



## **Interpretation of Railway On-Train Data of Electrical Train Service (ETS) in Malaysia**

**Muhammad Azki Hafizi Mohamed Nazul<sup>1</sup>, MZ Ngali<sup>1\*</sup>, Jalil Azlis-Sani<sup>1</sup>**

<sup>1</sup>Faculty of Mechanical and Manufacturing Engineering,  
Universiti Tun Hussien Onn Malaysia, Batu Pahat, 86400, MALAYSIA

\*Corresponding Author Designation

DOI: <https://doi.org/10.30880/rpmme.2022.03.01.058>

Received 15 Nov. 2021; Accepted 15 April 2022; Available online 30 July 2022

**Abstract:** On train data recorder (OTDR) widely use in rail transportation. Industry 4.0 is currently trend on extensive use of data. Railway transportation is complex Railway systems are technologically complex, including diverse human actors, organizations, and technology solutions. The utilization of OTDR currently has been used for the purpose of monitoring system. Form existing on board data this study discovers new method interpreting data for other purpose. The objective of this study is to interpret raw mechanical data from *Keretapi Tanah Melayu Berhad* (KTMB). This study focusses on pre-existing data with introducing different tools of interpreting from previous research. The method was used based on research methodology and finding on literature review to achieve research objective. A few recommendations were suggested to use future study with using OTDR as the main source of data.

**Keywords:** On-Train Data Recorder (OTDR), Electric Train Service (ETS), Data Interpretation

### **1. Introduction**

According to the Ministry of Transport (MOT), a statistic of rail transport 2020 in the first quarter over 800,000 passengers Electric Train Service (ETS) [1] have used the railway service to make a living. ETS had been qualified as a High-Speed Rail train service with a maximum speed of 140 km/h on an electrified meter gauge rail line. The rolling stock has been equipped with robust safety features to keep the dumping conditions and assets excellent. Other than that, Safety measures had been taken, ensuring each train has its recording device to monitor the driving style of the train driver with a view to proof in the event of an accident. Interrogation of OTDR data can develop Key Performance Indicators for specific trainsets, assisting in efficient maintenance resources. Therefore, this research will be an opportunity to see a different view of OTDR to understand better to give benefits for maintenance or other purposes. This paper presents the interpretation and analysis of railway on train data and contributes the further recommendation research On-Train Data Recorder (OTDR).

OTDR's are used to collect data about train operations and the state of different train systems throughout a voyage. These variables include the location of the power and brake controllers and the

---

\*Corresponding author: [zamani@uthm.edu.my](mailto:zamani@uthm.edu.my)

2022 UTHM Publisher. All right reserved.

[penerbit.uthm.edu.my/periodicals/index.php/rpmme](http://penerbit.uthm.edu.my/periodicals/index.php/rpmme)

driver's acknowledgment of signaling system alerts[2]. The previous accident reports used a train data recorder to investigate fatal accidents on pedestrian walkways by reading signal records involving two trains produced by Rail Accident Investigation Branch (RAIB)[3]. The following accident report reveals the continuous brake applied and sounding horn passenger's trains before hitting the car [4]. Balfe, in 2017, created a framework for human factors analysis by systematically analyzing OTDR application in human performance analysis [5]. Subsequent research represents a technique to draw out-train driver workload by creating measurement and investigation on an actual case study using OTDR[6]. As a result, OTDR can show a significant workload determined by the increasing number of actions per minute during arrival. The research continues with Geoghegan's 2020 investigation of the data generated by OTDR and its applicability in analyzing train driver and infrastructure performance[7].

According to Walker and Strathie, train recorder data are an underutilized but potentially critical data source for evaluating human performance and spotting hazards before accidents[8]. In the following study, they were analyzing on-train data recorders using link analysis and related graph theory. The study diagrammatically connected each control action taken by a driver to the following control action taken by the driver, allowing for an additional examination of driver styles and differences[9]. Moreover, there were continue their research that offers various proof-of-concept demonstrations to show that long-standing ergonomics approaches may be driven from big data and succeed in providing insight into human performance in a unique way[10].

