Innovative Systems Design and Engineering
ISSN 2222-1727 (Paper) ISSN 2222-2871 (Online)
Vol.9, No.4, 2018
www.iiste.org

Predictive Maintenance – Analysis of Seasonal Dependence of Vehicle Engine Faults

Patrick Pana[1*] Dennis Kijek[1] Christoph Ullrich[1] Adam-Alexander Manowicz[1] Fabian Seithel[2]
1.Hochschule Esslingen, Kanalstraße 33, 73728 Esslingen, Germany
2.Geotab Inc., 21-1075 North Service Road West, Oakville, Ontario, L6M 2G2 Canada

Abstract
This paper presents methods and results for the analysis of interrelationships between the occurrence of specific engine faults according to seasons. The issue of maintenance is substantial for the automotive industry and improvements are requested due to the enhancement of profitability. Findings of this paper are based on logged-vehicle data from 760.976 vehicles provided by the company Geotab. Utilization of such data gains importance for the automotive service sector with special regard of increasing importance of predictive maintenance. The visualization of the data of three different engine faults was realized with the free graphic and statistics program "Tableau Desktop 2018.1" as well as "IBM SPSS Statistics Subscription Trial for Mac OS". The result is that the tested interrelations are significant, leading to the conclusion that the engine faults of "vehicle battery has low voltage", "low priority warning light on" and "general warning light on" are dependent of seasons. This finding can be used to help car manufacturers and car service providers to reduce maintenance costs.
Keywords: automotive, big data, predictive maintenance, seasonal engine faults, vehicle error codes

1. Introduction

Most of the cars nowadays contain numerous sensors, collecting, processing and analyzing data permanently. Single portions of this data can be used to identify changes in engine running or other points in vehicle operation including future faults of particular vehicle components. The framework of this paper is predictive maintenance. It is the further development of the concept of preventive maintenance. In this case the replacement of vehicle parts occurs on the basis of simple parameters as time since last maintenance service or mileage. Parts replaced due to these simple parameters often do not show the degree of deterioration to justify the replacement (Mobley 2002, p. 3). In consequence preventive maintenance thus can be unnecessarily expensive for the Original Equipment Manufacturer (OEM) as well as for the customer. If the exchange of parts occurs during the warranty period, the OEM bears the expenses. Otherwise the customers cost of ownership rises. As a result of globalization, the stronger competition increases the sensibility of prices and hence puts a lot of pressure not only on the OEMs but as well on other smaller sized businesses. (Manowicz 2017, p. 8) Therefore, the interest in minimizing the replacement of still usable components is given on both sides. In addition, predictive maintenance related to the automotive industries is a far bigger challenge than in relation to immobile equipment and devices. Data has to be transferred over the air, but without worldwide standards of network coverage and data security there are many more difficulties to master (Prytz et al. 2015, p. 2)

We present a work dealing with vehicle engine faults as a function of different seasons. The data for this purpose are provided by the company Geotab which collects logged-vehicle data with the aid of portable OBD-Dongles and replace them at the disposal of different customers. The provided data refers to over 760.976 vehicles with different types of fault indications. We selected the most important fault indications (low priority warning light, general vehicle engine light, low battery warning light) which we in the course of this paper relate to seasons and analyze with statistical methods.

The paper is structured as follows: A compilation of related research gives an overview of the subjects of predictive maintenance in general and especially condition-based maintenance, estimation of remaining useful life and big data. This is followed by the chapter of methods used to obtain results and conclusions which are processed in the following sections.

2. Literature review

As of today, there are few publications using big data for predictive maintenance for the automotive industries. First, it is necessary to define predictive maintenance. There are many different definitions. One of them is to detect inconstancy in the vibration of a rotating system to anticipate changes to avoid breakdowns. A second one deals with conduction of infrared analysis in electrical components to spot developing issues. Summarizing, the general condition of predictive maintenance is that "regular monitoring of the actual mechanical condition, operating efficiency, and other indicators of the operating condition of machine-trains and process systems will provide the data required to ensure the maximum interval between repairs and minimize the number and cost of unscheduled outages created by machine-train failures." (Mobley 2002, p. 4) This definition is used in this work because the contemplation of engine faults is applied to all faults in general not only in a specific part such as vibration or monitoring electrical components.

