400 Hz Ground Power Utilization
A Case Study on San Francisco International Airport

Software Instruction Manual

July 15th, 2020

Pietro Achatz Antonelli
Chanan Walia
Jasenka Rakas

Table of Contents

Introduction

400Hz Ground Power can be provided to aircraft at the gate as an alternative to the expensive and polluting Auxiliary Power Units (APU), to supply electrical power. Although there is an incentive to reduce fuel consumption costs for airlines, there are also operational challenges that sometimes prevent the successful switch towards a better power source.

Although airports have a policy in place to ensure airlines avoid the overuse of APU's at the gate, they do not have a systematic way to monitor and enforce it. Human inspection is not sufficient to precisely and comprehensively evaluate the performance of thousands of gate operations every day. Airlines will not share their internal data, especially if it can be used to enforce policy against them. This gap in information needs to be closed to make sure ground power is used to its highest potential.

A software program that can evaluate 400Hz ground power use and estimate APU use is the proposed solution toward providing the necessary oversight. This idea was developed into a software prototype by UC Berkeley students in 2019 and 2020 with the support of the Airport Cooperative Research Program (ACRP) and San Francisco International Airport (SFO).

This manual is a guide toward understanding and implementing the monitoring system for ground power systems at an airport. It outlines the conceptual method, data collection process and instructions of a prototype analysis to be retrospectively performed on airport data.

The analysis of 400Hz ground power was conducted with Python code running on Jupyter Notebooks, which should be provided with this manual. The software was designed to receive a variable input, provided that it is in the correct format, and output the relevant results. The Python code is thoroughly documented so that whoever is handling it can understand what is happening, even without these instructions.

The sample data initially provided was provided by San Francisco International Airport for the time period between April 1st 2019 to January 31st 2020. This data and its results are sensitive and should not be publicly shared. If the data collection systems at San Francisco International Airport were not altered, the instructions can be followed closely without editing anything. If this analysis is being applied on different systems or a different airport, minor edits will likely be necessary.

For any further inquiries or explanations into the ground power analysis tool, please contact the creators Pietro Achatz Antonelli (p.achatzantonelli@berkeley.edu) and Chanan Walia (chanan@berkeley.edu).

Method

The centerpieces of the data analysis consist in data cleaning, connecting separate databases, forming a prediction model, and evaluating performance, as shown in the conceptual flow chart in Figure 1.

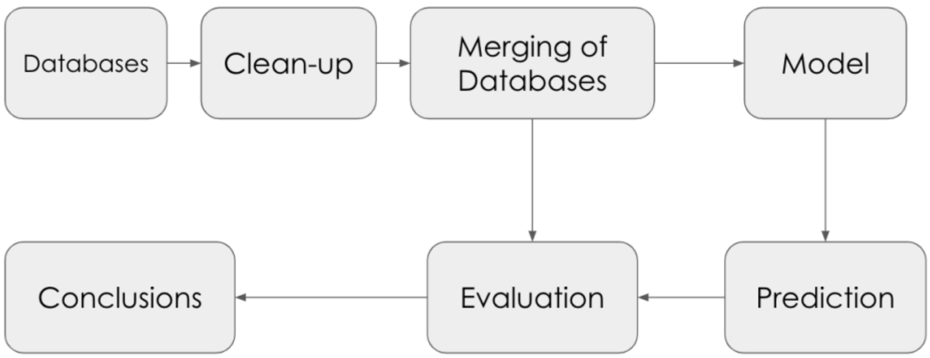

Figure 1: *Conceptual flow chart of the analysis*

The primary data consists in the gate metering database, which records energy consumption from the 400Hz ground power cables at SFO, and the aircraft operation data, which provides information on the characteristics of the aircraft and their movement in the apron area. This raw data has to be extensively processed to avoid incompatible formats, missing data or erroneous measurements.

By removing any data that is suspected as faulty, the scope of the analysis is reduced, as shown in figure 2. Therefore, the results generated by this tool are not a necessarily representative sample of the reality at the analysed gates and time frame.

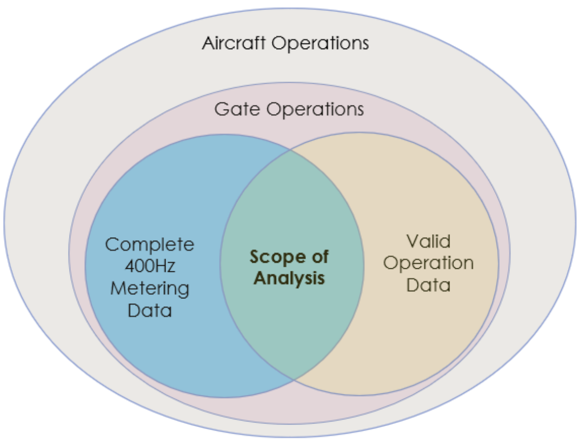

Figure 2: *Scope of analysis, limited by valid data*

Both the databases index their data based on their event time and gate at which the event occurs. By the mapping of gate names to their respective energy meters, aircraft operations at the gate and their energy use can be merged. In this manner, if power is being consumed, it can be associated with a specific aircraft reference number, carrier name, and aircraft carrier.

With this richness in attributes to the power consumption, a model is constructed to predict power consumption. In the current code, a simple linear regression with attributes such as aircraft subtype is used, but the model can be customized by implementing methods like tree modelling and by using more attributes. Figure 3 conceptually displays how the prediction model is formed.

