Reliable fatigue design of personal vehicle chassis parts from multi-input loads and unsupervised statistical analyses

Emilien Baroux, Patrick Pamphile, Benoit Delattre, Andrei Constantinescu, Laurent Rota

Available online at www.sciencedirect.com

ScienceDirect

JOURNALNAME 00 (2023) 1–??

Journal Logo

www.elsevier.com/locate/procedia

Reliable fatigue design of personal vehicle chassis parts from multi-input loads and unsupervised statistical analyses

Emilien Baroux[a,b,c,d], Patrick Pamphile[c,d,*], Benoit Delattre[a], Andrei Constantinescu[b], Laurent Rota[a]

[a]*Stellantis, 2 Boulevard de l'Europe, 78300 Poissy, France*
[b]*Laboratoire de Mécanique des Solides, Institut Polytechnique de Paris, CNRS UMR 7649, Ecole Polytechnique, 91120 Palaiseau, France*
[c]*Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay, CNRS UMR 8628, Bat 307, 91405 Orsay Cedex 9, France*
[d]*CELESTE team Inria-Saclay, 1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau, France*

Abstract

A vehicle's chassis plays a critical role in its reliability and durability. To ensure occupant safety and vehicle maneuverability, it is necessary to assess the fatigue strength of chassis components, i.e. their ability to withstand repeated loads during use. This assessment begins at the design stage, with the identification of operating conditions and associated loads. In the case of personal vehicles, various loads must be considered due to diverse road types (e.g., highway, city, ...) and driving styles (aggressive, sporty, ...). In this paper we use a multi-dimensional characterization of the damage caused by external multi-input loads on wheels during vehicle use, including load combinations between the left and right wheels of the front and rear axles. Field measurements are used to calculate pseudo-damages for each load and road type, creating multivariate data with hierarchical structure. Unsupervised statistical analyses are used to explore correlations between pseudo-damages and identify driving profiles, providing a multi-dimensional assessment of severity while avoiding overlearning. A multi-dimensional Gaussian mixture model is then fitted to damage-equivalent constraints. This probabilistic model extrapolates damage computing and simulates driving styles, providing design teams with a stress analysis tool for accurate, realistic fatigue design of chassis components in future vehicles.

Keywords: Automotive fatigue design, Multi-input loads, Operational loads, Damage accumulations, Unsupervised statistical analyses, Finite mixture distributions

1. Introduction

1.1. Fatigue design

The main objective of modern structural design is to reduce production costs while maintaining the reliability and durability of the structure. Particularly in the context of the automotive industry, which must meet the requirements of reducing fuel consumption by minimizing the weight of vehicles, while ensuring

*Corresponding author:
Email address: patrick.pamphile@universite-paris-saclay.fr (Patrick Pamphile)

a high level of safety. The chassis plays a key role in safety, supporting the vehicle and ensuring its stability and handling. As a result, its design is a critical step in the development of a new vehicle model. Material fatigue is one of the major failure factors for mechanical parts: chassis components are subjected to repeated stresses due to wheel-transmitted loads caused by road irregularities and driving maneuvers [1, 2, 3, 4]. These stresses can lead to premature localized cracking and failure of parts essential to the stability and handling of the vehicle.

The chassis is comprised of numerous components such as wheel suspension, springs, shock absorbers, steering mechanism, drive shafts, steering column, etc. These components account for approximately 20% of the total weight of a personal vehicle [5]. To ensure occupant safety and chassis durability in a future vehicle, the design office must assess the fatigue strength of the various components, that is, their capability to endure recurrent loads during service. This involves identifying the vehicle operating conditions and analyzing the resulting loads. When analyzing the loads for a personal vehicle, there is considerable variability. This variability comes from the different uses of the vehicle, such as daily commuting, weekly shopping, leisure and vacation trips, etc., and from the different driving styles of the drivers, who may have aggressive, sporty, or economical driving styles, etc.

Load analysis for a single input signal benefits from extensive documentation and well-established fatigue damage calculation methods [6, 7, 8, 3, 9]. When the vehicle passes over a bump, the vertical load applied to the front left wheel generates a tensor of local stresses on zones of the suspension components. The stress tensor is related to the geometry of the component and the type of loads it is subjected to (tension, compression, etc.). The components of the stress tensor are then combined using a multi-axial method to obtain an equivalent stress. This is subsequently used to determine the component ability to withstand this loading. In fact, the local stress on an zone of a component results from the superposition of forces acting on the component and, therefore external loads to which the vehicle is subjected. Therefore, it is necessary to consider the loads applied to the wheel in all three directions (vertical as well as lateral and longitudinal), applied to all four wheels of the vehicle, as well as the load combinations between the left and right wheels of the front and rear axles.

Accurate evaluation of these various loads is critical to the design office, as it allows them to predict the fatigue strength of the different chassis components of a vehicle during the design phase. To achieve this, the design office uses either standard loads corresponding to specific scenarios such as acceleration, braking or cornering, or loads recorded during field tests with customer-driven vehicles equipped with sensors [10, 11]. Specific scenarios are characterized by their precision in defining load levels and their frequency. However, they cannot reproduce actual driving conditions due to driver behavior (e.g., aggressive or relaxed driving, etc.), use (e.g., urban or rural driving, etc.), road imperfections (e.g., potholes, bumps, etc.), traffic congestion, or weather conditions. This may result in deviations from the actual loads that the vehicle will be exposed to during its operational life. Customer-measured loads provide authentic data from vehicles on

the road, allowing the natural diversity of driving conditions to be taken into account. However, collecting this data in the field can be costly in terms of time, resources and budget. In addition, it is important to note that data collected in this way only reflects conditions specific to the location where the vehicles were used, which may not be an exhaustive representation of all possible conditions.