Ahmad, et al. (2012, p. 9) outline the differences between time-based and condition-based maintenance and compare the challenges of implementation. A more detailed view on condition-based maintenance and hence machine prognostics is given by Peng, et al. (2010, pp. 3-11). Four categories of prognostic models are introduced: the physical model, the knowledge-based model, a combination model and a data-driven model which also will be used in this work. Dragomir, et al. (2009, p. 2) relate the prediction of faults to the condition-based maintenance and outline the different approaches. The model-based approach refers to mathematical and physical calculations to predict system faults. In comparison with this method the data-driven approach relates to real data produced by sensors with the intention to identify the level of deterioration of single components to generate a prediction for the performance of the overall system.

The Intelligent Predictive Decision Support System (IPDSS) developed by Yam, et al. (2001, p. 1) supports the conventional condition-based maintenance with an intelligent fault diagnosis. Thus the level of deterioration is detected. The system was tested on a power plant what generates the challenges to transfer knowledge from an immobile system to a mobile operating system like a vehicle. Medjaher, et al. (2012, p. 1) go even further. They generate a model for the prediction of the remaining useful life and applied it on bearings as critical mechanical objects. In comparison with this work Miao, et al. (2012, p. 1) describe a method of the prediction of remaining useful life of a critical electrical component. The lithium-ion battery which they deal with gets more and more important due to todays issues of electrification. With rising rate of registration of electrical vehicles the maintenance of large packs of battery becomes more significant for the OEMs but especially for small service places which do not want to miss the change and get left behind. Frisk, et al. (2014, p. 7) also study the remaining useful life of a battery. Their data-driven approach is related to lead-acid batteries. The fault of a low battery is also a part of this work. The approach used in this paper is different from theirs because they use the data to model the remaining useful life of the battery. In comparison we take the data of the battery faults and analyze them in relation to different seasons.

A comparison between patterns of collected on board and service record Data and the prediction of remaining useful life is given in the work of Prytz, et al. (2015, p. 2). They deal with predictive maintenance in the case of air compressors in heavy duty trucks. One step further is the process of decision-making which is covered by Karim, et al. (2016, p. 2). Their work explains four steps from the “Maintenance Descriptive Analytics” related to the past over the “Maintenance Diagnostic Analytics” describing the understanding of actions and the “Maintenance Predictive Analytics” predicting future actions to the “Maintenance Prescriptive Analytics” addressing future interventions. This work also treats the influences of predictive maintenance on the industry.

Data policy is an acute issue in the modern and globalized world. Dryde, et al. (2018, p. 2) give an overview of Big Data in general and especially in the modern transportation infrastructure. The key point is: The more vehicle data is collected the more knowledge about the vehicle condition but also about the customer is built up. Liang (2010, p. 3) analyzes customer value by integrating data mining approach in the automotive maintenance industry and hence showing the great value Big Data can have for the customer relationship management.

Reflecting that there are not only advantages to the term of big data, the abuse of data is a critical issue in politics but also for the use of predictive maintenance. Not everyone is credited with the collection of their data. Therefore specific parties in the automotive industries have to look not only for their advantages but also on the legal and ethical consequences. In conclusion, yet there are no published examples in which collected data of vehicle engine faults is studied in relation to the different seasons with the presented availability of vehicle data and the extensive use of vehicles in the transportational infrastructure we claim that it is a very important research field.

3. Methology

The statistical analysis of specific, seasonal dependent error codes, deposited in the onboard diagnostic memory, are based on the dataset the company Geotab as industry partner provided for this project in terms of a CSV-file shown in Table 1.

Id	Make	Model	Year	Vehicle Type	Class Label	Fault Date Time	Description	Odometer Value
41570	Mack	CL	2000	Truck	Mixed	2017-10-18 22:56:45.027 UTC	Low priority warning light on	547254.3
41570	Mack	CL	2000	Truck	Mixed	2017-09-21 12:30:56.637 UTC	Low priority warning light on	540845.6
41570	Mack	CL	2000	Truck	Mixed	2017-11-13 20:27:48.438 UTC	Low priority warning light on	548664.7

Table 1: Geotab CSV-file. (Source: Authors)

This file contains the exact designation of the error codes with the associated date of occurrence. This is a spectrum of 760.976 error codes of commercial vehicles that were read out from the OBD-II error memory and also resigned in big data by means of the Geotab GO-7 device, in 2017. It also contains different irrelevant parameters regarding seasonal statistic analysis which were in consequence excluded from further analysis. The aim in the first place was the preparation of a goal-oriented parameter separation of the codes to obtain suitable data for statistical analysis. In the further course for the statistic analysis, “Tableau Desktop 2018.1” and “IBM

SPSS Statistics Subscription Trial for Mac OS" were used. Primarily the effortless incorporation in these programs, furthermore the various useable visualization possibilities for our issues combined with the opportunity of the statistic calculation, were the main reasons for the use of these two software solutions.