Indicated of safety operational, OTDR one of the cheap sources for monitoring train safety systems. This approach determines a driver's level of compliance with specific safety system regulations without disturbing the driver. El Rashidy in 2016[11] three safety systems had been selected, which are Emergency Bypass Switch (EBS), the Train Protection and Warning System (TPWS), and the Driver's Reminder (DRA) had used as an initial analysis of OTDR. Later on, a study proposed using train driver competency performance indicators (DCPIs) to assess driver competency exploring the utilization of OTDR. Some drivers displayed a wide range of reaction times, while others consistently responded quickly. Although the data does not support it, this consistently fast response could indicate an automatic response to the warning, implying that the driver does not process and understand the notice consciously. According to El Rashidy, such data can be used to create an automated system to monitor driver competency[2].

## 2. Methods

This study aims to interpret the raw mechanical data into a meaningful understanding graph based on the objectives of this study. This part explained the tools and the flow of the process to achieve the aim of this study.

### 2.1 Methods Flowchart

Generally, the function of the flowchart is to provide information on the work process. Hence, a methodology flowchart was used to briefly explain the process of interpreting data to achieve this study's objectives. Based on the flowchart shown in figure 1, the process begins with obtaining the raw data from KTMB as the core subject for this study.

The pre-processing to get as much as possible the information of interpretation process on the previous study. Then, the process will continue from raw data until it can be a representable graph. Second, editing and coding data is the challenging part of this study. We are using MATLAB to achieve the primary process. Next, after completing the immediate process, data could be processed into the graph and do some analysis before deriving the equation for the secondary process. Last but not least, giving the recommendation research On-Train Data Recorder (OTDR).

## 2.2 On-Train Data Recorder

OTDR is a device that can collect data about train operations systems such as train speed, braking, brake pressure, powering, etc. Most of the signals are digital and in nominal/binominal (0/1), with exceptions to train speed (in km/h) and brake pressure (in bar) [7]. The raw data containing a maximum data matrix of 65,554 samples (row) with 31 parameters (column) on the single file of record. In this study, there were two data had been given. The first data was recorded from 15th January 2020 until 17th January 2020, and the second data is from 1st September 2019 until 30th September 2020. This study was using TELOC 2500 and Hasler 1500 to acquire the raw data.

## 2.3 MATLAB

Visualize data with new bubble and swarm graphs and configure graphs with a new title, mark, and axis boundary choices. Diff and use the MATLAB Comparison Feature to combine Software Creator applications and add unique figure icons and template elements to your MATLAB software. For this study, we were using MATLAB as computer programming software for analyzing OTDR. This study is the first research using MATLAB to decode data to visualize the graph.

## 2.4 Pre-Processing Phase

Pre-processing is essential for this study before initiate the interpretation process. This process is the first step in interpreting data to avoid any overlap data and error data. Tavel P said we must remain vigilant when selecting data dimensions to visualize to avoid losing intriguing patterns[12]:

- i. Identify the appropriate parameter of this study which is there are seven parameters are commonly used in previous research.
- ii. Get essential information of raw data to obtain as much detail as possible to pre-concept this research before elucidating data for any information such as timetable, type of train, departure station, and any train characteristic that can contribute comprehension analysis data.
- iii. Recognize the ETS route using Google Maps to get the exact location based on real-time situations and estimate the time to reach each station.
- iv. Discover suitable visualization tools for extensive data analysis, which is a comparison between MATLAB and Excel.

