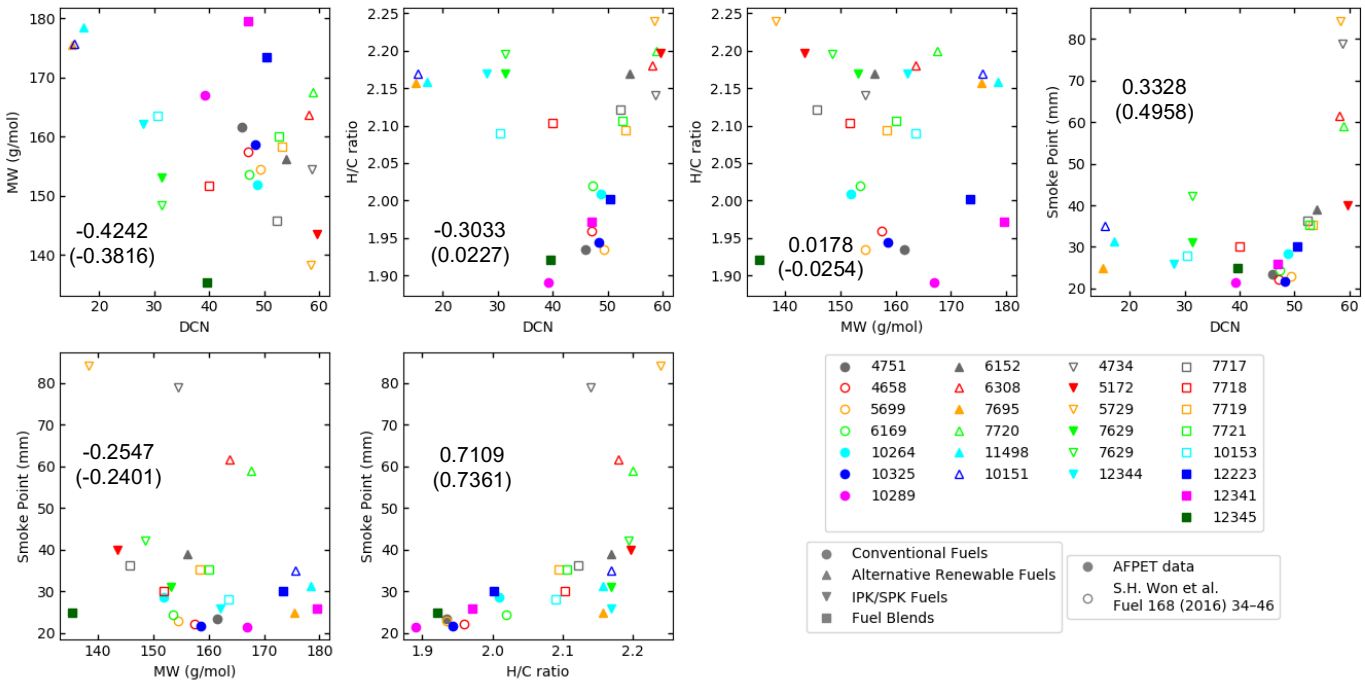

Figure 8. MLR Analysis

Milestone(s)

Proposed (3 Month): At the 3 month mark, we should have decided on a standard non-relational data format in collaboration with the JETSCREEN program in Europe. Plans should be underway to convert database structure into a non-relational format.
Achieved: A JSON format for integrating all data was selected (in discussion with JETSCREEN) and the conversion of all of the data in the database was initiated. Special routines were programmed using an assortment of codes (mostly in PYTHON) to facilitate the transition.

Proposed (6 Month): At the 6 month mark, data conversion to the JSON format should be well underway. We should also have started the integration process of additional GCxGC data from AFRL as well as airport data from Metron.
Achieved: All the data was gathered from AFRL and Metron and specialized routines were set up for conversion of all this data to JSON format. We also decided on a DynamoDB framework supported by Amazon to host the non-relational data.

Proposed (9 Month): At the 9 month mark, conversion of all data to JSON format should be nearing completion and preliminary analysis of data using the JSON format should be conducted.
Achieved: Nearly all of the data (>85%) was converted to the JSON format at this point and preliminary analysis of data was initiated using PYTHON codes. Correction of errant data from METRON and AFRL was also completed.

Proposed (12 Month): At the 12 month mark, all data to date should be converted to JSON format. Discussions with the School's IT office should be underway for integration of JSON format data into the DynamoDB framework and integration of this into the database website.
Achieved: All data conversion to was JSON completed. Discussions with school IT office of integrating the database analysis into the website were initiated.

Major Accomplishments

Conversion of Data to JSON format (non-relational digital format)
A total of 24,386 fuel records have been converted to the JSON format, which can easily be accessed by most programming language. GCxGC data have been obtained for 87 AFRL fuels and added to the JSON schema. This is the first step in the establishing a foundational fuel database that would enable researchers to more easily analyze surrogate models and statistical correlations for alternative jet fuels. This format will also allow us to link the database with others including JETSCREEN.

Integration of New Data
We have included a number of new data into the database in year IV including the GCxGC data from AFRL, actual airport data from Metron Aviation, and the PQIS database. All the data integrated into the database has been converted into JSON format for full analysis.

Preliminary Analysis
Analysis of fuel data have demonstrated the efficacy of the database in identifying erroneous data and outliers. Insightful compositions graphs and distillations plots highlight the most salient features of alternative and conventional fuels were also produced. As more fuel data is obtained and added to the database, more meaningful and statistically significant analysis would be possible.

Publications

(In Progress) Oldani, Anna. "Alternative Jet Fuel Variation and Certification Considerations." 2018.

Outreach Efforts

None

Awards

Anna Oldani (Graduate Student): DOT Student of the Year Award
Anna Oldani (Graduate Student): Society of Women in Engineering (SWE) Award for Research Excellence

Student Involvement

Two graduate students (listed above) have participated in this project on a rotational basis to address various aspects of the project. They have surveyed the data, interacted with the data sources and created strategies to integrate the data into the database. They developed the web-based portal for the actual implementation of the web interface. They have also conducted a statistical analysis of the available data to evaluate property variance. They continue efforts to update the database with relevant alternative jet fuel test data as it is made available.

Plans for Next Period: Start of Analysis and Integration of Database

Year V for the database project will see the full extended evaluation activities of the data using the new non-relational structure. We anticipate that a publication or report will result from the full analysis.

- Evaluation and analysis of all GC x GC, Metron, PQIS data
- Full integration of the JSON format into the DynamoDB infrastructure
- Integration of the database structure (DynamoDB) to a web interface on aljetfuels.illinois.edu for public access.
- Inclusion of NJFCP data: continued integration of the NJFCP data to the website for fuel certification activities
- Integration of efforts with European JETSCREEN program to collaborate fuel screening and property evaluation under NJFCP program (Meeting Scheduled in December of 2018)