Monitoring train driver fatigue through operational performance data

Rebecca Kearnes[1], Mark S. Young[1], Nora Balfe[2] & Megan Thurgood[2]

[1]Transportation Research Group, University of Southampton, [2]Irish Rail

---

SUMMARY

Fatigue presents a significant risk for safety-critical systems such as the railway. Reaction time, which is crucial in tasks such as train driving, offers a valid proxy for fatigue. This study analysed driver reactions to vigilance alerts from on-train ('black box') data to show that both temporal and workload factors can affect driver performance, demonstrating the feasibility of this method as an objective means of detecting early indicators of driver fatigue.

KEYWORDS

fatigue, reaction time, train driving

Background

In safety-critical systems such as the railway, fatigue presents a significant risk as it can reduce alertness, impair motor control and increase reaction times (Caldwell et al., 2019; Völker et al., 2016). In particular, train driving demands continuous vigilance; even slight increases in reaction time can result in safety-related consequences, as drivers might fail to respond in a timely manner to alarms and alerts, resulting in a brake intervention. Fatigue-related response delays have been well documented in both simulated and real-world scenarios (Dorrian et al., 2006; Sallinen & Kecklund, 2010). These impairments are particularly evident during periods of cumulative fatigue or circadian disruption (Gander et al., 2017; Xu & Hall, 2021).

Reaction time is increasingly recognised as a valid proxy for fatigue. Research shows that reaction time reflects not only acute fatigue but also its progression across a single duty period or multiple consecutive shifts (Gurubhagavatula et al., 2021).

The level of workload while on shift also contributes to fatigue in two ways. High workload, as a result of excessive task demands overwhelming attentional resources, leads to active fatigue. Meanwhile, passive fatigue is a result of under-stimulation, such as during continuous low-demand train driving (Ma et al., 2018). This latter form can gradually deplete attention without the driver necessarily feeling overloaded (cf. Young & Stanton, 2002), and has been attributed as a factor in fatigue-related accidents (RAIB, 2020). The combination of active and passive fatigue is consistent with the inverted-U performance curve, whereby either too little or too much demand is detrimental to performance (McWilliams & Ward, 2021).

In common with other safety critical industries, fatigue risk management (FRM) is applied to various degrees in the rail industry (Gander et al., 2011). The most visible element of FRM in UK rail is limits on working hours (manifested through the so-called 'Hidden limits', which are now seen as out of step with current fatigue science) and the use of related biomathematical tools for roster assessment, typically through the Fatigue and Risk Index calculator (FRI; see e.g., Folkard et al., 2007). Education and awareness raising, as well as medical screening for sleep disorders, are also common. But methods and technology to monitor fatigue risk remain less utilised across the sector. In south London, Tram Operations Ltd. successfully implemented a camera-based driver monitoring system to detect fatigue and distraction episodes, alongside other interventions as part of a wider FRM system (see Townsend et al., 2023). Its driver monitoring system used technology more familiar in the road haulage sector, but the application of such technology in heavy rail is unproven and, consequently, is met with challenge from drivers and trade unions. Furthermore, because the technology only identifies sleep onset, and not the earlier indicators of fatigue, the benefits are largely in terms of identifying fatigue trends over time. Therefore, it is worth exploring other methods could be both more acceptable and more useful in identifying fatigue earlier, facilitating proactive intervention to control the associated risk.

In the rail industry, recent work (Balfe, 2017) has explored using on-train data recorders (OTDRs, or 'black boxes') to monitor driver performance in real time. These metrics offer the potential to capture fatigue-related performance degradation in operational contexts, offering advantages over subjective (self-report) data (which may be influenced by organisational culture or other biases; Young & Steel, 2017) and other objective methods (such as physiological monitoring) which can be intrusive. However, further development and validation work is needed to determine which OTDR measures relate to fatigue and how fatigue influences such performance.

This study focuses on reaction time metrics from the OTDR with the aim of detecting early indicators of fatigue and, ultimately, using driver performance data as a practical approach for fatigue monitoring in real-world rail operations. The key research question concerned whether reaction time varied as a function of elapsed journey time, time of day, and workload.

Method

OTDR data were collected from 13 return journeys across four routes, covering a period between May and July 2025. Each journey represented an outbound and return trip (i.e., 26 trips in total), with the average trip duration being 2 hrs 37 mins. Drivers completed an outbound trip and then undertook the return following a scheduled break, which averaged between 1 hr 8 mins and 1 hr 53 mins, depending on the destination. All journeys took place at broadly similar times of day, departing between 13:30-14:00 hrs, which helps to control for time of day effects in the analysis.

