FIRST ANALYTICS®

Analysis of the Energy Reduction Benefits of MPL NatureBlend Locomotive Wheel Flange Lubricant

NOVEMBER 2020

FIRST ANALYTICS®

Contents

FIRST ANALYTICS®

Introduction

NatureBlend™ is an environmentally friendly solid polymer friction modifier to help reduce mechanical and energy forces. The product is offered to railroads by MPL Innovations, Inc. as a locomotive wheel flange lubricant.

Suggested benefits include fuel savings by reducing wheel/rail friction, decreasing flange wear and wheel turning, and increasing rail and wheel life. This study focuses specifically on the energy reduction (and thus, fuel) benefits.

The data for this study comes from two separate tests conducted at the Association of American Railroads' Transportation Technology Center (TTCI) in Pueblo Colorado. Tests were conducted on different track loops in 2012 and 2014. Each test procedure consisted of baseline, or "dry" laps, as well as "lube" laps where the NatureBlend™ formulation was applied.

The objective of this study is to quantify the difference between the lubricated versus dry conditions. As will be described in the methodology section, a statistical analysis brings rigor into the study in order to (1) provide a formal statistical test as to whether there is any effect of the claimed energy benefits and (2) if there is a benefit, provide a statistically determined estimate of its likely range (verses a point estimate).

FIRST ANALYTICS®

Executive Summary of Findings

Though this section of the report is a summary of findings, it is recommended that the detail sections regarding data, methodology, and detailed results, be reviewed. In particular, there is discussion about the strength of a pooled analysis combining two TTCI tests, and the power that statistical tools bring to support the findings.

Key Findings

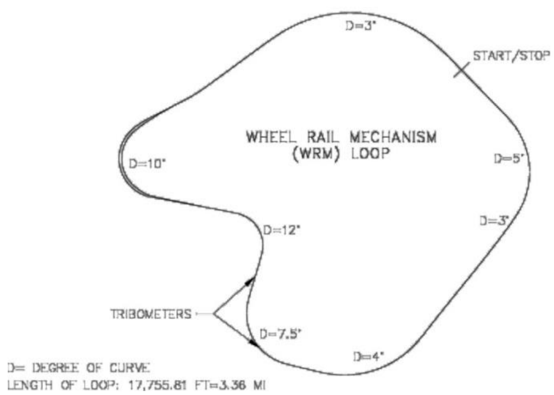
- The impact on energy savings of NatureBlend™ is estimated to be 3.2%.
- Two separate statistical tests were performed to evaluate the hypothesis that NatureBlend™ has no measurable effect on energy consumption. Both tests rejected this hypothesis at the 99% confidence level. In other words, we are 99% confident that NatureBlend™ has a real impact on energy consumption.
- The model provides a range of possible effect outcomes, providing a more enhanced view of the effect than a simple point estimate. The model suggests we should be 90% confident that the true effect is between 1.8% and 4.3% reduced energy consumption when using NatureBlend™.
- To provide a secondary view in addition to direct energy measurements, throttle position profiles were examined.
    - Runs where NatureBlend™ was being applied spent 7.1% less time in T8.
    - More time was spent in T4 and T5.
- Again, two separate statistical tests were conducted to evaluate the hypothesis that there is no difference in throttle position due to NatureBlend™. These tests rejected the hypothesis at the 90.1% and 90.0% level, respectively, indicating that railroads will experience a change in throttling.
- As a second view into throttling, a separate statistical model estimates that the use of NatureBlend™ reduces the odds of being in T8 by 4.8%. This estimate is significant (non-zero) at the 87.1% level.