1.2. Objectives of the article

To ensure the reliable design of chassis components, it is imperative to thoroughly and accurately consider all potential load configurations that the future vehicle could be subjected to. This article begins by taking a holistic and realistic approach to modeling the loads applied to the vehicle chassis. This modeling takes into account the multi-dimensional nature of the load history by simultaneously analyzing the vertical, horizontal and lateral loads measured at each wheel, as well as the load combinations between the left and right wheels of the front and rear axles to account for the correlations between these multiple loads. In addition, we introduce the concept of "local context" to improve our understanding of potential damage that could occur in different zones of a part. We conducted a field measurement campaign on different types of roads representative of actual customer use. Using the load histories, we calculated a pseudo-damage for the local context and road type. Thus, we obtained multi-dimensional data with a hierarchical structure created by taking into account road types. Therefore, this methodology makes it possible to analyze damage caused by multi-input loads without having to refer to a specific chassis component to which these loads are applied. We then applied unsupervised statistical analyses adapted to the hierarchical structure of these data. These analyses allowed us to identify the main sources of variation in the different damages and to reduce sampling noise to make our analyses more robust. This allowed us to characterize driving profiles that are likely to cause damage in different zones of the structure. Secondly, the damage-equivalent stresses calculated for a reference road consisting of a mixture of road types were examined. Based on these data, a multi-dimensional Gaussian mixture model is fitted to estimate the fatigue strength of stress concentration zones for local contexts.

In section 2, we first recall the classical procedure for damage calculation from a history for a given load direction. A multi-dimensional characterization of chassis damage is presented in section 3. In Section 4, we will present the measurement campaign from which we will implement our methodology. Section 5 details the unsupervised multivariate statistical analyses performed on the damage measurements by road types. Probabilistic stress modeling for fatigue design of chassis components is discussed in Section 6. Section 7 concludes this paper.

2. Fatigue design based on load history

The analysis of the durability and reliability of a chassis part requires the calculation of the fatigue life of its components. External loads are forces or combinations of forces that are applied to the chassis and are

considered to be ”causes” of the fatigue phenomenon in the parts. Local stresses or strains in the parts are thus the consequences of the external loads and are considered to be ”effects” of the fatigue phenomenon.

We adopt the framework of Baroux thesis [12], assuming that the parts under study exhibit linear elastic behavior. For zones of the chassis where theses assumptions does not apply, such as the attachment points where the vehicle suspensions are attached to the chassis, specific analyses are performed to evaluate the actual stresses [13]. In addition, we assume that the loads considered are representative of normal use, excluding accidental situations, which implies that the structure is resistant to deformation.

2.1. *From load history to load spectrum*

Sensors installed on the wheel collect data that provide a load history at regular intervals, $(t_i; F(t_i))$. Mechanical fatigue is the progressive deterioration of a material in response to repeated loads. The first step is then to identify and counting the load cycles present in the history (see [9] Chapter 3). Rainflow counting is the most commonly used for this purpose. This provides an accurate representation of repetitive load cycles $(k_i, \Delta F_i)$, where k_i is the occurrence of the load amplitude ΔF_i. Load spectrum model can be fitted to the cycle counting [10, 14, 15]. The load spectrum represents the distribution of load amplitudes over the load history. The use of a load spectrum model facilitates storage and, more importantly, allows the data to be interpolated to obtain load values for periods or cycle amplitudes not directly observed in the raw data. It is also a probabilistic model that can be used to simulate loads [16, 17].

2.2. *Cumulative Damage*

Fatigue life is then evaluated by counting the number of load cycles to failure. If the cycles have constant amplitudes, then it is common to use a Wöhler curve that relates the number of cycles to material failure (N) to the amplitude of the stress cycles (S). There are numerous fatigue models for this purpose, the most popular being the Basquin model (see [18] Chapter 7). In practice, however, cycle amplitudes are variable, in which case a cumulative damage calculation is used [3, 9, 19]. From a cycle counting $(k_i, \Delta F_i)$, we then proceed as follows:

- Wöhler curve modeling: the Basquin model is used to fit the Wöhler curve, relating N_i, the number of cycles required to cause part failure, to ΔF_i, the amplitude of the load cycle:

$$N_i \cdot (\Delta F_i)^b = C, \tag{1}$$

  where the Basquin exponent b represents the sensitivity of the material to the load amplitude: when subjected to alternating loads, a material with a Basquin coefficient equal to 8 (e.g. shot-peened zone) will have a longer life than a material with a Basquin coefficient equal to 4 (e.g. weld zone) [20]. Since the structure has a linear mechanical response, the constant C depends on the material and vehicle mass. In addition, the Basquin model generally takes into account the average load over a

complete cycle. In our case, the vehicle mass varies very slightly, so Rainflow cycles averages do not have a significant impact on marginal load induced damage (see [12]). Then, here we ignore the effect of cycle mean or cycle load ratio (see [21, 22, 23] for alternatives).

- Cumulative damage calculation: the damage accumulation hypothesis of Palmgren-Miner [24] is that each cycle with amplitude ΔF_i, uses a fraction $1/N_i$, of the total life:

$$D_i = \frac{1}{N_i} = \frac{1}{C} \cdot (\Delta F_i)^b.$$

Thus, cumulative damage is given by:

$$D(b, F) = \sum_i k_i \cdot D_i = \frac{1}{C} \sum_i k_i \cdot (\Delta F_i)^b. \tag{2}$$

The lifetime can then be estimated by assuming that the failure occurs when the damage reaches the value of 1.

Note that the Palmgren-Miner rule assumes that each stress cycle is independent of the other cycles. In other words, the damage caused by one cycle does not affect the response to subsequent cycles.