Reaction times were derived from the OTDR data as responses to the driver vigilance alarm, an onboard system intended to ensure that drivers are alert. For the rolling stock used in the present study, this is implemented in a speed-dependent way. At speeds below 120 km/h, an alarm sounds on a fixed cycle every 90 seconds, which the driver must respond to within 2.2 s by releasing and depressing a foot pedal, otherwise an automatic brake intervention is made (Balfe, 2017). At speeds of 120 km/h and above, the 90-second vigilance cycle is also reset through other driving actions (e.g., adjusting the power controller, sounding the horn), so the alarm might not necessarily sound if other controls are being used. Consequently, this analysis only included responses using the foot pedal.

The derivation of reaction time involved some additional processing. The OTDR records timestamps for vigilance alarm activation and reset (as a result of the driver releasing the foot pedal); the interval between these points reflects the driver's reaction time. However, the low resolution of these timestamps prevented precise measurement. Instead, a more accurate measure of time was calculated from the speed and distance channels on the OTDR (which were recorded to higher resolutions). Furthermore, four files contained an unusually low number of recorded vigilance alarm events. Upon investigation, these four files represented two round trips (i.e., outbound and return) by the same driver, whose apparent driving style involved frequently releasing the foot pedal as a means of pre-empting the vigilance alarm. As such, these journeys provided little insight into performance and so they were excluded from the analysis.

For each trip, speed profiles were developed to infer workload demands along the route. It was assumed that frequent or irregular changes in speed increased workload, whereas extended periods of constant speed reduced workload. Each route involved a similar number of station stops, but two of the routes involved more consistent speed profiles, whereas the other two were more variable. In each case, though, drivers rarely experienced prolonged periods of monotony in task demands; periods of constant speed were relatively brief (not exceeding 10 minutes). Nevertheless, previous research indicates that low task demands can induce a state of underload well within 10 minutes (Young & Stanton, 2002).

To define low workload sections of the route, three conditions were applied: consistent speed profile across all journeys, duration over five minutes, and speed variation within 15 km/h across the entire period. Based on these criteria, all except one route contained identifiable low-workload sections.

Results

To test whether reaction time varied systematically during low workload periods, a mixed-effects linear model was applied with reaction time as the dependent variable, elapsed time as a fixed effect and driver ID (anonymised) as a random intercept to control for baseline differences between individuals. This model showed that time into low workload period was not a significant predictor of reaction time ($F(1,154) = 1.314$, $p = 0.253$), with an estimated slope of effectively zero.

Random effects analysis, meanwhile, highlighted substantial between-driver variability (σ^2 intercept = 0.058, $SE = 0.030$) as well as residual within-driver variance (σ^2 residual = 0.044, $SE = 0.005$). The intraclass correlation coefficient (ICC) indicated that around 57% of total variance was attributable to stable individual differences, with the remainder reflecting within-driver fluctuations.

Thus there was no evidence that reaction times varied as a function of time into low workload periods. However, there were pronounced individual differences in baseline reaction times between drivers, demonstrating the importance of modelling repeated measures with random effects to control for baseline differences between individuals.

Although reaction times did not systematically vary within low workload periods, it is still relevant to compare workload at the broader level by distinguishing low from ‘non-low’ workload sections (i.e., those that did not meet the above definition of low workload – which does not necessarily mean that they were high workload sections, hence the deliberate terminology of ‘non-low’). Adding workload as a predictor into the model therefore allows for a direct test of whether reaction times differed between workload conditions, while controlling for elapsed journey time and time of day. These variables were modelled separately as fixed effects owing to a moderate positive association between them, as evidenced by a significant Pearson’s correlation ($r = 0.516$, $p < 0.001$), meaning that the two predictors cannot be considered fully independent.

The first model, which included elapsed journey time, revealed several significant effects. Reaction times were lower during low workload periods (estimate = 0.085 s, $F(1,2067) = 7.77$, $p < 0.01$) and also increased as the journey progressed ($F(1,2067) = 12.676$, $p < 0.001$), although the effect size was minimal (slope $\approx 1.7 \times 10^{-5}$ s per second). A significant interaction between workload and elapsed time was observed ($F(1, 2067) = 6.71$, $p < 0.05$), as reaction times tended to rise across non-low workload trips but decreased slightly in low workload periods (see Figure 1). Again, there was substantial between-driver variability, accounting for about half of the variance (ICC = 0.50).

