

**REGARDING THE IMPORTANCE OF THE MILEAGE
DISTRIBUTION FOR RELIABILITY PROGNOSIS IN
AUTOMOTIVE INDUSTRY**

**ROZKŁAD PRZEBIEGU A PROGNOZOWANIE
NIEZAWODNOŚCI W PRZEMYSŁE
SAMOCHODOWYM**

**Dirk Althaus¹, Andreas Braasch²,
Marco Schlummer³, Arno Meyna⁴**

University of Wuppertal
Gaußstr. 20, 42097 Wuppertal, Germany

(1) E-mail: dalthaus@uni-wuppertal.de (2) E-mail: braasch@uni-wuppertal.de
(3) E-mail: mschlumm@uni-wuppertal.de (4) E-mail: meyna@uni-wuppertal.de

Abstract: *In automotive industry there are few information about the load of a car in use. Therefore, the available data needs to be used and looked at carefully. Data that is nearly always available is the time in field, given by the date of registration and the date when failure occurred, and the driving distance. The driving distance can be used to describe the load of a car because in most cases the operating time is not known. This data can be used to calculate the mileage distribution of a car, which contains much additional information for safety and reliability analysis. Furthermore, it is useful to adopt testing routines and lifetime distances that are written down in specifications.*

Keywords: *mileage, reliability prognosis, failure behaviour*

Streszczenie: *Przemysł samochodowy nie posiada wielu danych o obciążeniu eksploatowanego samochodu, a zatem należy korzystać z tych danych, które posiadamy i dokładnie się im przyglądać. Danymi, które są zawsze dostępne, jest czas eksploatacji określony przez datę rejestracji i datę awarii oraz przejechany kilometr. Przejechana odległość może służyć do określenia obciążenia samochodu, ponieważ czas jego użytkowania jest w większości wypadków nieznan. Parametr ten może posłużyć do obliczenia rozkładu przebiegu samochodu, który zawiera wiele dodatkowych informacji przydatnych w analizie bezpieczeństwa i niezawodności. Ponadto ma on znaczenie w przyjęciu programów badań okresowych i formułowaniu założeń dotyczących trwałości/całowitego przebiegu.*

Słowa kluczowe: *przebieg, prognoza niezawodności, zachowanie awaryjne*

Introduction

One possibility of getting a reference parameter for the load in road vehicles, especially passenger cars, light and heavy good trucks, is using the driving distance. This parameter is also affected by the driving behaviour, e.g. frequent vs. infrequent drivers, but is not manipulated by the driver in general. As additional information the time in service (TIS), that means time between the first date of registration and date of defect, needs to be known. Using this data the annual mileage for a car can be calculated. For passenger cars the mileage usually follows a lognormal distribution whereas sometimes for commercial vehicles a normal distribution can be used.

Knowing the mileage distribution can have benefit for many areas in automotive industry. First use case is to define the car specification. Today it is common to fix the car lifetime at 10 years and 300,000 km. When using the 95%-percentil of the mileage distribution of an equal or parent car as the maximum driving distance a more realistic lifetime distance can be calculated. For several cars in field it will be much more or less than 300,000 km. This value has also a direct impact on testing routines. When having a lower lifetime mileage the cost-intensive long time testing can be reduced to save money and precious testing time. However, testing for cars with higher mileage can be extended for safety reasons.

The mileage has also an impact on reliability prognoses when using the reliability prognosis model, invented by the University of Wuppertal. As a first check, an uncommon mileage distribution is an indicator for systematic failure modes. In further calculation of the model, the mileage is used to correct the recorded failures (correction of failure aspirants) and to derive the time-dependent failure probability. Therefore, “wrong” mileage distributions can have a significant effect on the achieved results.

1. The mileage database

The mileage database of the section Safety Theory & Traffic Engineering of the University of Wuppertal is based on warranty data, data of automobile clubs and data of general inspections of cars that are regulated in Germany. Over 3.5 million entries of vehicles are discharged. With a self-programmed software the data can be read out and analyzed immediately. Thus the database enables analyses of nearly all vehicles placed on the German market regarding the particular mileage of the car brand, the car type, the different motorization or the type of drive. Furthermore, it is possible to have a look at special time periods scaled by registration dates or the failure dates of the cars. Hereby anonymity of the driving behaviour of every car

owner is assured, because no vehicle identification number (VIN) is included.

