Reconstructibility Analysis and Vehicle Health Management Data

Risto Rushford

March 13, 2020

Contents

1 Abstract **2**

2 Introduction **2**

3 Methodology **3**

4 Analysis **8**

5 Results **11**

6 Discussion **14**

7 Future Research **15**

8 Conclusion **15**

9 Works Cited **16**

[]

1 Abstract

This study uses Reconstructability Analysis to develop a predictive model for aircraft engine health based on diagnostic sensor and conditional indicator information. The study examines variable looped and loopless searches as well as neutral systems to examine a set of sensor and vehicle state data provided from a NASA Ames Research study on aircraft turbo jet-engines. The variables of interest were systematically analyzed so that those with higher correlations to engine performance were identified and used for further refinement and analysis. Finally, a model was produced which could be used to predict engine status and health conditions. The outcome of this study will inform a Portland State student rocketry team on the development of an engine health management system for a sophisticated liquid fuel rocket engine project.

2 Introduction

Aircraft engine management is an important aspect of product life-cycle management in the aerospace industry to prolong the lifespan of aircraft and reduce the overall cost of ownership per vehicle. Traditional aircraft maintenance is accomplished via routine checks performed according to a maintenance schedule based on flight hours per vehicle. This approach neglects variations in parts quality and events that may happen over the vehicle lifetime that can result in additional wear and tear than is expected according to the schedule. There has been much interest in reducing the cost of aircraft maintenance through embedded sensors and predictive models using data-mining and machine learning tools (usually Bayesian Networks and related algorithms) to indicate precisely when maintenance and/or replacement parts are needed.

The data used in this discussion comes from a NASA Ames study on using data mining to improve Vehicle Level Reasoning Systems, or automated control mechanisms that would allow aircraft engines to self-regulate based on sensor data, and signal when maintenance and repair may be necessary. The NASA Ames study was performed in 2012 and used a Bayesian approach. This study will differ in that a different approach will be used called Reconstructability Analysis using a tool called OCCAM, with the aim of examining the strength of the relationships between time-series sensor data and conditional indicators (boolean values which return a 0 when all readings are nominal, or 1 if a failure mode is detected).

Reconstructability Analysis is a data mining method based largely on Information Theory and is "an approach to discrete multivariate modeling developed in the systems community. RA includes set-theoretic modeling of relations and information-theoretic modeling of frequency and probability distributions" (Zwick, *An Overview of Reconstructability Analysis*). OCCAM is a discrete multivariate modeling tool which uses the Reconstructability Analysis methodology. It was developed by System Science researchers at Portland State University (PSU) and hosted on PSU servers, accessible at http://dmit.sysc.pdx.edu/weboccam.cgi.