## 2.5 First Phase: Interpret the raw data on primary process.

Firstly, import one of the excel file into MATLAB by using the command window. This process takes a few seconds before it can fully enter MATLAB, depending on the data size. Secondly, declare a variable for parameters in the file with numbers, text, and raw data, including string import into MATLAB. Every parameter needs to cautious by giving their name for a variable to avoid confusion. This data will run into a loop which this data will be screening from top to bottom. Next, this data will go into two conditions. The first condition is separating the data by using the time parameter. After the state valid, it will save the previous input and start the following information. The next-day trips terminate this condition by converting the time parameter from hours into seconds. On the other side, another variable changes the Not a Number (NaN) value into zero. After the condition is proper, the NaN value will change into zero, and it carried forward until the finish iterate for all data. The last condition we set is all the data we place, it must be read, and if this data had not finished yet, it would return into variable until it ends. After completing all data, the program will set for reading back all the variables to generate graphs.

## 2.6 Second Phase: Derive Equation for Secondary Process

In the Secondary process, we identify appropriate parameters by utilizing raw and primary process data for extra comprehension data analysis on this study. For instance, there are two suitable parameters to be used as a reference for this study:

i. Acceleration

In Britannica[13], acceleration means the rate of velocity change with time in both speed and direction. In other words, acceleration is when an object is speeding and slowing from point A to B. An algorithm is developed to plot the train's acceleration against the relative journey of time to achieve our objective. Here the formula of acceleration:

$$Acceleration (ms^{-2}) = \frac{Current\ Velocity (V_2) - Previous\ Velocity (V_1)}{Current\ time (T_2) - Previous\ Time (T_1)} \quad Eq. 1$$

Referring to equation 1, current velocity and previous velocity can be obtained in the raw data based on train records concurrent with time. This acceleration will be plotting with respectively in the relative time of train journey graph.

ii. Average of Speed

From dictionary, the average speed divides the total distance travelled by the elapsed time. For example, the entire length for KL-IPOH is 208 km and the whole time taken for ETS from KL Sentral to Ipoh is 2 hours 45 minutes. That means the average speed is approximately 75 km/h. That's kind relatively easy if we have only one or two data. But if we have extensive data, it will take a lot of time to process this data. For this, we show you the formula we use to decode this problem by using computer programming.

$$Average\ of\ Speed \left(\frac{km}{h}\right) = \frac{Total\ Speed\ Record \left(\frac{km}{h}\right) \times Total\ Time\ of\ travelled (h)}{Total\ Time\ of\ travelled (h)} \quad Eq. 2$$

### 3. Results and Discussion

#### 3.1 Result of primary process interpretation

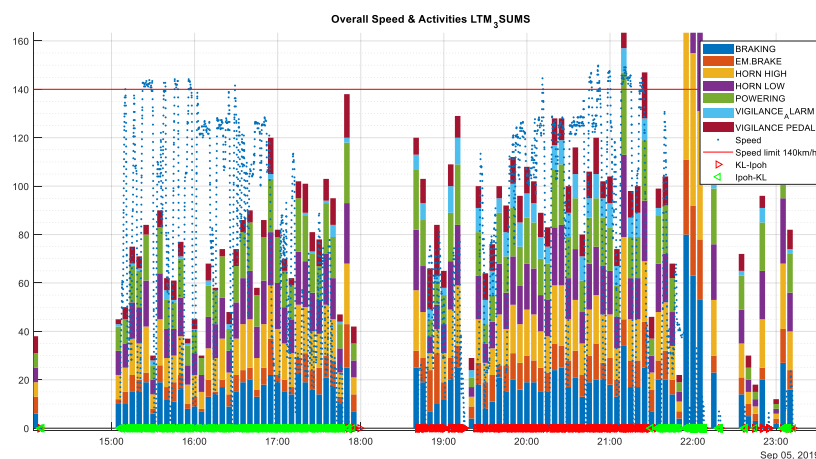
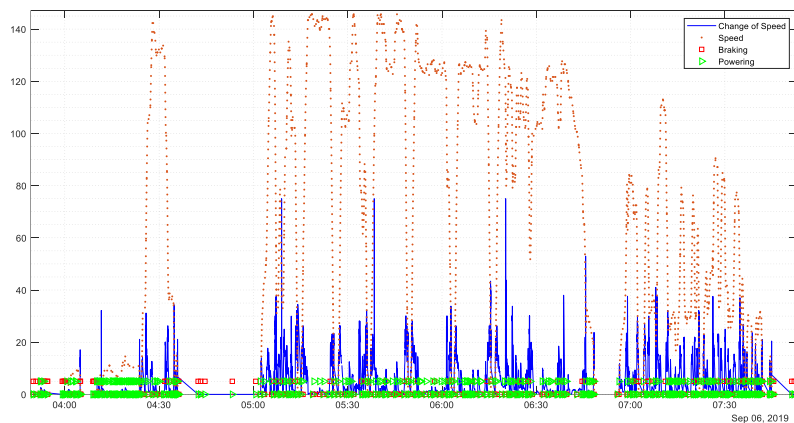