2. Mathematical background

The recorded data has no identical TID. Therefore each dataset needs to be scaled to an identical base. Using the annual mileage as base is meaningful, because the mileage can be reached out of nearly every dataset, checking if parameters are meaningful is very easy and adoption to other time scales is possible.

If the mileage per year with $a=365$ days is expected to be constant over the years of usage the annual mileage is calculated as follows:

$$s_1(i) = \frac{s_A(i) \cdot a}{\Delta t_{A,Z}(i)} \quad (1)$$

with:

$s_1(i)$: mileage of the car i within one year,

$s_A(i)$: mileage of the car i until failure [Tkm],

$\Delta t_{A,Z}(i)$: time in field [days] of car i until failure.

To achieve a theoretical distribution function for the mileage distribution the empirical data has to be fit by means of appropriate parameter estimations. In most cases for passenger cars the lognormal distribution with

$$L(s) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_0^s \frac{1}{\tau} \cdot e^{-\frac{(\ln \tau - \mu)^2}{2\sigma^2}} d\tau = \Phi\left(\frac{\ln s - \mu}{\sigma}\right) \quad (2)$$

has been proved to fit the empirical data quite good. The parameters are estimated by Maximum-Likelihood method in the way

$$\hat{\mu} = \frac{1}{n} \cdot \sum_{i=1}^n \ln s_1(i) \quad \text{and} \quad (3)$$

$$\hat{\sigma} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (\ln s_1(i) - \hat{\mu})^2} \quad (4)$$

with:

n : total number of data,

$\hat{\mu}, \hat{\sigma}$: estimated parameters of mileage distribution.

For commercial vehicles the normal distribution is often suitable.

The expected annual mileage $E_1(S)$, calculated from

$$E_1(S) = e^{\left(\hat{\mu} + \frac{\hat{\sigma}^2}{2}\right)}, \quad (5)$$

is often used to compare the driving behaviour of car types.

The mileage of a car can be received out of nearly every existing dataset, when there are enough entries (>25). As in general analysis gets better with more usable data.

The Weibull distribution can be used to describe the driving behaviour, too. When assuming a Weibull distribution the method of least squares gives satisfying results. But it can be shown, when analyzing a huge amount of data, that the lognormal distribution obtains better results than the Weibull distribution (Pauli, 1998). Beyond that parameter estimation for Weibull parameters can only be done numerically. By contrast parameters μ and σ of the lognormal or normal distribution can be calculated directly by most spreadsheets (e.g. MS Excel), which are available at nearly every used computer.

The following Fig. 1 shows the empirical mileage distribution (dots) and the fitted theoretical mileage distribution (line) for a passenger car (example 1), with estimated parameters $\hat{\mu}_{ex1} = 2.9638$ and $\hat{\sigma}_{ex1} = 0.4363$. The value of the expected annual mileage is $E_{ex1}(S) = 21.3 \text{ Tkm}$.

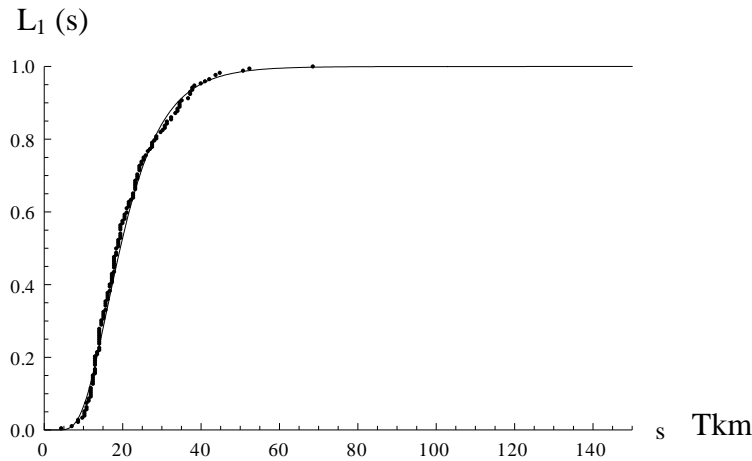


Fig. 1 Empirical (dots) and theoretical (line) mileage distribution

As can be seen in Figure 1, the theoretical function fits the empirical data very well. This qualitative aspect can be approved quantitatively by the coefficient of determination, which is derived from

$$B = \left(\frac{\sum_{i=1}^n x_i \cdot y_i}{\sqrt{\left[\sum_{i=1}^n x_i^2 \right] \cdot \left[\sum_{i=1}^n y_i^2 \right]}} \right)^2 \quad (6)$$

with:

x_i : i^{th} empirical distribution value,

y_i : i^{th} theoretical distribution value.

Guide value for the coefficient received out of many analyses is $B > 0,996$. For the example with 153 datasets the value is $B=0.9986$, which tends to be suitable for the analyzed car type.

When assuming a time constant driving behaviour over lifetime $t_L = c \cdot t_1$ the parameters for the lifetime mileage $L_L(\mu_c, \sigma_c)$ can be calculated directly by a linear transformation of the annual mileage $L_1(\mu_1, \sigma_1)$, with:

$$\mu_c = \mu_1 + \ln c \quad \text{and} \quad (7)$$

$$\sigma_c = \sigma_1. \quad (8)$$

As will be shown in the following chapter, the mileage distribution has an impact on many decisions and specifications.

3. Usage of mileage

The mileage of a car contains a lot of information that are used for different purposes.

4.1 Distance over lifetime

In specifications for passenger cars lifetime generally is defined to be $t_L = 15 \text{ years}$. A common assumption for the averaged driving distance per year is $s_1 = 20,000 \text{ km}$ and hence the distance over lifetime is $s_L = 300,000 \text{ km}$. This value is often used as reference value for many purposes. Nearly every testing routine is build up on this value. But as expected it can be shown, that the mileage differs a lot between different car types. In the following Table 1 there are typical values for different car classes (Stiels 2008). The given parameters are the quantity of datasets, the averaged days in field (TIS: time in service), the parameters μ_1 and σ_1 of the annual mileage distribution, the annual expected value $E_1(S)$ and the coefficient of determination B.

Table 1: Parameters of different car classes

Class	Quantity	$\bar{\varnothing}$ TIS	μ_1	σ_1	$E_1(S)$ [Tkm]	B
City cars	45,322	2,715	2.4638	0.4598	13.1	0.9980
Mini	64,983	3,410	2.4196	0.4761	12.6	0.9978
Subcompact	65,263	3,676	2.6063	0.4589	15.1	0.9984
Compact	65,312	3,372	2.7853	0.5131	18.5	0.9977
Mid-size	65,338	2,865	2.9637	0.5627	22.7	0.9986
Luxury	14,110	3,350	2.9162	0.5138	21.1	0.9986

Each dataset contains of many different car types with no time limit. But nevertheless the parameters are representative for their classes. The parameters of the car from example 1 match the parameters of the mid-size or luxury cars and indeed the car is classified as a mid-size car.

When extrapolating the annual expected values over lifetime 300,000 Tkm seems to fit lifetime distance very well, also mid-size and luxury car exceed the value. But the question came up, if this is the right value. In Table 2 the expected value for 15 years $E_{15}(S)$ and the distances for the 95% and 99%-percentil for one and for 15 years are given.

Table 2: Parameters for different car classes

Class	$E_{15}(S)$ [Tkm]	$s_1(90\%)$ [Tkm]	$s_1(95\%)$ [Tkm]	$s_{15}(90\%)$ [Tkm]	$s_{15}(95\%)$ [Tkm]
City cars	196.5	21.2	25.0	317.7	375.5
Mini	189.0	20.7	24.6	310.4	369.0
Subcompact	226.5	24.4	28.8	365.9	432.3
Compact	277.5	31.3	37.7	469.1	565.3
Mid-size	340.5	39.8	48.9	597.6	733.1
Luxury	316.5	35.7	43.0	535.2	645.1

When having a look at the 95% driver it is obvious, that over lifetime in all classes the value is overrun. The value for the 95% driver seems to be too high for reference value and should not be taken into account for decision making. When having a look at the original data there are a lot of cars reaching 300,000 km before aged 15. The assumed value for lifetime

distance tends to be best expressed by the 90% driver and should be set, according to the results, independently for each car.

As can be seen the 90% values for compact, mid-sized and luxury cars are up to 100% higher than the assumed value. For the components of the car and especially the safety related ones this can denote that testing procedures are not sufficient for lifetime.

4.1 Usability for reliability and safety calculations

Reliability and safety analyses are typically based on field failures that occurred while the car was in use by a customer. The failure data, recorded by the automobile manufacturer or their suppliers, is in general very dependable. A disadvantage of the database is the low amount of data when setting filter criterions as e.g. the car type and the date of production. This often leads to datasets with less than 30 entries. But nevertheless first reliability approaches are possible.

Assuming that failures occur stochastically and not systematically they could appear at every car, regardless if driven by frequent or infrequent drivers. The derived mileage distribution out of this data should be the same as for the overall car type (Pauli 2001). But often the failure behaviour is systematic e.g. trough early failure behaviour or wearout. Accordingly failures mainly occur at cars of frequent drivers, because they first reach a distinct mileage. The mileage distribution of these cars scaled to one year leads to very high mean values with a wide variance and doesn't fit the real distribution. Conclusion is that failures are systematically and should be analyzed in more detail.

Such data can be analyzed by a reliability prognosis model, which was invented by the University of Wuppertal and a major German automotive supplier. In (Pauli 1998), (Meyer 2002) and (Althaus 2009) an elaborate explanation is given.

The procedure of the reliability prognosis model is shown in the following Fig. 2.

The prognosis model uses the field failure data, which were recorded at every guarantee case. The useful data consist of the registration date, the date of failure and the driving distance until failure, because the operating time is usually not recorded in automotive applications. The sales quantity for the component is also necessary.

First step after data screening is to determine the mileage distribution like described in chapter 3.

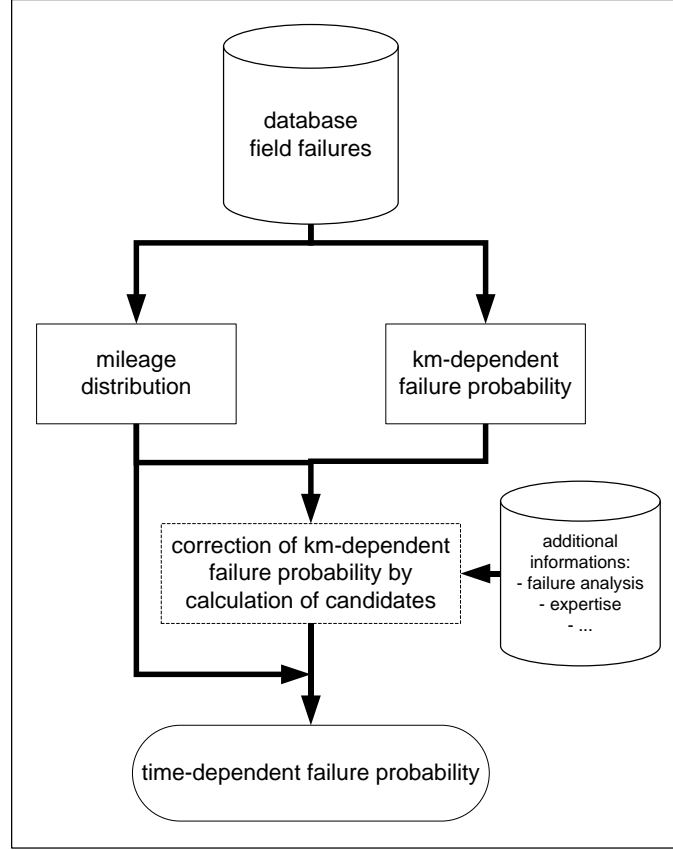


Fig. 2 Procedure of the reliability prognosis model

Next step is to determine the km-dependent failure probability. Due to the fact that commonly only data of the guarantee time is available, the empirical km-dependent failure probability is corrected by the calculation of candidates,

$$\tilde{F}_k(s) = \frac{n_k(s)}{n_0} = \frac{1}{n_0} \cdot \sum_{\zeta \leq s} n_k(\zeta) = \frac{1}{n_0} \cdot \sum_{\zeta \leq s} \frac{n_g(\zeta)}{1 - L_g(\zeta)} \quad (9)$$

with:

- $\tilde{F}_k(s)$: corrected km-dependent failure probability;
- $n_k(s)$: corrected number of failures at distance s ;
- n_0 : total sales quantity,
- $L_g(\zeta)$: mileage distribution for the guarantee time,
- $n_g(\zeta)$: failures in guarantee time at distance ζ .

As shown in formula (9) the mileage distribution has an extensive impact on the correction by candidates.

To receive a theoretical corrected km-dependent failure probability $F_k(s)$, which fits the empirical data, parameter estimation using the method of least squares is advised.

Next step after exceeding the km-dependent failure probability is to calculate the time-dependent failure probability. By using the theoretical failure density $f_k(s)$ and the mileage distribution for guarantee time the time-dependent failure probability $F(t)$ can be numerically calculated by:

$$F(t) = \int_0^{\infty} f_k(s) \cdot \left(1 - L_G \left(\frac{s \cdot t_G}{t} \right) \right) ds. \quad (10)$$

The mileage distribution is used here, too. It is thus needed twice in the model. It is obvious that using a wrong mileage distribution results in heavy biased values for candidates and for the time-dependent failure probability. In the next figures there are shown the km-dependent original and corrected results of an example, where on the hand only the warranty data is used (Fig. 3) and on the other hand independent data (Fig. 4) is used to calculate the mileage distribution.

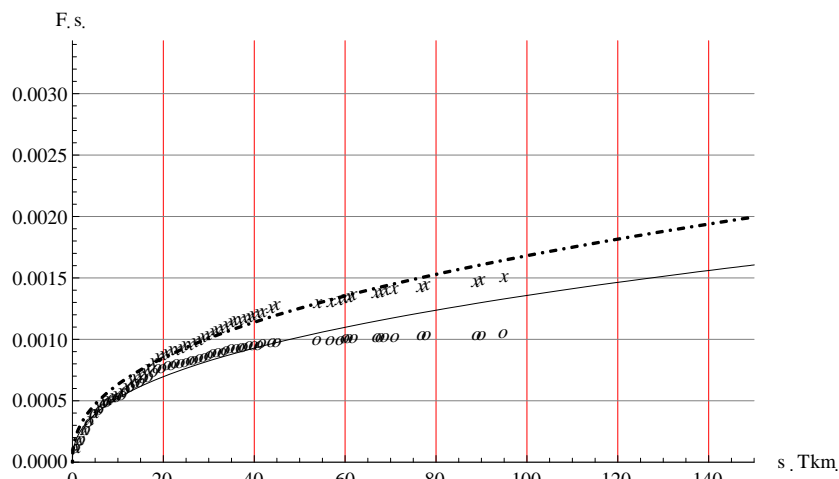


Fig. 3 Original (“o”) und corrected (“x”) km-dependent failure distribution using warranty mileage data

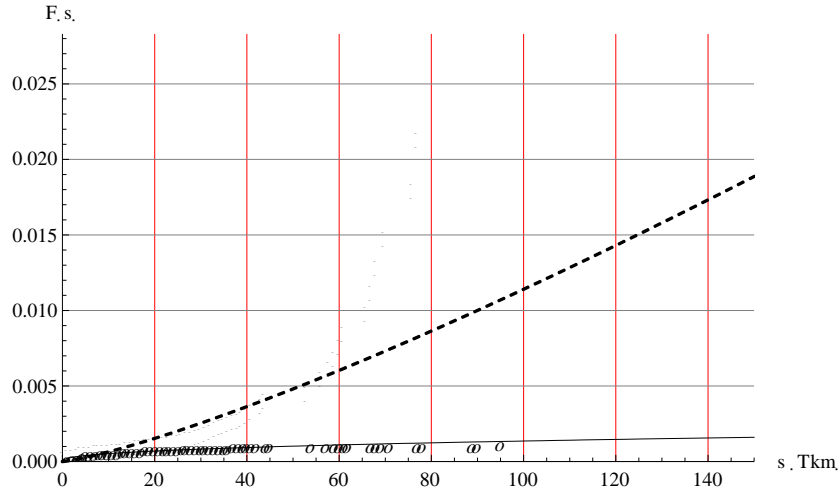


Fig. 4 Original ("o") and corrected ("-") km-dependent failure distribution using independent mileage data

The estimated parameters for each data set are given in Table 3.