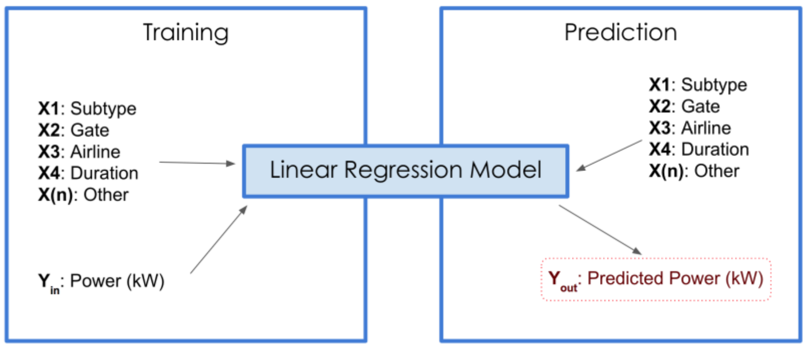

Figure 3: *Prediction model made with linear regression*

Only the data points with power readings above a certain threshold power (2kW) are used to form the model because the model should only predict the power demand of aircraft that are actually drawing power, and not when the gate is idle. Figure 4 represents the selection of the data accepted for the model training.

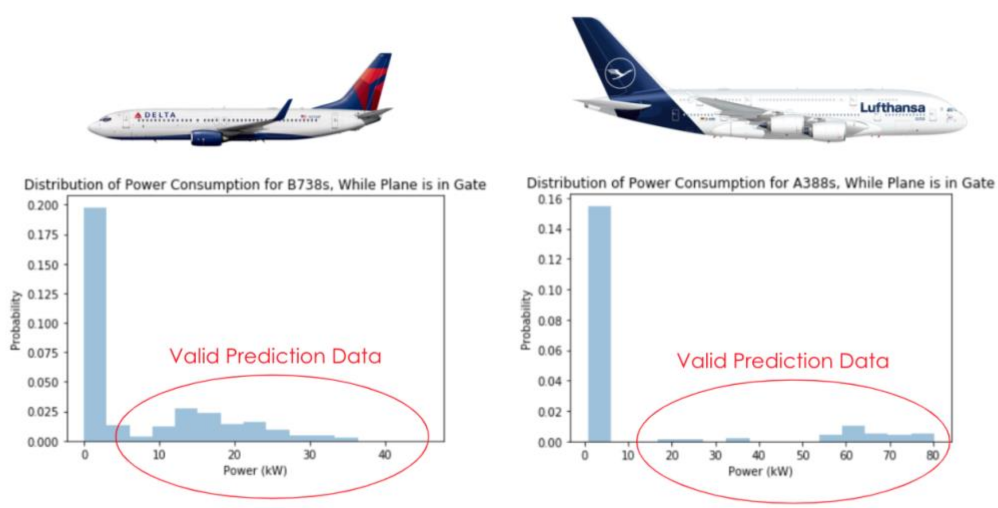

Figure 4: *Selection of valid prediction data, exemplified by different aircraft subtypes*

This prediction model is then used on both the data it was trained with and new data.

Overfitting can be a significant problem that is currently ignored, and with larger data sets, cross validation is recommended. In such a manner, for every second an aircraft is at the gate, it is possible to read the actual power being drawn and compare it to the expected power demand of the aircraft. This is exemplified for a single operation in figure 5 below.

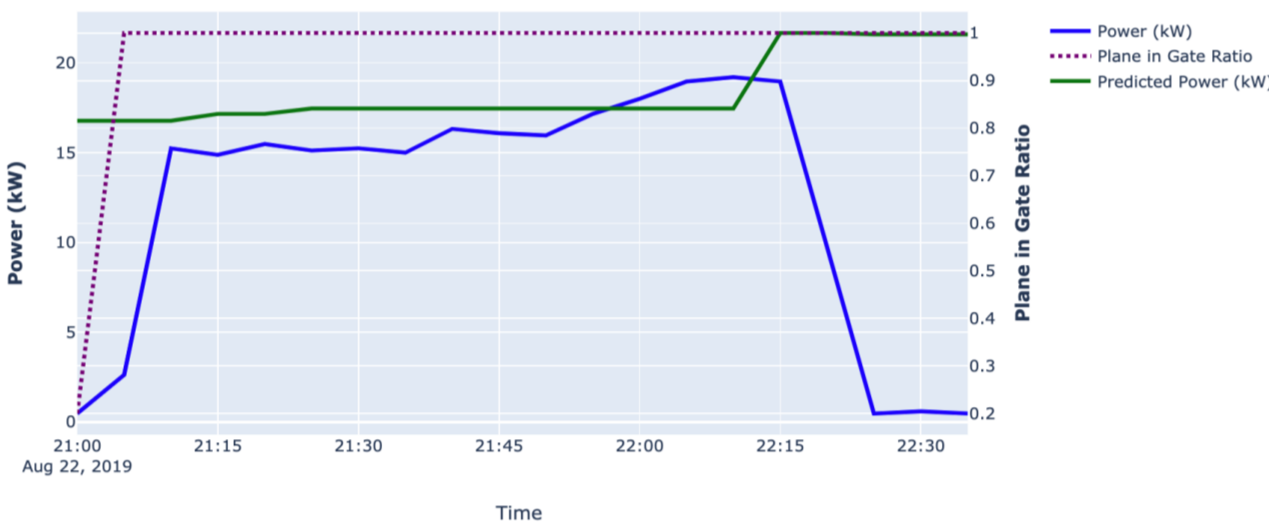

Figure 5: *Time graph of predicted and actual power for a sample B738 gate operation*

Even if the data granularity is insufficiently detailed (5 minutes), it is possible to make estimates on the exact time the 400Hz ground power started to be used by observing the ratio between actual power and predicted power. For example, if the expected power consumption during a 5-minute interval was 20 kW but instead was measured as 8 kW, one could estimate the power that was actually being drawn for 40% of the time, or 2 minutes.