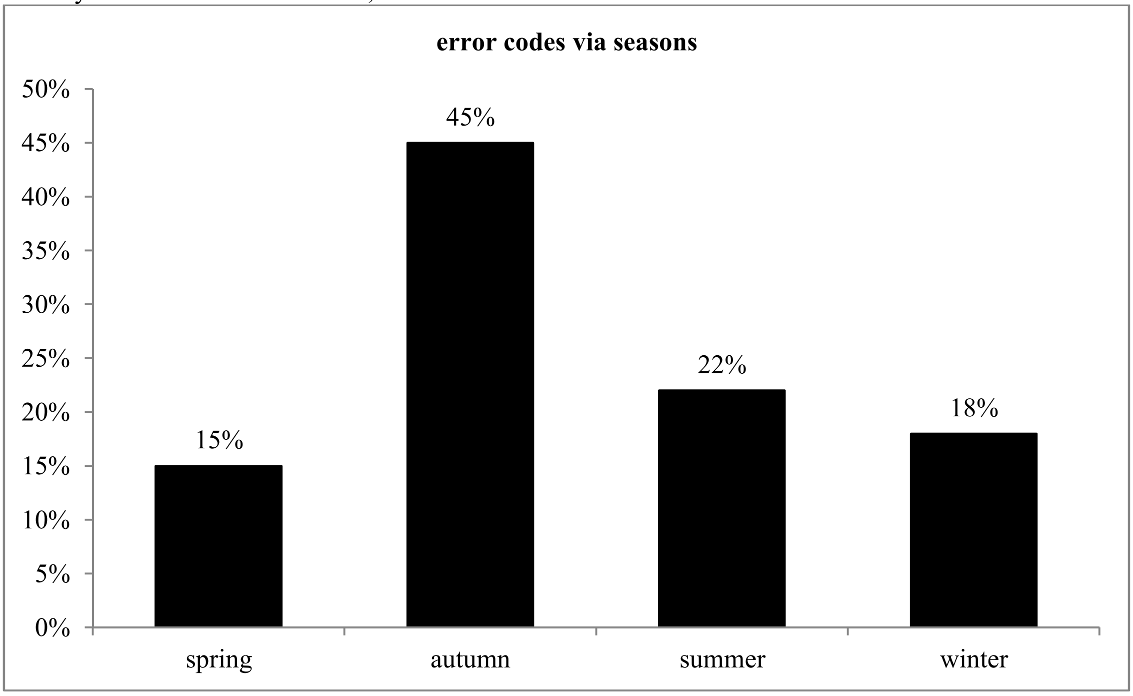

Fig. 1: all error codes via the seasons. (Source: Authors)

3.1 Graphical analysis of the datasets

The first step was to graph the data by using "Tableau Desktop 2018.1" in order to be able to evaluate and check for seasonal dependencies. The main advantage of this program is that even large amounts of data can be edited without necessary programming skills nor a extensive training. Furthermore, statistical relations are quickly and easily recognizable. We were able to apply all 760.976 error codes over the months and the four seasons. Moreover we were able to visualize the percentage dependency of individual codes. However a statistically meaningful statement must be made, as to whether the dependencies differ significantly.

3.2 Statistical analysis

The target in SPSS consisted the statistic analysis of the before visualized errors in "Tableau". We decided to single out the three most common occurred errors.