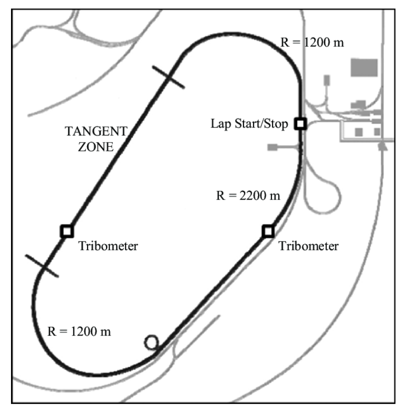
FIRST ANALYTICS®

TTCI Tests and Data

The Tests

Tests were conducted at TTCI in 2012 on the Wheel Rail Mechanism (WRM) loop and in 2014 on the Transit Test Track (TTT). These tracks have different profiles with respect to curves, and to some extent, elevation changes. Figure 1 below depicts the tracks and aspects of the tests.





2012 Wheel Rail Mechanism Loop (WRM)

- Target Speed: 30 MPH
- Consist: Two locomotives and thirty hopper cars
- Trailing Tonnage: 4696
- 3.5 mile loop
- 12, 10, 7.5, 5, 4, and two 3 degree curves

2014 Transit Test Track (TTT)

- Target Speed: 50 MPH
- Consist: Two locomotives and thirty hopper cars
- Trailing Tonnage: 4625
- 9.1 mile oval loop
- Mostly tangent rail, with one 0.5 degree and two 0.5 degree curves

Figure 1: TTCI Tests

Data for Analysis

The locomotives were equipped with a data collection system to measure volts and amps. From those measurements, electrical and mechanical energy was computed and converted to kilowatt hours (KWHR) which is our target metric for estimating energy savings. A GPS system provided the location (latitude and longitude) of the locomotive on the track. This allows us to augment the core energy data with:

- **Train speed**. Speed does vary slightly from the target speed, especially on the WRM which has more curves and grade changes.
- **Heading**. This allows us to examine potential correlations of curvature with energy consumption.
- **Elevation**. Latitude and Longitude information was used to obtain elevation information from government sources, allowing us to examine potential correlations of grade with energy consumption.
- **Mileage Traveled**. Though this may not be a modeled measure, it does help us in the computation of measures such as speed and grade change.

Measurements are at the one-second level (one-hertz data). Information about the lap number and the lubricating condition (lube or dry) is provided. Some filtering was applied to exclude records (laps or portions of laps), that were designated for "warm-up" or other calibration reasons, or to account for the fact that the train was accelerating or decelerating to or from the target speed. The goal in scoping the data was to have a crisp, contrasting picture of the lubricating versus the dry condition.

The data was harmonized and combined across the two tests with consistency in the way the measures were computed. Table 1 provides a summary of the two data sources.

Table 1- Data Included in the Analysis

Test	WRM 2012	TTT 2014
Number of Records Used for Analysis	5,307	15,329
Dry Laps	4	10
Lubricating Laps	9	12

Data Quality

Good statistical practice dictates that extensive time be spent checking the data for anomalies and outliers. Virtually no data source is perfect and that is true of measurement systems such as the one used at TTCI. Among the issues discovered include:

- At times, the GPS system would cease recording. We were provided with a GPS status flag to identify these records. In these cases, we had energy readings but incorrect measures on speed (e.g. speed = 0 with positive energy), heading, elevation, etc.

Additional checks were performed, looking for these conditions independent of the GPS status flag. In all, only 50 records were affected.

- Rarely, the energy management system would reset. Since energy consumption is computed as the differential from the previous measurement, care was taken to make sure that the energy reading was from the measuring equipment consistently measuring (e.g. no negative energy due to a reset)
- Occasionally, there were multi-second gaps (the longest being 22 seconds) between readings. The records immediately following these gaps were eliminated.

FIRST ANALYTICS®

Methodology

Statistical Tools

It is nearly universally common in these kinds of studies to arrive at conclusion based on a comparison of averages, or means, between a test condition and a “control” condition (e.g. the base laps). While our study also looks at the means, we apply formal statistical tools to bring more precision and confidence to the results. Specifically, we apply tools such as Analysis of Covariance (ANCOVA), tests-of-hypotheses, confidence levels, and regression modeling (including mixed effects empirical Bayesian regression).

While studying raw means can be productive, there are potential pitfalls that can be mitigated using such statistical tools.