In order to make the fatigue analysis independent of the vehicle type, we will use the pseudo-damage (see [3] Chapter 3).

Pseudo-damage is a factorization of the damage:

$$d(b, F) = \sum_i k_i \cdot (\Delta F_i)^b. \tag{3}$$

It should be noted that while the linear mechanical response assumptions may be restrictive (and prohibitive for a vibration fatigue study), it is conservative, which means that the pseudo-damages obtained will always be overestimated (see [12]).

If the pseudo-damage does not make it possible to calculate the lifetime, its calculation does not require knowledge of the material constant C and nevertheless makes it possible to compare the relative impact of different load cycles. This makes it a practical choice for our future analyses.

Often pseudo-damage is difficult to interpret because the unit is not physically comprehensible. It is therefore preferable to use the "damage magnitude" $DM(F)$:

$$DM(b, F) = \left(\sum_i k_i \cdot (\Delta F_i)^b \right)^{\frac{1}{b}}. \tag{4}$$

The damage magnitude is proportional to an "equivalent stress" (see [3, 25]). It corresponds to the amplitude of a single load cycle under given one-dimensional loading conditions that would produce the same pseudo-damage as in the local context. This equivalent stress is a "metric of severity" that can be used to compare different load histories or to characterize a population of drivers in terms of the cumulative damage done to the part.

It is also important to note that values of b equal to 4 and 8 introduce a high degree of variability in the values of the normalized pseudo-damages, which can affect data interpretation and statistical analyses. Using the damage magnitude, the b-th root of the pseudo-damage, reduces this difficulty.

Note that a load history F depends on the length of the road $L(F)$. Therefore, we will use "damage intensity", which is the b-th root of the pseudo-damage per kilometer:

$$DI(b, F) = \left(\frac{1}{L(F)} \sum_i k_i \cdot (\Delta F_i)^b \right)^{\frac{1}{b}}. \tag{5}$$

2.3. Stress-Strength Interference

The main formalism implemented in the industry for the fatigue design of automotive structures is the probabilistic "Stress-Strength Interference". This method was first introduced by [26] and is widely used in the design of electrical systems and structures subjected to mechanical loads. Stress is a random variable that represents the potential amount of damage caused to a part by load applied to the structure (e.g. forces, moments, vibrations, etc.) [9]. Stress can thus be used to compare two loads or to characterize a population of drivers. Fatigue design then involves determining the strength of parts, controlling the risk of the least strength part being installed on a vehicle used by the most severe driver. Strength therefore quantifies the highest level of stress that can be applied for a given risk of failure. In the automobile industry, the application of the stress-strength method for fatigue design involves the following steps [27, 11, 28, 29]:

1. Identify the Stress distribution: this stress distribution must be representative of the iterations and load levels experienced by the structure. Therefore, based on the loads measured on the wheel, the damage magnitudes are an excellent indicator of the stress that a part is subjected to;
2. Define the reliability requirements: this requirement can be defined for a part or the structure and is generally expressed in terms of failure risk;
3. Calculate the Strength: for a given material, the strength is determined by considering the stress distribution and the reliability requirement; the risk of the expected stresses exceeding the strength must be less than the defined reliability requirement. To ensure high reliability and durability, the fatigue strength of a part is set so that the risk of failure is very low (e.g. less than 0.01%) under normal vehicle operating conditions, even when driven by a "severe" driver in term of fatigue. This "target customer" is often defined based on the 99th percentile of the stress distribution (see Fig. 1).

Local stresses result from the forces or combinations of forces applied to the vehicle by its tires as it as it travels along the roadway. Therefore, it is necessary to determine the external loads applied to the chassis during vehicle use [2, 4]. These external loads have multiple inputs (longitudinal, lateral and vertical and are applied to all four wheels). In the multi-input case, the stress-resistance method is then applied separately to each of the loads, resulting in a set of specifications for a target conductor, even under different operating

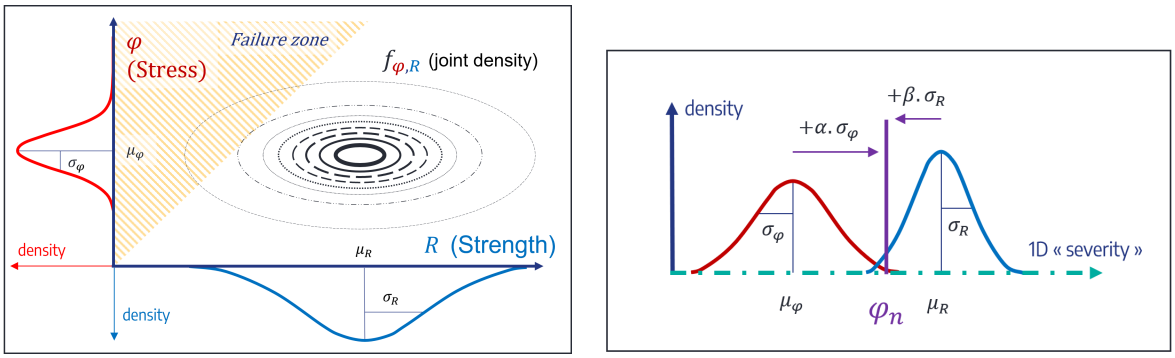


Fig. 1: 1D Stress-Strength interference. Strength is distributed over units/issues of a component/part/system. Load distribution must be representative of the iterations and load levels experienced by the structure. The objective is to determine the fatigue strength of parts, while controlling the risk of the least strength part being installed on a vehicle used by the most severe customer: φ_n is equal to the 99th percentile of the Stress distribution. The risk of failure is set to 0.01%.

conditions [29]. Selecting a limited number of "relevant" loads can reduce the time required for the design phase. Therefore, the first step is to pre-select the relevant loads to simplify the analysis while trying to maintain a high level of completeness.