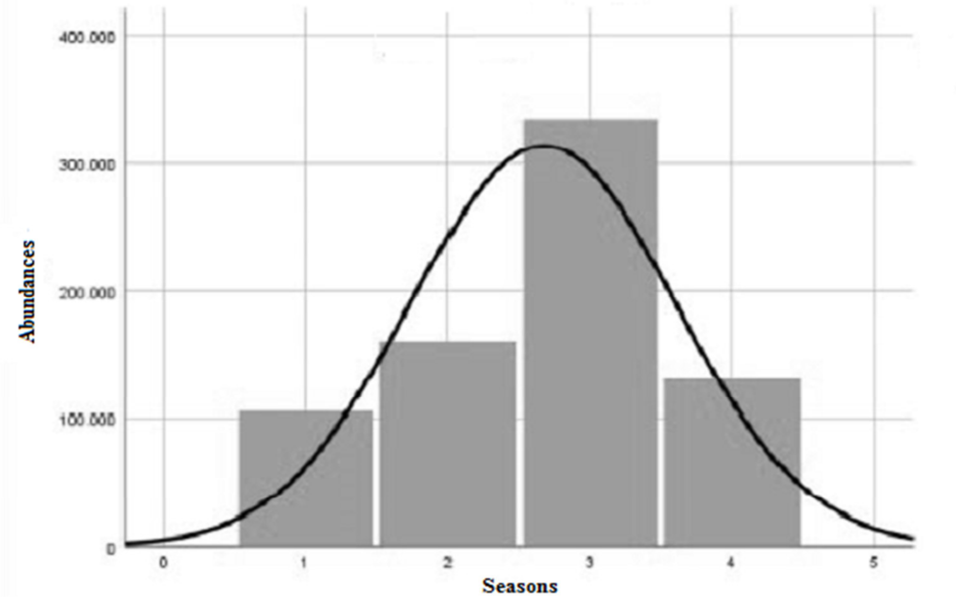

Fig. 2: Gaussian distribution curve. (Source: Authors)

The frequency of power train coherent error codes over the seasons is displayed in Figure 2. The errors were checked if they show any indication of normal distribution, which is a mandatory requirement to perform the needed statistics analysis. For better visualization, a Gaussian curve is also applied.

				95% confidence interval of the difference	
	Significance (2-digits)	mean difference	standard error of the difference	low	high
variances are equal	,000	,146	,032	,083	,209
Variances are unequal	,000	,146	,032	0,83	,209

Table 2: t-test analysis/significance. (Source: Authors)

Table 2 contains the evaluation from the t-test of the three selected errors. As expected the data is normal distributed. A t-test is executed. Later on, the attention turns to the significance.

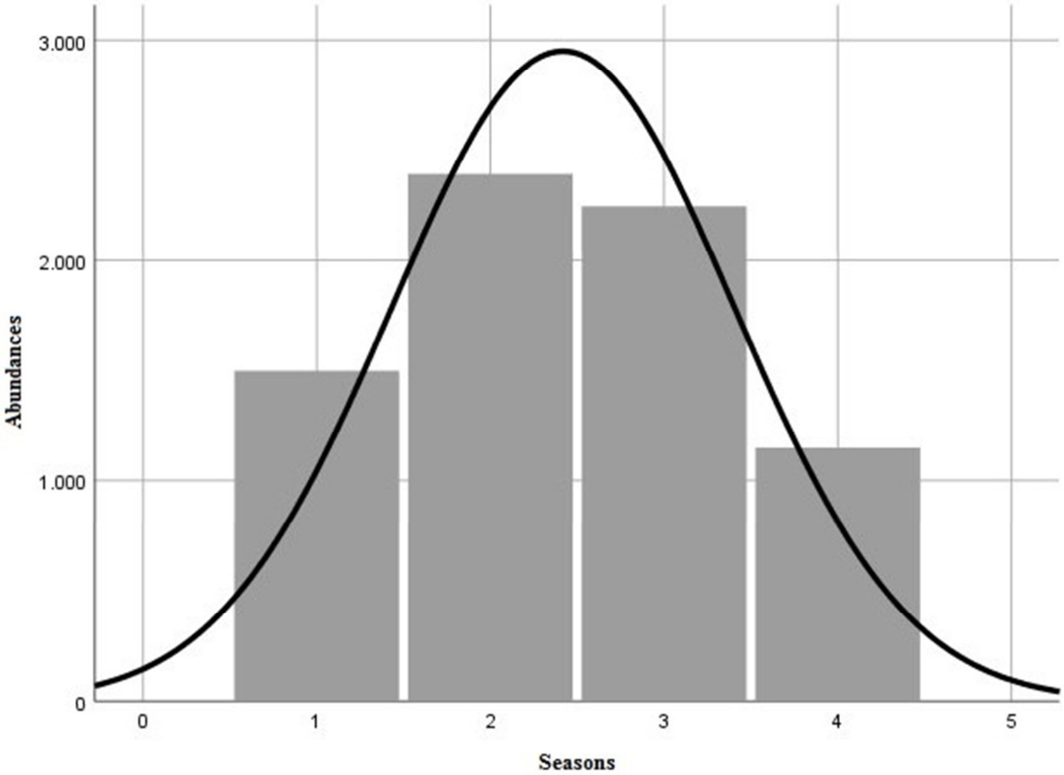

Fig. 3: Gaussian distribution curve. (Source: Authors)

Further we accomplished an examination of seasonal dependence on additional two errors, that are not in direct connection to the powertrain. Here as well, the necessary condition that a normal distribution of the errors is present, was examined. Another t-test was performed on the two supplementary non-powertrain relatet errors.

				95% confidence interval of the difference	
	Significance (2-digits)	mean difference	standard error of the difference	low	high
variances are equal	,696	-,021	,054	,126	,084
variances are unequal	,700	-,021	,054	-,128	,086

Table 3: t-test of non-powertrain connected error. (Source: Authors)

4. Results

Using the graphical representations of Tableau and SPSS, plus the various statistical tools, we were able to tget detailed knowledge of the seasonal error codes. After selecting three error codes from Tableau with the help of the graphic evaluations, we examined them for significance. As we have already assumed, they show a suspected seasonal dependency. In the case of two additional faults, which have no technical connection with the powertrain and have been tested as a counter-test, we were able to prove that there is no dependency of the season.

In the further course the results of the three seasonal errors will be considered. Figure 4 clearly shows the seasonal differences of the error: Vehicle battery has low voltage. In total, the error occurred 86,673 times, accounting for 11.3% of the total volume. While in the spring (12%), in the summer (16%) and in the winter (20%) the differences are only small, it is clearly visible that the autumn with 52% presents a big discrepancy. The level of significance here was well below 1%.This confirms the previous hypothesis of seasonal dependence.

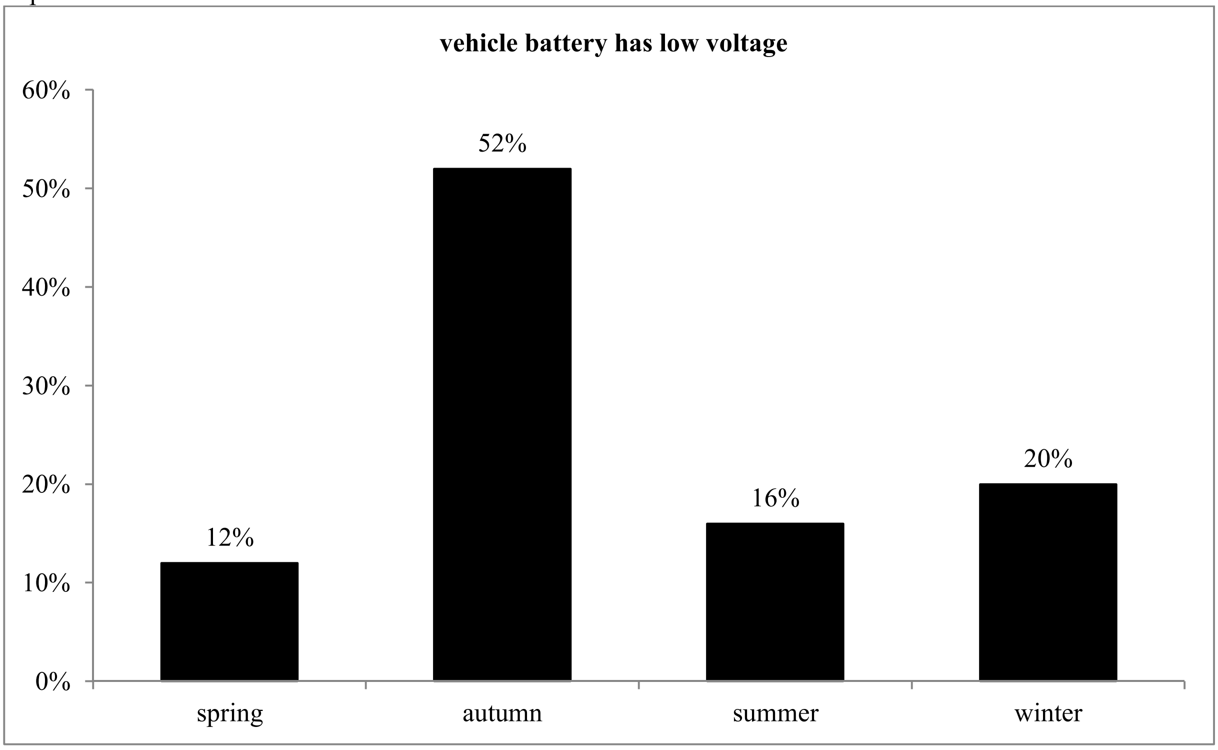

Fig. 4: error code "vehicle battery has low voltage" via the seasons. (Source: Authors)

	abundances	percent	valid percent	cumulative percent
spring	10112	11,7	11,7	11,7
autumn	45362	52,3	52,3	64
summer	14005	16,2	16,2	80,2
winter	17194	19,8	19,8	100
total	86673	100	100	

Table 4: error code: „vehicle battery has low voltage" abundances in percent via the seasons. (Source: Authors)

The next diagram (Fig. 5.) shows the error: Low priority warning light on. The error occurred a total of 425,473 times and has therefore the largest of all shares at 55.91%.

The distribution between spring (15%), summer (22%) and winter (18%) is nearly the same. It is clearly shown, that in autumn 45% of the volume showed the most abnormalities. Due to the significance on the 1% level tested in SPSS, a seasonal dependency can be confirmed.

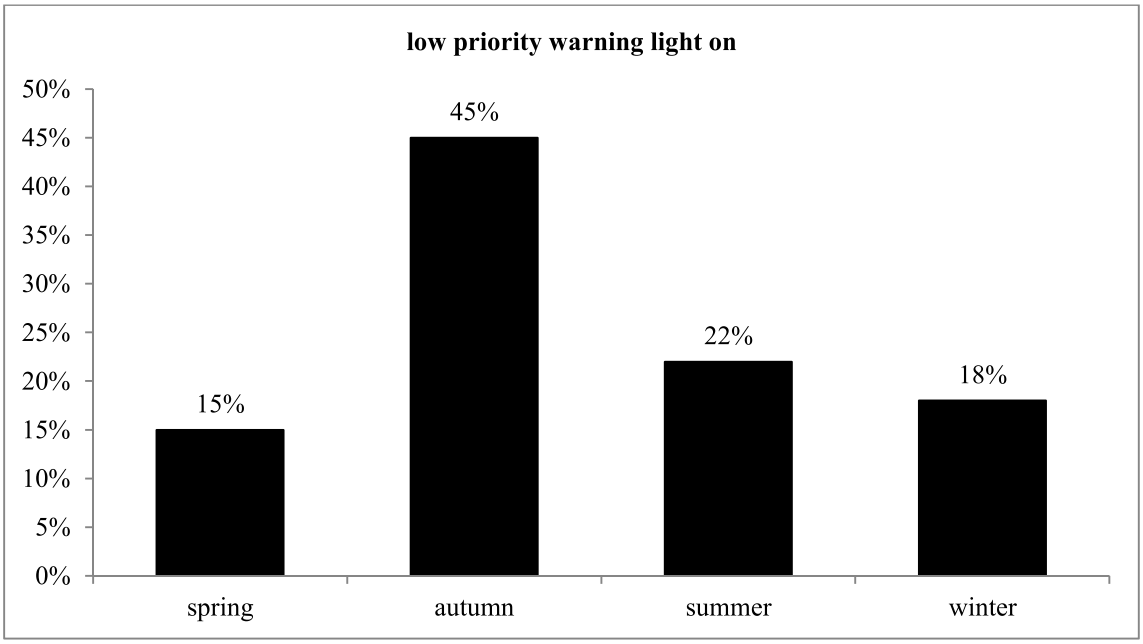

Fig. 5: error code „low priority warning light on" via the seasons. (Source: Authors)

	abundances	percent	valid percent	cumulative percent
spring	63307	14,9	14,9	14,9
autumn	188318	45,3	45,3	60,2
summer	98497	22,2	22,2	82,4
winter	75351	17,6	17,6	100
total	425473	100	100	

Table 5: error code: „low priority warning light on" abundances in percent via the seasons. (Source: Authors)

The third error (Fig. 6.) we studied was: General vehicle warning light on. The bar graph generated in Tableau, also illustrates our hypothesis of seasonality. As shown, it is visible that the frequency of the occurring error in the summer (22%), winter (18%) and spring (15%) are relatively similar. Autumn stands out with 45%, which in turn indicates a dependency.We were also able to use SPSS to determine the significance level, which was well below 1%.Thus, the hypothesis was confirmed here as well.

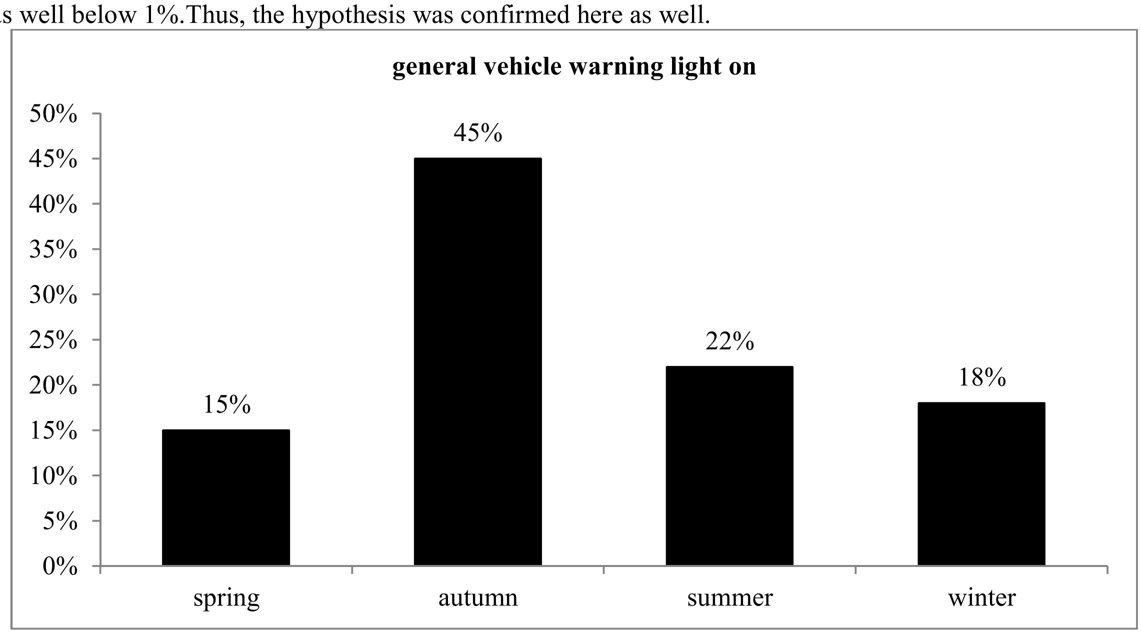

Fig. 6: error code: „general vehicle warning light on" via the seasons. (Source: Authors)

	abundances	percent	valid percent	cumulative percent
spring	32103	14,5	14,5	14,5
autumn	101135	45,7	45,7	60,2
summer	49018	22,1	22,1	82,3
winter	39239	17,7	17,7	100
total	221495	100	100	

Table 6: error code "general vehicle warning light on" abundances in percent via the seasons. (Source: Authors)

In summary, the question of a seasonal dependence can be answered with "yes". Especially powertrain specific fault codes can be assigned to a specific season. This suggests that an increased default rate occurs here. Exactly at this point, fleet operator can start with a predictive maintenance and remedy. This would significantly reduce the number of possible failures and thus contribute to a reduced time, in which they cannot be used.

5. Conclusions

Vehicle maintenance is a relevant field for OEMs, car service providers as well as for the owner of the vehicle. Proper maintenance reduces the risk of engine faults and failures. However changed vehicle components with inadequate deterioration are as expensive as faults in the long-term range. Predictive maintenance is the solution for this issue.

We present a data driven method to relate three different engine faults with different seasons. This analysis was realized with logged-vehicle data provided by Geotab Inc. These data sets were never used for this purpose before.

This work was focused on three main types of engine faults: the low battery voltage, the low priority warning light and the general vehicle warning light. The method applied thus can be used to analyze every recorded engine fault. The ability of editing large amounts of data are the main advantages of "Tableau Desktop 2018.1" as well as "IBM SPSS Statistics Subscription Trial for Mac OS".

As shown in the results, all three of the analyzed faults are related to season which could be a significant result for the automotive industry. With the presented knowledge for example a heavily used battery could be changed by the end of summer because the risk of a battery failure is over 50% higher in autumn than in the other seasons. This could reduce seasonal engine faults substantially and with that the overall cost of the vehicle ownership as well as the warranty costs for unnecessary changed components for the OEM and also for the service shop. But with all the advantages the minimized costs always have to be seen in relation to the cost of the implementation of recording and analyzing vehicle data.

Finally to say is that predictive maintenance is a topic which gets more and more important in todays research and practical discussions. Not only the increasing amount of recorded data will push the significance of predictive maintenance but also the ascending numbers of all sorts of transportational vehicle. This research therefore marks a suitable starting point for further investigations.

6. Literature

Ahmad Rosmaini, Kamaruddin Shahrul (2012): An overview of time-based and condition-based maintenance in industrial application. In: Computers & Industrial Engineering, February 18, 2012, Malaysia, pp. 135 – 149.

Dragomir Otilia Elena, Gouriveau Rafael, Dragomir Florin, Minca Eugénia, Zerhouni Noureddine (2009): Review of prognostic problem in condition-based maintenance. In: European Control Conference ECC'09, September 21, 2009, Budapest, pp. 1585 – 1592.

Dryde Ian L., Hodge David J. (2018): Journeys in big data statistics. In: Statistics and Probability Letters, February 28, 2018, Nottingham, pp. 121 – 125.

Frisk Erik, Krysander Mattias, Larsson Emil (2014): Data-Driven Lead-Acid Battery Prognostics Using Random Survival Forests. In: Conference Paper of the annual Conference of the prognostics and health management society 2014, October 2, 2014, Fort Belvior, pp. 92 – 101.

Jardine Andrew K. S., Lin Daming, Banjevic Dragan (2005): A review on machinery diagnostics and prognostics implementing condition-based maintenance. In: Mechanical Systems and Signal Processing, February 24, 2005, Toronto, pp. 1483 – 1510.

Karim Ramin, Westerberg Jesper, Galar Diego, Kumar Uday (2016): Maintenance Analytics – The New Know in Maintenance. In: IFAC-PapersOnLine, December 16, 2016, Lulea, pp. 214 – 219.

LiangYi-Hui (2010): Integration of data mining technologies to analyze customer value for the automotive maintenance industry. In: Expert Systems with Applications, May 12, 2010, Ta-Hsu Hsiang, pp. 7489 – 7496.

Manowicz A.-A.. Connected remote services - a new approach of connecting customers with porducts and dealers. In: Proceedings of the 8th Biannual CER Comparative European research conference, October 30-31, 2017, London, pp. 8 – 11.

Medjaher Kamal, Tobon-Mejia Diego Alejandro, Zerhouni Noureddine (2012): Remaining Useful Life Estimation of Critical Components With Application to Bearings. In: IEEE Transactions on Reliability, April 26, 2012, pp. 292 – 302.

Miao Qiang, Xie Lei, Cui Hengjuan, Liang Wei, Pecht Michael (2012): Remaining useful life prediction of lithium-ion battery with unscented particle filter technique. In: Microelectronics Reliability, December 25, 2012, Chengdu & Maryland, pp. 805 – 810.

Mobley R. Keith (2002): An Introduction to Predictive Maintenance. Butterworth Heinemann, 2002, Woburn.

PengYing, DongMing, Zou Ming Jian (2010): Current status of machine prognostics in condition-based maintenance: a review. In: The International Journal of Advanced Manufacturing Technology, January 06, 2010, pp. 297 – 313.

Prytz Rune, Nowaczyk Slawomir, Rögnvaldsson Thorsteinn, Byttner Stefan (2015): Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data. In: Engineering Applications of Artificial Intelligence, March 05, 2015, Göteborg & Halmstad, pp. 139 – 150.

Yam R.C.M., Tse P.W., Li L., Tu P. (2001): Intelligent Predictive Decision Support System for Condition-Based Maintenance. In: The International Journal of Advanced Manufacturing Technology, February, 2001, Hong Kong & Christchurch, pp. 383 – 391.