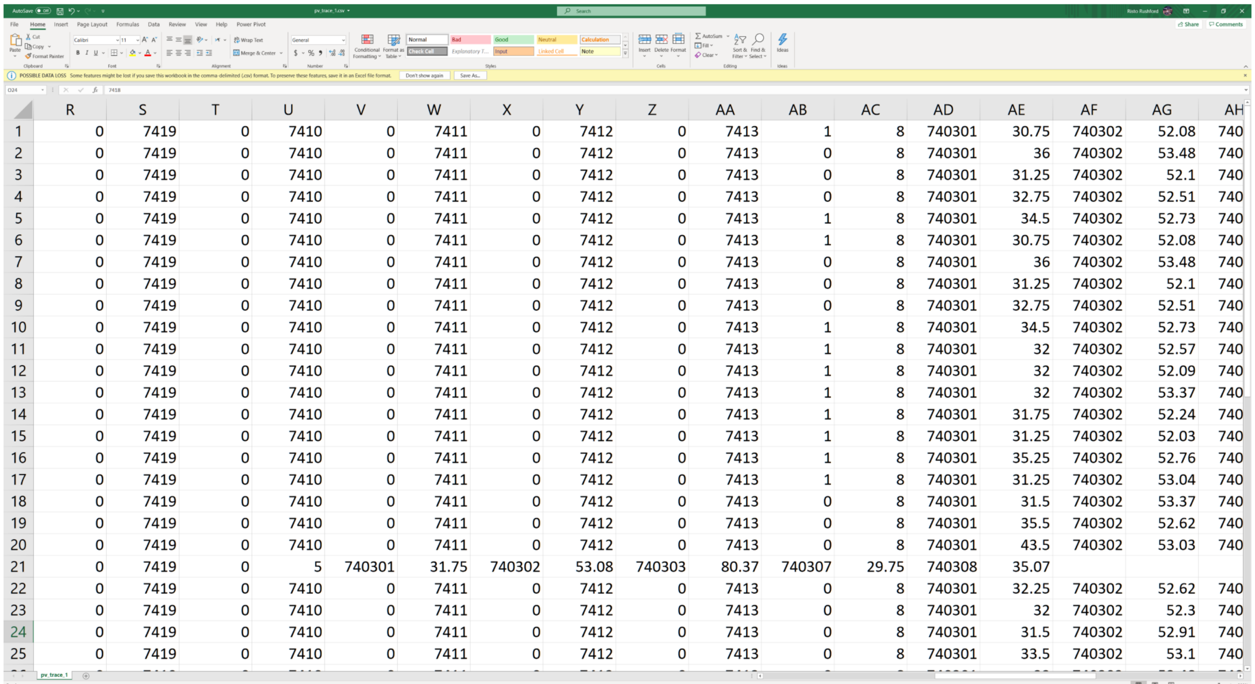
3 Methodology

This study used a dataset provided by NASA Ames Research Center in a 2012 study which used Tree Augmented Naive Bayesian Networks to analyze the sets of sensor readings and conditional indicators for diagnostic evidence of failure modes in four simulated jet aircraft engines. A further breakdown of the data is provided in the following subsection.

3.1 The Data

The data used in this study was generated using the **Modular Aero-Propulsion System Simulations (C-MAPSS)** software. The software itself is a Matlab Simulink tool of high fidelity developed by the NASA Glenn Research Center and is currently only available for U.S. Government agencies and approved contractors, but a number of generated datasets have been released to the public for study and machine learning competitions. The particular dataset used was one of a few available generated by C-MAPSS, and the only one with an explanation for each of the variables in the Readme file.

The data is in a rectangular form in a .csv file, with data from 15 monitors/sensors in each of 4 engines, 8 conditional indicators (nominal or state information), and 47 samplings from each engine. Information about this dataset can be found under the Vehicle Level Reasoning System-VLRS project at nasa.gov/dashlink or by a number of papers available online. There are some missing values in the dataset, as well as some lines of data with anomalous readings that don't seem to correspond with the rest of the dataset. The dataset was cleaned of these anomalies where it seemed necessary, but the missing values were kept and put through the binning utility as well, which simply adds one additional level of cardinality to the variables affected.



	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
1	0	7419	0	7410	0	7411	0	7412	0	7413	1	8	740301	30.75	740302	52.08	740
2	0	7419	0	7410	0	7411	0	7412	0	7413	0	8	740301	36	740302	53.48	740
3	0	7419	0	7410	0	7411	0	7412	0	7413	0	8	740301	31.25	740302	52.1	740
4	0	7419	0	7410	0	7411	0	7412	0	7413	0	8	740301	32.75	740302	52.51	740
5	0	7419	0	7410	0	7411	0	7412	0	7413	1	8	740301	34.5	740302	52.73	740
6	0	7419	0	7410	0	7411	0	7412	0	7413	1	8	740301	30.75	740302	52.08	740
7	0	7419	0	7410	0	7411	0	7412	0	7413	0	8	740301	36	740302	53.48	740
8	0	7419	0	7410	0	7411	0	7412	0	7413	0	8	740301	31.25	740302	52.1	740
9	0	7419	0	7410	0	7411	0	7412	0	7413	0	8	740301	32.75	740302	52.51	740
10	0	7419	0	7410	0	7411	0	7412	0	7413	1	8	740301	34.5	740302	52.73	740
11	0	7419	0	7410	0	7411	0	7412	0	7413	1	8	740301	32	740302	52.57	740
12	0	7419	0	7410	0	7411	0	7412	0	7413	1	8	740301	32	740302	52.09	740
13	0	7419	0	7410	0	7411	0	7412	0	7413	1	8	740301	32	740302	53.37	740
14	0	7419	0	7410	0	7411	0	7412	0	7413	1	8	740301	31.75	740302	52.24	740
15	0	7419	0	7410	0	7411	0	7412	0	7413	1	8	740301	31.25	740302	52.03	740
16	0	7419	0	7410	0	7411	0	7412	0	7413	1	8	740301	35.25	740302	52.76	740
17	0	7419	0	7410	0	7411	0	7412	0	7413	1	8	740301	31.25	740302	53.04	740
18	0	7419	0	7410	0	7411	0	7412	0	7413	0	8	740301	31.5	740302	53.37	740
19	0	7419	0	7410	0	7411	0	7412	0	7413	0	8	740301	35.5	740302	52.62	740
20	0	7419	0	7410	0	7411	0	7412	0	7413	0	8	740301	43.5	740302	53.03	740
21	0	7419	0	5	740301	31.75	740302	53.08	740303	80.37	740307	29.75	740308	35.07			
22	0	7419	0	7410	0	7411	0	7412	0	7413	0	8	740301	32.25	740302	52.62	740
23	0	7419	0	7410	0	7411	0	7412	0	7413	0	8	740301	32	740302	52.3	740
24	0	7419	0	7410	0	7411	0	7412	0	7413	0	8	740301	31.5	740302	52.91	740
25	0	7419	0	7410	0	7411	0	7412	0	7413	0	8	740301	33.5	740302	53.1	740

Figure 1: The C-MAPSS Raw Data

The readme file that came with the dataset explains it like so:

1. The first entry is the flight timestamp. It is not a real number, but just a monotonically increasing index

2. An integer number indicates the number of diagnostic monitors used for the flight in question

3. Monitor states are listed as either 0, 1 or -1 indicating on, off, or failure (note, none of the data in this particular set include a "-1" reading).

4. An integer represents the number of conditional indicators used for the flight

5. The value of the conditional indicator at the time of the sampling

Of particular interest is how the sensor data will relate to the conditional indicators, and can translate toward developing models for similar predictive diagnostics for the engine management systems of other aerospace systems. Some amount of data preparation was necessary to get this set ready for OCCAM, but since it is already in a rectangular format, most of the preparatory work was trivial (combining four data files, perhaps relabeling some of the variables, and adding the OCCAM front matter). The goal for using this data set is to gain insight into existing engine management systems such as those used in conventional aircraft such as produced by Boeing and Airbus.

3.1.1 Binning

The C-MAPSS dataset includes sensor data (8 variables) and conditional indicators (CI, 7 variables) for four simulated turbine jet engines from a single aircraft. An additional variable was also created to represent which engine the other data is associated with, as these were originally located in four separate files. Initially the data was binned into a cardinality of 12 or less based on recommendations in a Data Mining with Information Theory (DMIT) course. However an experimental run was made on several sensor variables with higher cardinality for the sensor variables (some variables had a much higher range) which returned better results. See **Tables 1 & 2** of the variables including the cardinality used in the analysis.

3.1.2 The Variables

Here I will list the variables in two subcategories: the sensor data (independent variables) and the conditional indicators (dependent variables).

3.1.3 Sensor Data

Since these are time-series data recordings of continuous variables, it seemed that higher cardinality would be better since some of these have fairly high ranges. The conditional indicators, having boolean values, were also binned via an excel-based binning utility provided by the DMIT course. *Note: sensor variables with an odd number of bins indicates there were missing values.*

Table 1: Sensor Variables

Variable	Sensor Range	Cardinality	Description
startTime	23 - 44.75 seconds	8	Time for engines to start up, measured in seconds
idleSpeed	51.05 - 65.71 % RPM	10	Rotational speed for each turbine at idle, measured in % RPM (rotations per minute)
peakEGTC	73.67 - 98.9 % °C	12	Maximum engine exhaust gas temperature, measured in % °Celsius
peakN1TKO	86.75 - 95.91 % rpm	11	Peak fan speed at take-off, % RPMs
tkoN1	86.66 - 95.85 % rpm	11	Average fan speed during take-off, % RPMs
tkoEGTC	73.5 - 92.1 % °C	13	Exhaust gas temperature at takeoff, % °Celsius
rollTime	5 - 38 seconds	9	Time duration of engine's roll-down phase, seconds
resdTemp	33.01 - 67.97 % °C	13	Engine exhaust gas temperature at the end of the engine roll-down phase, % °Celsius.

3.1.4 Conditional Indicator Variables

Table 2: Conditional Indicators

Conditional Indicator	Occurrences	Probability	Description
phOneDwell	8	4.17 %	Indicates the phase 1 starter cut-off speed too low
starter	6	3.14 %	Indicates trouble starting the engine
ignitor	1	0.52 %	Indicates engine ignitor error
highRollEGT	1	0.52 %	Indicates alarmingly high engine exhaust gas during engine run
medTempMargin	11	6.01 %	Indicates marginal engine temperature during take-off
lowTempMargin	88	48.09 %	Indicates an unusually low marginal engine temperature during take-off

The dataset also included several conditional indicator variables that were never triggered and so were cleaned from data used in the OCCAM input file.

If anyone is interested in reading a paper that performed a Bayesian analysis on a similar dataset which uses many of the same variables, here is a link to a conference paper titled *Deriving Bayesian Classifiers from Flight Data to Enhance Aircraft Diagnosis Models.*

3.1.5 Bin Analysis

Bin Distribution
(frequency)

bin #	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15
1	48	23	20	17	18	19	16	23	16	184	185	190	190	172	95
2	48	24	19	15	18	18	15	27	16	8	6	1	1	11	88
3	48	25	19	16	21	19	15	21	16						
4	48	17	17	17	19	18	15	21	19						
5		33	21	15	17	19	16	23	14						
6		21	19	16	19	17	15	29	15						
7		26	20	16	17	19	16	26	16						
8		23	19	16	17	21	16	21	16						
9			19	16	20	16	15		16						
10			19	16	18	18	15		16						
11				17			16		16						
12				15			14		15						

Figure 2: Frequency Distribution of Binned Variables

The above graphic shows the frequency distribution of the binned variables (up to cardinality 12, though the sensor variables are of higher cardinality). Note that the first variable was created in this study to indicate which jet engine is associated with the sensor data. Variables 2-9 are the sensor measurement data which were binned using the binning utility in Excel, and variables 10-15 are the boolean conditional indicators.

Here we will look at the sensor variables with the highest and lowest ranges. Variable #9 in the table above is resdTemp (residual temperature). This variable has the highest range, representing the temperature of the engine exhaust gas after the roll down phase. The actual range for this variable is in % °Celsius, from 33.01 to 67.97. Despite the large range, this variable is mostly distributed evenly across its range, although there is a slight increase in frequency of values in the 40-50 cardinality range. The bins in this variable data shows a high point in bins 5, followed by a lower frequency in bins 60-69. Bins 40-50 correspond to a measurement of around 38.77, while bins 60-69 correspond to measurements of around 39.20. This suggests that there may be a mechanical or material property of the fan (nominal roll speed, or specific heat of the material?) corresponding to the higher frequency of temperature at the one point versus the slightly higher value.

3.1.6 Data Exclusions

There were several conditional indicator variables which were excluded from this analysis because they were never triggered during the engine simulation runs. Since these variables are boolean and never switch from 0 to 1, there would never be any relationships to derive from them.

4 Analysis

```
:action
search

:nominal
EngineNumber	,4,1,A
startTime	,8,1,B
idleSpeed	,10,1,C
peakEGTC	,12,1,D
peakN1TKO	,11,1,E
tkoN1	,11,1,F
tkoEGTC	,13,1,G
rollTime	,9,1,H
resdTemp	,13,1,I
phOneDwell	,2,1,Z
starter	,3,1,Y
ignitor	,3,1,X
highRollEGT	,3,1,W
medTempMargin	,3,1,V
lowTempMargin	,3,1,U

:no-frequency

:data
#A	B	C	D	E	F	G	H	I	Z	Y	X	W	V	U
1	4	3	9	9	9	8	4	10	1	1	1	1	1	2
1	8	10	8	4	5	5	3	9	1	1	1	1	1	1
1	5	3	8	3	3	4	5	11	1	1	1	1	1	1
1	6	6	9	4	4	6	5	12	1	1	1	1	1	1
1	7	7	10	7	7	8	4	7	1	1	1	1	1	2
1	4	3	9	9	9	8	4	10	1	1	1	1	1	2
1	8	10	8	4	5	5	3	9	1	1	1	1	1	1
1	5	3	8	3	3	4	5	11	1	1	1	1	1	1
1	6	6	9	4	4	6	5	12	1	1	1	1	1	1
1	7	7	10	7	7	8	4	7	1	1	1	1	1	2
1	5	7	8	8	8	11	3	10	1	1	1	1	1	2
1	5	3	5	10	10	7	2	7	1	1	1	1	1	2
1	5	9	6	9	9	10	3	9	1	1	1	1	1	2
1	5	5	8	8	9	10	3	8	1	1	1	1	1	2
1	5	2	8	9	9	7	4	10	1	1	1	1	1	2
1	8	8	9	10	10	11	3	8	1	1	1	1	1	2
1	5	9	10	6	7	7	4	7	1	1	1	1	1	2
1	5	9	2	5	5	2	1	4	1	1	1	1	1	1
```

Figure 3: OCCAM Input File, Neutral Search

Once the data has been cleaned and binned, one can begin to analyze it using OCCAM. The OCCAM input file is .tsv format with specific header information to provide instructions to the OCCAM program regarding the data. Note in **Figure 3** the name of each variable, followed by a tab, and then the cardinality, another number, and a letter. The second number tells OCCAM how to treat the variable. In the input file shown, each variable is being used in a neutral search, and this is indicated by the "1" value. Another version of the input file (see **Figure 4**) tells OCCAM to perform a directed search, and only the lowTempMargin conditional indicator variable is turned on, but with a "2" value instead of a "1".

```
:action
search

:nominal
EngineNumber    ,4,1,A
startTime       ,8,1,B
idleSpeed       ,10,1,C
peakEGTC        ,12,1,D
peakN1TKO       ,11,1,E
tkoN1   ,11,1,F
tkoEGTC ,13,1,G
rollTime        ,9,1,H
resdTemp        ,13,1,I
phOneDwell      ,2,0,Z
starter ,3,0,Y
ignitor ,3,0,X
highRollEGT     ,3,0,W
medTempMargin   ,3,0,V
lowTempMargin   ,3,2,U

:no-frequency

:data
#A  B   C   D   E   F   G   H   I   Z   Y   X   W   V   U
1   4   3   9   9   9   8   4   10  1   1   1   1   1   2
1   8   10  8   4   5   5   3   9   1   1   1   1   1   1
1   5   3   8   3   3   4   5   11  1   1   1   1   1   1
1   6   6   9   4   4   6   5   12  1   1   1   1   1   1
1   7   7   10  7   7   8   4   7   1   1   1   1   1   2
1   4   3   9   9   9   8   4   10  1   1   1   1   1   2
1   8   10  8   4   5   5   3   9   1   1   1   1   1   1
1   5   3   8   3   3   4   5   11  1   1   1   1   1   1
1   6   6   9   4   4   6   5   12  1   1   1   1   1   1
1   7   7   10  7   7   8   4   7   1   1   1   1   1   2
1   5   7   8   8   8   11  3   10  1   1   1   1   1   2
1   5   3   5   10  10  7   2   7   1   1   1   1   1   2
1   5   9   6   9   9   10  3   9   1   1   1   1   1   2
1   5   5   8   8   9   10  3   8   1   1   1   1   1   2
1   5   2   8   9   9   7   4   10  1   1   1   1   1   2
1   8   8   9   10  10  11  3   8   1   1   1   1   1   2
1   5   9   10  6   7   7   4   7   1   1   1   1   1   2
1   5   9   2   5   5   2   1   4   1   1   1   1   1   1
```

Figure 4: OCCAM Input File, Directed Search

With the input files ready, the following procedure was used:

1. Run a loopless search on each conditional indicator (see the directed search file example above). This is to determine if there are any discernible relationships between the CIs and the sensor readings, with the null hypothesis that there is no relationship. For the variables highRollEGT, ignitor, and medTempMargin, OCCAM returned the independence model for all search variations, indicating that there was no meaningful relationship between these indicators and the sensor data within this dataset.

2. Run additional searches on the more promising variables. Due to the relatively small sample size of the data, there was no need to ever select a starting model. For several searches, looped models were searched with a wide scope without any penalty in OCCAM. This allowed for interesting results which will be discussed in the next section.

3. Run State-Based searches on promising relationships.

4. Run a neutral search with all variables turned on. This was successfully done with a loopless search which provided interesting results. However OCCAM was consistently unable to provide results for a looped neutral search as the model quickly grew too much in complexity.

5 Results

5.1 Directed Searches

5.1.1 phOneDwell

This conditional indicator was especially interesting because the OCCAM model **IV:BZ** suggest that the time to start up an aircraft engine is predictive of the starter cutting off too soon. This seems to follow from logic, and is a promising first result for this study.

ID	MODEL	Level	H	dDF	dLR	Alpha	Inf	%dH(DV)	dAIC	dBIC	Inc.Alpha	Prog.	%C(Data)	%cover
Best Model(s) by dBIC:														
1*	IV:Z	0	7.6265	706717439	66.5108	1	0	0	1413434811	3715563598	0	0	95.8333	100
Best Model(s) by dAIC:														
10*	IV:BZ	1	7.4883	706717432	29.7202	1	0.55315282	55.3153	1413434834	3715563598	0	1	95.8333	100
Best Model(s) by Information with all Inc. Alpha < 0.05:														
10*	IV:BZ	1	7.4883	706717432	29.7202	1	0.55315282	55.3153	1413434834	3715563598	0	1	95.8333	100

Table 3: OCCAM Search Results, Directed Search phOneDwell

5.1.2 lowTempMargin

Low temperature margin was the most interesting of the conditional indicators to be examined. Several different OCCAM search parameters suggested different relationships between this variable and the sensor data. When this conditional indicator is triggered in the simulation, it means that the engine is "in the red" and about to overheat. It appears that there are several contributing factors as we will discuss:

ID	MODEL	Level	H	dDF	dLR	Alpha	Inf	%dH(DV)	dAIC	dBIC	Inc.Alpha	Prog.	%C(Data)	%cover
Best Model(s) by dBIC:														
3*	IV:EU	1	7.6256	20	259.8214	0	0.79680179	79.6802	219.8214	154.6715	0	1	90.1042	100
Best Model(s) by dAIC:														
7*	IV:AU:GU	2	7.5206	30	287.7706	0	0.88251451	88.2515	227.7706	130.0457	0.0166	5	95.3125	98.0769
Best Model(s) by Information with all Inc. Alpha < 0.05:														
10*	IV:AU:FU:HU	3	7.4539	42	305.5141	0	0.93692903	93.6929	221.5141	84.6993	0.0009	4	99.4792	29.0404

Table 4: OCCAM Search Results, lowTempMargin with Loops

A run through OCCAM using width of 5, depth of 9, and allowing for loops returned the best fit model by Inc. Alpha **IV:AU:FU:HU**, suggesting the the engine number, the average fan speed during takeoff, and the amount of time needed for an engine to "roll down" at the end of flight are predictive of whether or not the engine temperature is "in the red". The best model by ΔAIC IV:AU:GU also predicts engine number as well as the exhaust gas temperature at take-off, indicating that the temperature problem begins very early on in the flight time. Finally, the best fit by ΔBIC IV:EU suggests that the peak fan speed at take-off is predictive of a low temperature margin. What's interesting is that logically, each of these factors would seem to co-contribute to this problem, yet OCCAM predicted them separately under different selection criteria.

Since the model IV:AU:FU:HU is the only first chain model to appear in in the OCCAM results, a DO-FIT and a State-Based search were performed to glean more information about this set of variable relations.

Conditional DV (D) (%) for each IV composite state for the Relation AU.												
(For component relations the Data and Model parts of the table are equal so only one is given.)												
IV	|	Data										
	|		obs. p(DV|IV)									
A	|	freq	U=.	U=1	U=2	rule	#correct	%correct	E(DV)	MSE	p(rule)	p(margin)
1	|	48	4.167	58.333	37.5	1	28	58.333	0	2.083	0	0.476
2	|	48	6.25	47.917	45.833	1	23	47.917	0	2.312	0	0.875
3	|	48	4.167	41.667	54.167	2	26	54.167	0	2.583	0	0.517
4	|	48	4.167	50	45.833	1	24	50	0	2.333	0	0.985
	|	192	4.688	49.479	45.833	1	101	52.604	0	2.328		
	|	freq	U=.	U=1	U=2	rule	#correct	%correct	E(DV)	MSE	p(rule)	p(margin)
--												
Conditional DV (D) (%) for each IV composite state for the Relation FU.												
(For component relations the Data and Model parts of the table are equal so only one is given.)												
IV	|	Data										
	|		obs. p(DV|IV)									
F	|	freq	U=.	U=1	U=2	rule	#correct	%correct	E(DV)	MSE	p(rule)	p(margin)
.	|	8	100	0	0	.	8	100	0	0	0	0
1	|	19	0	100	0	1	19	100	0	1	0	0
2	|	18	0	100	0	1	18	100	0	1	0	0
3	|	19	0	100	0	1	19	100	0	1	0	0
4	|	18	0	77.778	22.222	1	14	77.778	0	1.667	0	0.051
5	|	19	0	73.684	26.316	1	14	73.684	0	1.789	0.001	0.107
6	|	17	5.882	58.824	35.294	1	10	58.824	0	2	0.028	0.683
7	|	19	0	5.263	94.737	2	18	94.737	0	3.842	0	0
8	|	21	0	0	100	2	21	100	0	4	0	0
9	|	16	0	0	100	2	16	100	0	4	0	0
10	|	18	0	0	100	2	18	100	0	4	0	0
	|	192	4.688	49.479	45.833	1	175	91.146	0	2.328		
	|	freq	U=.	U=1	U=2	rule	#correct	%correct	E(DV)	MSE	p(rule)	p(margin)
--												
Conditional DV (D) (%) for each IV composite state for the Relation HU.												
(For component relations the Data and Model parts of the table are equal so only one is given.)												
IV	|	Data										
	|		obs. p(DV|IV)									
H	|	freq	U=.	U=1	U=2	rule	#correct	%correct	E(DV)	MSE	p(rule)	p(margin)
.	|	1	0	100	0	1	1	100	0	1	0.368	0.6
1	|	23	21.739	69.565	8.696	1	16	69.565	0	1.043	0.001	0
2	|	27	0	74.074	25.926	1	20	74.074	0	1.778	0	0.036
3	|	21	0	23.81	76.19	2	16	76.19	0	3.286	0	0.018
4	|	21	9.524	28.571	61.905	2	13	61.905	0	2.762	0.012	0.13
5	|	23	0	78.261	21.739	1	18	78.261	0	1.652	0	0.02
6	|	29	6.897	55.172	37.931	1	16	55.172	0	2.069	0.005	0.642
7	|	26	0	42.308	57.692	2	15	57.692	0	2.731	0.001	0.319
8	|	21	0	9.524	90.476	2	19	90.476	0	3.714	0	0
	|	192	4.688	49.479	45.833	1	134	69.792	0	2.328		
	|	freq	U=.	U=1	U=2	rule	#correct	%correct	E(DV)	MSE	p(rule)	p(margin)

Figure 2: Low-Temp Margin Model IV:AU:FU:HU Fit

ID	MODEL	Level	H	dDF	dLR	Alpha	Inf	%dH(DV)	dAIC	dBIC	Inc.Alpha	Prog.	%C(Data)	%cover
Best Model(s) by dBIC:														
9*	IV:FU2:U	2	7.1769	10	188.7648	0	0.58575087	57.889	168.7648	136.1898	0	3	86.9792	100
Best Model(s) by dAIC:														
9*	IV:FU2:U	2	7.1769	10	188.7648	0	0.58575087	57.889	168.7648	136.1898	0	3	86.9792	100
Best Model(s) by Information with all Inc. Alpha < 0.05:														
9*	IV:FU2:U	2	7.1769	10	188.7648	0	0.58575087	57.889	168.7648	136.1898	0	3	86.9792	100

Table 5: OCCAM Search Results, lowTempMargin State-Based

A state based search starting with the previously given model IV:AU:FU:HU was narrowed down **IV:FU2:U** by all three metrics of best fit. This seems to agree with the best fit data, which states that the FU relation has the highest % correct.

5.2 Neutral Searches

While the directed searches were interesting, the results seem questionable as most of the models were assigned a negative ΔBIC value by OCCAM. We see very different results when performing a neutral search:

ID	MODEL	Level	H	dDF	dLR	Alpha	Inf	dAIC	dBIC	Inc.Alpha	Prog.
Best Model(s) by dBIC:											
45*	IVI:AB:AC:AD:AH:EF:EU:GU:XW:VU	9	23.5492	343464675489	4304.6285	1	0.33103808	686929346673	1805763937534	0	38
Best Model(s) by dAIC:											
46*	IVI:AB:AC:AD:AH:BZ:EF:EU:GU:VU	9	23.5049	343464675486	4292.847	1	0.332869	686929346679	1805763937530	0	39
Best Model(s) by Information with all Inc. Alpha < 0.05:											
46*	IVI:AB:AC:AD:AH:BZ:EF:EU:GU:VU	9	23.5049	343464675486	4292.847	1	0.332869	686929346679	1805763937530	0	39

Table 6: OCCAM Search Results, Neutral

This search shoes that several of the conditional indicators may be co-predicting. Due to limitations of time in this current study, we will not explore this model any further, however it is interesting to see the predicted relations and this may be a subject of further study later on.

6 Discussion

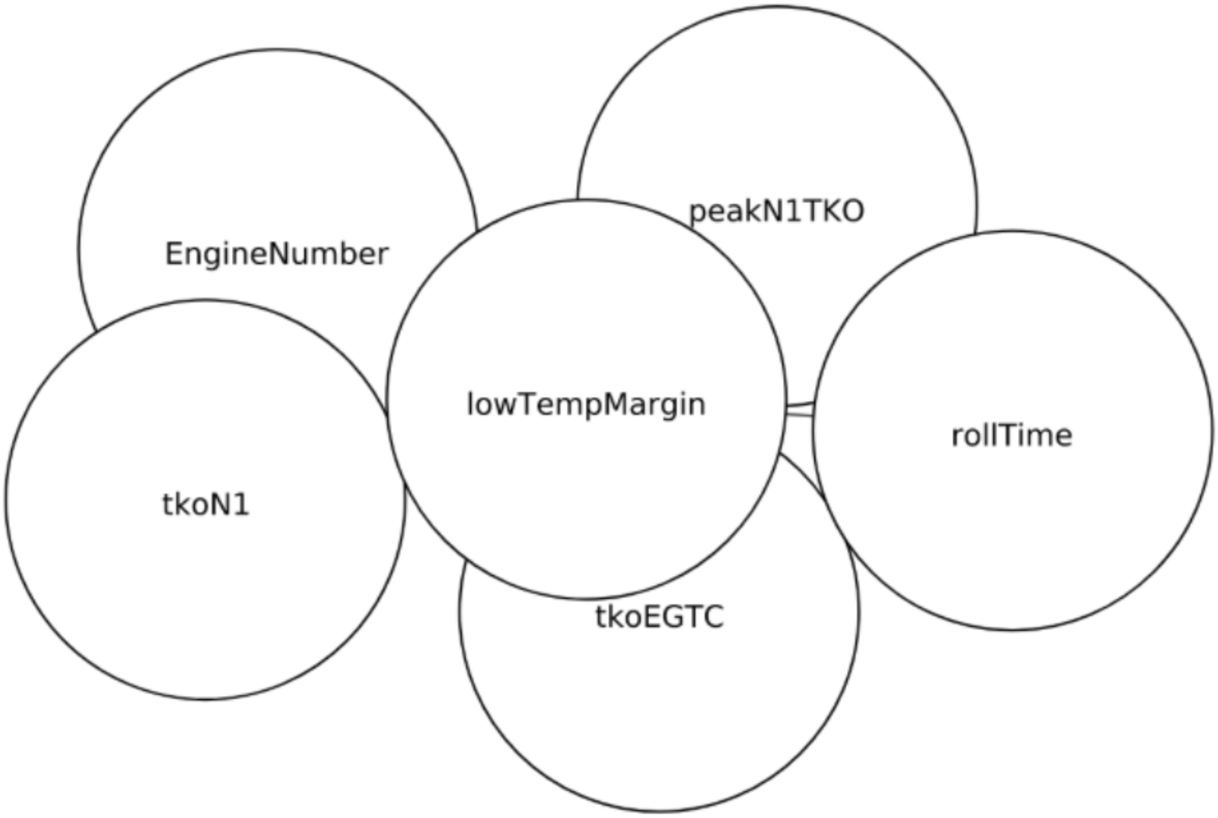


Figure 3: Hypergraph of Low-Temp Margin Model IV:AU:FU:HU:GU:EU

Due to current limitations in OCCAM development, hypergraph functionality is limited (part of why we aren't continuing discussion of the Neutral system at this time), but with a sufficiently simple model we can observe the nature of this one set of relationships. We can see that some of the sensor variables are better predictors of the low temperature margin than others.

An observation of note is that the engine number is a strong predictor of low temperature margin. This may be due to specific engines being the main culprits of triggering that particular conditional indicator over others. It's also somewhat surprising that the only temperature related sensor with an identified relationship via this study is exhaust gas temperature at takeoff. One would otherwise expect that the other temperature sensors would have played a larger role in creating this condition. This puzzle is largely why this particular set of relationships is of interest to the author.

7 Future Research

The lessons learned from this research project will be used in the near future to analyze sensor data from a liquid fuel rocket engine project being undertaken by an aerospace research organization in the Portland State School of Engineering. While there will differences in how the variables relate to each other due to the nuances of design and interactions of parts, the methodology used in this study will be invaluable to that future effort. Data Mining with Information Theory will be a powerful tool in the tool-chest of data science and engineering methodologies that are driving PSU space technology development forward.

8 Conclusion

While this study produced some very interesting and unexpected results, it was hampered by a lack of data and real world application at the time of the study. Aerospace engine data of any sort is difficult to find on the internet except for simulated datasets like the one used in this paper. Nevertheless DMIT has shown that it can be useful in analyzing aerospace systems and will see further use in this area in the near future.

9 Works Cited

Zwick, Martin. "*An Overview of Reconstructability Analysis*," Portland State University Systems Science Faculty Publications, 2004. https://pdfs.semanticscholar.org/5ce4/b93e04891e659f195e98f950

Mylaraswamy et. al, "*Using Tree Augmented Naive Bayesian Classifiers to Improve Engine Fault Models*," NASA NRA NNL09AD44T, 2011, Uncertainty in Artificial Intelligence: Bayesian Modeling Applications Workshop. https://www.isis.vanderbilt.edu/sites/default/files/UAIWS201

A. Saxena and K. Goebel (2008). "*Turbofan Engine Degradation Simulation Data Set*", NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/project/prognostic-data-repository), NASA Ames Research Center, Moffett Field, CA

Kjelgaard, Chris. "*What is Jet Engine Maintenance?*". AV Buyer, 4 Sept. 2019, https://www.avbuyer.com biz-av/what-is-jet-engine-maintenance-112549