

# bradscholars

## **A reliability inspired strategy for intelligent performance management with predictive driver behaviour: A case study for a diesel particulate filter**

Item Type	Article
Authors	Doikin, Aleksandr;Campean, Felician;Priest, Martin;Lin, C.;Angiolini, E.
Citation	Doikin A, Campean IF, Priest M et al (2021) A Reliability Inspired Strategy for Intelligent Performance Management with Predictive Driver Behaviour: A Case Study for a Diesel Particulate Filter. In: Proceedings of the International Conference on Engineering Design (ICED21), Gothenburg, Sweden, 16-20 August 2021.
DOI	<a href="https://doi.org/10.1017/pds.2021.20">https://doi.org/10.1017/pds.2021.20</a>
Rights	© 2021 The Authors. Published by Cambridge University Press for the Design Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence ( <a href="http://creativecommons.org/licenses/by-nc-nd/4.0/">http://creativecommons.org/licenses/by-nc-nd/4.0/</a> ), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.
Download date	2025-05-29 02:46:12
Link to Item	<a href="http://hdl.handle.net/10454/18706">http://hdl.handle.net/10454/18706</a>

# **A RELIABILITY INSPIRED STRATEGY FOR INTELLIGENT PERFORMANCE MANAGEMENT WITH PREDICTIVE DRIVER BEHAVIOUR: A CASE STUDY FOR A DIESEL PARTICULATE FILTER**

**Doikin, Aleksandr (1);**  
**Campean, Felician (1);**  
**Priest, Martin (1);**  
**Lin, Chunxing (2);**  
**Angiolini, Emanuele (2)**

1: University of Bradford;  
2: Jaguar Land Rover

## **ABSTRACT**

The increase availability of operational data from the fleets of cars in the field offers opportunities to deploy machine learning to identify patterns of driver behaviour. This provides contextual intelligence insight that can be used to design strategies for online optimisation of the vehicle performance, including compliance with stringent legislation. This paper illustrates this approach with a case study for a Diesel Particulate Filter, where machine learning deployed to real world automotive data is used in conjunction with a reliability inspired performance modelling paradigm to design a strategy to enhance operational performance based on predictive driver behaviour. The model-in-the-loop simulation of the proposed strategy on a fleet of vehicles showed significant improvement compared to the base strategy, demonstrating the value of the approach.

**Keywords:** Reliability, Machine learning, Case study, Simulation, Diesel Particulate Filter

## **Contact:**

Campean, Felician  
University of Bradford  
School of Engineering  
United Kingdom  
F.Campean@bradford.ac.uk

**Cite this article:** Doikin, A., Campean, F., Priest, M., Lin, C., Angiolini, E. (2021) 'A Reliability Inspired Strategy for Intelligent Performance Management with Predictive Driver Behaviour: A Case Study for a Diesel Particulate Filter', in *Proceedings of the International Conference on Engineering Design (ICED21)*, Gothenburg, Sweden, 16-20 August 2021. DOI:10.1017/pds.2021.20

# 1 INTRODUCTION

The increase availability of operational data from the fleets of cars in the field offers opportunities to deploy machine learning to identify patterns of driver behaviour. In turn, this can provide insight to contextual intelligence that can be used to manage the performance of the vehicles to deliver optimal performance including compliance with stringent legislation, as well as optimal vehicle health management. The integrated vehicle health management (IVHM) framework is regarded as state of the art for complex systems health management (Prajapati & Roy, 2018). The recently introduced aerospace and automotive recommended practice JA6268 (SAE International, 2018) provides a six-point reference scale to define a prescriptive vehicle health capability guide, illustrated in Figure 1, alongside the levels of integration of the system, from components to system of systems. The health management approaches evolve alongside the systems levels and hierarchy, towards self-adaptive health management.

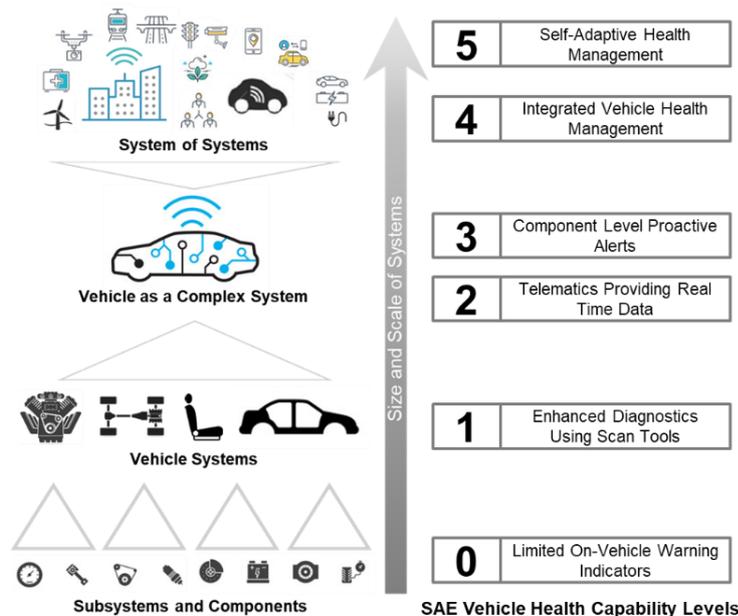


Figure 1. SAE JA6268 IVHM capability levels (Campean et al., 2019)

Increasingly, the vision of Health Management is that we should seek to optimise the performance of the system and maintain this performance around the optimal levels and in real time, and not only to predict and mitigate actual faults and failures. In practice, the implementation of this vision relies on diagnostics and prognostics modelling, underpinned by the on-line and off-line analysis of data streams from the vehicles. In this work, we illustrate via a case study approach, the application of machine learning to real world automotive data, in conjunction with a reliability inspired performance modelling paradigm to design a strategy to enhance operational performance based on predictive driver behaviour.