Table 3. Prognosis model parameters for warranty and independent data

Parameter		Warranty data set	independent data set
Mileage	μ	2.8398	2.7853
	σ	1.1372	0.5131
	B	0.9848	0.9936
km-dependent	α [Tkm]	238.3E-6	36.0E-6
	β	0.4244	1.2514
	B	0.9968	0.7156
time-dependent	α [year]	893.3E-6	1.458E-3
	β	0.4243	1.2445

The mileage distribution of the independent data is taken out of Table 1, compact car, because it is known, that the analyzed car belongs to this group. The parameters of the km-dependent failure probability are very different in those cases, best seen by the extremely different variances. The estimated failure behaviour changes from early failure behaviour to wearout.

Both cases should be used very carefully. With additional information, here e.g. that the component is an electronic control unit (ECU), that normally has early failures, the results could be checked, if they are plausible or not. Anyhow, the prognosis model is suitable for reliability predictions. If possible the estimated parameters out of the warranty data should be checked against data from independent sources.

4. Conclusion

The data that is available out of warranty claims in automotive industry is often improper. The load of the car and its components is not recorded in most cases. Therefore, an alternative needs to be found, to do load calculations. In many analyses it could be shown that the driven mileage is useful to describe the load.

An easy to estimate mileage distribution contains a lot of information which are yet often not used.

Testing routines are defined for nearly every car with 15 years and 20,000 km per year. Thus lifetime distance is set to be 300,000 km. But, as to be expected, the mileage is very different from car to car. A small car is usually used for low range distances such as in city traffic meanwhile large cars are often used by representatives, taxis, rental cars or people, who need to drive a lot.

Estimating the mileage distributions of cars, like shown in the article, gives a wide range of parameters. The 90% percentile of the mileage is in every case higher than 300,000 km and can exceed 500,000 km. Conclusion must be to look at the specifications and the defined testing routines for cars with regard to their real mileage.

Furthermore, the mileage is used in reliability and safety predictions for automotive components. In the reliability prognosis model, invented by the University of Wuppertal and major German automotive supplier the mileage is used two times: to correct the failure candidates and to transform the km-dependent failure probability into a time-dependent failure probability. Often, the warranty data is influenced by a special group of drivers, e.g. frequent drives, which can lead to biased mileage distributions. Therefore a second mileage out of independent data can be estimated, to check the given data. As shown in an example with warranty and independent data, the mileage can have such an extensive impact on the prognosis model that the forecasted failure behaviour is totally different (early vs. wearout failure behaviour). As conclusion the received mileage distribution out of guarantee data should be checked against an estimated distribution out of independent data.

References:

1. Althaus, D.: Ein praxisorientierter empirischer Ansatz zur Bestimmung des Ausfallverhaltens konventioneller Bremssysteme in Personenkraftwagen. Doctoral Thesis, University of Wuppertal, 2009.
2. Althaus, D.; Meyna, A.; Braasch, A.: Zuverlässigkeitsprognosen mit unabhängigen Fahrleistungsdaten. VDI-Bericht 2065, ISBN 978-3-18-092065-8, VDI Verlag GmbH, Düsseldorf, 2009.
3. Althaus, D.; Braasch, A., Schlummer, M., Meyna, A.: Comparison of the failure behaviour of the same function in different systems. ESREL'09, Taylor & Francis Group, London, ISBN 978-0-415-55509-8, 2009.
4. Braasch A.; Althaus D.; Meyna A.: Influence of the mileage distribution on reliability prognosis models. ESREL'08, ISBN 978-0-415-48513-5 Taylor & Francis Group, London, 2009.
5. Meyer, M.; Meyna, A.; Pauli, B.; Grodzick, P.: Reliability Prediction for Automotive Components in Field use. SAE Technical Paper 2002-01-2241, Society of Automotive Engineering, Inc., Warrendale PA, USA, 2002.
6. Pauli, B.: Zuverlässigkeitsprognosen für elektrische Steuergeräte im Kraftfahrzeug: Modellbildung und praktische Anwendung. Doctoral Thesis, University of Wuppertal, 1998.
7. Pauli, B.: Statistical Method for Recognizing the Failure Behaviour of Motor Vehicle Components in Field Use. ATZ Worldwide, Volume 4, Vieweg Verlag, Wiesbaden, 2001.
8. Stiels, M.: Auswertung unabhängiger Kraftfahrzeugdaten bezüglich der Fahrleistung und systematische Zusammenstellung zur weiteren Verwendung in Zuverlässigkeitsprognosen. Bachelor Thesis, University of Wuppertal, 2008.