The second model used time of day as the predictor. Although there was no significant main effect of time of day on reaction time ($F(1, 2067) = 0.216$, $p = 0.642$), reaction times did again differ significantly by workload, but now they were higher during low workload periods (estimate = 0.461 s, $F(1,2067) = 13.872$, $p < 0.001$). A significant interaction between workload and time of day was observed ($F(1, 2067) = 13.072$; $p < 0.001$); reaction times decreased slightly later in the day across non-low workload periods, whereas they remained stable in lower workload periods (see Figure 2). Once more, there was substantial individual variability (ICC = 0.49).

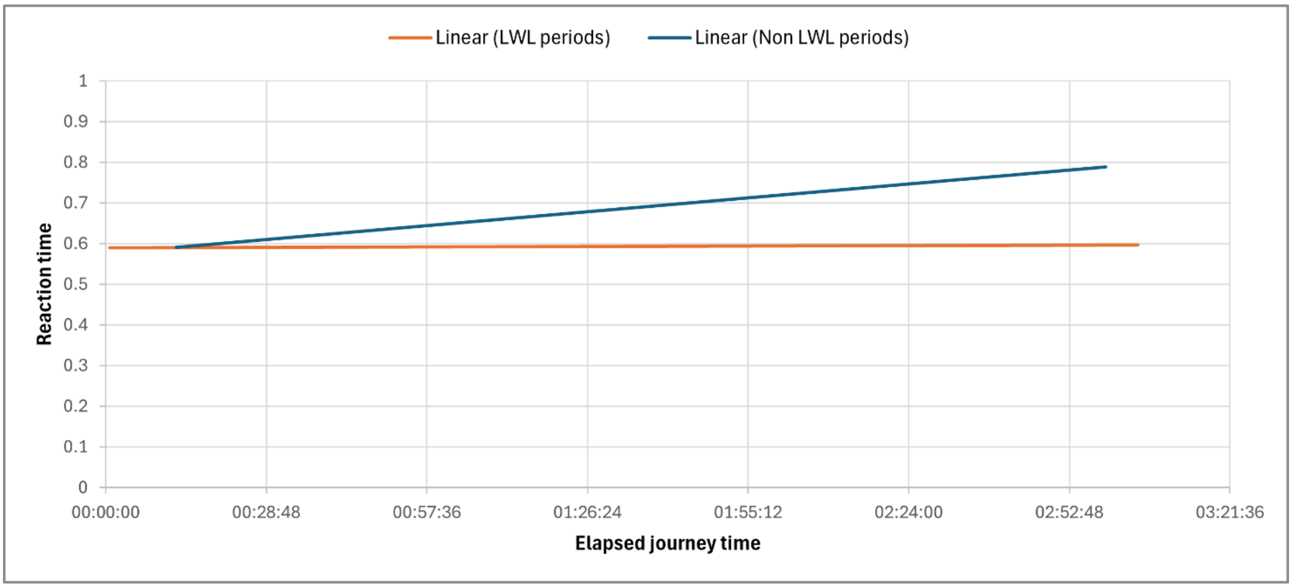

Figure 1: Reaction time in low workload (LWL) vs. non-low workload (Non LWL) periods as a function of elapsed journey time.