The case study considered in this paper relates to an automotive emissions aftertreatment system, i.e. a Diesel Particulate Filter (DPF). A technical operational challenge with the DPF system (similar to any other physical filtration system) is to ensure an optimal trade-off between fulfilment of the main of trapping solid soot particles from the gas flow, and minimising the detrimental effect of reduced gas flow due to the change in filter structure as it becomes loaded with particles. The required periodic adjustment, referred to as DPF regeneration, is akin to a "maintenance" action, and has an operational cost associated with it. Both soot accumulation in DPF and the regeneration process is dependent on the dynamics of the drive cycles and driver behaviour. The approach taken in this work is to firstly formulate the DPF control strategy for operational management as a reliability / health management problem, and then to define and implement a data driven prognostics strategy where machine learning (ML) models of driver behaviour are used to provide predictive insight for control strategy enhancement.

The paper is structured as follows: section 2 provides a review of related work; section 3 describes the research methodology based on the DPF case study; section 4 provides a summary of the results and analysis of the ML driver behaviour modelling and performance evaluation for the proposed intelligent health management strategy, followed by discussion and conclusions.

## 2 REVIEW OF RELATED LITERATURE

Maintenance is essential to keep the product operating efficiency within its designed functional requirements. According to [Kobbacy and Murthy \(2008\)](#), the main goal of maintenance is to achieve a desirable level of system dependability whilst minimising cost. BS EN 13306 (2017) standard defines maintenance as a combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in or restore it to, the state in which it can perform the required function. However, with growing system complexity, reliability modelling and further optimisation for system health management becomes more challenging. Preventative maintenance, which is still widely adopted in the automotive industry, has one significant drawback – uncertainly associated with over-maintaining and under-maintaining. This uncertainty becomes more significant if life distribution is not statistically based ([Howell, 2017](#)), which often results in replacing components before the actual end of useful life ([Ellis, 2008](#)). On the other hand, predictive maintenance, recognised as condition-based maintenance (CBM), presumes monitoring the actual condition of equipment and apply maintenance actions when the need arises ([Alaswad and Xiang, 2017](#)), e.g. if there are signs of decreased performance or impending failure. An open system CBM architecture was developed back in 2001 as a layered approach ([MIMOSA, 2010](#)) to move process flow from data acquisition to advisory generation (or decision-making). Its implementation relies on diagnostic, which is associated with the current health state, and prognostics, which extends evaluation to the remaining useful life (RUL) estimation ([Eker et al., 2012](#)).

In general, diagnostics is quite a mature field in comparison to prognostics, which remains to be the most challenging area of study. However, since prognostics also relies on the diagnostic output ([Sikorska et al., 2011](#)), challenges can be discussed within uncertainty context, which affect the robustness of RUL estimation, as illustrated in Figure 2 ([Campean et al., 2019](#)). Types of uncertainties have been widely discussed in the literature ([Saha and Goebel, 2008](#), [Atamuradov et al., 2017](#), [Nguyen et al., 2019](#), [Tamssaouet et al., 2020](#)) including modelling, measurement/sensor, system parameters/process uncertainty, operating environment, future load, failure threshold. [Sankararaman \(2015\)](#) discussed uncertainty related issues from a modelling perspective and combined them into four main categories:

- Present uncertainty – related to the precise estimation of the component/system current health state based on the direct damage measurements or estimation using filtering techniques.
- Future uncertainty – probably one of the most significant sources of uncertainty coming from unknown future operating, including loading and environmental conditions.
- Model uncertainty - relates to the uncertainty associated with model parameters - both current and future; if model parameters have a bias or are not accurate (underfitting or overfitting), then there will be discrepancies between the model predictions and actual RUL.
- Modelling method uncertainty – describes the combined effect of all uncertainties together. Even with the assumption of all the above uncertainties can be precisely calculated, their combined effect needs investigation. This creates additional uncertainty as precisely deriving the overall uncertainty effect is a significant challenge.

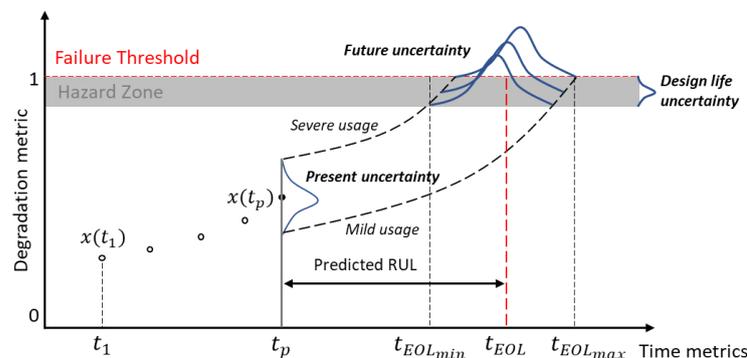


Figure 2. Uncertainty quantification in prognostics modelling

In engineering applications, present uncertainty is often treated as a significant challenge in applications where damage cannot be inferred directly. To overcome it, indirect sensor measurements of other physical properties around the system are correlated with the accumulated damage ([Petryna et al., 2012](#), [Javed et al., 2017](#), [Larsen et al., 2017](#)). However, even in these cases, the robustness of the current health

state estimation is affected by sensor fidelity (Tamssaouet et al., 2020), where measurement noise can be caused by electrical interference, digitisation error, sensor bias, dead-band, backlash, and nonlinearity in the response (Saha and Goebel, 2008). In the previous work (Doikin et al., 2020), a knowledge-enabled data-driven framework was presented to increase the trust level to sensor measurements.