Since SFO has ground power cables installed at all gates, and alternative ground power unit use is rare, it is reasonable to assume that whenever ground power is not being used, the aircraft must be using its APU. The assumption is not valid when an aircraft is entirely shut down, but that does not occur with short turnaround time, so longer turn-arounds are excluded from the analysis.

Based on that assumption, it is possible to estimate the APU use time. By combining it with the prediction on power demand, it is also possible to estimate the energy use from the APU. Additionally, it is possible to estimate how many times the APU was initiated, which is rarely more than once per operation.

With emission rates specific to different aircraft categories, the tailpipe emissions and the wasteful fuel consumption of the APU use at the gate are estimated. These estimates include both the emissions from the "normal" mode of operation of the APU and the "start" mode of the APU, which refers to the time the APU needs to power up (3 minutes). This estimate does not capture the full cost-benefit assessment or life-cycle assessment of the gate operation, but it does provide a valuable insight into the most significant emissions that occur at the gate.

To assess the performance of each operation with regards to ground power use, two metrics were created to evaluate every valid gate operation:

- The Operation Success Ratio, which refers to the ratio of time the ground power was used versus the total time the aircraft was at the gate
- Total APU time, which simply refers to the number of minutes, both in normal and start mode, that the turbine was operating.

All the information from the gate operations that were analyzed is then aggregated into a single database. This database can be filtered on attributes such as gate, aircraft subtype and airline, to provide a valuable big picture on emissions and performance of a specific set within the data.

The final results are indicative of the sample that is analyzed and not the entire population. It would be incorrect to draw conclusions on a wider population without sufficient representation. However, results in the sample should be used as indicators of potentially larger patterns in the data.

In future iterations of this analysis, the possible improvements are:

- Improving the mapping of gate meter to gate name by filling in missing links
- Increasing the granularity of the data
- Adding more features to construct a more comprehensive prediction model
- Improving the detail of the emission rates with regards to different APU engines
- Adding more metrics to assess performance of operations

Instructions

These instructions indicate the procedure to perform the data analysis retrospectively, from the data collection on-site to running the Python code on a remote computer.

The software was split into 6 separate parts, named System 1 through 6. Each system can be run independently and has its own inputs and outputs. The structure of how the systems interact with data is best summarized through the flowchart in figure 6 below.