Figure 2: Result of primary process

Figure 2 shows the result of primary process interpretation from raw data into this graph. This graph consists of seven parameters:

- i. **Speed**  
This figure shows the driver tends towards higher speeds before it stops at the Ipoh Station. It starts with low velocity and a lot of fluctuation at the beginning of the journey, while in the half of the trip, it starts to exceed more than 120 km/h. We can determine the distance for each station by calculating the area under the graph.
- ii. **Braking**  
Train drivers apply the brake to reduce the train to stop at a specific station based on the route that had been set. There are two bar graphs indicating braking: the blue color for the service brake and the orange color represents the emergency brake.
- iii. **Emergency Brake**  
For safety reasons, the emergency brake could be applied when the driver exceeds 140 km/h, and the Emergency Bypass Switch (EBS) will prevent the train from overreaching the safe speed limits[11]. Emergency brakes could damage some of the wheel protection if they are not using wisely. This data shows the frequency of trains experiences the emergency brakes every five minutes.
- iv. **Horn Low/Horn High**  
Horn is a compulsory activity for a driver to ensure the driver always stays awake and cautions with his driving. Train drivers sound the train horn at certain places across the network and frequently when they arrive or exit a station.
- v. **Powering**  
Powering is a significant signal to show the increase of velocity. In other words, powering supplies an electric source to a moving train in continuous running along the rail.
- vi. **Vigilance Alarm/Vigilance Pedal**  
Vigilance alarm and pedal are safety devices in the event of the engineer's incapacitation, and this system takes over. It's a cross between vigilance control and a dead switch. The fundamental flaw in the basic dead-control man's design is the risk of the functioning device being kept in place continuously, either consciously or unconsciously.

### 3.2 Result of secondary process interpretation



**Figure 3: Result of secondary process interpretation**

- i. **Acceleration**  
The effect from powering worked on acceleration graph—this graph shows the rate of velocity change with velocity versus time. Acceleration and deceleration will go for the same output because we are setting for absolute value in programming.
- ii. **Average of Speed**  
Average speed has the same purpose of knowing driving behavior for the whole journey. It accommodates researchers to identify the minimum rate of moving with the punctuality on stop schedule.

### 3.3 Discussion of overall process of interpretation

Discovery of valuable findings of OTDR data from both process interpretation data. Hence, three results valuable to this study:

- i. **Terrain and urban density**  
The uneven terrain impacts a personal driving style. In areas of high population density, urban areas are less suitable for carrying trains at high speed due to the nearby stopping stations.
- ii. **Human Factor and Driving Behavior**  
In earlier chapter 2, most researchers used OTDR for Human Factor study or driving behavior study. As we get the pattern of the graph, almost 90 % followed with their data interpretation. Most of the results for this study towards human factors use every minute on a train journey to analyses the number of actions[6]. In comparison, use OTDR as the example check and balance with the safety system[11]. As for utilizing OTDR data to unveil the possibility of reducing risks of human error with build-up workload profiling[14].
- iii. **Effects on wear and tear**  
Humans and machines are related if they use or handle it. The impact of driving behavior towards wear and tear for equipment is deniably. Human factors and wear and tear are correlated with each other, but we cannot prove it with any relevant data or statement in this research. The following study may include maintenance data and driver schedules to determine driving behavior's effect on wear and tear.