The challenge presented in this paper focuses on prognostics modelling, which is affected by future uncertainties. Various methodologies were applied in aerospace, railway and energy sectors (Rebello et al., 2018), however, automotive operating conditions are subjected to much more transients, which vary from duty cycles as well as depend on individual driver behaviour habits. From a practical application viewpoint, such uncertain future operating conditions often result in operational optimisation limitation. The example presented in this paper reveals around data-driven modelling for the DPF regeneration, where accurately estimated future usage can significantly improve the overall system performance.

### 3 RESEARCH METHODOLOGY

#### 3.1 Case study background

The Diesel Particulate Filter is an emission solution, which traps soot particulates released by the engine as a result of fuel combustion. Once the volume of the soot particles (or mass) retained in the DPF reaches a critical value, removal of particulates is needed, known as filter regeneration. Failure to deliver this function on time may result in a soot overloading failure mode (Kotrba et al., 2011) triggering limp mode activation, as any further attempts to regenerate can damage the catalyst as well as the other systems (Chen et al., 2011). Therefore, soot mass is continuously inferred from the combination of sensing (using the pressure drop in the gas flow across the DPF) and emission physics-based modelling. Figure 3 illustrates the DPF soot accumulation and regeneration cycles based on ECU data from real-world driving, where observable patterns define several phases. Firstly, soot accumulation (i), where the soot mass stored in the DPF grows with time at a rate dependent on the engine operating conditions. The soot mass can also decrease (ii) if the exhaust gas temperatures are high enough to cause a passive regeneration (also referred as “spontaneous”) (Bai et al., 2016). Active regeneration (Tong et al., 2018) is triggered by raising the exhaust gas temperature using fuel post-injection. If current driving conditions (i.e. combination of vehicle speed and emission temperature) do not allow for a complete DPF regeneration, further attempts will be made to complete the process when conditions allow. A complete regeneration (iii) restores the DPF to a condition of minimal soot mass stored. However, a successful regeneration requires around 10-15 minutes of driving under operating conditions with a sufficient gas mass flow rate at the required temperature passed over the DPF. If the journey is not long enough, this causes an interrupted regeneration process, which is illustrated by the (iv) pattern annotated in Figure 3. This illustration provides an example that the ability to complete the active regeneration strongly depends on user behaviour – in terms of the driving style, the type and the length of the journeys.

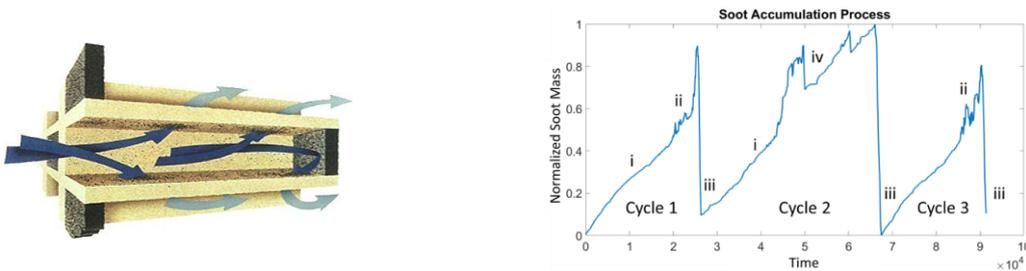


Figure 3. Wall-flow DPF (Majewski, 2005) & real-world soot accumulation cycles

#### 3.2 Reliability based paradigm for DPF operational management

From a reliability perspective, the soot accumulation process in the DPF, illustrated in Figure 3, can be regarded as a “damage accumulation” process, which can be described using the Eq. 1:

$$m = m_0 + \sum_i \bar{m}_i \cdot d_i \quad (1)$$

where,  $m$  is the current soot load (in grams) in the DPF,  $m_0$  is the soot mass at the beginning of the cycle (initial soot mass),  $\bar{m}_i$  is the average soot rate for the journey  $i$ , and  $d_i$  is the duration of the

journey  $i$ . However, the collected particulates eventually increase back pressure (BP) in the exhaust flow, which negatively affects the engine operation (Singh et al., 2009). In functional terms, increased back pressure (BP) can be regarded as the degrading performance characteristics; however, the system is repairable in the sense that backpressure (and hence the soot mass) can be restored through thermal soot oxidation, known as DPF regeneration. Nevertheless, the cost function of this process is expressed in fuel consumption penalties. Therefore, minimising the cost function can be regarded as a reliability/maintenance optimisation problem.