- Looking at raw means alone, in a univariate sense, does not account for other ancillary factors that may also be driving the difference of the means. Statistical modeling using multivariate methods allow us to quantify those ancillary factors and adjust the means accordingly to have a more balanced comparison.
- Means are a single number – a point estimate. Statistical modeling can provide an estimate of the range of potential outcomes.
- Statistical models can assign a degree of confidence that the true effect (energy savings) is non-zero.

Our firm, First Analytics, has studied fuel conservation technologies for a number of years. In 2016 we performed a meta-analysis of 29 reports of mobile top-of-rail friction modifying technologies. The studies were conducted by railroads, a technology vendor (not MPL), TTCI, and First Analytics. We found a wide variation in reported results, ranging from 0% to a couple of extreme cases of over 10% savings. It was clear that this wide variation was due to examining means alone, and not accounting for confounding factors.

Data Pooling

Rather than look at the WRM and TTT tests individually and in isolation, we pool the data. Even though there is variation in the conditions of the study (e.g. track profiles and speed), it is this kind of variation that aids statistical modeling in more fully understanding the effects of the phenomenon. Additionally, more data points and observations can lend strength to the analysis. Thus, the upcoming results consolidate information from both tests into one answer.

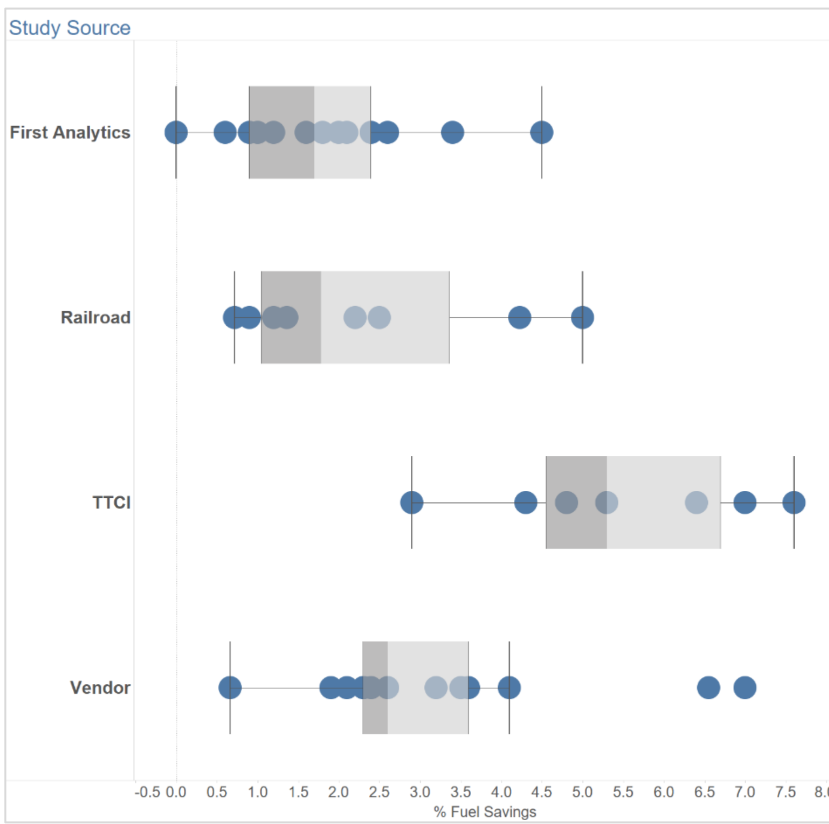
Projectability

We recognize that track profiles of the WRM and TTT test tracks differ from the usual network profiles of each railroad. We also recognize that the test train consist profiles differ from "the real world". It can be challenging to project the results from controlled test to every unique situation. For this our primary focus is to provide statistical support for a positive energy benefit, if it exists, and to provide a perspective on the range of what the savings benefit may be.

