Appendix 1. Previous study into wheel slide prediction

Title:

Predicting wheels slides; traditional and future methods

Authors & affiliations:

H J Parkinson, G Bamford †,*
**Digital Rail Ltd, hjparkinson@digitalrail.co.uk † Advisory & Risk Management Solutions, Gary.Bamford@advisory-ltd.com*

Introduction

Wheel slides cost the railway millions of pounds every year in engineering costs and delays due to wheel flats, and seem to be caused by a multitude of events including train performance, weather conditions, time of year, train condition, track quality, track cleaning approaches, leaf-fall data, level crossing proximity, train driving policy and more. Taken together these multiple data streams would constitute big data and now is the era of big data. We have introduced the ELBowTie [1] big data concept of using multiple data streams to provide enhanced management of risk.

Data streams require various analytics and we are currently investigating Natural Language Processing, Autonomous Vision, and Neural Networks to analyse safety processes, track defects, and passenger risk, respectively [1]. There are many detectors and sensors on trains and track providing masses of data with the majority of this is in the form of time series data points. There have been significant advances already in the asset management arena utilising conventional analysis of this data. which in general simply analyse each individual data source in isolation. The question is 'why bother with anything beyond this'? Techniques such as machine learning and AI are being considered but are these just a fad or would they take the analysis to a new level.

This work describes how we are seeking to make sense of a wide variety of data sources to make real savings on wheel flats. We examine key parameters and analyse the data using conventional statistical methods to look for insights that could help to manage the problem. We investigate the limitation of this approach and suggest a method for improving the approach.

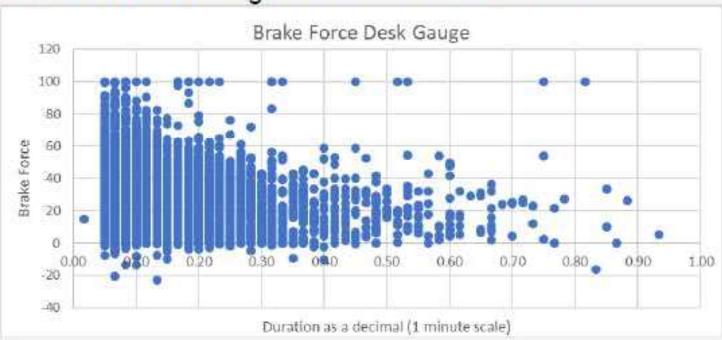
Methods

We have taken data kindly supplied by Siemens who maintain rolling stock in the North of England. The data includes train position (from GPS), braking and power data, station dwell data, wheel slide data, sanding application data, speed, acceleration and so on. The focus is on wheel slides. The first question to address was, what variables are important in this analysis i.e., what drives wheel slides, by applying a traditional approach, could we find the root cause of these events (without any advanced analytics). We used a relatively simple focus on the outcome of wheel slide and a traditional approach based on correlations between data streams. The rational being that correlations imply some form of causality between data streams which can then be investigated and mitigated.

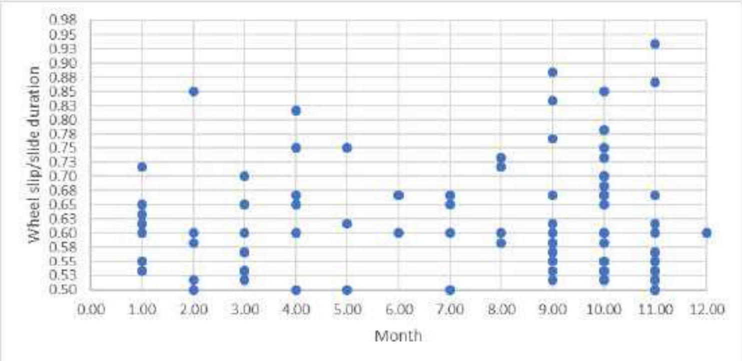
We begin by evaluating correlations between what would be thought of as the main contributors. The data we analysed covered: braking force, time of year, time of day, rainfall, temperature and locations. The data set was 12 months of train data. In the results section below, we show comparisons of brake force versus duration of the slide, wheel slide duration versus time of year and then the against time of day. We then built up a route map of slide black spots 30 seconds in duration and looked to see if any general patterns were visible. This analysis took into account the known timetables on these routes in order to normalise the data.

Results

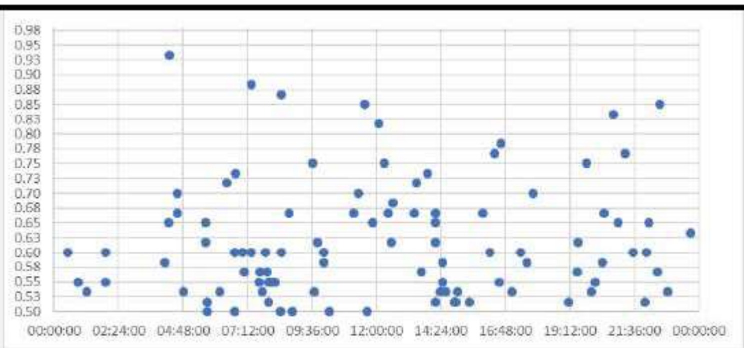
The graph below shows all data received, for all wheel slides. There are some anomalies in the data set which we have not had time to clean. There does not appear to be an increase in wheel slide duration resulting from increased brake force.



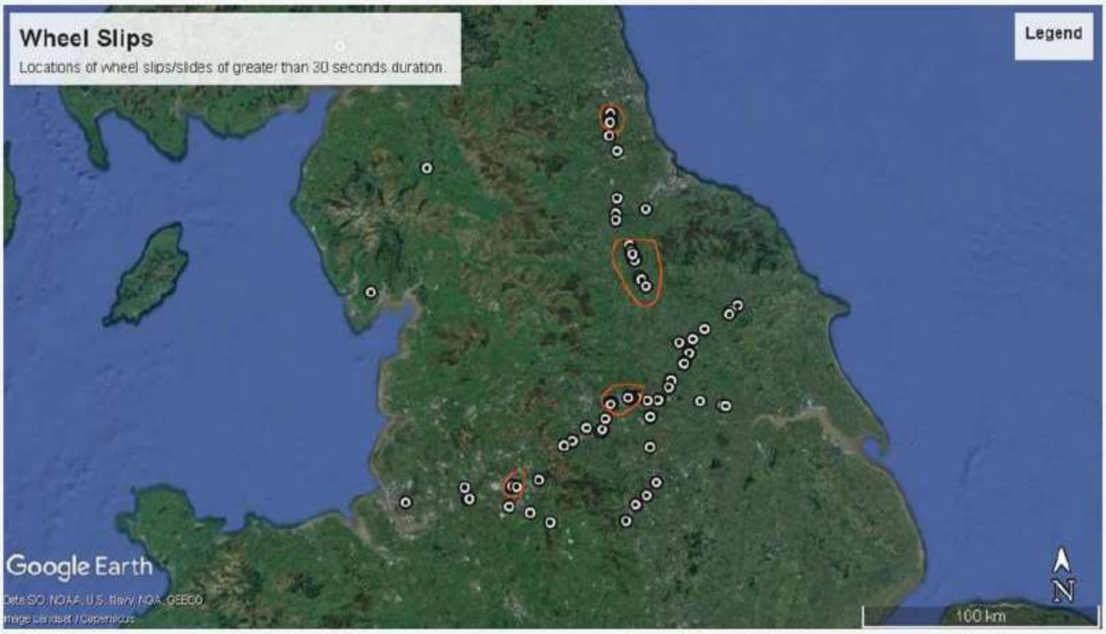
To narrow down our data set we decided to look into those wheel slides which are over 30 (0.5) seconds in duration, which we decided would be classed as significant, which we then compared against the time of year. All the figures from now on related to slides over 30 seconds. The graph below shows that there appears to be more occurrences during winter months.



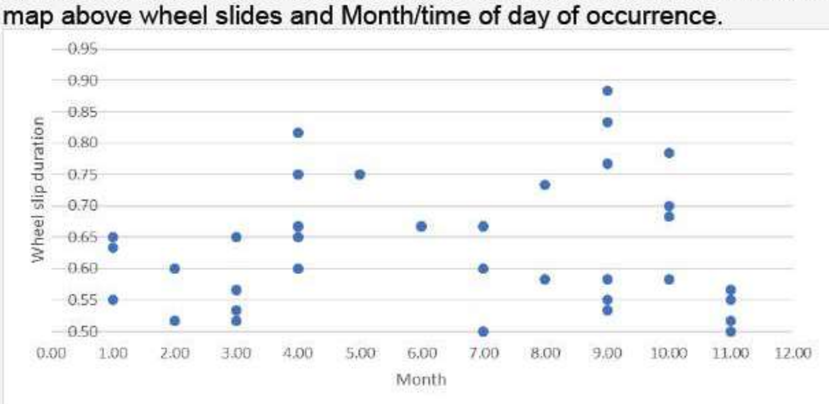
The graph below shows wheel slide duration compared to time of day of occurrence.

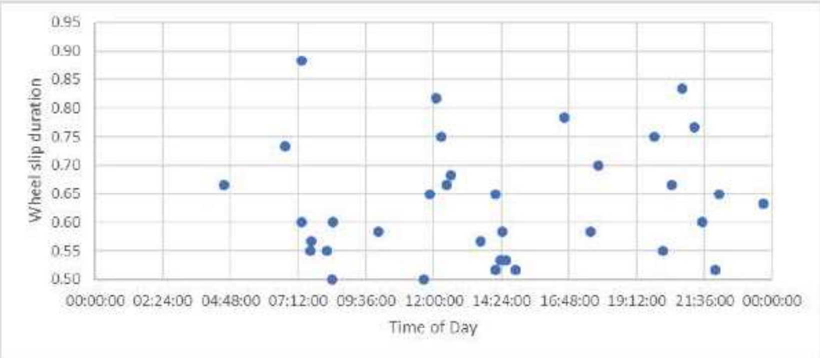


The below map shows all wheel slides over 30 seconds in duration. We can see very few wheel slides over 30 seconds occur on the west coast line. The map has been highlighted to show the areas where the highest number of wheel slide/slides have occurred over 30 seconds. These tend to be on the approach to cities or at stations that have long unobstructed approaches.



We looked see if there was any correlation between these slides in the 4-area s indicated in the map above wheel slides and Month/time of day of occurrence.





The analysis shows;

- Wheel slides occur at all braking force applications, very large wheel slides occur at low braking forces!
- For the whole data set here are no discernible correlations with time of day, temperature or rainfall
- Very large slides (over 30s) also show no correlation with time of day, temperature or size of braking force
- Very large slides do however show some location specific correlations. Some other factors are driving these beyond the environmental factors analysed above. These could be for example, geography, train configuration, track layout, other operational issues, dirty track, type of traction.

Conclusions and Contributions

The question is when does this traditional analysis end? We cannot keep on adding a data set and guessing what data may or may not be useful. Hence the need to use advanced analytics techniques. If we do not use a traditional approach, what would be needed to set up a more complex analysis and how wide would we need to cast our data analysis?

Utilising the ELBowTie framework we are currently instigating the following approach. Identifying network conditions that lead to known causes of wheel slide to provide an exhaustive list of parameters to create a data-driven, multi-layer probabilistic model for predicting system conditions affecting railway systems. A data sandbox will be developed to enable automated analysis and modelling of the network in real time and to provide recommendations to mitigate conditions that cause slides or means of control. The approach will be to predict and avoid wheel slides and reactionary delays in the rail network by cross-correlation of large heterogeneous data sources.

This paper has shown the analysis of the correlations between the data sets and wheel slides is complex and contributes to the existing body of work as it:

- It provides insight into the difficulties of analysing what appears on the surface to be a simple problem
- Demonstrates that traditional analysis techniques are shown to be of limited value, time consuming and do not get to the heart of the problem
- It demonstrates the need for more advanced analytics and presents these approaches
- It proposes a way to begin the analytics of this data.

1. Parkinson H J, and Bamford G, Big data and the virtuous circle of railway digitization, Springer, 20 Oct 2016: Advances in Big Data: Proceedings of the 2nd INNS Conference on