Figure 4 depicts the state-based representation of DPF “damage” cycles, which describes system operation in accordance with the default control strategy:

- “DPF empty” – denotes the new or after renewal state, with initial soot mass  $m_o$  ( $m_o = 0$  for a new DPF)
- “DPF partial load” –  $m_o < m < m_{max}$ , where  $m_{max}$  denotes the level of mass when delivering regeneration (or maintenance action) is needed and  $\dot{m}$  is a soot rate which governs soot mass growth process.
- “DPF full load” – when this state is reached, the system should be repaired as soon as suitable conditions are met (dependent on the engine operating mode). The repair process is governed by a negative soot rate  $\dot{m}$ , and can be complete (successful), or incomplete (interrupted).
- “DPF critical load” – denotes a critical state that is reached if repair action is not delivered in time. The system will move in a malfunction warning state, where the regeneration is “managed” with an action request on the driver.

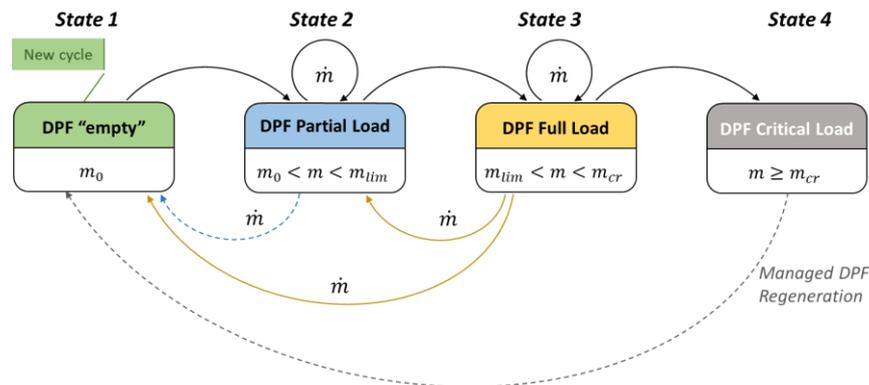


Figure 4. DPF state-based representation

In operation, maintenance action (i.e. regeneration) is needed once the DPF State 3 is reached, and research efforts are targeted to deliver further system optimisation to avoid unnecessary DPF regenerations due to “under-loading” (Barba et al., 2017). For example, a recent study from General Motors presented a methodology focussing on a more reliable engine-out soot mass simulation approach (Barba et al., 2017). Dawei et al. (2017) proposed a new method for DPF regeneration, underpinned by mathematical model for more accurate regeneration timing and temperature control. Castellano et al. (2013) introduced a novel adaptive control strategy for temperature control using fuel post-injection, which reports improvement of regeneration quality. However, the reviewed research work mainly revolves around diagnostic, while benefits associated with prognostics modelling are not fully discussed. From DPF health management viewpoint, the success of the maintenance process to restore DPF loading depends on the type of the journey, which, in simple terms, from a user perspective can be associated with the journey duration that needs to be long enough to allow complete soot burning. Shorter journeys cause process interruption resulting in partial soot burning. In practical terms, this reduces the system efficiency as regeneration occurs more frequently, which has negative effects on fuel consumption and CO<sub>2</sub> emissions, and also increases the oil degradation rate due to oil dilution with fuel (Ito et al., 2019). Therefore, the DPF maintenance efficiency can be enhanced if regeneration action is only triggered in conjunction with a long journey. However, the length of the current/next journey cannot be known by the system a priori. The approach proposed in this work is to use supervised machine learning (ML) modelling to historic journeys of a specific driver / vehicle pair (referred to here as “driver behaviour modelling”), to develop a prediction model for the journey duration, which could then be used to optimise the maintenance efficiency.

Figure 5 summarises the proposed framework for DPF regeneration efficiency enhancement with predictive driver behaviour insight. Driver behaviour modelling can be based on historical vehicle usage data, specifically the details of journeys of a driver / vehicle, with ML models capturing the patterns of usage, and be used to predict the likelihood of a long journey to be observed in the current / next journey. As shown in Figure 5, historical data is processed in the “driver behaviour modelling” block, where data is subjected to cleaning, pre-processing and transformation depending on the use case scenario and also requirements for the ML training algorithms. Vehicle trip history database is regularly updated with new records, which allows maintaining data-driven model up to date depending on the current usage profile.

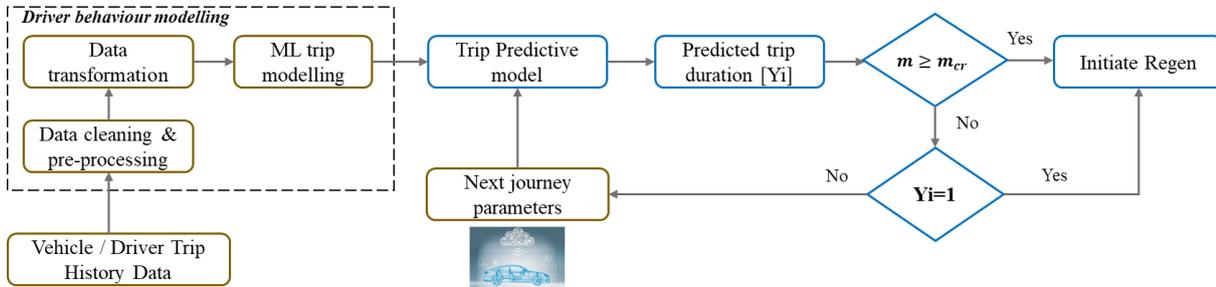


Figure 5. Data-driven modelling in DPF health management

### 3.3 Framework for Driver Behaviour Modelling

The specific aim for the data-driven model is to predict the trip duration to satisfy DPF maintenance/regeneration requirements. For simplicity, the assumption is applied that time required to restore the filter to the empty state is fixed ( $X_{lim}$ ) for all cases. Therefore, a binary transformation can be applied to the prediction variable  $Y$ , as shown in Eq. 2:

$$Y'_i = \begin{cases} 0, & \text{if } Y_i < X_{lim} \text{ minutes} \\ 1, & \text{if } Y_i \geq X_{cr} \text{ minutes} \end{cases} \quad (2)$$

where  $X_{cr}$  is a critical value defining the time required to complete DPF repair process.