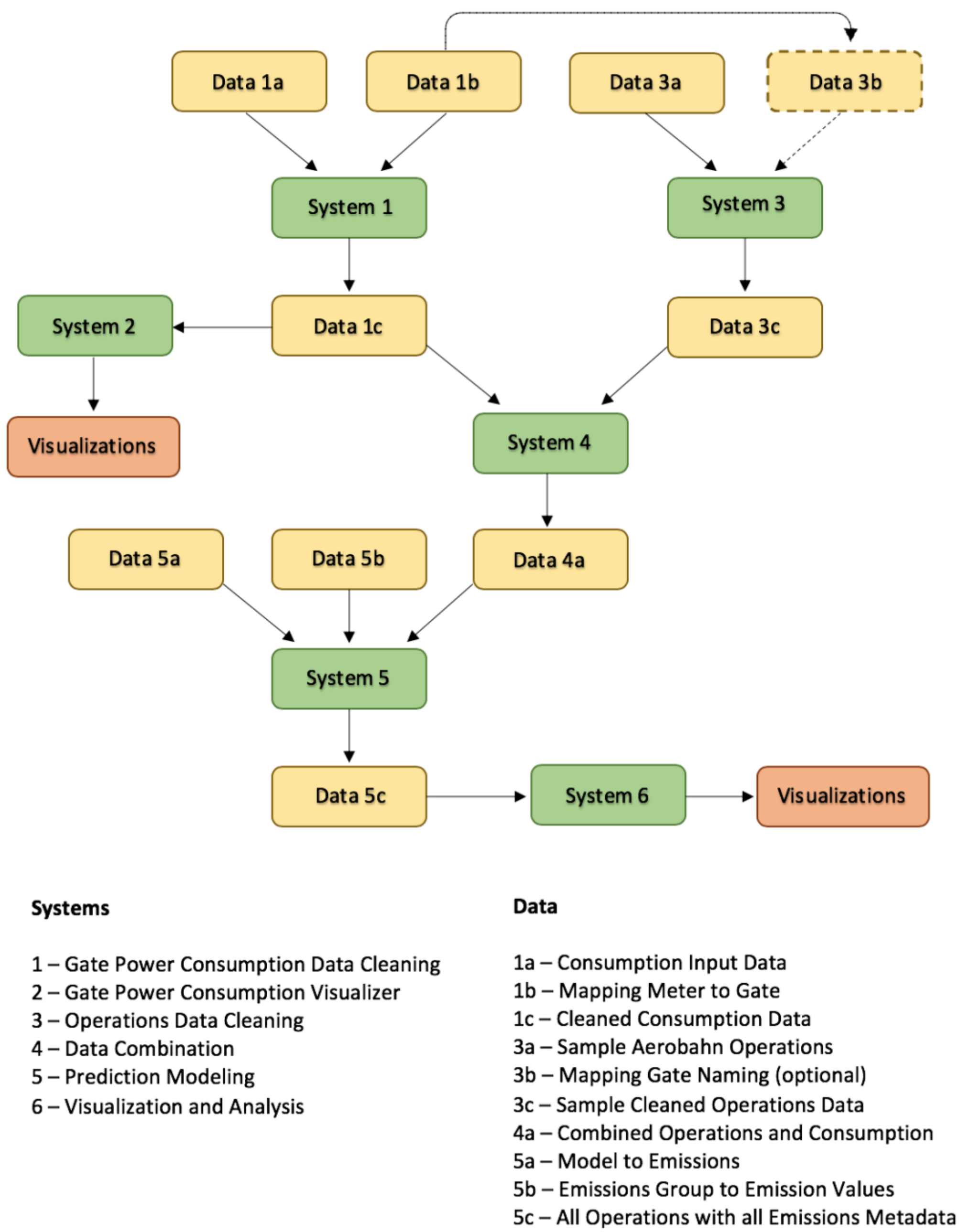

Figure 6: *Organizational chart for running the tool*

Data Collection and Formatting

First, the temporal scope of the analysis has to be defined. Select a continuous interval of time, ideally between 3 months and 1 year. It is recommended that the time interval does not include times in which the system was modified, such as significant changes in naming, infrastructure or data collection systems. Prolonged extraordinary circumstances in aviation such as the COVID-19 pandemic should be taken into consideration when defining the scope of analysis. Take note of the exact date and time of the beginning and end of the interval.

Then, the inputs of the analysis must be obtained from the airport for the selected time interval. In the absence of an API, the data must be collected in person at the airport with a computer directly connected to the databases. In the past, a hard drive and Google Drive have both been used without issues to store data from SFO. The five critical datasets are described:

Gate Mapping

The first database that is needed is a simple map between gate name and the respective 400Hz Ground Power meter. This information could be found in airport documentation or by inquiring with electrical engineers responsible for the 400Hz systems. Try to include as many gate-meter pairs as possible, but do not worry if some are missing. The file can be made with Excel to look exactly like Figure 7 below. The correct naming of the gate and meter are critical towards the functioning of the code. Make sure to have the same column names and to store the final file in .csv format named 1b_mapping.csv .

gate	meter	new_gate
Gate_D50B	ALASKA_AIRLINES.	Gate_D2
Gate_D51A	ALASKA_AIRLINES.	Gate_D3
Gate_D51B	ALASKA_AIRLINES.	Gate_D4
Gate_D52	ALASKA_AIRLINES.	Gate_D5
Gate_D53	ALASKA_AIRLINES.	Gate_D6
Gate_D54A	ALASKA_AIRLINES.	Gate_D7
Gate_D54B	ALASKA_AIRLINES.	Gate_D8
Gate_D55	ALASKA_AIRLINES.	Gate_D9
Gate_D50A	ALASKA_AIRLINES.	Gate_D1
Gate_D56A	AMERICAN_AIRLINE	Gate_D10
Gate_D56B	AMERICAN_AIRLINE	Gate_D11
Gate_D57	AMERICAN_AIRLINE	Gate_D12
Gate_D58A	AMERICAN_AIRLINE	Gate_D14

Figure 7: *1b Mapping meter to gate*

In the case of SFO, which changed the gate naming system on 10/16/2019 at 00:00:00, an additional column "new_gate" was necessary to understand the mapping between the two naming systems. If the naming system did not change, this step can be avoided and a 2-column table is sufficient.

Gate Metering Data

Schneider Electric collects raw data on energy consumption from meters at 400Hz Ground Power Cables at SFO. Unfortunately, it does so in a way that is strikingly incompatible with Python code, meaning that a lot of handling and formatting is necessary.

First, set up the data extraction from the Schneider electric database. Identify all of the 400Hz ground power meters that are named exactly in the previously described Gate Mapping database. Set the data collection time period to the selected interval. Select "Real Energy Total (kWh)" in intervals of 5 minutes as the desired output variable. Select .xls as the output format.

The extraction tool will only let you select up to 30 columns at once and will crash otherwise. If you have more then 30 gates, or you would like to collect multiple data columns per gate, you will need to use the extraction tool multiple times. Name each file produced in a unique manner. Each file, if opened with Excel, should look like figure 8.

Schneider Electric

26

9/1/2019 12:00:00 AM - 2/1/2020 12:00:00 AM (Server Local)

Gap found between '9/9/2019 3:20:00 AM' and '9/9/2019 3:25:00 AM': BA_G.400HZ_Gate_102B Real Energy Total (kWh).
Gap found between '9/26/2019 3:20:00 PM' and '9/26/2019 3:25:00 PM': BA_G.400HZ_Gate_102B Real Energy Total (kWh).
Gap found between '10/4/2019 3:10:00 PM' and '10/4/2019 3:15:00 PM': BA_G.400HZ_Gate_102B Real Energy Total (kWh).
There are 434 unlisted warnings for: BA_G.400HZ_Gate_102B Real Energy Total (kWh).