## 4. Conclusion

In conclusion, interpretation of railway on-train data was successful to decode the big data into presentable graph. This study was using two processing phases in order to obtain refined graph. This study shows the potential interesting OTDR in big data research and practice. Other than that, this research can be more vital if there have a comprehensive data including maintenance and worker schedule.

## Acknowledgement

This research was made possible by funding from research grant number H896 provided by the Universiti Tun Hussien Onn Malaysia with collaboration Keretapi Tanah Melayu Berhad (KTMB). The authors would also like to thank the Faculty of Mechanical and Manufacturing Engineering, Universiti Tun Hussein Onn Malaysia for its support.

## References

- [1] M. of T. Malaysia, “Statistic of Rail Transport (Q1-Q4) 2020,” 2020. .
- [2] R. A. H. El Rashidy, P. Hughes, M. Figueres-Esteban, and C. van Gulijk, “Automated train driver competency performance indicators using real train driving data,” *Saf. Reliab. - Safe Soc. a Chang. World - Proc. 28th Int. Eur. Saf. Reliab. Conf. ESREL 2018*, pp. 3071–3076, 2018, doi: 10.1201/9781351174664-384.
- [3] R. A. I. B. (RAIB), “Rail Accident Report: The fatality at Barratt’s Lane No.1 footpath crossing 21 November 2005,” 2006.
- [4] R. A. I. B. (RAIB), “Rail Accident Report: Passenger train collision with a road vehicle at Swainsthorpe level crossing, Norfolk 13 November 2005,” 2006.
- [5] N. Balfe, “A framework for human factors analysis of railway on-train data,” vol. 4959, 2017, [Online]. Available: [www.hfes-europe.org/wp-content/uploads/2016/11/Balfe2017.pdf](http://www.hfes-europe.org/wp-content/uploads/2016/11/Balfe2017.pdf).
- [6] N. Balfe, K. Crowley, B. Smith, and L. Longo, “Estimation of train driver workload: Extracting taskload measures from on-train-data-recorders,” *Commun. Comput. Inf. Sci.*, vol. 726, pp. 106–119, 2017, doi: 10.1007/978-3-319-61061-0\_7.
- [7] N. Balfe and S. Geoghegan, “Human factors applications of On-Train-Data-Recorders,” pp. 152–161, 2020.
- [8] G. Walker and A. Strathie, “Combining human factors methods with transport data recordings,” *Adv. Hum. Asp. Transp. Part*, vol. 2, 2014.
- [9] A. Strathie and G. H. Walker, “Can Link Analysis Be Applied to Identify Behavioral Patterns in Train Recorder Data?,” *Hum. Factors*, vol. 58, no. 2, pp. 205–217, Dec. 2015, doi: 10.1177/0018720815613183.
- [10] G. Walker and A. Strathie, “Big data and ergonomics methods: A new paradigm for tackling strategic transport safety risks,” *Appl. Ergon.*, vol. 53, pp. 298–311, 2016, doi: 10.1016/j.apergo.2015.09.008.
- [11] E. L. Rashidy *et al.*, “University of Huddersfield Repository DRIVER COMPETENCE PERFORMANCE INDICATORS,” 2016.
- [12] P. Tavel, “Modeling and Simulation Design. AK Peters Ltd,” in *Proceedings of the Human Factors Society 36th Annual Meeting, A usability and diary study assessing the effectiveness of call acceptance lists*, 2007, pp. 216–220.
- [13] “acceleration | Definition, Facts, & Units | Britannica,” *Encyclopædia Britannica*. 2021, [Online]. Available: <https://www.britannica.com/science/acceleration>.
- [14] E. L. Rashidy *et al.*, “University of Huddersfield Repository Operational Safety Indicators Using Real Train Driving Data,” 2017.