From DPF health management perspective, pre-trained data-driven model can predict the upcoming trip duration ( $Y_i$ ) based in the initial journey parameters (such as start time and day of the week) available in real-time. The regeneration is initiated only if predicted trip is long enough, otherwise regeneration is postponed until the next journey where the same logic is repeated until  $Y_i$  evaluates to 1.

Figure 6 illustrates the set-up of the driver behaviour ML modelling. This reflects the partition of predictor variables into “journey parameters” and “driver behaviour”. The second category of predictors capture the way in which the vehicle is driven – which is dependent on the driver choice or driver specific behaviour, for which a range of trip statistics can be used as data features for the machine learning model (see for example Doikin et al, 2020). The classification accuracy is then evaluated using the confusion matrix, which calculates the number of correctly classified classes versus the total number of attempts. Feature importance factor is another important representation of the model, which allows investigating which features have a greater impact to characterise driver behaviour and have more impact on the model decision.

## 4 RESULTS AND ANALYSIS

### 4.1 ML for driver behaviour modelling

A sample of historical usage collected from real-world driving was available for the case study. The data covered the journeys of 100 individual vehicles over a period of 3 months of driving, including the history of journeys as a statistical summary for each trip, which included start time and averaged driver behaviour statistics. It is worth to say that study is carried out without the access to GPS coordinates, however, this can be used as another input to achieve higher classification accuracy. The sample size was variable in terms of number of recorded journeys in the observation period, in average containing a history of 500 to 700 trips per vehicle.

The data preparation for the ML process flow included several steps. Firstly, data was subjected to cleaning and pre-processing to satisfy data quality requirements. This included identification of missing values, as well as outliers - as values outside the expected range. In the next step, the binary

transformation was applied for the variable of interest from a prediction point of view (i.e. journey duration), based on Eq. (2). What was observed is that this transformation led to an imbalance in the data, which is known to cause difficulties for ML training algorithms. This drawback was overcome by using a boosting technique specifically designed for imbalanced data sets (Seiffert et al., 2009). Classification efficiency can be enhanced through the conversion of multiple weak learners into a single composite robust classifier (Mounce et al., 2017). Therefore, RusBoosted Ensemble was selected as a primary modelling choice, but other common ML methods (such as support vector machines, SVM, and feed-forward pattern recognition neural networks) were also considered for the study.

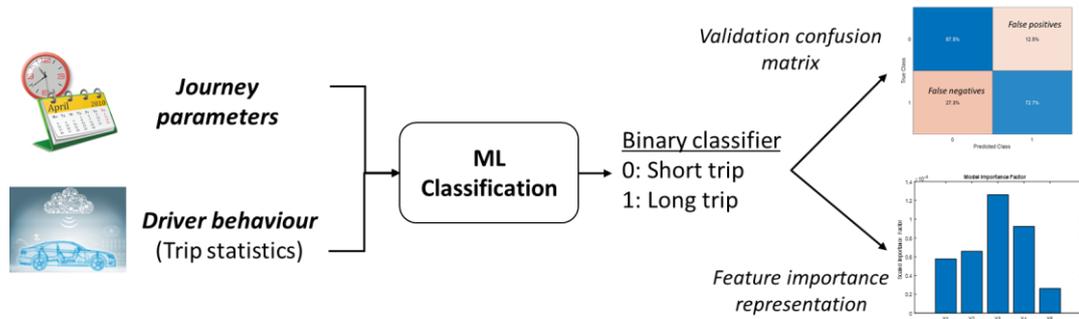


Figure 6. Setting up a data-driven model

Given the data imbalance, standard accuracy measure to evaluate algorithm performance is inefficient, as it is biased towards the majority class (Torgunov et al., 2019). Normally,  $F_\beta$  measure is applied as a better measure of accuracy, defined in Eq. 3 with  $\beta = 1$  (Shaojun et al., 2012):

$$F_1 = 2 \times \frac{TP}{TP + FN + FP} \quad (3)$$

where true positives (TP), false negative (FN) and false positives (FP) represent model prediction confusion matrix components. However, since there are 100 separate models and the rate of imbalance for each vehicle is different, which can also be towards both classes,  $F_1$  is also not an efficient measure to allow fair comparison between vehicles. Therefore, classification accuracy was inferred by taking the average between the rate of true positives and true negatives, also known as balanced accuracy:

$$Accuracy\ measure = \frac{TP + TN}{2} \quad (4)$$

The average classification accuracy results for 100 vehicles are summarised in Table 1. The training was performed using first 2 months of the available history, where accuracy is based on k-fold cross-validation metric with chosen  $k = 5$ , and the remaining data was left for testing on the unseen data.