Timestamp	BA_G.400HZ_Gate_100B Real Energy Into the Load (kWh)	BA_G.400HZ_Gate_100B Real Energy Total (kWh)	BA_G.400HZ_Gate_101A Real Energy Into the Load (kWh)	BA_G.400HZ_Gate_101A Real Energy Total (kWh)	BA_G.400HZ_Gate_101B Real Energy Into the Load (kWh)	BA_G.400HZ_Gate_101B Real Energy Total (kWh)	BA_G.400HZ_Gate_102A Real Energy Into the Load (kWh)
9/1/2019 12:05:00 AM		628.041,20		264.581,46		297.825,79	
9/1/2019 12:10:00 AM		628.041,26		264.581,54		297.825,87	
9/1/2019 12:15:00 AM	628.041,31	628.041,34	264.581,60	264.581,61	297.825,94	297.825,95	64.408,83
9/1/2019 12:20:00 AM		628.041,41		264.581,69		297.826,02	
9/1/2019 12:25:00 AM		628.041,48		264.581,76		297.826,09	
9/1/2019 12:30:00 AM	628.041,55	628.041,56	264.581,83	264.581,84	297.826,17	297.826,17	64.408,90
9/1/2019 12:35:00 AM		628.041,63		264.581,91		297.826,24	
9/1/2019 12:40:00 AM		628.041,70		264.581,99		297.826,31	
9/1/2019 12:45:00 AM	628.041,75	628.041,78	264.582,05	264.582,06	297.826,40	297.826,40	64.408,99

Figure 8: *Schneider electric raw data*

Unfortunately, despite having a nice aesthetic, these Excel documents have a user-unfriendly format. They have unnecessary text cells, invisible rows, and missing data. To clean it up, run the macro "Format_Interp_Interval_ECAM_SuperMacroUltra.xlsm", which was designed by an intern at SFO and attached in the software files for this project. Once the macro has run on each file, save each file as a .csv . Each file should look like Figure 9 below.

Timestamp	ALASKA_AIRLINES. Real Energy Into the (kWh)	ALASKA_AIRLINES. Real Energy Total (kWh)	ALASKA_AIRLINES. Real Energy Into the (kWh)	ALASKA_AIRLINES. Real Energy Total (kWh)	ALASKA_AIRLINES. Real Energy Into the (kWh)	ALASKA_AIRLINES. Real Energy Total (kWh)	ALASKA_AIRLINES. Real Energy Into the (kWh)	ALASKA_AIRLINES. Real Energy Total (kWh)
9/1/19 0:05		924,224.58		1,030,057.56		1,017,489.33		1,030,180.44
9/1/19 0:10		924,225.66		1,030,058.72		1,017,491.05		1,030,182.35
9/1/19 0:15		924,226.75		1,030,059.45		1,017,492.79		1,030,184.06
9/1/19 0:20		924,227.84		1,030,059.85		1,017,494.51		1,030,185.77
9/1/19 0:25		924,228.91		1,030,060.26		1,017,496.19		1,030,187.47
9/1/19 0:30		924,229.99		1,030,060.66		1,017,497.90		1,030,189.17
9/1/19 0:35		924,231.07		1,030,061.08		1,017,499.62		1,030,190.89
9/1/19 0:40		924,232.15		1,030,061.48		1,017,501.33		1,030,192.59
9/1/19 0:45		924,233.25		1,030,061.88		1,017,503.06		1,030,194.27
9/1/19 0:50		924,234.33		1,030,062.29		1,017,504.85		1,030,195.99
9/1/19 0:55		924,235.41		1,030,062.70		1,017,507.17		1,030,197.70
9/1/19 1:00	924,235.08	924,236.51	1,030,060.51	1,030,063.11	1,017,507.19	1,017,509.61	1,030,192.50	1,030,199.41

Figure 9: *1a Consumption Input Data*

Finally, collect all the .csv files and place them in a folder named "1a_consumption_input_data.csv". Make sure to record the names of each .csv file because they will need to be called in System 1.

Aircraft Operations Database

Aerobahn collects data on the operations of aircraft at SFO. For each arrival or departure, Aeroban records the metadata of the flight and the times at which specific operations were performed. Data for every operation at SFO for the specified time interval needs to be collected.

Whether multiple Excel files or a single one are obtained, the final product should be a unique .csv file named "3a_sample_aerobahn_ops_data". The column names of this file should correspond exactly to those called out in System 3. A screenshot of this file is shown in figure 10 and 11.

Carrier Group	Call Sign	Registration	Model	Operation	Origination Airport	Destination Airport	Gate Assigned (Aero	Gate	Runway Assigned (A	Runway
United Airlines	UAL257	N769UA	B772	Arrival					28R	28R
United Airlines	UAL385	N57439	37K	Arrival	IAH	SFO	Gate_F73	Gate_F73	28R	28R
Alaska Airlines	SKW3395	N404SY	E75	Arrival	PSP	SFO	HardStand_2_A13_	HardStand_2_A13_	28R	28R
Delta Air Lines	DAL2416	N115DU	BCS1	Arrival	SLC	SFO	Gate_C42	Gate_C42	28R	28R
United Airlines	SKW5557	N788SK	CR7	Arrival	SBA	SFO	Gate_F84D	Gate_F84D	28R	28R
United Airlines	UAL351	N587UA	75K	Arrival	BOS	SFO	Gate_F86	Gate_F86	28R	28R
Other	TAI560	N769AV	A20N	Arrival	SAL	SFO			28R	28R
American Airlines	AAL680	N977UY	A321	Arrival	PHL	SFO	Gate_D57	Gate_D57	28R	28R
American Airlines	AAL2798	N953AN	B738	Arrival	ORD	SFO	*HGR	Gate_D59	28R	28R
United Airlines	UAL1105	N77871	75E	Arrival	DEN	SFO	Gate_F88	Gate_F88	28R	28R
Southwest	SWA1093	N455WN	B737	Arrival	LAX	SFO	Gate_B14	Gate_B14	28R	28R
Jet Blue	JBU915	N929JB	A321	Arrival	JFK	SFO	Gate_B7	Gate_B7	28R	28R