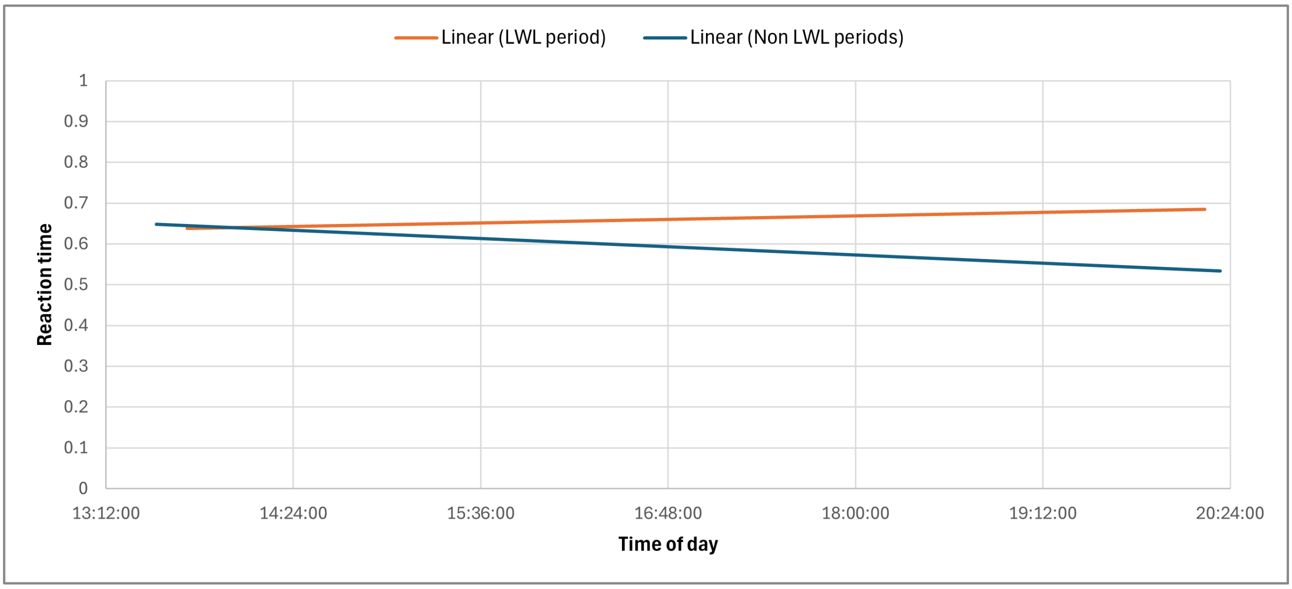

Figure 2: Reaction time in low workload (LWL) vs. non-low workload (Non LWL) periods as a function of time of day.

Discussion and conclusions

The time of day model proved to be the best statistical fit, slightly outperforming the model based on elapsed journey time, suggesting that circadian timing may explain more overall variance. This model found no main effect of time of day but reaction times were slower during low workload periods. There was also an interaction with workload, as reaction times decreased later in the day only in non-low workload sections. This result appears counterintuitive at first glance but, when considering the time period that the journeys covered (mid-afternoon onwards), ‘later in the day’ in this sense reflects emergence from the mid-afternoon dip in circadian alertness. Thus we may intuit that greater task stimulation (i.e., non-low workload sections) help to recover performance from that dip, since the decrease in reaction time with time of day was not evident in low workload sections (cf. Young & Stanton, 2002).

Meanwhile, the model based on elapsed journey time produced apparently conflicting results, this time showing faster reaction times during low workload along with gradual increases in reaction time as the journey progressed, but only for the non-low workload sections. Again, though, this finding is consistent with expected fatigue effects, indicating that higher workload may exacerbate the impact of time-on-task related fatigue. As a caveat to all of this, though, it is worth noting that there was substantial between-driver variability in both models, accounting for about half of the variance.

One limitation of this study lies in the data reduction approach to derive reaction times. By only including foot pedal responses to the vigilance alarm (which was a pragmatic decision based on the available data on the OTDR), the analysis is effectively constrained to those sections where speed was less than 120 km/h. Not only does this limit the size of the dataset, but it also means that potentially higher workload scenarios (taking place at higher speeds and with more driving task activity involved) are omitted. This could have an impact on the statistical modelling by reducing the contrast between low and non-low sections of the route.

Another related area of discussion, to consider for future research in this domain, surrounds the definition of low workload periods. Alternative speed profiles, such as segmenting by time at constant speed, or varying the threshold for the length of time at constant speed, may yield different results. Moreover, using speed profiles altogether may oversimplify workload; alternative approaches, based on a more qualitative assessment of route features or driver self-report measures, could enhance sensitivity. Refining these definitions of workload may therefore reveal patterns not captured in the current modelling framework. Reaction time modelling could also be extended using non-linear representations of time of day to better reflect circadian rhythms, including the mid-afternoon dip observed in this dataset, which may be obscured by the linear specifications used here.

Overall, though, this study shows that operational data can provide valuable insights into fatigue-related performance decrements not easily captured by more intrusive measurement methods. Together, the two models show how both temporal and workload factors can affect driver performance. Moreover, they demonstrate that reaction time, as measured on the OTDR, is sensitive to both journey progression and time of day influences, with the relative effect being dependent on the workload context. Future work should aim to scale this analysis methodology up to larger datasets, to enable deeper exploration of individual differences and to support development of driver-specific fatigue profiles.

References

Balfe, N. (2017). A framework for human factors analysis of railway on-train data. In D. de Waard, A. Toffetti, R. Wiczorek, A. Sonderegger, S. Röttger, P. Bouchner, T. Franke, S. Fairclough, M. Noordzij & K. Brookhuis (Eds.), *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2016 Annual Conference* (pp. 69-78). Santa Monica, CA: Human Factors and Ergonomics Society.

Caldwell, J.A., Caldwell, J.L., Thompson, L.A. & Lieberman, H.R. (2019). Fatigue and its management in the workplace. *Neuroscience & Biobehavioral Reviews, 96*, 272-289. https://doi.org/10.1016/j.neubiorev.2018.10.024

Dorrian, J., Roach, G.D., Fletcher, A. & Dawson, D. (2006). The effects of fatigue on train handling during speed restrictions. *Transportation Research Part F: Traffic Psychology and Behaviour, 9*(4), 243-257. https://doi.org/10.1016/j.trf.2006.01.003

Folkard, S., Robertson, K.A. & Spencer, M.B. (2007). A Fatigue/Risk Index to assess work schedules. *Somnologie, 11*, 177-185. https://doi.org/10.1007/s11818-007-0308-6

Gander, P., Hartley, L., Powell, D., Cabon, P., Hitchcock, E., Mills, A. & Popkin, S. (2011). Fatigue risk management: Organizational factors at the regulatory and industry/company level. *Accident Analysis & Prevention, 43*(2), 573-590. https://doi.org/10.1016/j.aap.2009.11.007

Gander, P.H., Wu, L.J., van den Berg, M., Lamp, A., Hoeg, L. & Belenky, G. (2017). Fatigue risk management systems. In M. Kryger, T. Roth & W.C. Dement (Eds.), *Principles and Practice of Sleep Medicine (Sixth Edition)* (pp. 697-707). Philadelphia, PA: Elsevier. https://doi.org/10.1016/B978-0-323-24288-2.00073-8

Gurubhagavatula, I., Bargar, L.K., Barnes, C.M., Basner, M., Boivin, D.B., Dawson, D., … Van Dongen, H.P.A. (2021). Guiding principles for determining work shift duration and addressing the effects of work shift duration on performance, safety, and health: guidance from the American Academy of Sleep Medicine and the Sleep Research Society. *Journal of Clinical Sleep Medicine, 17*(11), 2283-2291. https://doi.org/10.5664/jcsm.9512

Ma, J., Gu, J., Jia, H., Yao, Z. & Chang, R. (2018). The relationship between drivers' cognitive fatigue and speed variability during monotonous daytime driving. *Frontiers in Psychology, 9*, 459. https://doi.org/10.3389/fpsyg.2018.00459

McWilliams, T. & Ward, N. (2021). Underload on the road: Measuring vigilance decrements during partially automated driving. *Frontiers in Psychology, 12*, 631364. https://doi.org/10.3389/fpsyg.2021.631364

RAIB (2020). *Overturning of a tram at Sandilands junction, Croydon, 9 November 2016*. Report 18/2017, v2.2. Derby: Rail Accident Investigation Branch.

Sallinen, M. & Kecklund, G. (2010). Shift work, sleep, and sleepiness: Differences between shift schedules and systems. *Scandinavian Journal of Work, Environment & Health, 36*(2), 121-133. https://doi.org/10.5271/sjweh.2900

Townsend, J., Spencer, C., Wallace, A., Jennings, B. & Groome, B. (2023). Tram Operations Limited: Insight from our journey to improve fatigue and wellness management. In C. M. Rudin-Brown & A.J. Filtness (Eds.), *The Handbook of Fatigue Management in Transportation: Waking Up to the Challenge* (pp. 403-424). Boca Raton, FL: CRC Press. https://doi.org/10.1201/9781003213154-35

Völker, I., Kirchner, C. & Bock, O.L. (2016). Relation between multiple markers of work-related fatigue. *Safety and Health at Work, 7*(2), 124-129. https://doi.org/10.1016/j.shaw.2015.11.003

Xu, S. & Hall, N.G. (2021). Fatigue, personnel scheduling and operations: Review and research opportunities. *European Journal of Operational Research, 295*(3), 739-754. https://doi.org/10.1016/j.ejor.2021.03.036

Young, M. S. & Stanton, N. A. (2002). Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Human Factors, 44*, 365-375. https://doi.org/10.1518/0018720024497709

Young, M. S. & Steel, T. (2017). Rail worker fatigue: Identification, management and countermeasures. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 231*(10), 1098-1106. https://doi.org/10.1177/0954409716675383