Table 1. Average performance evaluation of ML algorithms

Algorithm tested	Accuracy (training)	TP (training)	TN (training)	Accuracy (testing)	TP (testing)	TN (testing)
Support Vector Machine	68%	57.4%	79.4%	63%	55.5%	70%
Pattern Recognition NN	63%	46%	80.8%	60%	44%	76%
RusBoosted Tree Ensemble	73%	71%	75%	71.5%	70%	72.5%

The RusBoosted Tree Ensemble model performed best for journey classification. Support Vector Machine and Neural Network demonstrated significantly lower accuracy. Moreover, both methods failed in true positive classification accuracy, which also indicates difficulties to overcome data imbalance challenge. Figure 7 plots RusBoosted method validation accuracy for all available 100 vehicles, where an average accuracy of over 71% was achieved. The results confirm that driver behaviour can be modelled from the historical data with an acceptable accuracy. The analysis also points out that some vehicles are more predictable than others. However, considering the mixture of drivers and the overall nature of future uncertainty problem complexity, results are very promising. The main outcome is that driving patterns can be fairly quickly learnt from the historical data, while the presence of some errors is acceptable. From a reliability perspective, this assumption is tested in the next step, where validation is extended to the practical example.

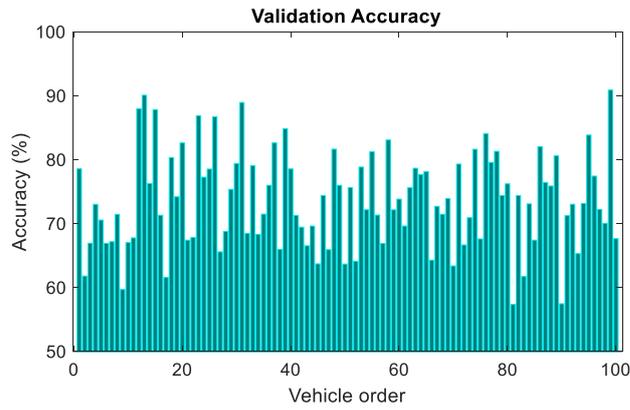


Figure 7: RusBoosted model validation accuracy on unseen data

#### 4.2 Performance evaluation of the predictive DPF health management strategy

To evaluate the efficiency of the proposed reliability / maintenance inspired framework with driver behaviour insight, a simulation experiment was set up based on the journey history for the 100 vehicles, and using a surrogate soot accumulation mean value model. The average soot mass rate for each trip was evaluated using the soot model and the relevant engine parameters available from the summarised trip statistics, and the trip duration, as described in Eq. 1. The soot accumulation in the DPF was estimated based on the sum of soot mass in the journey sequence. Similarly, a surrogate model for the soot burning rate was used to simulate the regeneration process, and evaluate the soot mass at the end of the regeneration, or the reduction in soot in the case of an interrupted regeneration. For regeneration process, the assumption was applied that each minute is proportional to burning fixed amount of soot.

In order to evaluate the gain in efficacy from the proposed health management strategy, its performance was compared to the benchmark solution, which assumes that regeneration is attempted at any journey once the soot mass stored in the DPF reaches State 3. The proposed strategy is based on predictive insight in the driver behaviour, with prediction for the current trip duration obtained from the ML model, as illustrated in Figure 5. Thus, if the predicted trip duration is 0 / "short", regeneration will not be automatically attempted in State 3. However, it was recognised that this strategy could induce a higher risk of reaching State 4, which has a potential cost to the user. Therefore, the proposed intelligent strategy was deployed within a soot loading window limited by a threshold  $m_{lim}$  lower than  $m_{cr}$ , i.e.

$$m_{max} \leq m \leq m_{lim} < m_{cr}$$

This means that if the cumulated soot mass in the DPF reaches  $m_{lim}$  due to a sequence of short trips or trip duration misclassification errors after reaching State 3 /  $m_{max}$ , regeneration will be attempted for each trip (i.e. revert to the benchmark strategy).

The evaluation of performance for the proposed strategies was based on the total number of interrupted regenerations across the journeys recorded during the unseen observation period (1 month), based on the simulation using the surrogate soot model. Due to the data limitation, validation period was looped to 3 months to allow sufficient number of soot cycles and more robust evaluation, which considered performance across the fleet of 100 vehicles in the study.

The simulation results showed a significant improvement from the proposed health management strategy, with reduction of incomplete regeneration cases from 1,176 to 574, i.e. 51.2% improvement. In practical terms, this could be translated to fuel economy / CO2 efficiency gain, with a more detailed simulation model. However, it is noteworthy that while a significant improvement in the overall fleet performance was demonstrated, with an average of 6 interrupted regens avoided per vehicle, for some vehicles the proposed strategy resulted in a slight decrease in efficiency - i.e. marginal increase in the number of interrupted regens. On a closer examination it was seen that these cases were associated with cases where the driver behaviour model was of poor accuracy - below 60% in Figure 7.

## 5 DISCUSSION, CONCLUSION AND FUTURE WORK

This paper illustrated a reliability-based approach for an intelligent control strategy to manage the operational performance of a system, with predictive insight on operational demand. The methodology was illustrated with a case study for a DPF system, but approach of considering a cumulative

degradation and renewal model as operational performance model of the system and to derive alternative governing control strategies could be applied to many other systems, e.g. electric batteries. The distinguishing feature of the proposed approach is the integration of data-driven ML modelling in a contextual intelligent system for health management. The deployment of framework was illustrated on automotive case study, where the vehicle usage and driver behaviour patterns were successfully learnt from historical journey records. The journeys were classified based on the trip duration as feature, based on the DPF system requirement for a complete regeneration. A RusBoosted Ensemble ML model demonstrated satisfactory journey classification accuracy for most vehicles in the study, and superior to other ML modelling techniques attempted. This is expected given the known ability of this ML modelling technique to overcome data imbalance. An immediate benefit is that this model can be employed as the universal solution across the whole fleet of cars - i.e. as more vehicle journeys data becomes available for the study, there the RusBoosted Ensemble can be employed as the ML model of choice. There could be many reasons for the lower accuracy seen in the ML models for some of the vehicles; in all cases this was related to higher "noisiness" in the signal - possibly reflecting multiple users for the same vehicle. However, the signal sequence was not long enough to support a deeper analysis using other ML techniques that are more adept at identifying concept drift in time series. Predictive driver behaviour modelling - based on the ML model for the journey duration, facilitated the definition of a DPF regeneration strategy that takes advantage of the contextual intelligence in the form of predictive journey duration, as condition for a successful regeneration. Overall this strategy was shown to be highly efficient, leading to a significant (close to 50%) reduction in the interrupted regenerations across the fleet, based on a simulation model. The strategy could be further improved, e.g. by optimising the value of the threshold  $m_{lim}$  either for each individual vehicle or across the fleet. Given that the proposed strategy did not show performance improvement in a small number of cases, highlights the fact that the robustness of the ML models needs to be studied further and improved. This could include consideration of different ML modelling methods (if larger data sets are available), or defining an accuracy index at which sufficient confidence in the model is achieved, and the intelligent health management adaptive strategy can be adopted. Other more detailed reliability-based formulations of the maintenance optimisation problem can be developed, e.g. by taking into account the details of the cost functions involved, including both the operational costs (fuel / CO<sub>2</sub> emissions), and the physical degradation of the DPF and the engine system. This could lead to the deployment of multidisciplinary optimisation strategies, which could also be adapted to evolving conditions - e.g. degradation of components and sub-systems.

## REFERENCES

- Alaswad, S. & Xiang, Y. 2017. A review on condition-based maintenance optimization models for stochastically deteriorating system. *Rel Eng & System Safety*, 157, 54-63. doi: <https://doi.org/10.1016/j.res.2016.08.009>.
- Atamuradov, V., Medjaher, K., Dersin, P., Lamoureux, B. & Zerhouni, N. 2017. Prognostics and health management for maintenance practitioners-review, implementation and tools evaluation. *International Journal of Prognostics and Health Management*, 8, 1-31.
- Bai, S., Tang, J., Wang, G. & Li, G. 2016. Soot loading estimation model and passive regeneration characteristics of DPF system for heavy-duty engine. *Applied Thermal Engineering*, 100, 1292-1298. doi: <https://doi.org/10.1016/j.applthermaleng.2016.02.055>.
- Barba, F., Vassallo, A. & Greco, V. Estimation of DPF Soot Loading through Steady-State Engine Mapping and Simulation for Automotive Diesel Engines Running on Petroleum-Based Fuels. 2017. SAE International. doi: <https://doi.org/10.4271/2017-24-0139>.
- Bsi Standards. BS EN 13306:2017: Maintenance. Maintenance terminology. 2017. British Standards Institute.
- Campean, F., Neagu, D., Doikin, A., Soleimani, M., Byrne, T. & Sherratt, A. Automotive IVHM: Towards Intelligent Personalised Systems Healthcare. *Proc Design Society: International Conference on Engineering Design*, 2019. Cambridge University Press, 857-866. doi: <https://doi.org/10.1017/dsi.2019.90>.
- Castellano, J., Chaudhari, A. & Bromham, J. Adaptive Temperature Control for Diesel Particulate Filter Regeneration. 2013. SAE International. doi: <https://doi.org/10.4271/2013-01-0517>.
- Chen, K., Martirosyan, K. S. & Luss, D. 2011. Temperature gradients within a soot layer during DPF regeneration. *Chem Eng Science*, 66, 2968-2973. doi: <https://doi.org/10.1016/j.ces.2011.03.037>.
- Dawei, Q., Jun, L. & Yu, L. 2017. Research on particulate filter simulation and regeneration control strategy. *Mechanical Systems and Signal Processing*, 87, 214-226. doi: <https://doi.org/10.1016/j.ymsp.2016.05.039>.
- Doikin, A., Campean, F., Neagu, D., Priest, M., Soleimani, M. & Lin, C. Knowledge-Enabled Machine Learning for Predictive Diagnostics: A Case Study for an Automotive Diesel Particulate Filter. *Proceeding 30th ESREL Conference and 15th PSAM Conference*, 2020.

- Eker, Ö. F., Camci, F. & Jennions, I. K. Major challenges in prognostics: study on benchmarking prognostic datasets. 1st European Conf Prognostics and Health Mt Society, 2012 Dresden, Germany. PHM Society.
- Ellis, B. A. 2008. Condition Based Maintenance. The Jethro Project, 1-5.
- Howell, M. A., Fadi S. 2017. Different Maintenance Types and The Need for Energy Centered Maintenance. Energy Centered Maintenance - A Green Maintenance System. Fairmont Press, Inc.
- Ito, T., Kitamura, T., Kojima, H. & Kawanabe, H. Prediction of Oil Dilution by Post-injection in DPF Regeneration Mode. 2019. SAE International. doi: <https://doi.org/10.4271/2019-01-2354>.
- Javed, K., Gouriveau, R. & Zerhouni, N. 2017. State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels. Mechanical Systems and Signal Processing, 94, 214-236. doi: <https://doi.org/10.1016/j.ymssp.2017.01.050>.
- Kobbacy, K. a. H. & Murthy, D. P. 2008. Complex system maintenance handbook, Springer Science&Business.
- Kotrba, A., Bai, L., Yetkin, A., Shotwell, R. & Gardner, T. DPF Regeneration Response: Coupling Various DPFs with a Thermal Regeneration Unit to Assess System Behaviors. 2011. SAE International. <https://doi.org/10.4271/2011-01-2200>.
- Larsen, J. M., Christodoulou, L., Calcaterra, J. R., Dent, M. L., Derriso, M. M., Hardman, W. J., Jones, J. W. & Russ, S. M. 2017. Integrated Systems Health Management: Enabling Technology for Effective Utilization of Air Vehicle Systems. Materials Damage Prognosis. TMS (The Minerals, Metals & Materials Society).
- Majewski, W. A. 2005. Wall-Flow Monoliths [Online]. DieselNet Technology Guide. Available: [https://dieselnet.com/tech/dpf\\_wall-flow.php](https://dieselnet.com/tech/dpf_wall-flow.php) [Accessed 10/05/2020].
- Mimosa 2010. OSA-CBM V3.3.1 UML Model. Normative Information Specification.
- Mounce, S., Ellis, K., Edwards, J., Speight, V., Jakomis, N. & Boxall, J. 2017. Ensemble decision tree models using RUSBoost for estimating risk of iron failure in drinking water distribution systems. Water Resources Management, 31, 1575-1589. doi: <https://doi.org/10.1007/s11269-017-1595-8>.
- Nguyen, V., Kefalas, M., Yang, K., Apostolidis, A., Olhofer, M., Limmer, S. & Bäck, T. 2019. A Review: Prognostics and Health Management in Automotive and Aerospace. International Journal of Prognostics and Health Management, 10. doi: <https://doi.org/10.36001/ijphm.2019.v10i2.2730>.
- Petryna, Y., Link, M. & Künzel, A. Modeling and Monitoring of Damage in Grouted Joints. 6th European Workshop on Structural Health Monitoring, 2012.
- Prajapati, A. K. & Roy, B. K. A State of Art Review of Integrated Vehicle Health Management System. 2018 3rd International Conference for Convergence in Technology (I2CT), 2018. IEEE, 1-5. <https://doi.org/10.1109/I2CT.2018.8529590>.
- Rebello, S., Yu, H. & Ma, L. 2018. An integrated approach for system functional reliability assessment using Dynamic Bayesian Network and Hidden Markov Model. Reliability Engineering & System Safety. <https://doi.org/10.1016/j.ress.2018.07.002>.
- Sae International. Design & Run-Time Information Exchange for Health-Ready Components. 2018. [https://doi.org/10.4271/JA6268\\_201804](https://doi.org/10.4271/JA6268_201804).
- Saha, B. & Goebel, K. Uncertainty management for diagnostics and prognostics of batteries using Bayesian techniques. 2008 IEEE Aerospace Conf, 1-8. <https://doi.org/10.1109/AERO.2008.4526631>.
- Sankararaman, S. 2015. Significance, interpretation, and quantification of uncertainty in prognostics and remaining useful life prediction. Mechanical Systems and Signal Processing, 52, 228-247. <https://doi.org/10.1016/j.ymssp.2014.05.029>.
- Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J. & Napolitano, A. 2010. RUSBoost: A hybrid approach to alleviating class imbalance. IEEE Trans Systems, Man, and Cybernetics-Part A: Systems and Humans, 40, 185-197. <https://doi.org/10.1109/TSMCA.2009.2029559>.
- Shaojun, H., Jin, C., Ruixu, G. & Guijun, W. The Capability Analysis on the Characteristic Selection Algorithm of Text Categorization Based on F1 Measure Value. 2nd Int Conf on Instrum, Measurement, Computer, Communication and Control, 2012. IEEE, 742-746. <https://doi.org/10.1109/IMCCC.2012.180>.
- Sikorska, J., Hodkiewicz, M. & Ma, L. 2011. Prognostic modelling options for remaining useful life estimation by industry. Mech sys & signal proc, 25, 1803-1836. <https://doi.org/10.1016/j.ymssp.2010.11.018>.
- Singh, N., Rutland, C. J., Foster, D. E., Narayanaswamy, K. & He, Y. Investigation into Different DPF Regeneration Strategies Based on Fuel Economy Using Integrated System Simulation. 2009. SAE International. <https://doi.org/10.4271/2009-01-1275>.
- Tamssaouet, F., Nguyen, K. T., Medjaher, K. & Orchard, M. E. 2020. Degradation Modeling and Uncertainty Quantification for System-Level Prognostics. IEEE Systems <https://doi.org/10.1109/JSYST.2020.2983376>.
- Tong, D., Zhang, J., Wang, G., Yang, B., Cai, K., Liu, S., Abdalla, A. & Shuai, S.-J. Experimental Study and Numerical Interpretation on the Temperature Field of DPF during Active Regeneration with Hydrocarbon Injection. 2018. SAE International. <https://doi.org/10.4271/2018-01-1257>.
- Torgunov, D., Trundle, P., Campean, F., Neagu, D. & Sherratt, A. Vehicle Warranty Claim Prediction from Diagnostic Data Using Classification. UKCI Workshop, 2019. Springer, 483-492. [https://doi.org/10.1007/978-3-030-29933-0\\_40](https://doi.org/10.1007/978-3-030-29933-0_40).