3. A Multi-dimensional Fatigue Characterization from Multi-input Load Measurements

These external loads applied to the chassis during vehicle use result from various factors, such as road conditions (bumps, potholes, etc.), vehicle maneuvers (cornering, braking and acceleration), and aggressive driving (sudden acceleration and braking, frequent overtaking, high speed, etc.) (see [3] Chapter 1). Therefore, the most exhaustive possible knowledge of the loads applied to the vehicle, during normal use, and a thorough understanding of the stresses generated in the chassis parts, are essential from the earliest stages of the design process [11, 28, 3].

3.1. Reference loads

Load histories are given in the three directions, longitudinal, lateral and vertical (X, Y, Z), with different phase angles for the left and right wheels of the front and rear axles of the vehicle as shown in Fig. 2.

For the three axes $I \in \{X, Y, Z\}$, the two axles $J \in \{\text{Front (f), Rear (r)}\}$ and the phase angles $\alpha \in \{0°, 45°, 90°, 135°\}$, we define the loads:

$$FI\_J\alpha = \cos(\alpha) \cdot F_{I,J,l} + \sin(\alpha) \cdot F_{I,J,r}\,. \tag{6}$$

These different load cases will cause chassis deformation (see [1]). For instance:

$$FX\_f0 = F_{X,f,l},$$

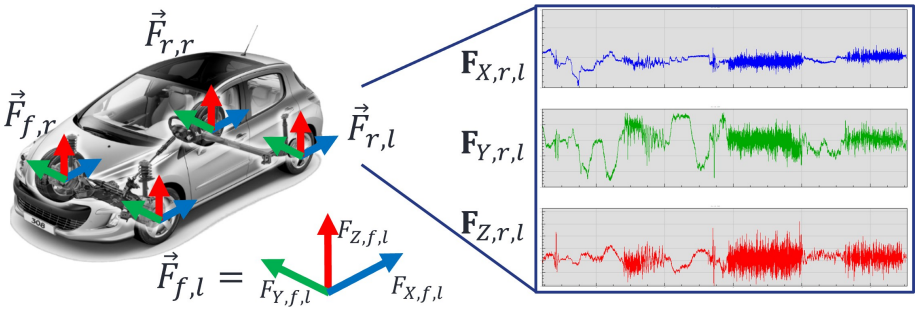


Fig. 2: Multi-axial four-wheel loads and load histories

represents the longitudinal load on the left wheel of the front axle and induces bending around the Z-axis of the left front suspension parts;

$$FY\_f90 = F_{Y,f,r},$$

represents the lateral load on the right wheel of the front axle and induces shearing along the Y-axis of the right front suspension parts.

We also use linear combinations of two of the measured loads to generate new loads resulting from particular situations:

$$FZ\_r45 = \cos(45°) \cdot F_{Z,r,l} + \sin(45°) \cdot F_{Z,r,r},$$

represents the load on the rear axle when the vehicle passes over a bump and induces bending under in-phase loads;

$$FZ\_r135 = \cos(135°) \cdot F_{Z,r,l} + \sin(135°) \cdot F_{Z,r,r},$$

represents the load on the rear axle as the vehicle passes over a pothole, causing torsion under out-of-phase loads. In addition,

$$FXY\_fl45 = \cos(45°) \cdot F_{X,f,l} + \sin(45°) \cdot F_{Y,f,l},$$

represents coupled deformation of left-hand front suspension arm.

In this way, we enrich the family of measured loads on left and right wheels ($\alpha = 0°, 90°$), with linear combinations ($\alpha = 45°, 135°$).

For durability analysis, this enriched load family will allow us to:

- model more driving conditions: evaluate the mechanical response of the chassis under more varied conditions such as cornering, acceleration, braking, potholes, bumps etc. This provides a better understanding of the mechanical response of the chassis to specific roads and driving situations. It also increases confidence in the analysis performed during validation testing;

- optimize design: by analyzing the response of chassis parts to specific loads, it is possible to identify zones that may be susceptible to repeated torsional or bending stresses. These observations can then

be used to guide the design by reinforcing these zones or opting for stronger materials, especially at the level of parts such as suspensions, dampers or springs, to better meet chassis durability and performance requirements.

3.2. Local Contexts

It's important to note that different types of load, such as cornering or driving over a bump, will cause damage to specific zones of the part. Therefore, we will focus on zones that are potentially susceptible to damage from reference loads. These are not only zones that are critical to the operational safety of the part, but also zones where weight reduction or geometry changes could be considered by the design office. For this reason, we will refer to the "local context" as the given Basquin coefficient b and load history F. For each pair (b, F), we compute a damage magnitude $DM(b, F)$, see Fig. 3.

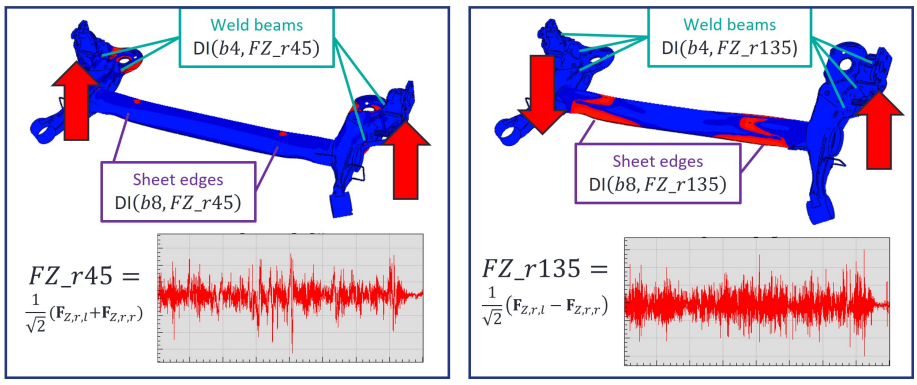


Fig. 3: Loads and local contexts

Consequently, we have a set of reference loads representing different contexts, see Table 1. During a field measurement campaign, several itineraries are studied, each of which is used by several drivers. As a result, the research department has a data set of damage intensities calculated for each context, itinerary and driver. This allows us to analyze damage caused by multi-input loads without having to refer to a specific chassis part to which those loads are applied.

