LOCOMOTIVE RESISTANCE FORMULAE - 4th July 2017

Reply by John Knowles to Letter from Doug Landau of 7th March

This is the first stage of my reply to Doug Landau's letter of 7th March. As usual Doug's criticisms are laced with at least as many insults as science, plus in this case calling on several great men most of whom had nothing to do with the subject of the Rugby test plant or LR. In addition he calls on repeatability as a criterion for acceptability or accuracy of data, when all the repeated data can all be wrong. The matters he presents require a great deal of answering. I intend to do that in three parts – first, here, (i) the accuracy of data, statistics and regression, and the form of argument he has adopted, (ii) the great men, and (iii) other matters, including the Rugby plant.

A list of abbreviations used is given at the end.

1 What I am accused of and Regression Analysis

In his final paragraph, he says:

In summary the supposed shortcomings of the Rugby Test plant, its designers and operators are groundless. The available experimental data demonstrates consistent repeatability over time and circumstance. Repeatability is a key indicator of metrological integrity. That is not to say everything is perfect and falls in place in place like a jig saw. Given the understood limits of experimental error, however small, and the random nature of scatter, the real world is more complicated. Exactly the same problems obtain when reconciling the data from road tests. Road tests have however confirmed the differences in test plant MR in the case of the Crosti and standard 9Fs. In other words the empirical evidence derived by different methods remains consistent. A key test of scientific proof is that its claims are consistent with the empirical evidence. The powers of the regression statistical process used by John Knowles fails the empirical test significantly and is thus unsound, supposed statistical integrity notwithstanding.

He has not shown any of his claims made in this conclusion, ie the conclusions come out of the air unsupported by the content of the paper. He has not shown anything to be wrong with regression, and what criterion he has employed to reach his astonishing conclusion about it. He does not appreciate that repeatability is an insufficient criterion for acceptability of experimental data – the repeated data can be all wrong. He does not show repeatability to exist in the Rugby data – I find precious little of it. He gives no reference for the claimed confirmation of TSR by road tests for the Crosti and standard 9Fs, nor explained how he reconcilied what are essentially different measurements – TSR given on the test plant and LR on the road. Given the lack of repeatability in the Rugby data, he does not say which 9F data among the non-repeating 9F data he picked for his own use as the resistance of the 9Fs. The doubts about the test station results are far from groundless, his assertion notwithstanding.

I have answered much the same points in my previous letters on the Society webpage on this subject. As he pronounces further on the subject with no more evidence of knowing much about scientific analysis, and in particular about testing data and regression, there will be repetition in this reply.

He has not explained what he means by his statement that is not to say everything is perfect and falls in place in place like a jig saw, and that given the understood limits of experimental error, however small, and the random nature of scatter, the real world is more complicated. It is all very well to claim there is scatter in data, that is random and that it cannot be avoided, but scatter is lack of repeatability, and its extent and pattern gives the probability of the data yielding sound results. Indeed, what appears to be scatter could be "good" in revealing important aspects of behaviour, which were not previously appreciated. Randomness, in the sense of absence of bias, is an essential feature in experimentation and in analysis of data.

Does he mean that if the data do not fit precisely what he is looking for, the random scatter has to be treated in some way to make it amenable? That is precisely where statistics, as a science accepted by millions of practitioners worldwide, has its place. Simply drawing a line through data, or fitting an equation to data by trial and error, with a self-chosen criterion of acceptability of the relationship implied by the line is no proof that accuracy or acceptability of data has been established, quite the contrary. Further, where there are two or more determining variables, or the relationship posited is complex (eg it changes over the range of the data, or there is variation with powers, including fractional powers, in one or more of the determining variables, it is impossible to fit a relationship to data without regression. The supposed deficiencies of regression are mostly the result of Doug Landau's lack of knowledge of the process and what it can achieve. He is decrying regression because it can show deficiencies in data and/or methods and/or relationships which he wants to claim are satisfactory, that the Rugby data in his hands can be declared to be satisfactory, and is declaring often, apparently in the hope that if the declarations are made often enough, they will eventually be accepted, especially if he can deprecate my explanations and remarks sufficiently. I say that because he has done nothing to show the data to be satisfactory. As for deprecating, see the net paragraph also.

Whatever is the basis of his claim that the powers of the regression statistical process I used fails the empirical test significantly and is thus unsound, supposed statistical integrity notwithstanding? This conclusion is not even discussed, ie he gives no basis for it. The conclusions are therefore not based on a scientific approach or discussion. There is no reference to the small difference problem (SDP). Nor any appreciation that data can exist but can be not good enough for any sound result to emerge; or that any analysis or conclusions require testing the data, choosing the right form of analysis, ie the right form of equation, and applying well known and easily available tests of the probability of the results being acceptable. In other words, the nearness to fitting the jigsaw or some other criterion says whether the data really say anything worthwhile.

Conclusions of a paper follow from its content. In this case they do not. Doug Landau's supposed conclusions do not follow from the content. These are broad statements of his beliefs not supported by the content of the paper, and without any references to other literature which do support them. His approach amounts to false argumentation, false accusation, especially in relation to things I have said. In other words, anyone quickly reading the conclusions could be led to believing the paper had cogent argument about regression and the soundness of the Rugby data (among other things) whereas it does not even remotely do that. What are his motives for such action? Is he hiding that he has no supporting arguments, or trying to put readers off what I have said?

Further, I should say Doug Landau is not in a position to judge on the matters just mentioned, or the conclusions he drew. Consider two examples of "analyses" he performed, which are simply not right. First, he wanted to establish the TSR for 9F 92050 at 30 mph. He chose seven observations from a Rugby test of that engine, and obtained a trend line from a computer program (Excel) in the form of a quadratic equation ($aX^2 + bX + c$) for each of IHP and WRHP (at Rugby this was DPHP) against Q, the steam rate. The results were:

$IHP = -1Q^2/10^6 +.1148\,Q – 463.45$
$WRHP= -9Q^2/10^7 + .1064Q – 440.41$ (this WRHP is DPHP)

From these trend lines, it follows that
$IHP – DPHP\ (= TSRHP) = -Q^2/10^7 + .0084Q – 23.04$ by subtraction,

And TSR = $-12.5Q^2/10^7 + .105Q - 288$, multiplying by 12.5 to convert HP at 30 mph to a force. From that,

For Q of 14,000, TSR = -245 + 1470 – 292 = 933
For Q of 21,000 (ie plus 50%), TSR = -551 + 2205 – 292 = 1362 (plus 46%)
For Q of 28,000 (ie plus 33%), TSR = -980 + 2940 – 292 = 1668 (plus 22%)

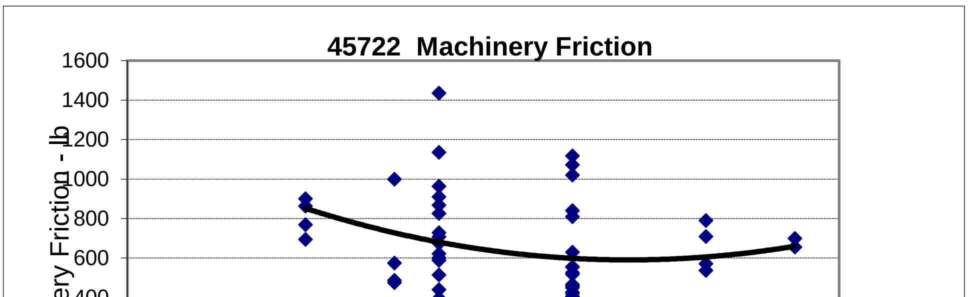
This exercise was supposed to show that TSR was constant at 30 mph (like a dog following its master on a lead he claimed – see *Backtrack*, April 2014, p 253). It does the exact opposite. It shows TSR supposedly varying with Q, but not as fast, and at a declining rate, to high levels.

But this is inappropriate analysis. There are only seven observations, out of 191 for all non-Crosti 9Fs tested. It is unscientific to select only some data from the total without a good scientific reason. Why were not all observations at 30 mph pooled, or indeed all 191, and the effect of speed tested as well? With only seven observations, the chance of finding sound results is much reduced. With the considerable range usually found in Rubgy TSR values under similar circumstances (as exemplified below) that is a considerable failing – it is not known how reliable the answers are. Nor is there any examination of the data and these results in relation to the Small Difference Problem (SDP), nor any testing of the data, to see if it is sensible.

Why was a quadratic chosen? Q has its effect on ITE (not in direct proportion, because SSC varies across the range of Q). Q^2 however is not known to have an effect on ITE, especially when its value is in millions (steam rate Q is expressed in lbs/hr, which occurs in thousands). Presumably the idea was to obtain something resembling the quadratic form of the VR element of LR, in the hope that the TSR and VR could be added together. That results in a minute coefficient on Q^2 as would be expected, but as the values of Q^2 are in millions, they are still large. In any case, the unit squared, Q, is not the same as the unit squared in the VR, ie V. No statistical tests are available, a considerable failing, for they would have shown the fallibility of the reasoning and analysis.

The basis of the analysis is incorrect in using Q at all. IHP is dependent on Q, but not as a straight line (as is clear from any curve of SSC). But DP is not dependent on Q. It is dependent on ITE and TSR (and the components of TSR), not on Q or Q^2.

Second, he is in the habit of using inappropriate trend lines to draw conclusions. See my previous post, in which I pointed out that a trendline of TSR against speed, and only speed, cannot be the right relationship to examine. The six vertical lines obviously contain the real determinant of MR, with speed a lesser factor. The proper approach would have been to use the data at each speed separately (look at the number of observations at both 35 and 50 mph), and test the various possible explanations, of which PTTE is likely to be the best, because it is the major source by far of MR, and to fit regressions rather than trend lines.



These trendlines are not regressions. As immediately above, there is no discipline to them – Doug Landau has used them here to obtain relationships which do not exist in physics or mechanics. They can be done without any of the tests possible with regressions.

Doug Landau's statement that a key test of scientific proof is that its claims are consistent with the empirical evidence is certainly not satisfied by either of these cases, by observation. In the graph above, the line claimed by the relationship ignores most of the data, because the supposed relationship is not valid. At each speed, TSR (his vertical axis) is shown dependent on speed. But TSR is little dependent on speed, which is why his supposed relationship ignores most of the data. TSR varies mostly with other things, on which see below.

The usual logic applied in scientific investigation is formulating hypotheses which from first principles might be relevant to the subject in hand, gathering data which enables the hypotheses to be tested and new ones to emerge (ie almost everything which can be measured about the subject should be measured), testing the data through physical and statistical tests, forming relationships from the tested data to show whether the hypotheses can/should be accepted, including to what degree the acceptability applies. The data has to agree with the theoretical, scientific and/or common-sense expectations, there has to be enough of it, and it has to be sufficiently exact. The empiricism is only part of the process.

For the kinds of claims he makes, he should appreciate that things have moved on since he was a boy, that for decades the data used in deriving a relationship is tested in advance for its soundness, and subject to various forms of analysis, of which regression is the most common, that analysis subject to tests of goodness of fit, whether it differs sensibly from alternative values (including zero), and tests of alternative explanations. With some education in the subject, he would learn that regression is often *the* empirical test, or the most important and useful empirical test – ie part of testing the data for soundness, for formulating explanations of the data, and saying how sound any explanations tested by regressions are. That would save him having to offer weak excuses, such as, to quote, the understood limits of experimental error, however small, and the random nature of scatter, and the real world being more complicated.

Further, on his idea that a key test of scientific proof is that its claims are consistent with the empirical evidence. This puts the cart before the horse. The empirical evidence might be wrong, very poor in itself, subject to the SDP, or untested for its reliability. Then he has to test the relationships, ie establish scientific proof. Doug seems to believe the data are sacrosanct, apparently perfect, or if not perfect (a real world situation?) they are as good as can be obtained in the real world, and are not to be questioned. Not so, as should be clear from almost everything I have written so far. He should be aware of a good example in locomotive testing in this country. The overall BR testing system was badly flawed in the principles guiding it because it depended on an unjustified assumption that a constant blast pipe pressure (BPP) ensured constant Q, at all speeds, and on the plant and on the road. That is why, in general, it is not possible to take the ITE from the plant (where it was usually measured), and deduct EDBTE from road tests for the same Q and V, EDBTE corrected for ind conditions, and to claim that the difference between ITE and EDBTE (as shown in the BR Test Bulletins) gives LR. Only late in the testing was it discovered by simple consideration of the data, that for LR in this case, that such was not correct, that for a given pressure Q varied with speed (as seems obvious). Further, the Q provided by the boiler for a given BPP was different on the road from that on the plant, so my question to him about the 9Fs is crucial.

It is difficult to prove conclusively that experimental data are correct. As above, sheer repeatability is insufficient – all the data can be wrong. Doug uses Carling's belief that because the ITE results for the same test circumstances fall in a narrow band, the ITE data are acceptable, even accurate. Carling also believed that the results from the Farnborough indicator used at Rugby were much the same as those from mechanical indicators available to BR. Mechanical indicators were susceptible to lags and incorrect readings, however, on account of the multiplier in the working, and the small size of the indicator cards being difficult to measure. No proof there. Inserting the input data (pressure, Q, cut off, steam temperature) into the Perform program gives results a little higher than those from Rugby. Perform is by far the best way of approximating cylinder outputs, but itself requires some approximations to inputs, especially cylinder temperature at the beginning of a stroke. Very persuasive, but not absolutely a proof. The Rugby indicator results are highly consistent for a given engine when regressed against Q and V (which themselves determine cut off and steam temperature) in an equation of the form $ITE = cQ^aV^b$, a, b and c being constants, giving good equations and good test statistics. Again, not absolute proof, because the data could all be wrong.

Doug is a great advocate of the accuracy of the instrumentation proving something, eg the Amsler dynamometer, claimed to be accurate to within +/- 1%. That too says little, nay can be completely misleading, if what pull reaching the dynamometer is itself distorted or other factors he has not allowed for, or the SDP is present. (See equation below for the passage of energy from ITE to DP.) [The same Amsler was the source of the DP readings in the first two years of the operation of the Rugby plant, when DP typically exceeded ITE, ie that energy was added to TSR (ITE – DP) by processes in TSR which should all have absorbed energy, ie what was measured by the DP was impossible. This was said to have been cured, by taking oil out of the dashpot in the chain between ITE and DP and replacing it with air, ie replacing a high resistance (oil in the dashpot) in the chain by a lower one (air in the dashpot) resulted in energy being absorbed between ITE and DP, as it should have been. If the change of the medium in the dashpot is all that was done to the system, it is not an explanation for the change in the relativity of ITE and DP, and DP readings remain suspicious. If of course, other things never reported were done, that could well be different.]

The major test to use if there are none available for the data as data is to fit the relationships to which the data should conform, decided either from past research, or from first principles, as used in formulating hypotheses about the subject before the research started.

And if data fail tests, or no tests are possible, then no more use can be made of it. It cannot be used to prove anything, except how not to specify and conduct experiments, and whether it is possible to obtain TSR at all.

Doug Landau does not appreciate that the data are the real world, (see his remark above about the "real world" and things not fitting together like a jigsaw puzzle). Whether he likes it or not, in science, he cannot interfere with data. He might, with some statistical and technical analysis, show that is probable (even to a degree of probability) that the data would be useful for finding MR or TSR if such and such had been or not been done (I do some of this below), but he cannot impose anything on the real world.

Last, be it remembered that it was said in the *Locomotive Railway Carriage and Wagon Review* for December 1957, pp 233-4, in one of a series of articles in that journal during the second half of 1957 on *Locomotive Testing on the Rugby Plant, BR*, that it is *not* possible to measure the internal friction of a locomotive accurately on a test plant, only to confine its value within comparatively wide upper and lower limits. (As the data are so unsatisfactory, the confidence with which any declared upper or lower limit can be held must be low.) The articles were unattributed, but were almost certainly prepared by D R Carling, Superintendent of the Rugby Testing Station during its operations. Certainly, Carling did not refute the point. It is therefore extraordinary that Doug Landau, after all these years, claims to be able to judge the Rugby data better than Carling, and to want to do so without explaining how. That is the same as setting his face against regression results – nothing declaring against the Rugby results, specially by me, is to be tolerated.
I suspect too that he believes that scatter is evenly distributed and that the true answer lies in some sort of average of all the data. I fear not. The testing and consideration of the data requires consideration of the scatter, its extent and an examination for biases.

Simply declaring that the Rugby data are fit for providing TSR values avoids crucial steps in showing that it is fit. Declarations are empty if the steps have not been taken. Doug Landau has never shown that he has considered the data, so it follows his declarations are empty.

I have therefore turned to testing the data for their soundness. This involves going back to the first principles of the mechanics involved, analysing the forces involved, and considering from acceptable references the likely friction coefficients involved. I have found the data lacking.

2 Are the Data Sensible?

I have considered their "soundness" in four ways. First, they have been graphed against PTTE, for their consistency or repeatability. This has been done for every engine tested on the plant where there were at least a dozen observations at one speed. In some cases, more than one speed was available, with up to four speeds suited to this analysis. In no case were the data consistent or repeating. [Graphing is mostly sufficient to show this, but in one case (Duchess 46225) it was shown in addition by painstakingly listing and ordering the observations which are inconsistent with one another.]

Second, I considered the values of TSR obtained from ITE – DP (the experimental results) for their magnitude. Using the same data from the cases where there are at least a dozen observations at a single speed, from each TSR observation were deducted the CWBR and the items varying with speed squared (where relevant), both of which items should be constant at the speed concerned, to leave a residual, which ought to be the value of all items varying with piston thrusts. In analyses and comparisons of mine, these were found to be a ratio of .05 to .07 of PTTE (details available on request). In these Rugby TSR data, the ratio is much lower than .05 to .07. For the twelve engine-class/speed combinations considered, the vast majority result in ratios which on average are less than .025. Only the Jubilee at both speeds (40 and 50 mph) could be said to demonstrate coefficients approximately those expected, but still on the low side, but the Jubilee data are problematic in other respects. Some are very low indeed, and the value of the ratio is generally erratic.

Third, TSR was regressed against PTTE for the same twelve class/speed combinations, for each speed/class combination. The logic is that an equation in TSR at each speed should in those circumstances have a positive constant covering all items constant at that speed, and a positive coefficient on PTTE covering all items varying with PTTE, ie constant + xPTTE at each speed.

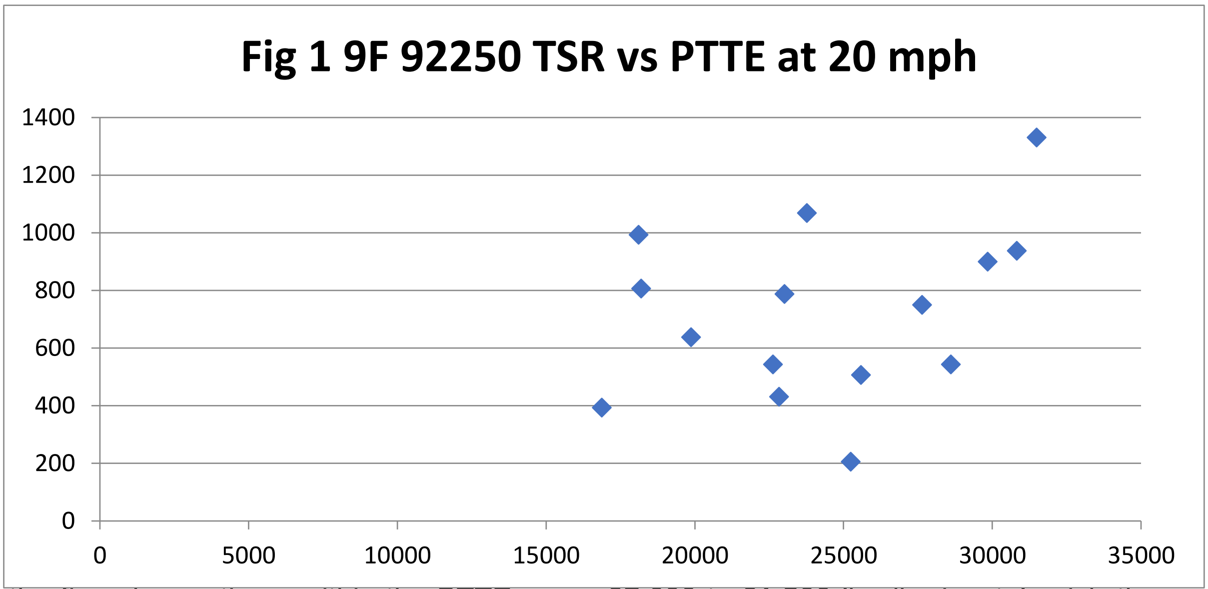
Fourth, Rugby data were also used to apply the input/output approach to MR for a couple of classes, as used in obtaining the approximate MR of internal combustion engines. These yield

MRs which are far too high. This is consistent with the low values of TSR. This however is incidental to the previous three approaches.

3 Consistency/Repeatability of Rugby Data

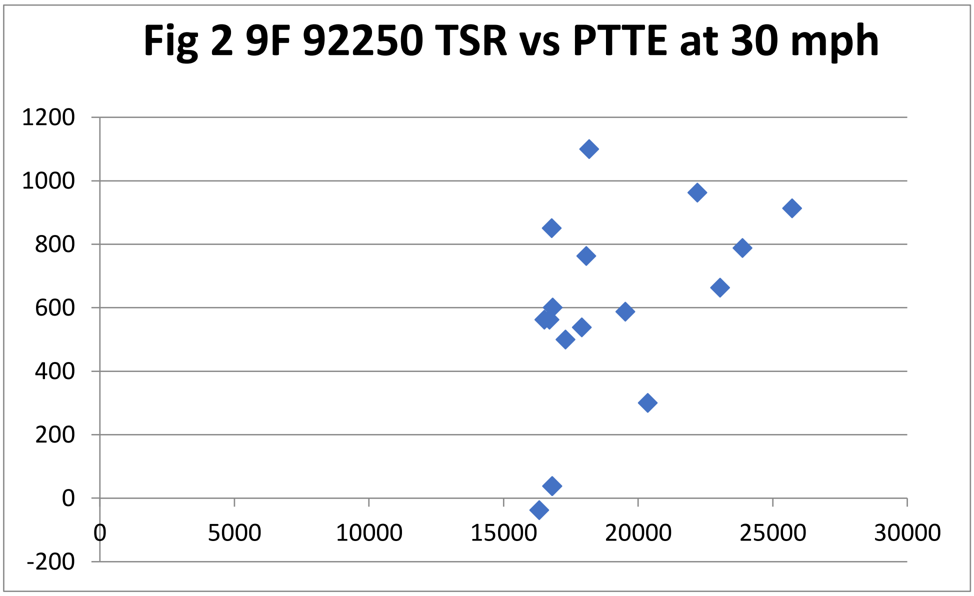
To exemplify the point about non-repeatability of the Rugby TSR data, I have chosen the data from 9F 92250 , the last steam engine tested on the plant. By then, practice on plant should have been as good as it ever was. In this case, the data are available for at least 12 observations for four speeds, 20, 30, 40 and 50 mph.

In all the figures TSR is on the vertical axis, PTTE on the horizontal.



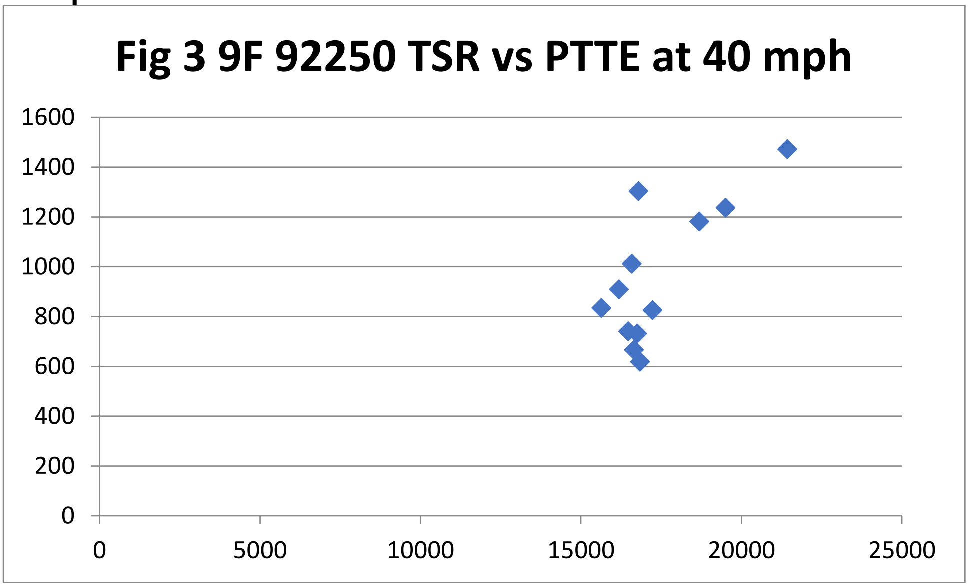
For the five observations, within the PTTE range 27,600 to 31,500 lbs (horizontal axis), the TSR range is 544 to 1331, the average TSR is 844, and its Standard Deviation 290.

30 mph



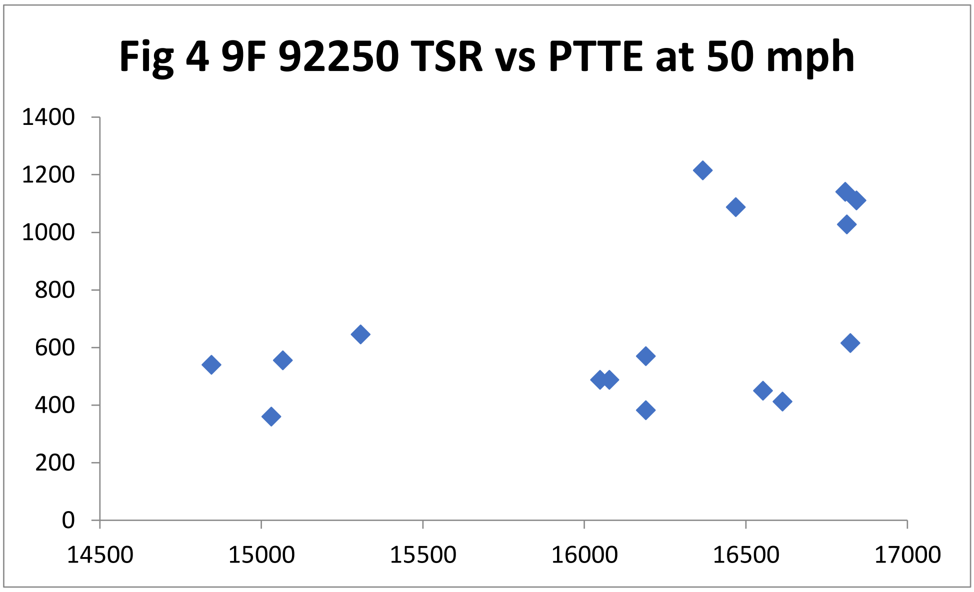
Twelve of the 19 observations fall in the PTTE range of 16,300 to 19,500 lbs, in which the TSR range is -38 to 1100 lbs. The average TSR of these 12 observations is 508, and their standard deviation 343.

40 mph



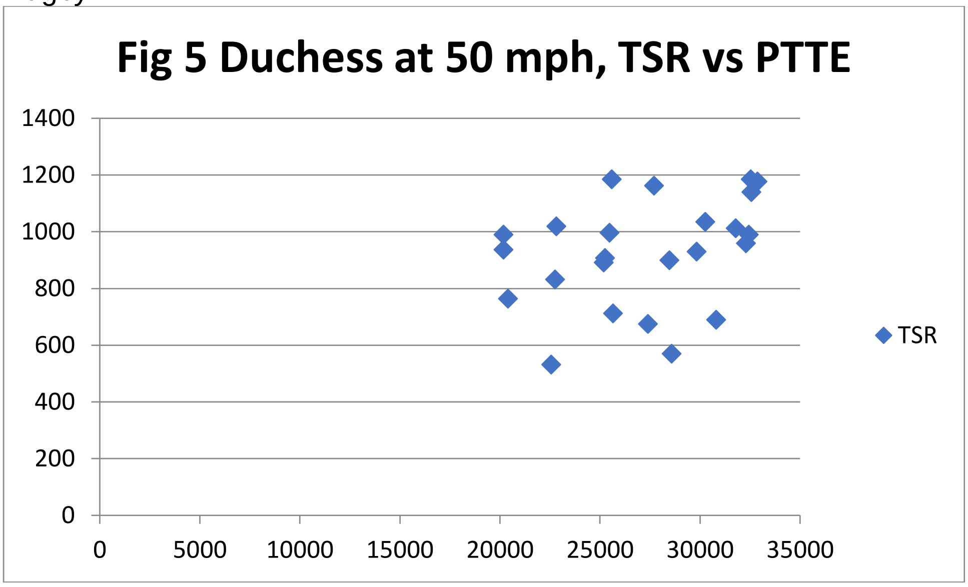
Of the 12 observations, nine are within the PTTE range of 15,600 lbs to 17,200 lbs. The TSR range of those observations is 619 to 1303 lbs, the average 849 and the standard deviation 209.

50 mph



The four observations at about 16,800 lbs PTTE contain TSR in the range 615 to 1140, for which the average is 973 and a standard deviation of 243. The four observations at about 16,500 lbs contain TSR in the range 615 to 1140, for which the average is 823. Given the circumstances of their origin (and the SDP), the three observations in the far top left of Fig 4 are as good as could be expected, but the fourth observation at 16,800 lbs demonstrates the lack of consistency, or repeatability.

In addition, Fig 5 gives the TSR and PTTE data for Duchess 46225 at 50 mph, for which here are 24 observations, the greatest number at any one speed for any single engine tested at Rugby.



At a PTTE of close to 25,000lbs PTTE, the five TSR values vary from 713 lbs to 1185 lbs, with an average of 939 lbs. At a PTTE in the range of 28,000 to 30,000 lbs, TSR varies from 570 lbs to 1163 lbs, with an average of 881 lbs. At a PTTE of 32,000 to 33,000 lbs, the six values of TSR vary from 960 to 1185 lbs, with an average of 1083 lbs, this being the only case of TSR values being even remotely close of all the PTTE ranges discussed here, there being two groups of three observations which could even be said to demonstrate repeatability, even though the two groups of three are about 200 lbs or 20% apart.

In all five cases, the spread of data is much greater than modest variations about what Doug Landau seems to consider the right value of TSR derived from the Rugby data, these modest variations being what he terms scatter, something he regards as unavoidable, but perhaps excusable. TSR is of course the subject of interest. The variation is in most cases indeed modest in terms of ITE or DP or PTTE, but in terms of TSR it is large, on account of the SDP. Far from showing that TSR is constant at a wide range of PTTE, the data characteristics show the opposite, that TSR varies a lot to a degree to which mechanics provides no basis, seen also in the large standard deviation in TSR. Further, considering the variability in relation to DP is not sound, because DP is simply a measurement of ITE less TSR, ie DP is a result of those other two items; or DP is the result of the effect of TSR. Furthermore, scatter is not something to be judged according to the ideas of Doug Landau. Statistics has methods for making this judgement in relation to the best fit to the data recorded, and the size and regularity of the deviations from the best fit, ie whether even the small amount of repeatability occurs by chance.

I did the same for every engine tested at Rugby for which there are at least a dozen observations at a given speed. It all shows similar characteristics. The data are available on application.

It is obvious that there is almost no sensible repeatability in most of these data. No doubt this will draw forth the cry that strict repeatability is impossible in most experimentation, and that there are some observations that are close enough to be regarded as the same. Where the observations are close, that is indeed what I expect. But I have considered narrow ranges of PTTE above and found a wide variation in associated TSR, in each case detailed under each Figure. The TSR data can be said to be no better than erratic. Further, a considerable number of observations are low, which raises the question of what value they should have. On that see the next two sections.

With the wide spread of TSR data at a given rate of working, given his criticisms of my remarks, it would be of interest to know what Doug Landau would consider to be the TSR of 92250 in the range of 20 to 50 mph based on Rugby data. Given his defence of these data, that seems a fair question to ask him to answer.

4 Implied Value of (TSR – CWBR – MR – (resistances varying with V^2))/PTTE

In this exercise, it is considered that TSR comprises CWBR, MR, resistances (friction and work) varying with V^2, DR and heat. The value of these constituents of TSR is not separately measured, but any DR for example will be included in TSR. If heat is lost, it is not included in TSR.

Using the same data from the cases where there are at least a dozen observations at a single speed, from each TSR observation were deducted the CWBR and the items varying with speed squared ($PTTEV^2$) (where relevant), both of which items should be constant at the speed concerned, to leave a residual, which ought to be the value of all items varying with piston thrusts. The deductions for CWBR and $PTTEV^2$ were obtained in my earlier analysis of MR from first principles (available on request), and are very reasonable values (the CWBR uses Cfs consistent with rolling stock resistances which emerged from Ell's researches into British rolling stock resistances (Ell was an officer in the locomotive testing on BR). In that analysis, the value of this ratio was found to be .05 (low) to .07 (high) of PTTE. Note that this .05 to .07 is not a coefficient of friction, but the proportion of the friction to the net forces involved in PTTE both at a common point, the CW rims. The actual Cfs occur at many locations (piston rings, glands, crosshead, and its guides, gudgeon pin, rod pins and the addition to the vehicle only CWBR from the PTTE forces); Cfs at particular points vary from .012 to 0.14. Amalgamated, these yield the ratio of .07. Lower illustrative values in some cases yield the .05.

The following tables are the results of applying this approach to 9F 92250.

In Tables 1 to 4, (a) represents net friction of rods on pins and work done working on unbalanced reciprocating masses; and residual (b) is column 3 – column 4 – column 5.

20 mph

1 Run	2 PTTE	3 TSR	4 CWBR	5 V sqd items (a)	6 Residual (b)	7 Residual/PTTE (c)
2237	16875	393.75	228	38	127.75	0.008
2251	18116	993.75	228	38	727.75	0.040
2168	18200	806.25	228	38	540.25	0.030
2229	19879	637.5	228	38	371.5	0.019
2243	22626	543.75	228	38	277.75	0.012

2249	23016	787.5	228	38	521.5	0.023
2164	22831	431.25	228	38	165.25	0.007
2226	23774	1068.75	228	38	802.75	0.034
2250	25240	206.25	228	38	-59.75	-0.002
2167	25585	506.25	228	38	240.25	0.009
2230	27644	750	228	38	484	0.018
2255	28613	543.75	228	38	277.75	0.010
2235	29851	900	228	38	634	0.021
2233	31496	1331.25	228	38	1065.25	0.034
2170	30822	937.5	228	38	671.5	0.022

Table 1 Ratio of residual (see text) to PTTES in Rugby Data for 9F 92250 at 20 mph
Average value of column 7, .019.

30 mph

1 Run	2 PTTE	3 TSR	4 CWBR	5 V sqd items (a)	6 Residual (b)	7 Residual/PTTE (c)
2146	16324	-37.5	228	86	-351.5	-0.021
2238	16515	562.5	228	86	248.5	0.015
2150	16801	37.5	228	86	-276.5	-0.016
2228	16825	600	228	86	286	0.017
2155	16790	850	228	86	536	0.032
2147	16808	37.5	228	86	-276.5	-0.016
2227	16703	562.5	228	86	248.5	0.015
2144	17302	500	228	86	186	0.011
2252	18073	762.5	228	86	448.5	0.025
2156	17908	537.5	228	86	223.5	0.012
2225	18179	1100	228	86	786	0.043
2145	19518	587.5	228	86	273.5	0.014
2231	20358	300	228	86	-14	-0.0006
2157	22200	962.5	228	86	648.5	0.029
2234	23049	662.5	228	86	348.5	0.015
2148	23877	787.5	228	86	473.5	0.020
2149	25718	912.5	228	86	598.5	0.023

Table 2 Ratio of residual (see text) to PTTES in Rugby Data for 9F 92250 at 30 mph
Average value of column 7, .024

40 mph

1 Run	2 PTTE	3 TSR	4 CWBR	5 V sqd items (a)	6 Residual (b)	7 Residual/PTTE (c)
2177	15639	834	228	153	453	0.029
2176	16189	909	228	153	528	0.033
2162	16474	741	228	153	360	0.022
2253	16657	666	228	153	285	0.017
2239	16748	731	228	153	350	0.021
2174	16795	1303	228	153	922	0.055
2163	16841	619	228	153	238	0.014
2175	16586	1013	228	153	632	0.038
2161	17232	825	228	153	444	0.026
2180	19504	1238	228	153	857	0.044

2160	18683	1181	228	153	800	0.043
2186	21428	1472	228	153	1091	0.051

Table 3 Ratio of residual (see text) to PTTES in Rugby Data for 9F 92250 at 40 mph
Average value of ratio (c) in column 7 .033

50 mph

1 Run	2 PTTE	3 TSR	4 CWBR	5 V sqd items (a)	6 Residual (b)	7 Residual/PTTE (c)
2244	14846	540	228	239	73	0.005
2183	15307	645	228	239	178	0.012
2241	15030	360	228	239	-107	-0.007
2169	15066	555	228	239	88	0.006
2246	16077	487.5	228	239	20.5	0.0013
2248	16190	382.5	228	239	-84.5	-0.005
2240	16048	487.5	228	239	20.5	0.0013
2165	16190	570	228	239	103	0.006
2247	16613	412.5	228	239	-54.5	-0.003
2242	16552	450	228	239	-17	-0.001
2182	16842	1110	228	239	643	0.038
2166	16807	1140	228	239	673	0.04
2245	16812	1027.5	228	239	560.5	0.033
2257	16823	615	228	239	148	0.009
2181	16366	1215	228	239	748	0.046

Table 4 Ratio of residual (see text) to PTTES in Rugby Data for 9F 92250 at 50 mph
Average value of ratio (c) in column 7 .013

The residual is often negative or very low. The value of the ratio in Col 7 in each table is far too low, meaning TSR is too low subject to the items in cols 3, 4 and 5 being correct. A ratio of .05 to .07 expected, for low and expected coefficients of friction. Only two runs in the 92250 data, in table 3, 40 mph, nos 2174 and 2186, satisfy this criterion, and then only at the lower expected value.

This approach relies for its conclusions on other analyses I have made in other contexts. I do not claim that the data could not be tested for this purpose in other ways. I do not accept the judgemental comment (made with no exemplification) of my critic that I have selected friction coefficients to justify my conclusions on this second or any other approach. I defend the values I chose from several sources.

5 Regressions of TSR data

TSR was regressed against PTTE for the same twelve class/speed combinations used in analyses above. Each regression was made at a particular speed. The logic is that an equation in TSR should in those circumstances have a positive coefficient on PTTES and that the rest of TSR should be included in a constant, the values of that coefficient and the constant emerging from the data, not imposed. The question then arises, what relationship should be sought?

Between ITE and DP are the components of TSR, plus BR and DPP. BR is braking resistance at the braked rollers and equal to WRTE. It is transmitted through the frames and CW bearings to the locomotive drawbar, where it emerges as drawbar pull DBP, equal to BR (although as a couple causing oscillation in a vertical plane about a horizontal axis, resulting from the differing heights above rail of the locomotive drawbar and the CW centres). The resistance of CWs rolling on the rollers of the braking dynamometer has been considered as part of the resistance against which the engine was working.

ITE – DP = TSR.
ITE – CWVBR – MR = WRTE, all terms measured at the CW rims
WRTE = BR = DBP
DBP – DR – Heat – fV^2 = DP, whence
DP = ITE – CWVBR – MR – WRTE + DBP – DR – Heat – fV^2
= ITE – CWVBR – MR– DR – Heat – fV^2
(Remember that in my approach, after resolution of the weight borne static load on the bearings and the forces of the mechanism on those bearings, the CWVBR is deducted from the resolved sum and the remainder (the extra resulting from mechanical action) is part of MR.) [Doug Landau appears to be unaware of the convention applying to the term static axle or bearing load. He thinks it means without the wheels turning. It applies to both circumstances. There are plenty of examples of the term static in the sense in which I have used it – see for example the paper by Cox on locomotive axleboxes, which he quoted, with the flavour that Cox's paper proves I am wrong in some way. If this still offends him, he can ignore the word static].
The only source of energy in this system is ITE. DP has no independent existence of its own. It is merely a pressure measuring device giving the pull resulting from ITE – TSR. If any ITE observation is wrong in fact, the unintended error will affect the MR, DR and Heat elements of TSR, and pass to DP. If the ITE is correct, but measured wrongly, then MR, DR, Heat and DP will be correct, or at least as correct as if there were no error in ITE, but TSR will be wrong. It is therefore highly probable that DP will be wrong and TSR wrong in consequence. The equality between WRTE and DBP must remain. The fV^2 term applies to any net forces and net work associated with the revolving masses on pins, and work done revolving unbalanced masses. Heat arises at the dampening, (the dashpot when filled with oil was water cooled), and to any other loss of heat between the CW rims and the DP. Multiplying throughout by (-1), dividing some terms into fixed (constant) and variable portions, and rearranging:
at any given speed, CWVBR and fV^2 will be constant, as will any constant in MR, which means
TSR = constants + bPTTE
There are no data of DR per se. The Belleville washers and dashpot will have reacted in proportion to the forces involved (the dashpot) or be fixed for the speed concerned (the Belleville washers), and partly in proportion to the effort, which effects should divide into constant and variable in a regression. Heat from any effect (the Belleville washers and dashpot) will be lost from measurement, so that measured DP will have been too low and measured TSR too high.
TSR = constants + bPTTE is therefore what is to be regressed. That would be followed by examining the results and the residuals for any sensible conclusions which can be drawn about the effect of heat, even from calculating its value from first principles. Alternatively, if the results can be obtained for several speeds, and are very good, they can themselves be analysed for the approximate values by elimination. It will be noticed that the relationships are a result of the data speaking for themselves – nothing is imposed.

It would be wrong to regress DP against Q. Q has already influenced ITE, at a rate varying with Q per se and V, and as seen in the Specific Steam Consumption. The same applies to regressing DP against ITE. That would not provide any relationship of any value, on account of the big number which each represents, ie that DP will be close to ITE. The difference between the two large numbers will be small, and it is to be expected that the two will be highly correlated, which can distort the results. Further, any such regression will as a result give a high value of r^2, which to many unpractised analysts is the be all and end all of regression or other approaches to obtaining relationships. But regression will also give how high are the probabilities that the terms and coefficients on them are close to being correct (or significant, meaning significantly different from something in the relationships (eg zero, or a close value; significant does not mean large). The coefficients on ITE would have definite high values of the t ratio, ie that the slopes of the relationship are high, indeed very high. The relationships however give very low t values for the constants, which means that it is not possible to fix the relationships with any certainty. TSR is the thing to regress. It would be expected that if the data are good, a well-established constant and coefficient on PTTE would emerge.

It would also be wrong to regress the equation DP = ITE – constants – bPTTE. The three terms PTTE, ITE and DP are close in magnitude and highly correlated, which can affect the answer. Secondly, unless the data are very good, it will be impossible to separate ITE and DP statistically and obtain a sensible coefficient, especially when DP is so erratically related to ITE (as the data in Table 1 shows). Further, any estimation should use actual data and then as parsimoniously as possible, ie without needless complication. TSR is the subject of interest, the matter under investigation. TSR is available in its own right, ie as (ITE – DP). It is not acceptable to "smooth out" ITE and DP, by fitting an equation to both separately. Even if such is done, it is necessary to show that ITE and DP are statistically separable with various degrees of confidence, ie are significantly different at some level of probability accepted for experimentation of this kind from zero. As such, it is not possible to show that, because the values of both ITE and DP have wide confidence intervals of their own.

6 Regressions of the TSR data.

6.1 Duchess 46225

Equations for TSR at 50 mph

There are 24 observations for this engine at 50 mph, the greatest number at one speed for any engine tested at Rugby, the number a useful characteristic in obtaining good results.
The result of fitting TSR = constants + bPTTE is
TSR = 522 + .015 PTTE. [1]

The observed data are very dispersed, as shown in Figure 5, with consequential low significance of the results. Equation [1] is the fitted equation of Fig 5, the best fit to the data using the above form of equation. The SEE is 183, Sigf F 0.114, t on constant 2.12, and on the coefficient 1.66, with R^2 0.111. On all possible grounds, this is unacceptable. Despite having the right signs, the coefficient on PTTE is double or more that expected, ie the slope of the relationship is much too high. The SEE puts a range of $^+/-$ 183 to give an answer 68% significant, and $^+/_-$ 365 one significant at 95%, as might be expected from Fig 2 below. At 68% the range on the coefficient on PTTE is .006 to .02. Everything about the fitted equation, the best fit to the data, is the uncertainty of the results, and their low value, ie that DP recorded high. The extent of the high reading being unknown, that is no help in obtaining MR generally.

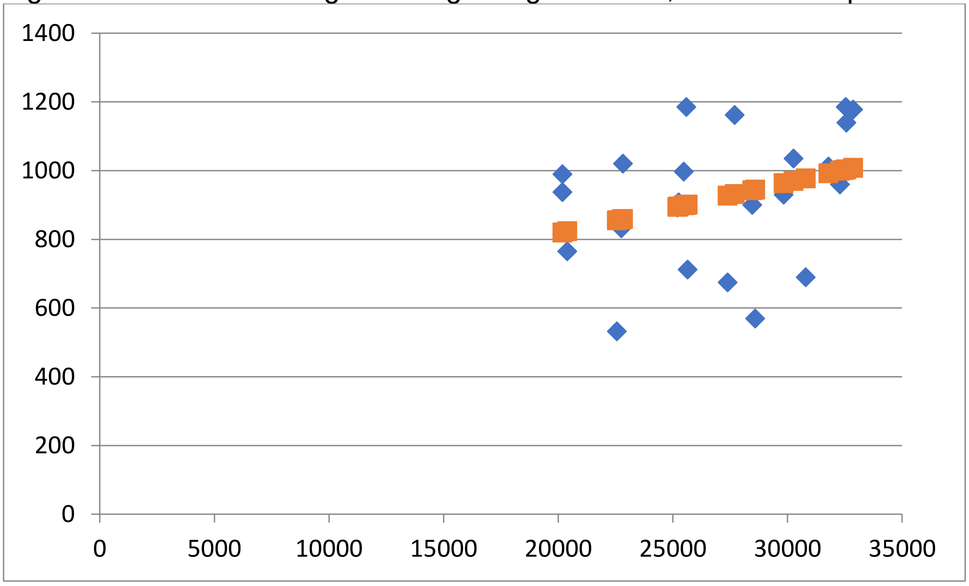


Fig 6 Observed and Fitted TSR (vertical axis) on PTTE (horizontal axis), blue observed, brown fitted by regression

From previous analysis, the expected constants in TSR for a Duchess would be some 228 for CWBVR, $.22V^2$ for speed related items or 550lbs at 50 mph, normal constant of MR 120, total constant about 900, to which .07PTTE has to be added. (Of this, the 228 of CWBR constant is not MR), all much higher than given by equation [1]. the data occur only at high values of PTTE. It is in that range that there is interest in TSR, but use of [1] to give them must be of even less reliability than and there is no professional way of formulating an equation in TSR which gives TSR values for lower values of PTTE.

See also Relating Input to Output, Willans Line approach to Determining MR directly for this engine, below.

6.2 9F 92250

This was the last steam engine tested at Rugby, in 1959. It could be said that procedures should by then have been such that the results were as good as they were going to be. On the other hand, the DP results were still problematical as shown below. The tests of this engine include both a double chimney arrangement and a Giesl ejector exhaust. Both these fittings should have resulted in lower back pressure, and slightly lower MR. Some of the tests involved use of slack coal, to test the ability of the Giesl ejector to allow satisfactory steaming with such. That should not of itself have affected MR. There is the considerable advantage in using the data for this engine because there are 60 observations, 15 at 20 mph, 17 at 30, 12 at 40 and 16 at 50 mph, a reasonable number at each speed for analysis, and for obtaining the effect of speed, although 12 observations at 40 mph is just sufficient. The equations in TSR are:

[2] 227 + .02PTTE at 20 mph.

15 observations, SEE 291, Signf F .02356, t values 0.56 and 1.24, r^2 0.106

[3] -436 + .053 PTTE at 30 mph.

17 observations, 299, 0.523, -0.9, 2.11, 0.23

[4] -1207 + .1246 PTTE at 40 mph.

12 observations, 195, .0058, -1.94, 3.50, 0.55

[5] -2774 + 0.215PTTE at 50 mph.

16 observations, 277, 0.277, 2.23, 1.51, 0.24

The data do not allow sensible explanations of TSR. The constant cannot be negative. The negative constants are compensating for the unduly high coefficients on the PTTE terms, at least at most values.

At all speeds together, ie all 60 observations, for TSR

[6] 433 +.0149 PTTE.

324, .136, 2.22, 1.51 and .038

[7] including V, -422 + .0373 PTTE + 12.17 V.

310, .017, -1.08, 2.87, 2.50 and 0.13

[8] including V^2 as well as V, -1292 + .044 PTTE + 56.9V - 0.61 V^2.

307, 0.174, -1.81, 3.21, 1,81, -1.44, 0.164

[9] 2.62 $\times 10^{-10}$ x $PTTE^{2.64}$ x $V^{1.35}$.

2.17, 2.69, 2.40 and .12

By solving [6] to [9] inclusive in turn for the four values of V, those equations can be converted to equations similar to [2] to [5]. They are, for TSR, first based on [7]

[2a] at 20 mph -179 + .037 PTTE,

[3a] at 30 mph -57 +.037 PTTE,

[4a] at 40mph 65 + .037 PTTE,

[5a] at 50 mph 187 +.037PTTE

Then based on [8]

[2b] at 20 mph, -130 + .044PTTE

[3b] at 30 mph, 360 + .044PTTE

[4b] at 40 mph 886 + .044PTTE

[5b] at 50 mph 1400 + .044PTTE

Then based on [9]

[2c] at 20 mph 110 $\times 10^{-10}$ x $PTTE^{2.64}$

[3c] at 30mph 258.5 $\times 10^{-10}$ x $PTTE^{2.64}$

[4c] at 40 mph 381 $\times 10^{-10}$ x $PTTE^{2.64}$

[5c] at 50 mph 515 $\times 10^{-10}$ x $PTTE^{2.64}$

It is obvious that the TSR data do not lead to any sensible explanations of TSR. Even [8] with satisfactory t values, has poor F and r^2 tests. This is not the fault of regression, but of the data

The TSR must have a positive constant, at least the 228 lbs or so expected value of the CWBR. [2], [3] and [4]) not only have unacceptable negative constants, but [3] and [4] also have coefficients on PTTE far too high, so the data have characteristics which by having these high coefficients, throw an increasing negative value on to the constants. The (c) set of equations require raising the PTTE to a power of 2.64, then multiplying it by a very small value coefficient, which is not sensible in principle. The t values on the coefficients are in many cases so low that the probability of the values given is too low to be acceptable. The values of r^2 are too low for the equations to be said to explain the data, that after all the relevant forms of analysis have been tried. All the coefficients on PTTE imply variation with PTTE well below the .05 which would be expected from engineering data, and there is no other term in which that friction appears. The lowest PTTE is about 15,000 lbsf, and the highest about 32,000 lbsf, as seen in the Figures, but that should not render the constant negative, and in some cases considerably so. Whatever, the negative coefficients in TSR equations, indicates an extra resistance between ITE and DP of at least 600 lbs. Equations [3]) and [4] for 40 and 50 mph respectively are not sensible at all, the high negative constants and the high coefficients on the PTTE not being credible.
If ITE is regarded as sensible, then the DP is too high generally, and behaves erratically with those features with which it should vary, in engineering terms. In other words, what was recorded at Rugby for DP was not worth recording, even at the end.

5.2 Other 9F

There is sufficient data at least at one speed to analyse the results for some other 9Fs, as follow:

[10] 92013, 1954, 14 observations at 25 mph.
TSR = 639 - .005PTTE.
241, 0.78, 1.79, -0.28, .007
[11] 92166, 1958-59, 15 observations at 30 mph:
TSR = 281 + .047PTTE.
175, .00387, -1.05, 3.5, 0.49

The ITE data for the 9F as a class in all tests combined are consistent, satisfying ITE = $13.24Q^{1.011}V^{-0.84935}$ with a r^2 of .99 and excellent statistical tests. The self-consistency does not mean they are perfectly measured. The poor TSR results have to result from odd behaviour of the constituents of the TSR, with, perhaps, low values of ITE.

The data cannot provide a sensible TSR for any engine of this class. Only [11] approximates what might be expected, and then with such low t value on the constant that it is clear that the vertical location of the curve (ie above zero PTTE line) cannot be fixed. Nothing consistent or conclusive results, the signs are completely inconsistent (those on the constant and the coefficient must both be positive, the statistical tests are almost all very poor, and most of the coefficients on PTTE which are positive are far too low.

5.3 Royal Scot 46165

In the 61 observations for this engine, 13 were at 40 mph and 20 at 50 mph. The remaining observations were at speeds from 20 to 80 mph, in small numbers at each speed. The modest numbers at 40 and larger number at 50 mph were regressed in the same way as for the Duchess and the 9Fs.

[13] At 40 mph, TSR = - 9386 + 0.702 PTTE.
The SEE is 345, the Sigf F 0.418, the t values -0.79 and 0.84, and r^2 0.06.

The high negative constant and high coefficient on PTTE are equally exaggerated, and the constant has the wrong sign. Such an equation simply shows that the data are so poor that an explanation of TSR is not possible.

[14] At 50 mph, TSR = -412 + 0.071 PTTE. The SEE is 309, Signf F .533, the t values -0.25 and 0.63, and r^2 0.02.

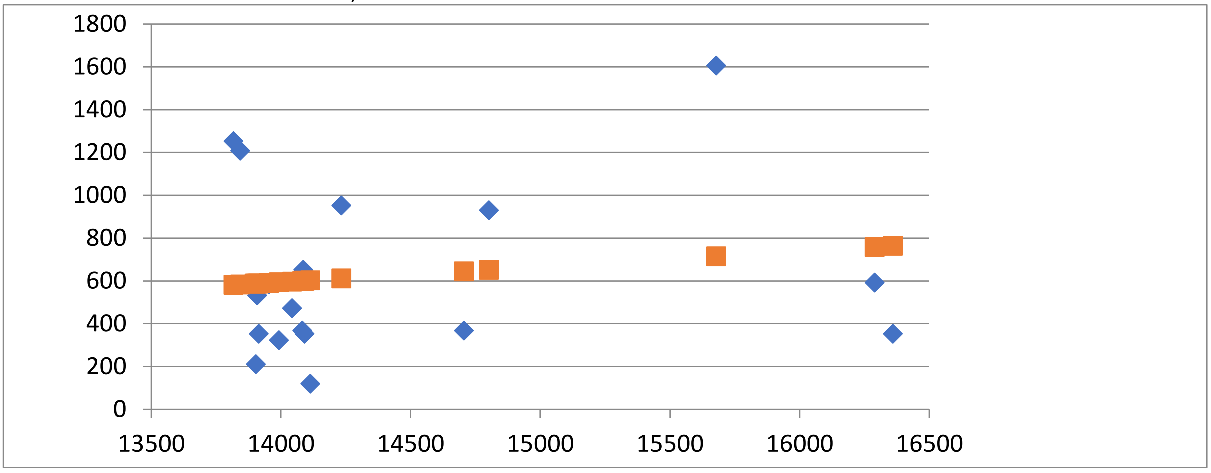


Fig 6 Observed and Fitted TSR (vertical axis) against PTTE, 46165 at 50 mph, blue observed, brown fitted, best fit

The PTTE data occur only in the range of ca 14,000 to 16,500 lbs. It is obvious why no satisfactory equation can be fitted to these data, given the wide dispersion. At ca 14,000 lbs PTTE and one speed (at which so many items are constant), the TSR should be close to constant, yet it is distributed from ca 100 lbs to 1200 lbs. It is low, given that TSR includes the constant for CWBR, and the value at 50 mph of the terms in MR which vary with V^2. It is slightly increasing with PTTE, at about the expected rate (but here with such a low t value that the value of that rate is not at all certain). That slope should continue back to zero PTTE, where it should have a positive constant. The result here of a negative is further indication that the values are all low. It cannot be argued that "something" would cause the fitted TSMR line to rise as it is projected back to zero PTTE, something not present in the data, or allowed for in the equation. That cannot be: as above there should be a considerable positive constant, all V and V^2 effects should be in the constant, and it is logical for the rate of variation of TSR with PTTE to be much the same at lower PTTEs as at higher.

Across <u>all speeds</u>, a regression of TSR against PTTE and V^2 gives a result of

[15] TSR = -376 + .074 PTTE – 0.165 V^2,

with the coefficient on PTTE significant at the 95% level of confidence. This is a better equation than those at 40 and 50 mph, SigF .0003, SEE 633, t values -0.2, 2.29 and -1.02 but with an r^2 of only 0.29. But a negative constant when the CWBR constant is 150 lbsf, and the negative coefficient on the V^2 term show that no relationships based on the technical first principles of MR emerge from these data.

There is no obvious pattern to the residuals. No interpretation can be placed on these results. The even effect of TF forces in a three cylinder engine with cylinders in line does not exist on this engine because the outside cylinders drive on to the second coupled axle, while the inside cylinder is forward, and drives on to the leading coupled axle, but that cannot explain the enormous negative constant.

5.4 Jubilee 45722

This engine was tested in 1956-57. There were 18 tests at 35 mph and 25 at 50 mph. The regression results were:

[16] 35 mph, TSR = -193 + .068 PTTE.
271, 0.143, -0.3, 1.54 and .13
[17] 50 mph, TSR = - 866 + .112 PTTE.

254, 0.316, -0.63, 1.02, .04

The same conclusions as drawn for the 9F apply in this case.

5.5 Standard 5 73030

There were 12 observations at 55 mph. The regression result was:
[18] TSR = -523 + .097PTTE.
343, 0.473, -0.34, 0.74 and .05.
The same conclusions as drawn for the 9F apply in this case.

Of course, it is possible to say that a negative constants are impossible, it is he absence of data below the observed values which are the reason for both the negative constants and high coefficients on the PTTE. That could of course be true had tests been conducted at lower efforts, but such data do not exist, and imposing values which make the data appear better, and at the same time removing the above deficiencies is not scientific. Further, the composition of the constant and PTTE terms are such that they should capture variation right down to low but positive values, ie the data should have such behaviour in it if the data were satisfactory. Further, it is at high values of PTTE that data will be observed because the experiments were conducted at outputs of interest to those testing the engines, and it is for values n about the same range for which TSR and LR will be needed. Whatever might be thought about the constants, the coefficients on PTTE cannot be judged other than being far too high.

Is the assumption that the relationship with PTTE is linear justified? I have not yet tested that, but do not expect any change in the conclusions.

7 Other Notes on Rugby Results

Some effort was devoted to the data for all classes across the whole speed range. Apart from finding consistency in the basis of ITE, no results of use emerged. In addition, the average MR of each class which emerged, MR here being TSR less CWBR, was analysed, with the following results:

Class	Average TSR lbs (a)	Calculated Constant of CWVBR lbs	Average recorded TSR lbs	In speed range, mph
9F	542	228	314	36 – 60
Duchess	953	227	726	50 - 85
Standard 5	640	151	489	45 – 75
Jubilee	681	150	531	50 - 85
Royal Scot	586	150	436	50 – 85
Crab	642	169	473	40 - 70

Fig 8 Comparison of Observed Average Apparent Resistances at Rugby for Five Classes
average TSR hides any variation with V^2, or more generally $(rpm)^2$. It differs from MR by CWBR

The Jubilee and Royal Scot differ mechanically essentially only in cylinder diameter. The latter has the larger diameter, with more circumference of piston rings to slide on the cylinder walls. Yet the average TSR of the Scot in the Rugby data is 18% lower than that of the Jubilee. The Crab and Standard 5 TSRs are also out of line. The Crab should have a higher average resistance than the 5, partly on account of its smaller CWs, partly on account of its bigger cylinder diameter. In that case, however, the lower pressures on the rings of the Crab will affect the comparison.

The average MRs for these engines are very low for the sizes of the engines, generally. Whatever might be considered about anything I have calculated, the correct average MR of the 5 of 489 is very low. The standard 5 should have much the same average MR as the Black 5 – its slightly larger cylinders are roughly balanced by its slightly larger CWs – instead of less than half. For an engine with such small CWs, the TSR of the 9F is very low. The third is that it cannot reasonably be expected that the MR should be constant over all outputs and speeds.

The results for the TSR regressions, however, are overwhelmingly disappointing, in terms of sense (ie behaviour and signs) and magnitudes, with wide standard errors of the estimate, low t scores on coefficients, high significance F values, and values of r^2 as low as 0.1. Neither the equation chosen, nor the basis of the analysis (regression) nor its application in this case, is at fault, it is the poor, inconsistent data. Further, given the remarks above about the ITE data being generally consistent when regressed against Q and V in ln form, while not necessarily accurate, (they appear a bit low when tested by the Perform program), the erratic TSR must therefore be the result of the erratic components of TSR or TSR as a whole (and that accepts that the DP measurement is accurate). With these results, no confidence can be placed in the Rugby ITE – DP (TSR) data and results for obtaining MR. Even where the constant and the coefficient are sensible, by sign and magnitude, the standard errors of the estimate are so high that the mean value is reduced to negative if two SDs are deducted from the mean.

The hypothesis can be put forward that the rapid to and fro movement on the Rugby plant distorted the results even after 1955. That fits with Chapelon's view that two-cylinder simple engines needed to be balanced to some 95% of the reciprocating masses to give acceptable results. At Rugby, a little extra reciprocating balance was added to a couple of classes where the proportion of reciprocating masses balanced was lower than average on some engines, but not all, and not to the extent of 95% suggested by Chapelon. Chapelon did not remark so far as I am aware about the balance of three and four-cylinder simple engines, but given different connecting rod lengths and drive on to different axles, they would have required reciprocating balance (GWR four cylinder engines had such), leaving some on a particular axle well below 100%, and subject to the same considerations as two cylinder engines. Or an hypothesis might be put forward that the to and fro forces were having a distorting effect, as implied in Chapelon's writings, but the origin thereof needs further thinking. Whatever, any TSR value will be subject to the SDP.

[Chapelon said quite clearly in five places that two cylinder engines did not give satisfactory results on testing stations on account of the recoil effect of the two and fro forces. (The sources for that are the Chapelon and Sauvage book *La Locomotive à Vapeur*, 1979 reprint, Section 77; his own book *La Locomotive à Vapeur*, 1935 edition, p 832; his 1952 paper *Conférences sur la Locomotive à Vapeur prononcées en Amérique du Sud* in 1952; and his comment p 137 of the Carling 1972/3 paper on Locomotive Testing Stations (Newcomen Society, Institution of Mechanical Engineers). He states that accurate answers for such locomotives on test plants required them to have 95% of the reciprocating masses balanced, which did not happen at Rugby. Note too that Carling did not explain why alterations to the plant in 1953 made the answers correct. I do not know any more about Chapelon's experience leading to these views. Further, the real problem which design and practice at testing stations in both France and the UK was avoiding resonant forces damaging the plant or its components, rather than achieving accuracy.]

Adrian Tester, who wrote a series of articles in *Backtrack* Vol 27 2013, about stationary testing plants, has informed me (personal communication) that Carling, superintendent of the plant, noted that the Amsler could record to +/- 1% for pull, and provided data within a +/- 1½% range for work done and +/- 2½% range for power (these are presumably at its own recording table, as might be expected from what these terms represent and the accuracy of the components. Only the pull, however, was recorded.

7 Relating Input to Output, Willans Line approach to Determining MR Directly, ITE made dependent on DP for 9F 92250 and Duchess 46225

Some Rugby data have been further analysed to test the idea that relating input to output can reveal the internal resistance between the input and output, in this case ITE to DP. This is not in terms of Q to DP, because on a steam locomotive, Q is first converted to ITE, and it is the relationship of ITE to DP which reveals TSR as used in this paper. As ITE is the independent variable in a relationship between ITE and DP, this case, performing a regression of ITE on DP is "back to front" in terms of the usual analysis based on cause and effect. The result is TSR, from which CWBR has to be deducted to give MR. For a 9F, CWBR by calculation is about 229 lbs.

The article by S J Pacherness, *A Closer Look at the Willans Line*, paper 690182, Society of Automotive Engineers, International Automotive Engineering Congress, January 1969, explains the underlying idea. If fuel is graphed as dependent linear variable against brake output of an internal combustion engine at a particular speed, as an increasing function, and projected back beyond the fuel line, the point where the graph line cuts the DP line, at zero fuel consumption, which occurs in the negative range of DP, represents, with the sign changed to positive, an approximation to the internal resistance of the motor. The slope of the line at any point is the specific rate of conversion of fuel to DP. If the graphed line at a particular speed is clearly a curve, ie Q is an increasing function or power function of DP, the tangent to the curve at any point projected back in the same way as the linear graph gives an approximation to the internal resistance of the motor at that speed and rate of working, and the slope to the tangent gives the specific fuel consumption at that speed and rate of working. Consistent derivatives can also be graphed. The fitting of the graph should be a regression in each case, but that is not said. In the automotive engine, the "friction" will include pumping losses and blowby. To result in correct MR, the engine must be working as it would be in use, and not be turned over by an external device. Numerous tests are said in the paper to give internal combustion MR of 6 to 8 psi.

For 9F 92250, using linear equations for each speed, this method yields an MR at 20 mph of 104, and at 40 mph of 18. At 30 and 50 mph, the constants in the relationship between ITE and DP are negative, which makes the method inoperative. All four equations, those for each speed, have excellent test statistics except that all have a low t score on the constant, which in turn leads to a high SEE, and inability to fix the location of the curve with any certainty.

For Duchess 46225, the equation to test this has been estimated in both linear and curved (power) forms (lnITE on lnDP).

A linear equation of ITE on DP is good statistically, ITE = 683.4 + 1.0199DP, signf F 3.08E-29, t on constant 4.51 and on coefficient 85.4, r^2 .997, standard error of the estimate 186.5. This results in a negative DP of –670 when ITE is zero. As there is a constant slope to the fitted line, that means MR + CWBR is 675 at all outputs at 50 mph, or MR alone is 446 lbs. Such constancy at all outputs should not be the case. The linear fit is based on observations of DP between 7373 and 17,085.

The curved form is ln ITE = c + b(lnDP), regressed on the 50mph data in ln form,

$\ln ITE = .650182 + 0.938868 \ln DP$, or $ITE = 1.9159DP^{0.938868}$ (a)

This is statistically a good equation, signf F 2.08E-28, t on constant 5.776 and on coefficient 78.3, r^2 .996. When DP is 0, ITE is 5, reflecting the problem of ln for 0 and 1. The differentiation of the curve to give the slope (dITE/dDP) reduces to 1.9159 x 0.938868 DP^-.061132, or 1.7988/DP^-.061132. The following shows the steps in obtaining MR for three trial values of DP within the data range at 50 mph:

DP lbs	Equival-ent ITE lbs (a)	DP^-0.061132	Slope dITE/dDP (b)	Equivalent horiznl of	MR+CWBR =(equiv hornzl of DP – DP) lbs	MR lbs

				DP = ITE/slope		
7,000	7806.7	0.5820	1.047	7456	456	228
10,000	10,910.6	0.5695	1.0248	10,647	647	419
16,000	16,962.3	0.5533	.9954	17,041	1041	813

by above equation (a), ITE = $1.9159DP^{0.938868}$
previous column x 1.7988
This method, although it can be applied, yields MR values which, by other analysis, are too low, by a considerable margin, in this case because DP measurement is not satisfactory. The slope of the curve of ITE on DP is too steep. In the range in which DP measurements occur, they are therefore too low, consistent with the conclusions above about the high values of DP and consequent low TSR.

8 Discussion

The Rugby data have been analysed in various ways, indeed all possible ways which reveal whether it is satisfactory, and what relationships are present in it, and in all cases, they are found wanting, being erratic and low for the circumstances. The bases of these conclusions have already been explained.

They only possible explanation of why that is so is the measurement system at Rugby. Scatter is unavoidable in investigations like these. But that does not mean that any old scatter is acceptable. The equations fitted are best fits – in probabilistic terms, nothing better is possible. If data are close to an expected or sensible relationship in physical terms, there will be a high probability that the relationships found are acceptable. Widely dispersed data do not do that. Scatter may be inevitable, but the more of it, or the less rational it is, the worse for the investigation. Just because data exist does not mean that they will be useful, and if they diverge widely from the expected relationship, and that relationship is correct, the worse for the data. All figures above show by observation that the data are far from conforming to the expected relationship; indeed, it has been a waste examining them as far as has been done here. They cannot be tampered with to “raise” them, by saying they are low only by reason of scatter. The experimental data have to be the basis of the investigation, not tampered figures.

To reemphasise, only the second approach depends at all on other analyses I have made, hence on my judgements of friction coefficients. All other of the four approaches rely on the data speaking for themselves.

There are several consequential comments.

Before I learned of the problems with the Rugby data, by examination and analysis, I earnestly hoped to obtain MR/TSR data from the tests done there, which data could be analysed to give reliable values for MR. I spent time at the NRM in 1988 extracting Rugby test data and reading files on the operations of the plant. I spent time since fruitlessly analysing those data, and from time to time testing out some new relationship or analytical approach to redeem those data, to no avail.

Doug Landau resolutely refused to declare is his letter what he does himself with Rugby data to convert it to MR, or LR, both for locomotives tested at Rugby, and for engines not tested there. Why the refusal, the sidestepping of the issue? What is he hiding? For this discussion to have been useful, other readers and I need to be informed how he drew conclusions on the Rugby data. Further, did he test the data, examine its soundness? How did he treat the SDP?

I shall write further comments on his letter, the great men approach and other, in due course.

10 Conclusion

Much of what Doug Landau has said about my previous letter amounts to unsupported declaration without analysis, tests or support. By example of what analyses he does, he is obviously not in a position to make these declarations. I consider his approach unscientific. I also consider his writing conclusions to his paper which have no relationship to the content of the paper to be dishonest. His motives for doing that are obviously not the purest. But I suspect they are to impress readers that his (unexplained) approach to obtaining TSR from the Rugby data is the correct one, and that such TSR values are good, and to deter readers from considering the matters put forward by John Knowles to be right.

Doug Landau's comments on regression are unsound, the fears expressed about consequence of its use groundless. They were made without any explanation of what are the supposed consequences of its use. Rather, regression is an essential tool in the analysis and explanation of experimental data. No one would now present a scientific paper relating to matters numerical, even less have it accepted for publication, without sections on examining or testing the data, and analysing the data for possible relationships in them, which analysis would be performed by regression (or similar). There are internationally accepted bases for drawing conclusions about the soundness of the results of analyses. It is not acceptable to deviate from them, and prefer Doug Landau's own (again unexplained), his attempt to justify use of data which to anyone else is not useful.

Given Doug Landau's stout defence of the Rugby data, his unwillingness to say how he uses the Rugby data to obtain MR and LR of locomotives generally is inexcusable in a scientific context.

What is said about repeatability of the Rugby test data is wrong and misleading. As is his comment that a key test of scientific proof is that its claims are consistent with the empirical evidence, when he accepts the empirical evidence without question or test.

From tests of the apparent soundness of the data, and relationships fitted the Rugby results on TSR are erratic, incapable of explaining the origin of TSR and on the low side.

I did and do not say that everything done at Rugby, its designers and operators, had shortcomings. That is Doug Landau putting words into my mouth. I certainly think that the measurement of the DP had shortcomings – it is hard to see otherwise. It might not have been possible for them to do better. By Carling's admission, elimination of the problems at Rugby would have required complete redesign of the plant, which was not done. At least Carling was clear that TSR at Rugby was not a sound measure of the internal resistance of the locomotive. It is also noteworthy that the DP data at Rugby were never published. What is known was obtained by me and a few others taking out the data at the NRM. I also think there were shortcomings in the whole philosophy and system of testing, even to having a Testing Station when there were the Mobile Testing Units, but especially the way of apparently or supposedly ensuring Q was a certain level on both the testing station and on the road. This is not the place to go into detail on these matters, but the same designers and operators, especially he senior ones, who Doug Landau is reluctant to see criticised, were involved in both the plant and road tests to some extent. But that criticism of Rugby is groundless is not right.

Last, this drawn out, often bad tempered, discussion on steam locomotive resistance has followed from a letter I wrote about the correctness of the second term in the usual formulae for railway vehicle resistance. It was Doug Landau who changed the subject to Steam Locomotive Resistance. Why did he do that? I find his motives questionable. In one sense, answering his erroneous notions is a waste of time, in another, it is useful if I can correct some of his ideas (as above). But no more than that. In my view he has not advanced the subject of steam locomotive resistance in these letters one jot. Overall, it would be better if this discussion were conducted in a peer reviewed scientific journal. For that to happen, he would need to learn about testing numerical data and scientific methods of analysing it.

Abbreviations

BPP	Blast Pipe (or Nozzle) Pressure.
BR	Braking Resistance
Cf	Coefficient of Friction
CO	Cut Off
CWVBR	Coupled Wheel Vehicle Bearing Resistance, without the wheels being powered
DBP	Drawbar pull (ontesting station)
DP	Dynamometer Pull
DR	Damping Resistance
EDBTE	Equivalent (to running on level track) Drawbar Tractive Effort
ITE	Indicated Tractive Effort
IHP	Indicated Horsepower
ln	In terms of Naperian logarithms
LR	Locomotive Resistance, basically VR plus MR
MR	Machinery Resistance, including the addition to CWVBR from the CWs being powered
PTTES	Piston Thrust Tractive Effort propulsive and compressive
$PTTEV^2$	Piston Thrust Tractive Effort forces from unbalanced reciprocating masses, dependent on speed squared
PTTE	The (net) sum of PTTES and $PTTEV^2$
Q	Steam Rate lbs per hour
SSC	Specific Steam Consumption, Q per Indicated Horsepower Hour
SDP	Small Difference Problem, as exists between two large numbers often or usually preventing exact measurement of the difference
SHM	Simple Harmonic Motion
SSC	Specific Steam Consumption (lbs per IHP hour)
TF	To and Fro (or Fore-and-Aft) Forces
TSR	Testing Station Resistance (ITE – DP)
V	Speed, mph
VR	Vehicle resistance
WRTE	Tractive effort (normal definition, cf PTTE) at coupled wheel rims
WRHP	WRTE as a HP

Descriptions of statistical tests are not given. (Standard Error of the Estimate, Significance F, t, r^2, Standard Deviation) can be found in Statistics texts.

John Knowles

4th July 2017