Figure 10: *3a Sample Aerobahn Data (first part)*

International or Dom	Event Time	Scheduled Off Block	Actual Off Block Tim	Movement Area Entr	Actual Take Off Tim	Actual Landing Time	Movement Area Exit	Scheduled In Block	Actual In Block Time	Total Taxi Time
	9/12/19 0:00					9/12/19 0:00				
Domestic	9/12/19 0:02	9/11/19 16:31	9/11/19 20:10		9/11/19 20:27	9/12/19 0:02	9/12/19 0:05	9/11/19 20:44	9/12/19 0:06	0:04:25
Domestic	9/12/19 0:04	9/11/19 20:40	9/11/19 22:32		9/11/19 22:57	9/12/19 0:04	9/12/19 0:08	9/11/19 22:10	9/12/19 0:13	0:09:16
Domestic	9/12/19 0:05	9/11/19 21:25	9/11/19 22:00		9/11/19 22:31	9/12/19 0:05	9/12/19 0:08	9/11/19 23:30	9/12/19 0:08	0:03:25
Domestic	9/12/19 0:07	9/11/19 17:18	9/11/19 23:09		9/11/19 23:24	9/12/19 0:07	9/12/19 0:09	9/11/19 18:30	9/12/19 0:12	0:05:08
Domestic	9/12/19 0:08	9/11/19 12:55	9/11/19 17:29		9/11/19 18:04	9/12/19 0:08	9/12/19 0:24	9/11/19 19:20	9/12/19 0:24	0:15:59
International	9/12/19 0:10	9/11/19 18:10				9/12/19 0:10	9/12/19 0:16			
Domestic	9/12/19 0:12	9/11/19 13:40	9/11/19 18:00		9/11/19 18:39	9/12/19 0:12	9/12/19 0:15	9/11/19 19:40	9/12/19 0:17	0:05:25
Domestic	9/12/19 0:14	9/11/19 15:00	9/11/19 19:11		9/11/19 19:57	9/12/19 0:14	9/12/19 0:16	9/11/19 19:39	9/12/19 0:20	0:05:59
Domestic	9/12/19 0:16	9/11/19 19:18	9/11/19 21:26		9/11/19 22:02	9/12/19 0:16	9/12/19 0:25	9/11/19 22:00	9/12/19 0:31	0:14:55
Domestic	9/12/19 0:18	9/11/19 21:50	9/11/19 23:10		9/11/19 23:26	9/12/19 0:18	9/12/19 0:23	9/11/19 23:10	9/12/19 0:29	0:10:18
Domestic	9/12/19 0:20	9/11/19 12:30	9/11/19 17:31		9/11/19 18:39	9/12/19 0:20	9/12/19 0:24	9/11/19 19:10	9/12/19 0:24	0:04:19

Figure 11: *3a Sample Aerobahn Data (second part)*

Model to Emissions Group

This data set serves to map specific aircraft subtypes to a broad emissions category, as exemplified in figure 12. It must be saved as a .csv named "5a_model_to_emissions_group.csv"

Subtype	Emissions_Grou	AAC	ADG
319	100-200 new	C	III
320	100-200 new	C	III
321	100-200 new	C	III
328	100-200 new	C	III
739	100-200 new	D	III
19C	100-200 new	C	III
19F	100-200 new	C	III
19G	100-200 new	C	III
19S	100-200 new	C	III
20C	100-200 new	C	III
20S	100-200 new	C	III
37K	100-200 new	D	III
3SE	100-200 new	C	III
73C	100-200 new	D	III
73H	100-200 new	D	III
73J	100-200 new	D	III
73Q	100-200 new	D	III
73R	100-200 new	D	III
73Y	100-200 new	D	III
75B	100-200 old	C	IV
75E	100-200 old	C	IV
75K	100-200 old	C	IV
75S	100-200 old	C	IV
76A	200-300	D	IV
76C	200-300	D	IV
77E	200-300	C	V
77G	200-300	C	V
77J	200-300	C	V
77M	200-300	C	V
77N	200-300	C	V
77Q	200-300	C	V
77U	200-300	C	V
77W	200-300	C	V
77X	200-300	C	V
77Y	200-300	C	V
78H	200-300	C	V
78J	200-300	C	V
78V	200-300	C	V
78Z	200-300	C	V
A20N	100-200 new	C	III
A21N	100-200 new	C	III
A319	100-200 new	C	IV
A320	100-200 new	C	III
A321	100-200 new	C	III
A332	200-300	C	V
A333	200-300	C	V
A343	>300 new	C	V
A359	>300 new	C	V
A388	>300 new	D	VI
B736	100-200 new	C	III
B737	100-200 new	C	III
B738	100-200 new	D	III
B739	100-200 new	D	III
B744	>300 new	D	V
B752	100-200 old	C	IV
B763	200-300	D	IV
B772	200-300	C	V
B77L	200-300	C	V
B77W	200-300	D	V
B788	200-300	C	V
B789	200-300	C	V
CR7	BJ/RJ	C	II
CRJ	BJ/RJ	C	II
CRJ9	BJ/RJ	C	III
E75	BJ/RJ	C	III
E75L	BJ/RJ	C	III
MD83	100-200 old	D	III

Figure 12: *5a Model to Emissions*

Emission Group to Emission Values

This dataset maps each emissions category to the respective emission rates of APUs for several criteria air pollutants. Both the “start” and “normal” mode of operations of the APU were considered because their emission rates are significantly different. Figure 13 below shows exactly how the table should look like. This table must be saved as a .csv named “5b_emissions_group_to_emission_vals.csv”

This data can be acquired from the literature or from additional experimentation. The original data was acquired from:

Winther, M., Kousgaard, U., Ellermann, T., Massling, A., Nøjgaard, J. K., & Ketzel, M. (2015). Emissions of NOx , particle mass and particle numbers from aircraft main engines, APUs and handling equipment at Copenhagen Airport. *Atmospheric Environment*, *100*, 218–229. doi: 10.1016/j.atmosenv.2014.10.045

APU Group	Start NOx (kg/hr)	Start NO2 (kg/hr)	Start CO (kg/hr)	Start HC (kg/hr)	Start PM (kg/hr)	Start Fuel (kg/hr)	Normal NOx (kg/hr)	Normal NO2 (kg/hr)	Normal CO (kg/hr)	Normal HC (kg/hr)	Normal PM (kg/hr)	Normal Fuel (kg/hr)
>300 new	1.21	0.414	1.486	0.18	0.031	235	2.892	0.989	0.149	0.078	0.032	240
>300 old	1.137	0.389	5.4	0.302	0.04	300	2.071	0.708	3.695	0.153	0.037	283
100-200 new	0.364	0.124	3.734	2.662	0.031	100	0.805	0.275	0.419	0.094	0.031	100
100-200 old	0.565	0.193	1.289	0.105	0.034	110	1.064	0.364	0.336	0.036	0.034	110
200-300	0.798	0.273	0.982	0.243	0.033	105	1.756	0.601	0.248	0.07	0.058	187
BJ/RJ	0.274	0.094	1.019	0.107	0.016	50	0.452	0.155	0.799	0.044	0.028	90

Figure 13: *Emission Group to Emission Values*

Store all the collected data in a folder named sample_data with the following organization:

```
> working_folder
    > sample_data
        > 1a_consumption_input_data
            > gate_data_1.csv
            > gate_data_2.csv
        > 1a_consumption_input_data.csv
        > 1b_mapping.csv
        > 5a_model_to_emissions_group.csv
        > 5b_emissions_group_to_emission_vals.csv
```

This folder will then be populated with more data once the Systems are run, but this is the only data necessary to begin.

Running the Systems

In the same working folder on JupyterHub as the sample_data folder, add:

- All the given code from the 6 .ipynb systems
- The given folder “screenshots”

Then, open the systems in their numerical order. After following the recommendations below, simply run the systems. For further details, read the extensive annotations in the code.

System 1 - Gate Power Consumption Data Cleaning

This system will:

- Assign the proper gate names to the meters
- Convert cumulative energy in kWh to average power
- Remove the data around the daylight savings time to avoid confusion
- Remove any negative or impossibly large power consumption

Before running the system, make sure that the input data is called correctly. Assign the variable “input_1a_consumption_data_csvs” to the proper file names inside the “1a_consumption_input_data” folder.

The limits of acceptable power consumption are at least -50 kW and at most +250 kW. Anything outside these bounds is considered to be a measurement error, which rarely occurs, but must be considered. If necessary, these boundaries can be edited.

System 2 - Gate Power Consumption Visualizer

Running this system is optional. This system has no defined output.

The first part identifies the different granularities within the consumption data. Some data are collected every hour, every 15 minutes or half an hour. However, in the case of SFO, the overwhelming majority of the data comes in 5-minute increments. In the later System(s), a single granularity should be selected for the purpose of comparison, so this section can be used to identify which one.

The second part serves as a practical way to visualize data from the Schneider Electric data. Select the gate and time interval to the graph. Once the plot is made, the graphing tool will allow you to zoom and scroll to better see the data.

The last part of this system identifies the mode of power consumption for all gates. Since the vast majority of the time the gate is not powering the plane, the model represents a small phantom load continuously drawing power from the 400Hz infrastructure. If the mode is 0 kW, the metering data for the gate is likely faulty or flatlining. If the mode is between 0.1 kW and 2 kW, that is likely the phantom load that is always occurring.

System 3 - Operations Data Cleaning

This system will:

- Combine arrival and departure operations into a single pair with matching metadata
- Delete any movement that is not within an operation pair
- Delete any pair with missing registration, gate, and time data
- Delete unnecessary data such as useless columns
- Delete any operation that lasted longer than 12 hours
- Convert a changing gate naming system into a consistent one (optional)

To have consistent naming that can then be used with the metering data, the function change_new_to_old_gate_scheme can be used. Set the date of the change and make sure to add data to a new database 3b, which is simply a table with the map from the new to the old gates. This step is optional, but in the case that naming does change in the selected time interval, do not neglect it.

Any operation that lasted more than 12 hours was removed because at that duration there is no certainty on whether the airplane remained at the gate consuming APU. Beyond the initial 5 hours, it is unlikely the APU is left on and it could be possible that the aircraft was completely shut down, invalidating a key assumption in the analysis. In this system, the maximum duration of an operation can be altered in the inputs, settings and toggles section.

Make sure that the column names called in the code correspond exactly to those in the data. If working with a new aircraft operations system or another airport, this is especially relevant and column names might have to be altered. Additionally, one may choose to include or not certain columns from the metadata of the operation. For example, "Runway Assigned (Aerobahn)" was excluded because it is not relevant to what occurs at the gate.

While running the system, you can see how much data is being deleted and for what reason. This data-cleaning process removes a large part of the data, making the scope of the analysis significantly smaller.

System 4 - Data Combination

This system will apply relevant clean power consumption data to the clean operation data.

This system will take a long time to run. Depending on the size of the input data, it could range from a few minutes to several hours. With a lot of data, leave this system running overnight.

System 5 - Prediction Modeling

This system will:

- Remove varying granularity in power consumption
- Train a model to predict 400Hz power consumption
- Estimate the exact time at which the switch between ground power and APU occurs
- Estimate the APU use in terms of duration and power demand
- Evaluate operation success for a given operation
- Estimate the APU emissions associated with each operation

For the purposes of consistent comparison and model training, it is important to set a unique granularity. Use the information from System 2 to decide which granularity is most appropriate and common across the dataset. In the case of the sample data from SFO, it is 5 minutes. Select the desired_granularity_mins in the settings of the system.

The threshold_kW is a critical arbitrary variable that determines whether a certain power reading is considered to indicate APU or ground power. It has to be set low enough to allow as many positive readings on the 400Hz cable as there are in reality, without allowing data noise and the phantom load from showing a false positive. In the case of the sample data for SFO, 2kW was selected as the threshold; which is around twice the phantom load that occurs when aircraft are not present. Based on context and info from System 2, modify this variable.

Later in the system, a linear regression model for predicting power consumption is trained. The training variables initially selected are limited to “subtype”. However, many other features can be added to formulate a better model. Making sure the attributes are in the metadata, add them to num_vars if they are numerical and to cat_vars if they are categorical. These attributes could be “gate”, “carrier”, “time of day” etc. If attributes are absent in the metadata, make sure to derive them for the entire data set.

In the final part of the system, num_starts is multiplied by 3 to determine the number of minutes that the APU is in start mode, which has its own emission rates. If a different amount of start time is assumed, change the multiplier to the expected number of minutes per APU start.

System 6 - Visualization and Analysis

This system has no defined data output. It serves as a way to visualize the results of the analysis.

Initially, this system provides the ability to merge several different outputs of System 5, to allow the scope to be changed.

The first part of the system includes an operation visualizer, which displays a graph and table that comprehensively describe an operation, its energy consumption and emissions. For each operation the following variables are shown:

- *Power (kW)*, which is the actual power measured by the meter in a time interval
- *Predicted Power (kW)*, which is the projected power demand for the time interval

- Threshold (kW), which is the minimum power consumption required for the 400Hz to be considered as being used
- *Gate-in ratio (out of 1)*, which indicates what fraction of the interval the airplane was present at the gate
- *400 Hz ratio (out of 1)*, which indicates the estimated fraction of the interval the airplane was powered with 400Hz ground power
- *APU ratio (out of 1)*, which indicates the estimated fraction of the interval the airplane did not use 400Hz ground power but was present at the gate.

The second part shows consumption patterns in the form of histograms and box-and-whisker plots cut by attributes such as gate, aircraft subtype and carrier.

The third part shows the performance, measured with respect to operational success. This data can be grouped by attributes such as gate, aircraft subtype and carrier.

The fourth part shows the performance, measured with respect to APU time. This data can be grouped by attributes such as gate, aircraft subtype and carrier.

The final part of this system aggregates all the data into a single representative table that includes the total 400Hz consumption, the total APU consumption and the respective emissions within the scope of the analysis.

Conclusion

This tool lays the foundations for an APU monitoring system that both evaluates operations and constructs a model to predict future energy demands of aircraft. With this working structure, additional data and systems can be connected to construct a more comprehensive and precise monitoring tool.

The software described is a fully working prototype of a monitoring system. It will be delivered to San Francisco International Airport for their use. By running it repeatedly, the airport will be able to not only observe issues relevant to each iteration of the analysis, but also understand how ground power performance varies through time.

This tool can therefore be used to observe the effectiveness of APU policies and to progressively restrict them. It can be used to evaluate the power demand and field performance of different aircraft subtypes. In addition, it can be used to inform future investments in 400Hz ground power infrastructure.

The primary limitations of the current tool is that they have a limited scope due to imperfect data and that the analysis depends on the assumption that 400Hz ground power is the only alternative to APU use. To overcome such limitations, improvements in the data collection systems and more integrated gate systems will be necessary. Although these would be challenging to implement in the short term at SFO, there are airports with these improvements already in place.

The core principles of data combination, prediction modelling and evaluation remain the guiding concepts on which a monitoring system should work. They can be adapted to fit different airport infrastructure with different data collection systems, by editing and rearranging the code.

Our team hopes this tool will inspire airports, airlines, operators, and GSE manufacturers to make use of the vast amount of data they collect, share their databases and ultimately implement the principles behind the "internet of things". With transparency, increased supervision, and informed decision making, many challenges across airport operations can be addressed, in addition to ground power monitoring.

Acknowledgements

The creation of this tool would not have been possible without the close collaboration, support and data provided by San Francisco International Airport. Specifically, our team thanks Benjamin Gould, Jonathan Kocher and Amy Nagengast for their incredible support.

Our team thanks all of the students who have previously worked on this project, namingly Adhitama Buana, Andy Guan, Grace Horvath and Ye Suk Kim.

Our team thanks the Airport Cooperative Research Program and the Virginia Space Grant Consortium, specifically Larry Goldstein and Mary Sandy, for the support and funding of this project.

Finally, our team thanks everyone at UC Berkeley for providing the tools, skill and environment in which this project could be completed.

References

Data100 class at UC Berkeley, Fall 2019, Python code for Linear Regression implementation

Rakas, J. et al (2019) ACRP, "Smart Gate System for 400Hz Power Monitoring at Airports". Web. Accessed 2020/4/12 13:24. https://vsgc.odu.edu/acrpdesigncompetition/wp-content/uploads/sites/3/2019/08/2019_ACRP_DesignCompetition_1st_Environmental.pdf

Winther, M., Kousgaard, U., Ellermann, T., Massling, A., Nøjgaard, J. K., & Ketzel, M. (2015). Emissions of NO x , particle mass and particle numbers from aircraft main engines, APUs and handling equipment at Copenhagen Airport. Atmospheric Environment, 100, 218–229. doi: 10.1016/j.atmosenv.2014.10.045