Table 1: The 26 local contexts for the suspension, for different solicitation directions and application points (front, rear, left, right)

Load case	Local context		Events	Zones	Basquin
Front longitudinal	b4_FX_f0 (left) b4_FX_f90 (right) b4_FX_f45 (in-phase) b8_FX_f45	4	Maneuver: braking	wheelpost, triangle	4, 8
Front lateral	b4_FY_f0 (left) b4_FY_f90 b4_FY_f45	3	Maneuver: cornering	gusset, triangle joints	4
Front vertical	b4_FZ_f0, b4_FZ_f90 b4_FZ_f45 (in-phase) b4_FZ_f135 (out-of-phase) b8_FZ_f45, b8_FZ_f135	6	Obstacles: speed bump, ditch, pothole	Cross member edges and joints, wheel posts	4, 8
Rear solicitations	b4_FX_r0, b4_FX_r90 b4_FX_r45, b4_FY_r45 b4_FZ_r0, b4_FZ_r90 b4_FZ_r45, b4_FZ_r135 b8_FZ_r45, b8_FZ_r135	10	Obstacles: speed bump, ditch, pothole	Rear suspension parts	4, 8
Triangle specific	b8_FXY_fl45 b8_FXY_fr45	2	Coupled maneuvers	Front triangle edge	8

4. Field Measurement Campaigns

Measurement campaigns with sensor-equipped vehicles provide authentic load data. However, we need to consider the many factors that explain load variability, focusing on those that are not intrinsically related to the structure of the future vehicle, such as its use [3]. While the usage of a vehicle is simply to get from point A to point B, its uses are very diverse and depend on the driving environment (road type, traffic, weather, etc.), the payload being transported, and the driver's driving style. Therefore, the choice of an itinerary for field testing must accurately reflect the actual conditions to which the vehicle may be exposed. This means taking into account the driving habits of future users, with particular emphasis on the type of roads used.

4.0.1. Road types

Generally, a distinction is made between:

- Highways: highways allow driving at high speeds and are characterized by a generally straight alignment, the presence of multiple lanes offering overtaking opportunities, and particularly high speed limits. They have very low occurrences of pavement imperfections such as potholes or bumps.

- Urban areas: urban roads require frequent maneuvers such as sharp turns, frequent stops and starts, and reduced speeds due to speed limits or general traffic density.

- Country roads: these roads are characterized by an higher occurrences of pavement defects. Traffic density is lower than in urban areas. Speed limits are higher than in urban areas, but still lower than on highways.

- Damaged road: this is a special case of a country road with numerous defects in the pavement.

Fig. 4: Road types in the test campaign

In the context of fatigue design, the focus is often on extreme load values, as these can be the most critical for the service life of a component or structure. It's worth noting that a vehicle used primarily in an urban area will not generate the same extreme multi-axial loads as a vehicle used primarily on country roads. For measurement campaigns, it may be useful to use itineraries with portions of all four road types, but at specific percentages.

4.0.2. Driving styles

Loads vary according to driver behavior. For example, loads differ between aggressive driving, characterized by hard acceleration and braking and frequent overtaking, and more relaxed driving. However, it is difficult to determine the driving style *a priori*. We therefore perform an *a posteriori* evaluation, considering the speed and also the correlations between the damage intensities.

4.0.3. Traffic and Weather

Since traffic density and weather conditions cannot be determined *a priori*, these two factors are not studied here. Therefore, any unpredictable event on a route (e.g., traffic jam, slowdown due to heavy rain, etc.) is considered and treated as sampling noise.

4.0.4. Payload

Variations in vehicle speed affect on the distribution of loads measured at the wheels due to various mechanical factors such as inertia forces, load transfer, etc. In this study, the payload will be identical for all test vehicles.

4.0.5. Specific vehicle characteristics

Specific vehicle characteristics, such as weight, engine power, type of suspension and shock absorbers, etc., influence the loads to which the chassis parts are subjected. We want to design the chassis of future vehicles whose specific characteristics we don't know *a priori*. That's why we use pseudo-damage, which doesn't take this information into account.

5. Analysis of data from a US measurement campaign

The measurement campaign took place in the US state of Michigan. A total of 44 drivers were asked to drive the same SUV, one per day, with the same payload. The reference itinerary is 240 km long and includes secondary roads (a rural road B1 and a suburban road B2), an urban road (C), a damaged road (D) and highways (H1 and H2).

The procedure for collecting road load data at Stellantis is described in detail in [30]. Prior to this measurement campaign, the same vehicle was tested on test tracks with accelerometers and wheel force transducers. In service, the vehicle was not equipped with the load cells. The in-service wheel axle loads are predicted from the acceleration channels using transfer functions learned from the proving ground measurements.

5.1. Road Types Analysis

Our data set consists of 6x26 variables corresponding to damage intensities (Eq. 5) calculated for the 26 different local contexts selected (Eq. 1) and the six road types. Each of these variables represents a unique combination of factors:

- a local context;
  - a Basquin coefficient (b=4 or 8), which defines the material fatigue behavior of a zone of a part;
  - an axis of the load measured on the wheel (for example, the longitudinal axis of the left front wheel), which represents the source of local stress in the zone.
- a type of road (e.g. city or highway), which characterizes the driving environment.

We also have 44 individuals corresponding to the drivers who use the vehicle on these specific road segments.

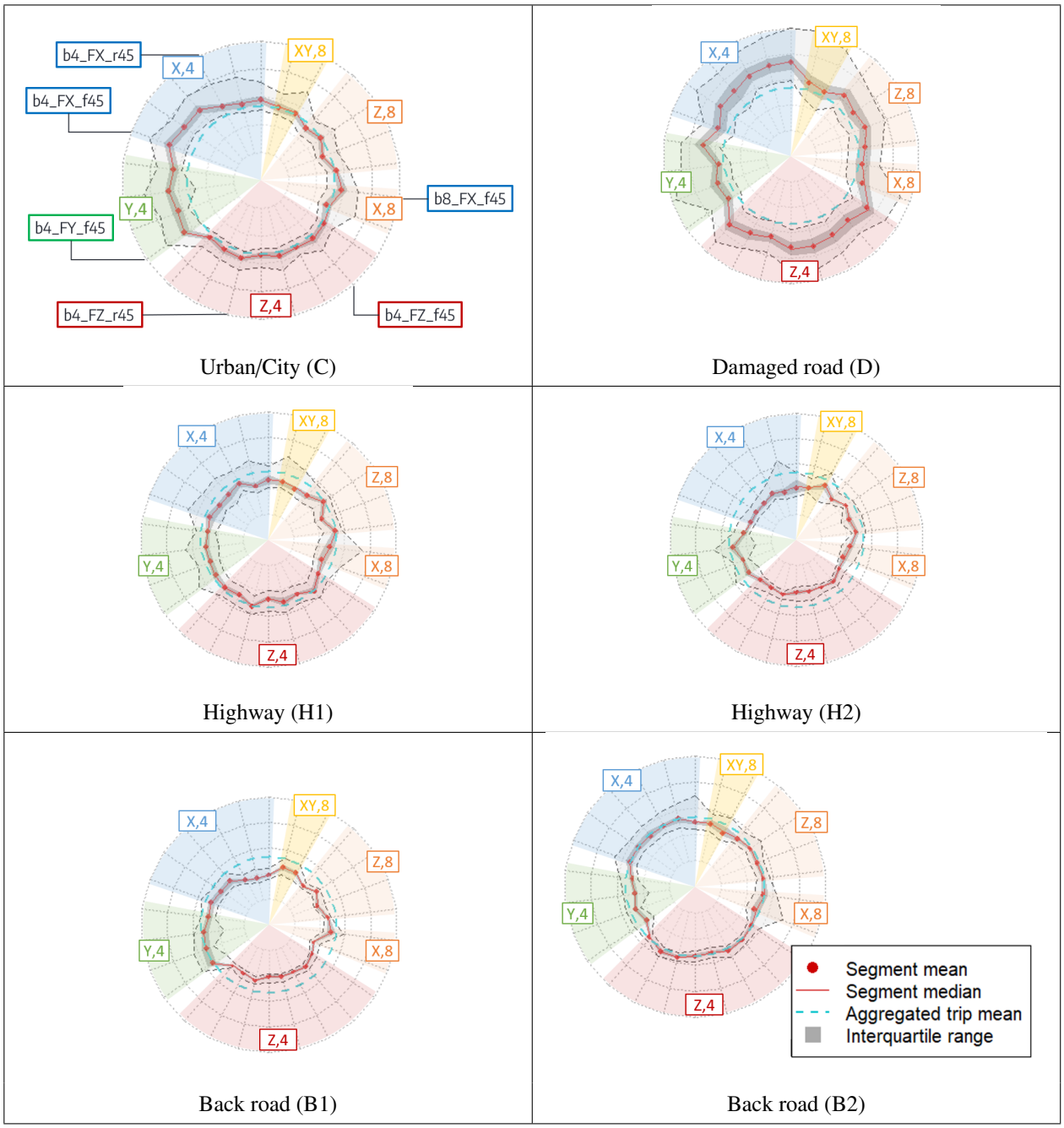


Fig. 5: Normalized damage intensity radar plot for each of the six road types. The light blue dashed circle is the average damage intensity for all drivers across all road segments. It is the basis for normalizing the damage intensity for each of the six road types. The red dots are the average of each damage intensity. The solid red line is their median. The dark gray area with blue borders is the interquartile range. The dotted limits of the light gray zones are the extremes of the road types.

5.1.1. Multivariate analysis

Thus, we have a multivariate data set containing 44 points in the space $\mathbb{R}^{26\times6}$, where each point represents a driver and each dimension corresponds to a damage intensity of a specific local context. We are then faced with a number of issues:

1. Variable correlation analysis: if two damage intensities are correlated, then the load axes have a mutual influence on the damage levels;
2. Drivers similarity analysis: if two drivers are similar, i.e., show proximity in the variables space, this indicates that they generate similar levels of damage intensities when using the same type of road in comparable local contexts;
3. Road types similarity analysis: two road types are similar, if two drivers are similar on one road segment, they are also similar on the other road segment.

But let's start with a descriptive analysis of the data. Fig. 5 shows radar plots of the 26 standardized intensities for each of the six road types. It can be seen that the C and D segments generally have higher damage intensities than the other ones, and also have large variations.

5.1.2. Correlations Structure Analysis

Our goal here is to characterize the different road types based on the observed interactions between damage intensities. The objective is to identify the relevant segments to reproduce fatigue critical load configurations. We have a data set with a large number of variables (i.e., 6×26 measures of damage intensity). Our first goal is to understand how these variables are related to each other in order to facilitate our analyses. Exploratory factor analysis (EFA) is used to analyze the correlation structure of a large number of variables using a reduced number of underlying factors independently of sampling noise [31]. Factors corresponding to the main latent sources of variability in the data. Fig. 6 shows the model with the optimal number of factors to explain the variability in damage intensities for each road type.

For C, H1, B1, and B2, we see a similar two-factor structure: the first factor explains the correlations between longitudinal and vertical intensities, corresponding to driving events such as braking or acceleration. The second factor explains the relationship between longitudinal and lateral intensities, corresponding to driving events such as cornering. The correlation between these two factors varies according to the road type. For segment H2, we also obtain a two-factor structure, one of which explains only the correlation between the load intensities of the front and rear axles, which could be explained by driving over potholes at high speed. For segment D, we again obtain a two-factor structure, one of which is related to the correlation between lateral and longitudinal intensities on the front axle, probably due to the presence of many sharp bends on this segment.

It is interesting to note that here we find results observed by [32] for a rear axle prototype subjected to a load history recorded during tests on a proving ground.

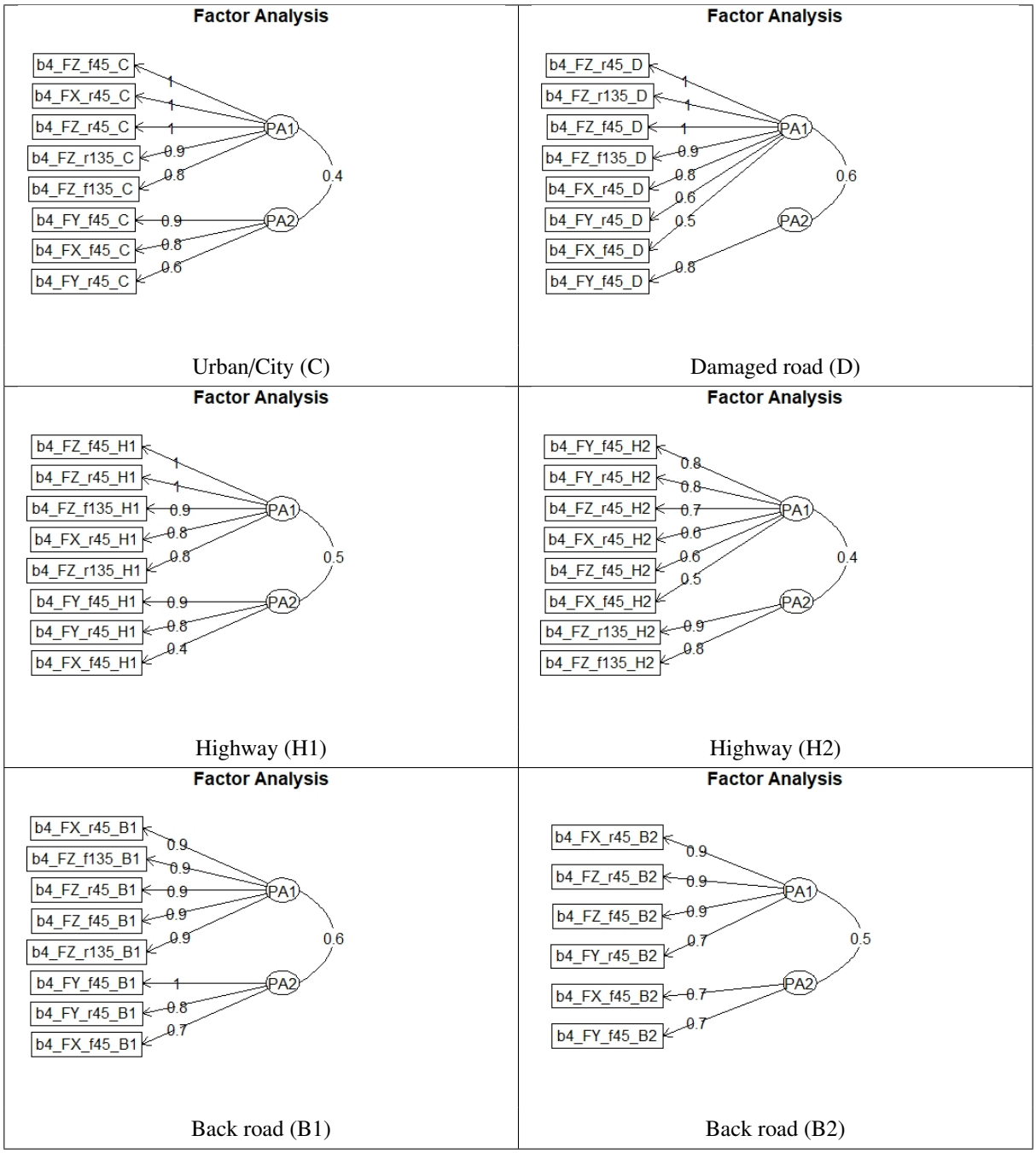


Fig. 6: Exploratory Factorial Analysis (EFA) identifies the main sources of variability in damage intensities for each road type. The weights on the arrows from factors to variables correspond to the percentage of variance in the variable explained by the factor. Only weights greater than 0.3 are displayed. The weights on the arrows between two factors correspond to the correlation between the two factors.

5.1.3. Driving profils analysis

Our goal is to identify which driving profiles are more or less damaging to chassis parts, i.e., to highlight a typology of pairs (driver × road type) in terms of damage. To do this, we used a clustering technique. Unfortunately, with only 44 drivers for 6×26 variables, we run the risk of facing the "curse of dimension" [33]. This situation, characterized by a large number of variables and a small number of observations, could lead to a problem of overlearning, where our analyses could be excessively influenced by the specific characteristics of the sample. The data collected may reflect fortuitous driving situations, such as a traffic jam when the vehicle is in use. This "unusual" variability in environmental conditions and vehicle use is considered to be "sampling noise", as it can distort the detection of truly significant trends. To overcome this difficulty and to obtain results that are robust to the random fluctuations induced by the sample, we can use Principal Component Analysis (PCA) to reduce the dimensionality of the data (see [31] Chapter 9).

Principal Component Analysis (PCA). The general principle of PCA is as follows: let $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_p)$, be p variables observed on n individuals. The data set is a matrix $Y[n \times p]$ with n rows and p columns. Without loss of generality, we assume that the variables are centered. PCA consists in creating p new features $\mathbf{PC} = (\mathbf{PC}_1, \ldots, \mathbf{PC}_p)$, called "principal components". These principal components are obtained by searching for linear combinations of the initial variables $\mathbf{Y}$ that are uncorrelated with each other and that maximise the total variance of the data at each stage. In practice, these components are computed by diagonalising the variance-covariance matrix of $Y[n \times p]$. The eigenvalues $(\lambda_1 > \lambda_2 > \ldots > \lambda_p)$ correspond to the proportion of variance explained by each component $(\mathbf{PC}_i)$. The result is the factorial model:

$$Y[n \times p] = PC[n \times p]\, A[p \times p] \tag{7}$$

where the $PC[n \times p]$ matrix corresponds to the coordinates of the n individuals in the space of the p principal components; the matrix $A[p \times p]$ is a transition matrix from one space to another.

PCA then allows us to do two important things:

- **Dimension reduction:** it is possible to reduce the dimension of the data, and thus tackle the "curse of the dimension", by approximating the original matrix $Y[n \times p]$ with the matrix $\tilde{Y}[n \times p]$:

$$\tilde{Y}[n \times p] = PC[n \times k]\, A[k \times p]. \tag{8}$$

  The number of principal components retained k, is obtained by selecting the first k principal components that explain 95% of the total variance. It is reasonable to assume that the remaining components are associated with sampling noise and can therefore be eliminated. Note that the matrix $\tilde{Y}[n \times p]$, is the best least squares approximation of $Y[n \times p]$ by a matrix of rank k (see the Eckart-Young theorem [34]). This allows us to reduce the size of the data (if $k \ll p$) while preserving the data structure as much as possible.

- **Data visualization:** PCA allows us to visualize $Y[n \times p]$, a scatter plot of n points in $\mathbb{R}^p$, by projecting it onto planes based on two principal components, such as the first plane $PC[n \times 2]$.

Here, damage intensities are obtained for each of the six road types. In the case of an *a priori* group structure of the variables, Multiple Factor Analysis (MFA) is used [35].

Multiple Factor Analysis (MFA). MFA is used as follows:

1. Partial PCA is performed for each road type;
2. For each road type, the damage intensities are then normalized by the largest eigenvalue of the partial PCA;
3. a global PCA is then performed on the overall normalized damage intensities.

As seen in the Fig. 5, segments C and D have high damage intensity values and may artificially contribute more than other road types during principal components construction. Step 2 aims to mitigate this unwanted effect in the analysis.

Fig. 7 shows the eigenvalue scree plot of the global PCA. It can be seen that the first 23 principal components explain 95% of the total variance of the 6×26 variables. The remaining components, representing the remaining 5% of the variance, can be considered as sampling noise.

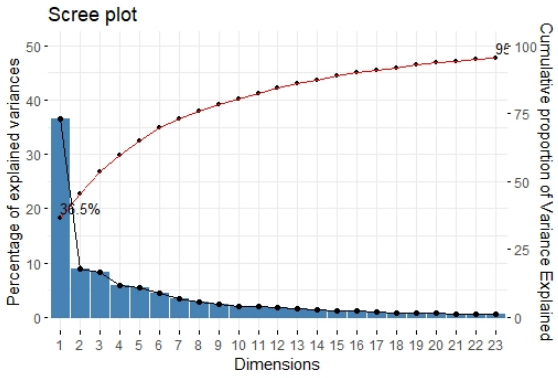
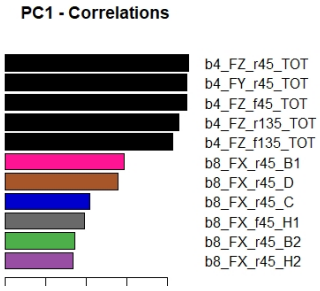
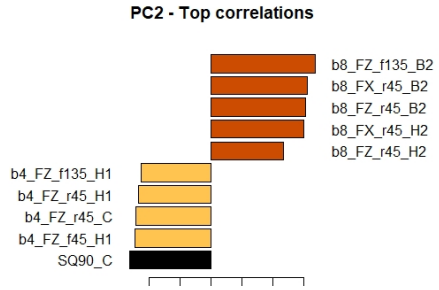


Fig. 7: MFA Scree Plot

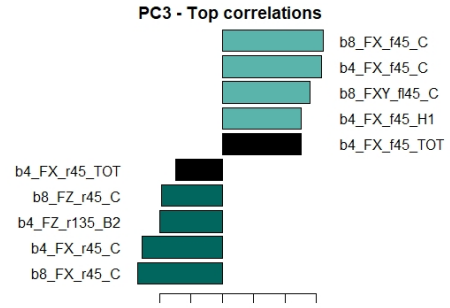
Speed is an aggravating factor in most driving events. It increases the acceleration of the wheel axles and therefore the load on the vehicle when maneuvering and negotiating obstacles. To make the MFA results easier to interpret, we have added the speeds (i.e., average value and S90, the 90% percentile) for each road type, as well as the damage intensity for the entire itinerary made up of the six road types (i.e., features marked with TOT). These intensities allow for a driving style that is unique to the driver. