Alignment detection using Internal Vehicle Signals and Measurements

Joshua McCready, M.S.E. ECE Student and Core Steering Mechatronics Engineer at Ford Motor Co.

***Abstract*—Improper tire wear due to wheel misalignment in passenger vehicles results in negative side effects for customers and the OEM. This paper proposes the beginning of passenger vehicle wheel misalignment prognostic strategy using internal vehicle signal for detection of degraded conditions in the absence of DTCs and without the need dynamic or kinematic models of the passenger vehicle chassis and road wheels. Identifying degraded conditions before they result in negative effects reduces waste, customer headache, and warranty cost to the automaker. The data in this study was collected in a longitudinal study of a 2017 fusion instrumented with a data collection system. The feasibility of implementing prognostics via machine learning classification of internal vehicle data using over the air data collection for offline classification shall be studied.**

I. Introduction

PER year 400 million tires worth an estimated $25 billion dollars are replaced before their tread wear rating resulting in waste equivalent to 35 million tons of carbon dioxide emissions [1]. Those figures are hardly startling considering that tires are the only thing keeping occupants and vehicle safely planted on the road. Not replaced these degraded tires result in poor braking and handing, reduced fuel economy, and tire failure. These costs are shared by the vehicle OEM or tire manufacturer, the customer, and the environment. It is in the interest of the vehicle OEM to curtail this waste, reduce warranty cost, and increase customer satisfaction.

The causes of premature tire replacement are varied. Because tires carry the lateral forces that accomplish cornering, road wheel misalignment or improper tire pressure can cause uneven tread wear. Eventually leading to reduced directional control, tire sidewall damage, and a flat or otherwise destroyed tire. Accelerated tire tread wear can result from an aggressive driving style, driving surfaces such as gravel, and pneumatic tires are susceptible to puncture. For some of the causes of degradation, there are technological advances that can prevent degradation from occurring.

Tire pressure monitoring systems have been mandated by the TREAD act in the U.S. and similar laws elsewhere. Recently, Michelin has developed non-pneumatic technology to eliminate puncture, blow out, and wear due to incorrect pressure [2]. Perhaps unintentionally, electric power steering input torque compensation features have corrected the customer observable effects of road wheel misalignment as they learn road crown. Torque pulling the steering wheel one way or another can be corrected as a learned offset value [3]. However, this torque compensation only reduces costumer observable steering torque and not wear. In this compensated case, the remaining observable effect to the driver is steering wheel angle offset which may not be perceived if slight. With reduced driver perception it is likely that the root cause will worsen the condition of the tires.

A. Data Source

As part of autonomous vehicle research at Ford Motor Company, a 2017 model year Ford Fusion was instrumented for collection of vehicle signals buses and internal control module signals. The data was collected with dSPACE Autobox real-time system. The test vehicle was modified with different tire and alignment conditions resulting degraded handling and safety. Test drives were performed on public roads throughout Southeast Michigan in both degraded and nominal vehicle condition. The resultant data comprises roughly 24.9 hours of time series data sampled at 100 Hz. The task of the work is to correctly identify misalignment of the road wheels using pattern recognition techniques in the presence of nominal operation and noise factors such as degraded tires, asymmetric weighted vehicle condition, and incorrect tire pressures. This work supports the development of prognostics and over the air diagnostics for future Ford products.

B. Literature Review

There are limited examples of misalignment detection in the literature, but generically using vehicle signal data to classify and predict degraded state has been successful. In [4] Poloni et. al. proposes an indirect prediction of tire wear using batch Ordinary Least Squares (OLS) for offline analysis and Kalhman Filter or Recursive Least Squares for online estimation. Their approach, like others relied at least somewhat on modeling of the feature being estimated - here tire effective radius. Historically, online estimation or compensation algorithms like that used by Badiru [3] are dynamic or kinematic in nature. This due in some part to the cost of computational resources, safety critical nature, and limited ability to update the software on chassis control modules.

While directly related to the task at hand, other authors Vasudevan et. al. [5] and Sun et. al. in [6] had success using machine learning techniques for driver drowsy state estimation and vehicle powertrain prognostics with internal vehicle signal data-sets. [6] expresses the difficulty of obtaining data for diagnostics or prognostics relative to faults in a power train control module and describes the challenges machine learning can help solve in this context. These decisions are out of scope to the project at hand, but are relevant for application ready use of machine learning in this context.

1) Decide when to record data
2) Decide what data is relevant to record
3) Root cause the the time series data

[6] also attempts to solve a problem with an inherently skew dataset. Likewise in the population of vehicles, there will be far fewer examples with alignment issues than without them. [6] presents two concepts that may be of use. The first is a system of majority voting of several different machine learning methods. Prognostics and diagnostics that are done offline and that gather data over the air would enable this kind of decision making. Second, is an assumption that vehicle components and they exchange are not independent, meaning some using reference signals the expected value of another signal can be inferred. This may be useful as the dataset in this research consists of the same kind of inter-dependant time series data as we will show later by studying correlation.

II. Database Description

A. Creating a Usable Database from Raw Data

The database was originally 366 hand labeled files recording various drives each with about 100 time series signals recorded from the Control Area Network (CAN) and from internal signals from different modules within the vehicle. The data stored in individual files had to be reconciled into a single database structure of X. The different data files did not originally contain the same CAN and internal module signals. Due to the structure of the data files, a great deal of processing was required to reconcile all the different signals and preserve most essential signals and minimize the number of data files that had missing or inconsistent data names. The resultant dataset X contained 50 time series signals, each labeled with a name and unit. The hand file labeling denoted the vehicle condition and sometimes drive location. Those hand labels were processed with regular expressions and the labels were translated such that each time series data entry. This created Y. Because missing signals or data inconsistencies, only 303 of the original data files have been included in X and Y

B. Features in the Database

Of the 50 time series signals, many are near duplicates. For instance, Vehicle Vertical Acceleration Rate Compensated and Vehicle Vertical Acceleration Rate Raw differ only because of signal conditioning and perhaps some filtering. While these differences are relevant to the software and control system, they may not be relevant in a stochastic sense. In Appendix B table of all 50 signals has been provided. We will use all of the signals in the exploratory data analysis. To better explain some of the concepts of these signals Figures 1 - 3 have been provided.

1) Electric Power Steering: Figure 1 describes the basic operation of steering system of a 2017 Ford Fusion.

Context Diagram.png

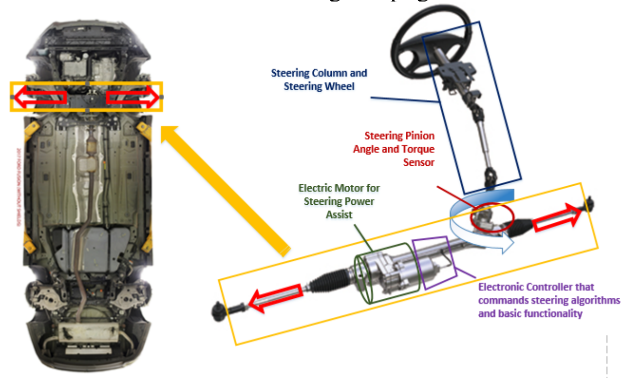

Fig. 1: A rack and pinion electric power steering system and its placement in the under body of a vehicle. The basic function of steering assist is to add gain to the drivers input torque and apply it to the wheels as a lateral force. This is accomplished by an electric motor that can be precisely controlled, a steering torque and angle sensor, and electronic control module to interpret signals, perform algorithms, and control the motor. Pull Drift Compensation (PDC) is an example of steering algorithms that improve driver experience. PDC adjusts the motor torque no driver torque in situation like driving on a road with crown or slant.

2) Anti-lock Braking Systems: Figure 1 describes the basic operation of braking system of a 2017 Ford Fusion and discusses its use in vehicle motion control systems.

3) Vehicle Body Motion: The vehicle body motion is best understood by understanding a 6 degree of freedom

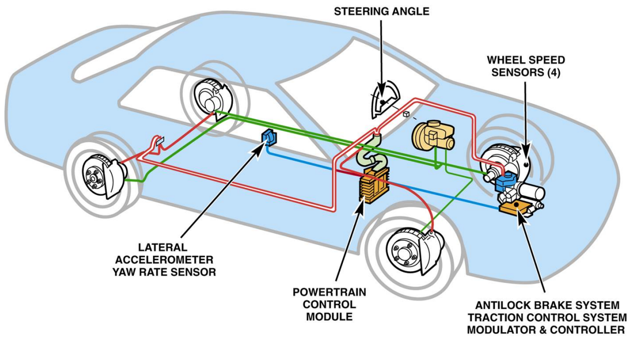

Fig. 2: Modern vehicle have anti-lock braking systems (ABS) that help to stop in conditions where the brakes lock the rotation of the tires, but the tires have broken loose from the road and slide without continuing to decelerate as required. Using a wheel speed sensor for each wheel the ABS can electronically actuate brake pressures to reduce or eliminate sliding. Braking pressure can also be described as torque on the road wheel. This torque is one the features in the dataset. Other sensors within the vehicle are shown in the diagram as ABS also contributes to traction and stability control algorithms employed by the vehicle to limit wheel slipping in uneven or low friction terrain and maintain vehicle yaw and roll within safe limits.

frame of reference attached to the body of a automobile, like the one shown in figure 3.

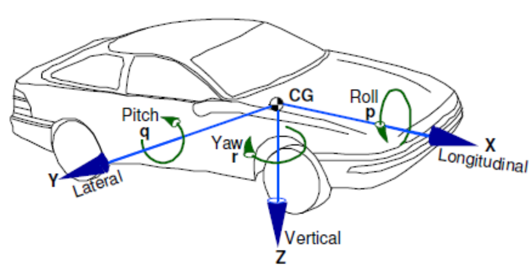

Fig. 3: SAE Vehicle Fixed Coordinate System with labels helps to explain the meaning of the Vehicle Body Motion Information signals. Gyroscopes and accelerometers within the vehicle measure these physical signals and are used in various vehicle motion control systems. All accelerations in the figure measured and two of the angular rates are captured in the dataset (roll and yaw).

C. Model Output

The output of interest is the classification alignment of road wheels relative to each other and the rest of the chassis. This machine learning problem is supervised and while the population is skewed the dataset is not. In general, there is relatively small population of vehicles with alignment problems. In our dataset the overall distribution of the output is shown in Figure 4, if we consider all the misalignment conditions together as one, the dataset will actually be skewed opposite of the true population.

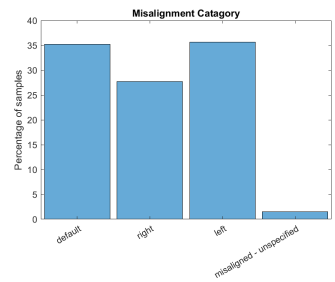

Fig. 4: Distribution various alignment conditions present in the data. ”default” (0): No known alignment issue, ”right” (1): Improper alignment with right bias, ”left” (2): Improper alignment with left bias, and ”misaligned” (3): Improper alignment with unknown bias.

In addition to a non-design state of the alignment, other noise factors have been included. These additional noise factors are tire condition: ”Nominal”, ”Bad”, and ”Conicity”, vehicle weight distribution: ”None”, ”GVWR Left”, ”GVWR Right”, and ”GVWR Centered”, and tire pressure. These noises, as well as environmental noises like sustained road crown, alter vehicle pull and steering wheel alignment will make the classification problem at hand more difficult. The overall fractions of the data with these other conditions present is shown in Figures 5 and 6. The percentage of samples with abnormal tire pressure is 37%.

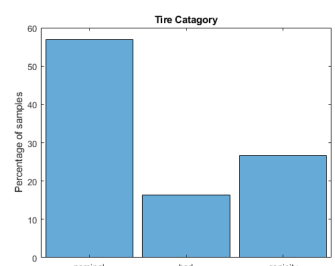

Fig. 5: Distribution various tire conditions present in the data.

1) Road Wheel Alignment: The most important parameters of alignment are the caster, camber, and toe of the right and left wheels and the total developed by their combination. Figure 7 - 9 have been provided for reference on these parameters. They can be set using a vehicle jig to precisely set them to their design. And they can be unset so to speak by suspension component wear or impact with harsh road surfaces.

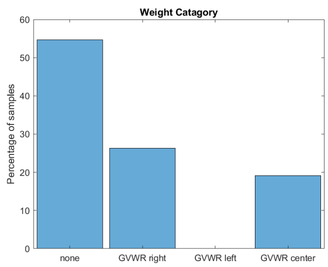

Fig. 6: Distribution of the various weighting conditions present in the data. No GVWR Left is present because the dataset has been reduced from the fully 366 files originally surveyed.

Alignment problems affect the ability of the vehicle to accelerate in a straight line, vehicle body motion reacting to driver inputs and road surfaces, or the steering performance in any number of ways. In general, alignment biased to the right or left creates steering torque pull or make the vehicle difficult to keep straight. The other noise factors, weight and tire condition can also affect the vehicle response to road and driver in similar ways.

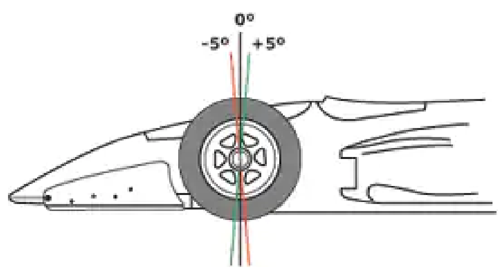

Fig. 7: Camber is the angle describing the inward or outward tilt of wheel from the vertical. Inward camber can help maintain heading under loss of traction, but an excess may decrease stability. Outward camber can reduce acceleration performance in a straight line. Ultimately camber is set to balance cornering performance and tire wear.

III. Exploratory Data Analysis

Some exploratory data analysis has already been completed. Some single valued signals in the dataset have been removed, and the original data files have been concatenated as to avoid having missing values in parts of the dataset. To improve memory consumption, the dataset has already been divided into categorical and numeric data by looking at the number of unique values contained in each signal. Signals are categorical if they contain less than 5 unique values. This delineation is shown in Appendix A. Our next step is to look deeper

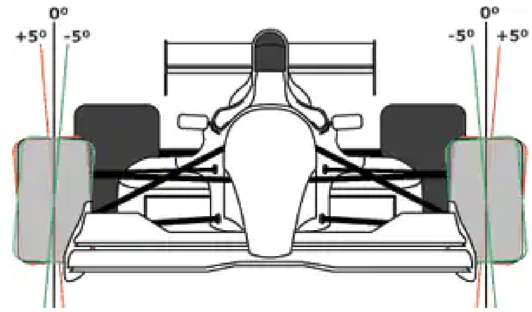

Fig. 8: Caster is the angle made with the vertical axis and the line between the steering's upper and lower pivot points. The choice of caster balances steering efforts, straight line tracking, and the ability to corner. Incorrect caster may effect any of these vehicle dynamics attributes. Conicity tires can mimic the effect that camber has, and tires can develop conicity from improper camber.

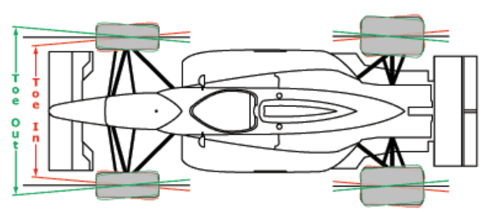

Fig. 9: Toe is the angle that either tire makes with the longitudinal direction. Toe has a great affect on under steer or over steer. Improper toe may negatively affect the ability to drive the vehicle in a straight line or corner and handling. Toe is the most critical factor for tire wear.

into the categorical and numeric parts of the dataset. We will review descriptive statistics for each signal and examine the correlation of the signals and determine what signals are important for predicting the output.

A. Numerical Feature Exploration

This dataset has 30 numerical features. My expectation is that the useful ones are directly related to: steering (torques and angles), vehicle motion (acceleration and gyroscope data and perhaps vehicle speed), and braking (braking torques and perhaps wheel speeds). I perceive steering as most useful. Steering torque is useful because the camber and caster aspects of alignment have direct effects on steering through torque pull which might be captured as an offset to steering column torque, pull drift compensation torque, any of the assist torque averages, or steering assist motor current. Similarly, improper toe leads to steering angle offset.

Secondary signals might also capture these effects. Vehicle motion is affected when there is misalignment - as we will see when we examine the correlation of the numeric signals. For instance, will the vehicle yaw rate signals be skewed to directional bias related to the steering pull? The increased pull in one direction would lead to the steering darting in that direction, something that should be measured by the yaw rate. A similar argument could be made with respect to signals related to roll. Lateral acceleration might be related to misalignment, but less so because it is sensitive to the the radius of a turn and small offsets one way might be washed out in noise around zero due to bumps or minor corrections. Because alignment fundamental changes how forces are distributed on the tire patch, total braking torque might increase under misaligned conditions. Less likely is that front wheel speed will shown any right to left difference. One might argue that improper toe could alter the curvature that that a tire travels along. Thereby making one tire rotate more over a distance and increasing its rotational speed. However, I think the change in distribution would be slight. However, [4] did use wheel speed and vehicle speed differences to judge tire wear.

The least useful signals will be related to throttle input, vehicle acceleration or longitudinal motion (”Accelerator Pedal Position Percent Rate”, ”Vehicle Longitudinal Acceleration”, ”Vehicle Longitudinal Acceleration Compensated”). Rear wheel speeds are not perceived useful as because we are not concerned with rear suspension misalignment (”Rear Left Wheel Speed” and ”Rear Right Wheel Speed”).

1) Numerical Feature Correlation: Figure 10 contains the full 30 by 30 correlation of the numeric signals. Many of the relationship that I've discussed can be seen. There are many correlations with high magnitude. Some of these represent natural relationships between different systems, such as how ”Total Braking Torque” is negatively correlated to to all the different measures of vehicle speeds. We can use the correlation here to judge if there is multicollinearity among any of the signals. There is a perfect linear relationship between wheel speeds and vehicle speeds. Without any feature engineering like that in [4], this relationship needs to be excluded by only including a single measure of vehicle speed. There are very high degrees of correlation between signals and their compensated counterparts. To reduce complexity, in a pair of a signal and it compensated counterpart, only one should be used.

2) Numeric Descriptive Statistics: The descriptive statistics for the numeric signals shown in Table I and can confirm some intuition about how vehicles operate and impose some questions on some of the signals. For example, ”Total Brake Torque” has 25, 50, and 75%tiles that are all 0, but its max value is 32 kNm. Quite a spread, but intuitively we know that most of the time a driver isn't braking, so the quartiles make sense, and sometime a driver has to slam on the brakes, so perhaps 32 kNm makes sense as a max value. ”Compensated Steering Wheel Angle” is expected to be distributed about zero, but the mean is near 6, and the 75%tile is above the absolute value 25%tile. Could this be evidence that there is a persistent steering wheel angle in some of the data? As for steering torque signals, it appears that ”Pull Drift Compensation Weighted Torque” and ”Steering Column Torque” both have distributions that skew negative, even though I would expect that they be symmetric.

B. Categorical Feature Exploration

A mapping of what all of the categorical signals are is not available to me, so consider the problem and how they might be useful in classifying the output for some of them. Others are more oblivious. In any case, there appear to be several signals involved with limiting the longitudinal acceleration, several that convey a particular system may have a fault (”Acceleration Torque Fault” or ”ABS Condition”), and several that demonstrate activation status of a vehicle feature (”ABS Active” or ”Traction Control System Percentage Activation”). Signals demonstrating feature activation are likely going to be the most useful, because they are responding to physical event, rather than a software state condition.

1) Categorical Feature Independence testing : Because there is not a correlation function for categorical data, we test whether or not the different categorical feature are independent or not with a Chi Squared test. For unique each pair of features [D_0, D_1] with categories[$C_0(m)$, $C_1(n)$] we tested a null and alternate hypothesis:

- H_0 : There is no statistically significant relationship between D_0 and D_1
- H_a : There is a statistically significant relationship between D_0 and D_1

The Chi-Squared test needs the following information:

$$\chi^2 = \frac{(observed - expected)}{expected} \tag{1}$$

Using *pd.crosstab()* we were able to calculate the frequency distributions of D_0 D_1. These were the observed values. Expected term is calculated by the *stats.chi2_contingency* function, but is easily written down as: Expected # elements of $C_0(m)$ in $C_1(n)$ =

$$\frac{\Sigma C_0(m) * \Sigma C_1(n)}{\Sigma D_0} \tag{2}$$

With this done, we can examine a heat-map that the p-values we calculated, Figure 11. When the p-value is

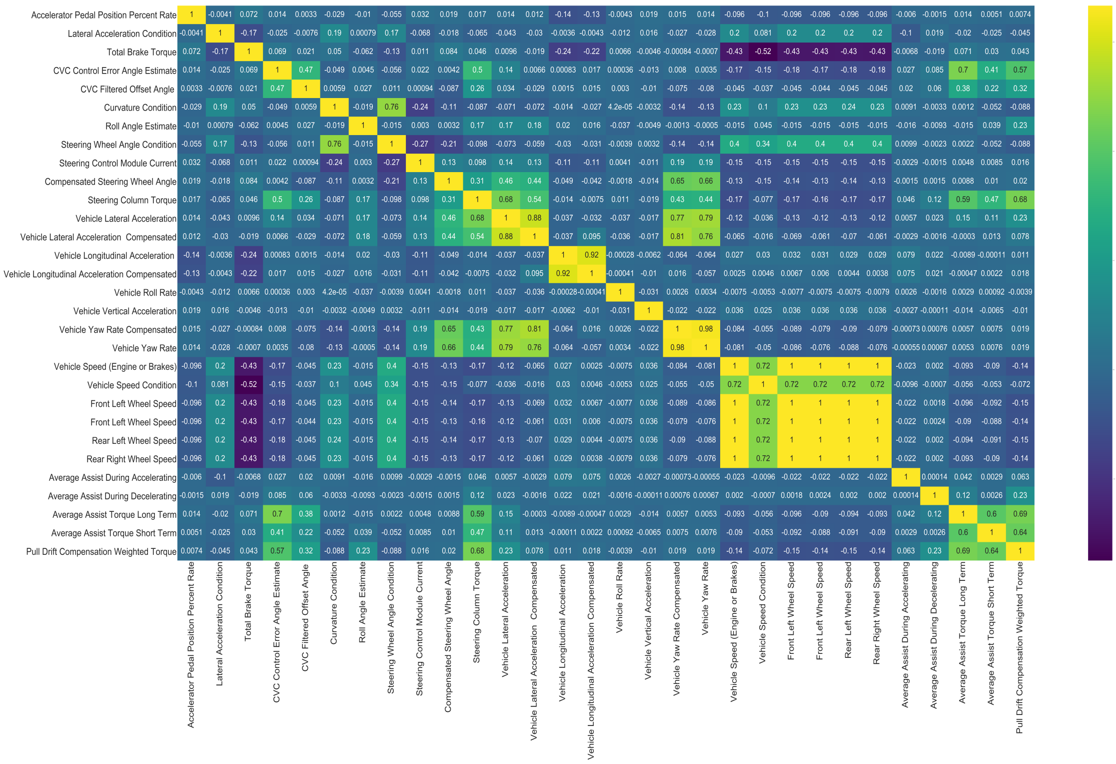

Fig. 10: Numerical feature correlation heatmap created using seaborns heatmap and pandas.corr functions. Yellow represents positive correlations and dark blue represents negative correlations.

	mean	std	min	25%	50%	75%	max
Accelerator Pedal Position Percent Rate	-0.004	0.018	-1.200	0.000	0.000	0.000	0.960
Lateral Acceleration Condition	0.957	0.195	0.000	1.000	1.000	1.000	1.000
Total Brake Torque	66.414	226.934	0.000	0.000	0.000	0.000	32760.000
CVC Control Error Angle Estimate	-4.310	3.988	-13.851	-8.526	-3.355	-0.359	4.564
CVC Filtered Offset Angle	-4.752	8.142	-481.023	-8.777	-5.019	-0.608	471.943
Curvature Condition	0.824	0.375	0.000	1.000	1.000	1.000	1.000
Roll Angle Estimate	-0.031	0.760	-23.043	-0.238	0.000	0.171	26.169
Steering Wheel Angle Condition	0.832	0.360	0.000	1.000	1.000	1.000	1.000
Steering Control Module Current	0.175	1.268	-14.500	0.000	0.000	0.050	67.350
Compensated Steering Wheel Angle	5.221	58.418	-492.800	-0.500	0.000	0.700	497.700
Steering Column Torque	-0.298	1.554	-8.000	-1.625	-0.375	1.000	7.750
Vehicle Lateral Acceleration	0.006	0.657	-7.290	-0.230	0.000	0.200	7.340
Vehicle Lateral Acceleration Compensated	0.005	0.673	-7.295	-0.155	-0.015	0.125	17.870
Vehicle Longitudinal Acceleration	-0.005	0.671	-7.880	-0.190	0.000	0.200	4.650
Vehicle Longitudinal Acceleration Compensated	-0.023	0.695	-8.625	-0.190	-0.015	0.160	17.870
Vehicle Roll Rate	-0.001	0.017	-0.908	-0.007	0.000	0.005	0.886
Vehicle Vertical Acceleration	9.755	0.387	0.000	9.600	9.750	9.900	21.340
Vehicle Yaw Rate Compensated	0.283	4.891	-51.850	-0.202	-0.018	0.165	74.963
Vehicle Yaw Rate	0.006	0.084	-0.904	-0.002	0.000	0.005	1.098
Vehicle Speed (Engine or Brakes)	69.155	40.239	0.000	38.480	68.530	106.280	138.920
Vehicle Speed Condition	0.823	0.371	0.000	1.000	1.000	1.000	1.000
Front Left Wheel Speed	59.348	34.567	0.000	33.000	58.840	91.240	119.200
Front Left Wheel Speed	59.378	34.519	0.000	33.040	58.840	91.240	119.640
Rear Left Wheel Speed	59.085	34.440	0.000	32.880	58.600	90.880	118.680
Rear Right Wheel Speed	59.141	34.421	0.000	32.920	58.640	90.920	118.880
Average Assist During Accelerating	0.003	0.144	-5.713	0.000	0.000	0.000	7.330
Average Assist During Decelerating	-0.003	0.518	-11.994	0.000	0.000	0.000	11.065
Average Assist Torque Long Term	-0.688	2.877	-10.490	-2.610	-0.564	1.671	12.313
Average Assist Torque Short Term	-0.178	1.610	-8.303	-0.097	0.000	0.000	7.631
Pull Drift Compensation Weighted Torque	-0.501	2.655	-10.000	-1.520	0.000	0.105	10.000

TABLE I: Descriptive statistics for numeric valued signals within the dataset, calculated with the help of Pandas *dataframe.describe()*.

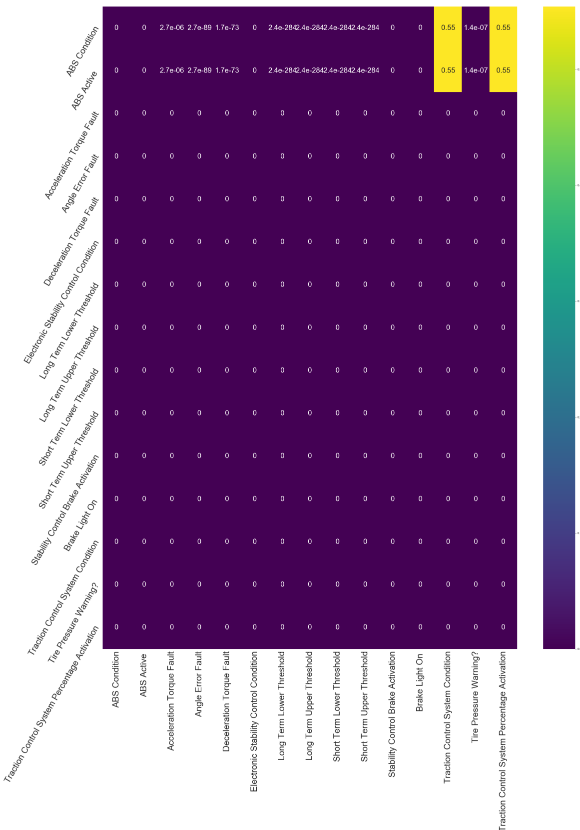

Fig. 11: Chi-Squared Independence test done for each pairwise configuration of the categorical signals. Annotations are the p-values, $p < 0.05$ means we can reject the hypothesis that the pair is unrelated. In the figure, $p > 0.05$ is more yellow or green.

below 0.05, we can reject the null hypothesis. Leading to the conclusion that few of the categorical data signals are unrelated because so many p-values in Figure 11 are small. This is important because the inclusion of many related categorical signals complicates the model at the expense of memory and computation time. This also validates the intuition that the most useful signals to include in a model will be "ABS Active" and "Traction Control System Percentage Activation" as they are the only signals for which we accept the null hypothesis.

IV. Proposal Description

The task is to classify the binary alignment response y using the features x.

$$y \in [0, 1] \tag{3}$$

$y = 1 \equiv$ {right, "left", 'misaligned - "unspecified"} and $y = 0 \equiv$ {"default"}. Lumping each of the misalignment categories together will create a skew dataset with more examples of misalignment than nominal alignment. Each x_j is labeled with y_i, so this is a supervised learning problem. The questions associated with task are:

1) What combination of supervised classification machine learning techniques will effectively predict misalignment in the presence of noise factors?
2) How will model performance be judged?
3) What is the minimum amount of data that can be used in training?
4) What data collection rate would be required to support classification using the minimum amount of data for the customer to see benefit?

A. Machine Learning Techniques

Binomial classification of supervised data can be done with increasing complexity using Naive Bayes classifiers, Logistic Regression, Support Vector Machines, and Neural Nets. [6] was effective at using these methods and majority voting to predict complex power train errors from vehicle communication better than individual methods on their own. Its likely that multiple methods will have to be combined to predict effectively.

1) Naive Bayes Classifier Definition: In Naive Bayes a binomial classification from the posterior probability can be made using knowledge of the likelihood, the prior probability, and the evidence from Bayes formula.

$$P(y = c_1|x) = \frac{p(x|y = c_1)P(y = c_1)}{p(x)} \tag{4}$$

The prior is simply the frequency of $y = c_1$, $p(x)$ is the evidence, $p(x|y = c_i)$ is the likelihood, and $P(y = c_1|x)$ is the posterior. The likelihood is given by a Bernoulli distribution, $Ber(x|\theta_{i,j})$, where $\theta_{i,j}$ is the probability that x_j is class c_i. The classifier decides that x belongs in $y = c_1$ or $y = c_0$ if x lies on either "side" of an equal probability boundary formed by $P(y = c_1|x)$ = $P(y = c_0|x)$.

2) Logistic Regression Definition: Logistic regression maps the hypothesis from a linear regression, $h(x) = w^{\intercal}x$, between a the values of 0 and 1. Where w are the weights found minimizing the residual sum of squares. There are different flavors of finding w, w can be found with ordinary least squares (no regularization), or it can be found with l_2 either l_1 regularization or a combination there of. The hypothesis $h(x)$ is passed through the sigmoid function to reduce its range between 0 and 1. Because this method requires minimization of some kind the number of iterations before a minimum is picked is a hyper parameter of this model. As are any basis functions used to transform the data for $h(x)$ or $h(\phi)$, and the degree of either l_1 or l_2 regularization.

3) Other methods: Other methods such as support vector machines and neural nets will be evaluated if there is time. Because we have not covered these topics in class I do not wish to state their mathematical models or the hyper parameters that are to be selected. Because of the success in the literature using methods beyond linear regression and Naive Bayes, I hope I can use them. I would also like to learn more about model combination methods.

B. Performance Evaluation

Because this is a classification problem, typical counting methods that go into precision and recall can be used and represented in a confusion matrix. F-score can also be used. It will also be important to specifically understand how the algorithm performs under noise conditions, it may be helpful in this regards to consider type I and type II error given the other labels that exist in the dataset. It might be helpful to understand what other labels type I and II errors have. This information will be helpful in quantifying how well the algorithm performs under a superposition of conditions.

C. Minimum Data for Model Training

Classifying a particular x value has misalignment can add value if can be done soon enough to notify a customer that their vehicle has some abnormal condition before wear occurs. With the adoption of wider bandwidth communication networks in the vehicle and the inclusion of 4G modems in most vehicles it will be increasingly possible to monitor vehicle health from afar. If an algorithm could be developed to use incremental data from a customer vehicle to predict if it had an underlying alignment error, it could save customer headache and the OEM warranty cost. Understanding how much data is needed to accurately (greater than 90% or so) predict would be key in learning sampling vehicle data telemetry to accomplishing this function. I believe this information can be learned by experimenting with the the test train division and with how the cross validation is performed.

References

[1] S. Carolina, "Michelin Explores The Truth About Worn Tires," no. C, pp. 1–6, 2019.

[2] M. Meutey Corinne and G. M. Nick Richards, "PRESS INFORMATION MICHELIN , GM TAKE THE AIR OUT OF TIRES FOR PASSENGER VEHICLES," 2019.

[3] I. A. Badiru, "Assessment of the Capability of EPS to Reduce Steering Wheel Pull and Vehicle Misalignment," no. 1, pp. 624–629, 2015.

[4] T. Poloni and J. Lu, "An Indirect Tire Health Monitoring System Using On-board Motion Sensors," *SAE Technical Paper Series*, vol. 1, 2017.

[5] K. Vasudevan, A. P. Das, B. Sandhya, and P. Subith, "Driver drowsiness monitoring by learning vehicle telemetry data," *Proceedings - 2017 10th International Conference on Human System Interactions, HSI 2017*, pp. 270–276, 2017.

[6] Y. Sun, Z. Xu, and T. Zhang, "On-Board Predictive Maintenance with Machine Learning," *SAE Technical Paper Series*, vol. 1, pp. 1–10, 2019.

	Time Series Name	Units	Data Type
1	ABS Condition	None	Catagorical
2	ABS Active	None	Catagorical
3	Acceleration Lower Threshold	Nm?	Catagorical
4	Acceleration Upper Threshold	Nm?	Catagorical
5	Acceleration Torque Fault	None	Catagorical
6	Angle Error Fault	None	Catagorical
7	Accelerator Pedal Position Percent Rate	%/ms	Numerical
8	Lateral Acceleration Condition	None	Numerical
9	Total Brake Torque	Nm	Numerical
10	CVC Control Error Angle Estimate	Degrees	Numerical
11	CVC Filtered Offset Angle	Degrees	Numerical
12	Curvature Condition	None	Numerical
13	Deceleration Lower Threshold	Nm?	Catigorical
14	Deceleration Upper Threshold	Nm?	Catigorical
15	Deceleration Torque Fault	None	Catigorical
16	Electronic Stability Control Condition	None	Catigorical
17	Long Term Lower Threshold	Nm?	Catigorical
18	Long Term Upper Threshold	Nm?	Catigorical
19	Roll Angle Estimate	Degrees	Numerical
20	Steering Wheel Angle Condition	None	Catigorical
21	Steering Wheel Rate Condition	None	Catigorical
22	Short Term Lower Threshold	Nm?	Catigorical
23	Short Term Upper Threshold	Nm?	Catigorical
24	Stability Control Brake Activation	None	Catigorical
25	Steering Control Module Current	Amps	Numerical
26	Compensated Steering Wheel Angle	Degrees	Numerical
27	Steering Column Torque	Nm	Numerical
28	Brake Light On	None	Catigorical
29	Traction Control System Condition	None	Catigorical
30	Tire Pressure Warning?	None	Catigorical
31	Traction Control System Percentage Activation	%	Catigorical
32	Vehicle Lateral Acceleration	m/s^2	Numerical
33	Vehicle Lateral Acceleration Compensated	m/s^2	Numerical
34	Vehicle Longitudinal Acceleration	m/s^2	Numerical
35	Vehicle Longitudinal Acceleration Compensated	m/s^2	Numerical
36	Vehicle Roll Rate	rad/s	Numerical
37	Vehicle Vertical Acceleration	m/s^2	Numerical
38	Vehicle Yaw Rate Compensated	rad/s	Numerical
39	Vehicle Yaw Rate	rad/s	Numerical
40	Vehicle Speed (Engine or Brakes)	kph	Numerical
41	Vehicle Speed Condition	None	Catigorical
42	Front Left Wheel Speed	rad/s	Numerical
43	Front Left Wheel Speed	rad/s	Numerical
44	Rear Left Wheel Speed	rad/s	Numerical
45	Rear Right Wheel Speed	rad/s	Numerical
46	Average Assist During Accelerating	Nm	Numerical
47	Average Assist During Decelerating	Nm	Numerical
48	Average Assist Torque Long Term	Nm	Numerical
49	Average Assist Torque Short Term	Nm	Numerical
50	Pull Drift Compensation Weighted Torque	Nm	Numerical

TABLE II: All signals available in the dataset after processing and removing signals were not present in the widest breadth of the individual data files. Signals have been given colloquial names and units have been provided where possible. Not sure/? entries have been labeled to express uncertainty about the units or signal type.

Appendix A
Appendix A: Full Dataset