Another aspect of projectability is the consideration that TTCI tests are conducted in a highly controlled environment and may overstate the effects. In the aforementioned meta-analysis of 29 studies of fuel technologies, we did indeed find that TTCI studies (vs. in-field) provided estimates that are higher than real-world studies. The average of TTCI vs. other studies was 3.3% versus 2.8%. While one might suggest that the result in this particular report may thus be discounted by 15.2% (2.8 / 3.3), we are not positing that this is necessary in this particular case – only that if such adjustment were argued, that we could provide one reference data point to inform that adjustment.

Figure two shows the spread of results in those 29 studies, broken out by the four study sources. Each dot represents the estimated savings from the study. The shaded areas represent the interquartile range (those studies falling within 25% and 75% of all studies. The vertical lines near the extreme ends represent 1.5 times the interquartile range.

Figure 2- Results from Multiple Fuel Savings Studies



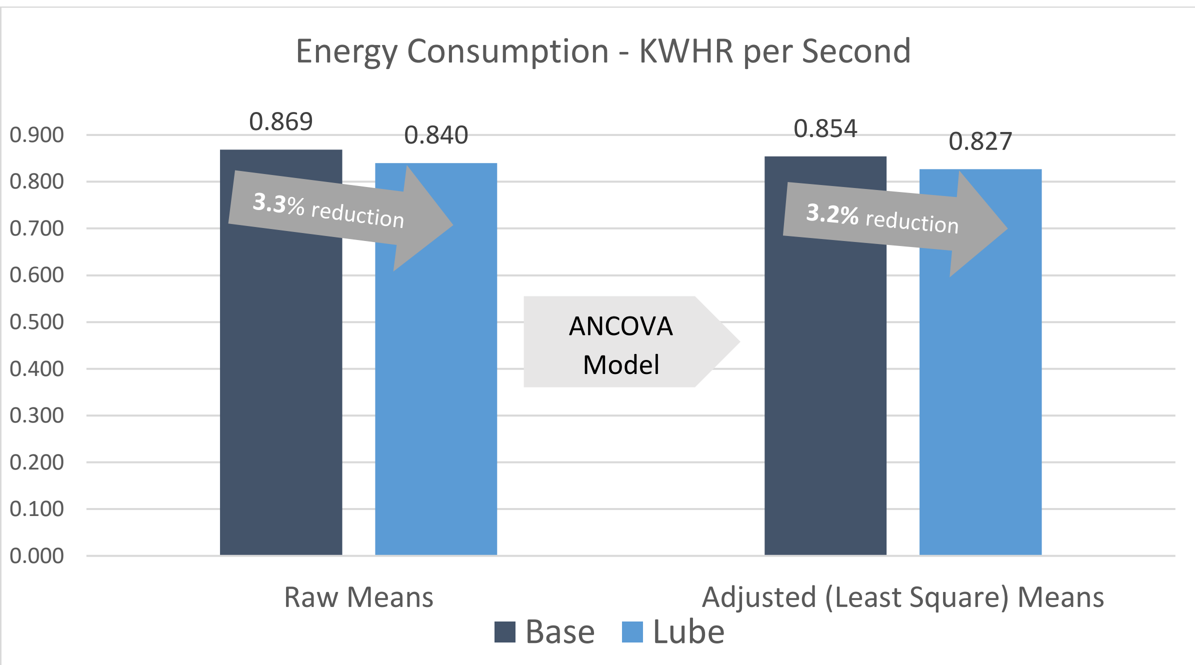
FIRST ANALYTICS®

Analysis

Overall Energy Savings

Overall energy savings of the NatureBlend™ lubricant, in kilowatt hours, is estimated to be **3.2%**. This represents a statistically adjusted estimate of the percent savings difference between the lubrication and base runs of the combined TTCI tests[1]. The results are *statistically adjusted* in the sense described in the methodology section. Analysis of Covariance (ANCOVA)[2] methods quantify potential other causal factors, such as speed, curvature, and elevation changes, to balance the comparison and isolate the effect due solely to the lubricant[3]. The raw mean difference between the groups was 3.3%. The fact that this was a very modest adjustment indicates that the tests were fairly well-executed to minimize any differences that might mistakenly be attributed to the lubricant.

Figure 3- Comparison of Means, Base vs. Lube



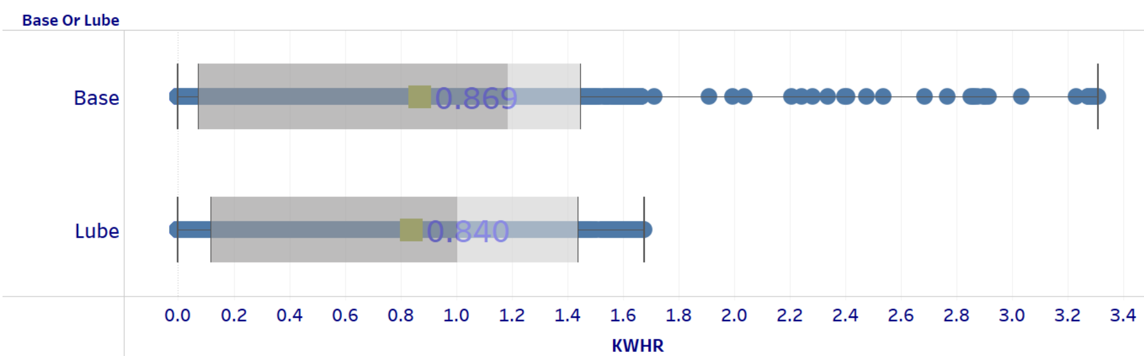
[1] These are known as the "least square means"
[2] Wikipedia article on "Analysis of Variance": https://en.wikipedia.org/wiki/Analysis_of_variance
[3] The model is a linear mixed effects model with KWHR being the dependent variable, the test condition (base or lube) as the main independent variable, and covariates for curvature, elevation change, and test (accounting for different speed targets). Curvature and elevation variables were allowed to vary across tests (a random coefficient shrinkage estimator model)

FIRST ANALYTICS®

Because of the very small adjustments to the means, we can, in fact, look at the dispersion of the raw, unadjusted data, to get some insight into the differences. Figure 4 is a box plot showing the spread of the measurements for the base versus lube conditions. For interpretation of this kind of chart, see the methodology section, figure 2. In the case of this visualization, the green box represents the mean, with the mean KWHR for Base being slightly higher than that for Lube.

We can also see that for the Base condition there are several measurements of high energy expenditure (the points above 2.0). While one may label these as outliers and conjecture that these alone are driving the difference in means, in reality, these represent a small number of observations among 8681 total observations for the Base runs. Furthermore, the ANCOVA approach allows us to study this variation as a whole in a way that accounts for these kinds of occurrences.

Figure 4- Distribution of KWHR Measurements



Condition	Number of Observations	Mean	Standard Deviation
Base	8,681	0.869	0.659
Lube	11955	0.840	0.636

FIRST ANALYTICS®

Statistical Tests of Significance and Confidence Intervals

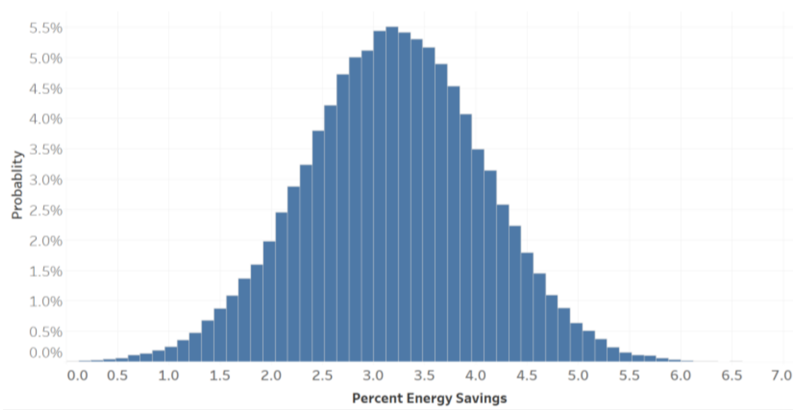
An output of the ANCOVA model is a statistical estimate of the hypothesis that the effect of the lubricant is different from zero[4]. In statistics parlance, this is the so-called "null hypothesis" – that of no difference between two measured phenomena[5].

Our formal statistical test indicates that we can *reject* the null hypothesis of no effect and can do so at the greater than 99% confidence level (t-value of 3.56). That is to say, there is an extremely small chance that there is no effect of NatureBlend™

A second, but less powerful test of differences is the so-called *two-sample t-Test*[6]. It is less powerful than ANCOVA in that it is a univariate approach comparing the simple means and variances, without accounting for other factors, such as those described previously, that could influence those means and variances. Again, the null hypothesis of no difference between the means was rejected at the greater than 99% confidence level (t-value of 3.14 using the Satterthwaite test of unequal variances).

In addition to these tests, we can use the same model to provide an estimate of the *range* of possible effects, versus a single point estimate such as the mean. This "histogram"[7] provides an extra dimension to our understanding of the effect. Figure 5 shows how the model simulates the spread of the possible effect, where the height bars indicate the relative probability of the effect being in the range shown in the horizontal axis. In other words, if we were to conduct the tests many thousands of times, the effects would fall into these "buckets." It is worth noting that the histogram does not cross zero. Additionally, we can say that 90% of the outcomes would fall within 1.8% and 4.3%. In other words, we are 90% confident that the *true* effect is within that range.

Figure 5- Range of Estimated Savings



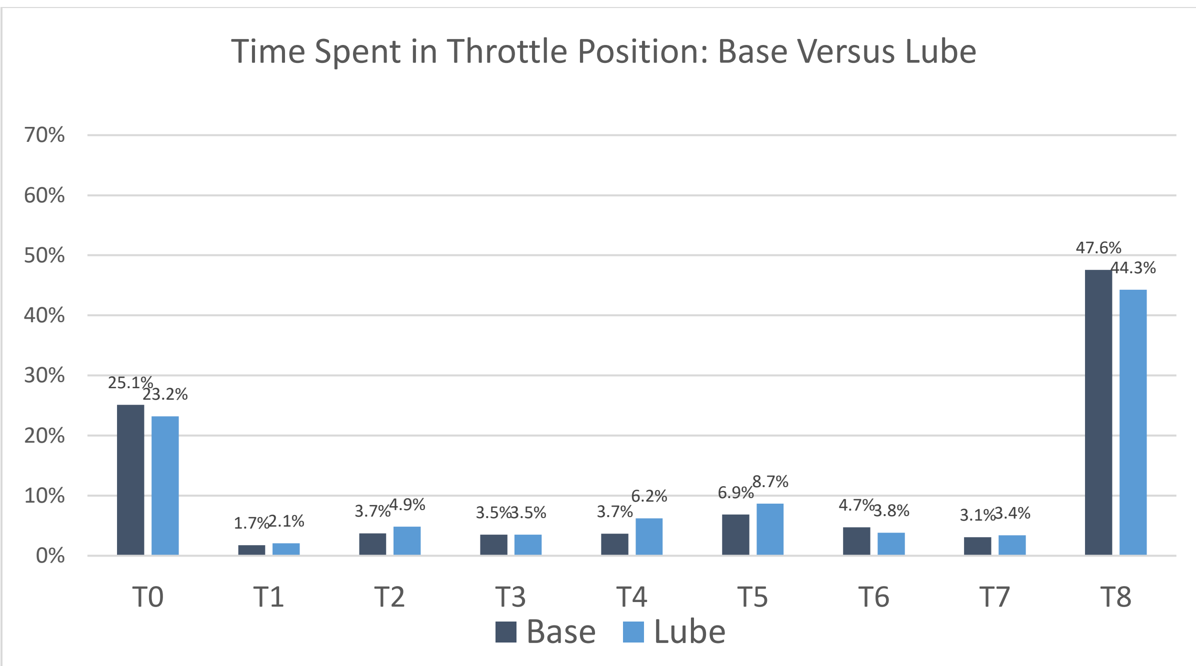
[4] Wikipedia article on "Hypothesis Testing": https://en.wikipedia.org/wiki/Statistical_hypothesis_testing
[5] Wikipedia article on "Null Hypothesis": https://en.wikipedia.org/wiki/Null_hypothesis
[6] Wikipedia article on "Student's t-test": https://en.wikipedia.org/wiki/Student%27s_t-test
[7] Wikipedia article on "Histogram": https://en.wikipedia.org/wiki/Histogram

Impact on Throttle Position

Throttle position is a proxy for energy consumption. Less time spent in higher notches is an indication of reduced energy. In our data we have an estimate of the throttle position, however, this is not the actual throttle position recorded by the locomotive. So, any analysis on throttle position will have a small amount of uncertainty associated with it.

Figure 6 shows the breakout of time spent in each throttle position, for each condition, base versus lube. Note that this was a controlled test, with the goal of maintaining target speed, so this throttle profile may not be similar to real-world profiles. But we can still use this data in a comparative sense. Additionally, consideration weight should be given to the impact in the highest notch, T8.

Figure 6 - Time Spent in Throttle Position



Visually, one can see that overall, more time is spent in some of the lower notches for the lube condition, particularly, in T4 and T5. This appears to be on offset from what normally would be T8. In fact, on a raw basis, the lube runs are operating in T8 44.3% of the time, versus the base runs with 47.6%

One might be tempted to say that "there is a 7.1% reduction (44.3 / 47.6) of time spent in T8." That may very well be true, but again we shall bring some statistical tools to the analysis to

(1) determine if there is a true difference in the throttle profiles and (2) whether we can provide an adjusted estimate, isolating it from other confounding factors.

We performed two statistical tests on the hypothesis that there is no difference between the lube and base condition. The first test, called a Mantel–Haenszel Chi-Square test[8], is specifically designed to compare two variables with responses grouped into categories. This is what we have in our throttle profile data; the test examines whether there is a statistical difference in the time spent in each throttle position as a group of categories. This test rejected the hypothesis of no difference at the 90.1% level (M-H Chi-Square statistic value of 2.73).

A second, less powerful test was a simple t-Test on the mean throttle position between groups. Very similar to the first test, the hypothesis of no difference was rejected at the 90.0% level (t-value of 1.65 using the Satterthwaite test of unequal variances). Both of these tests indicate that there is a difference in throttle operations due to the lubricant.

Similar to the ANCOVA approach for statistically adjusting the estimated percent savings, we built a model for time spent in position T8. This model differed, however, from the model used on energy (KWHR). In this case, we treated T8 in each record as an indicator flag (Yes/No). The nature of this binary data requires a variation of a linear regression model called a Logistic Regression[9]. These models implicitly assume binary responses, and calculate the impact of potential causal variables, such as being in a base run or lube run, as a probability of increasing the odds of being in the "Yes" (or "No") condition. Together with other factors such as curvature and elevation change, the model can isolate the impact of the variable in question.

The model estimates that the use of NatureBlend™ reduces the odds of being in T8 by 4.8%. Although not the same metric, one might compare this to the 7.1% difference in time spent in T8 (see above). This 4.8% takes into account some of the differences between runs in factors that are not associated with the lubricant, and rebalances the results based on those factors, so that the comparisons can be made on an equal footing.

This 4.8% (called the "odds ratio") is statistically significant (non-zero) at the 87.1% confidence level (Wald Chi-Square statistic of 2.31). In other words, we again reject the hypothesis that NatureBlend™ has no reductive impact on throttle position.

[8] Wikipedia article on "Cochran–Mantel–Haenszel statistics": https://en.wikipedia.org/wiki/Cochran%E2%80%93Mantel%E2%80%93Haenszel_statistics

[9] Wikipedia article on "Logistic Regression: https://en.wikipedia.org/wiki/Logistic_regression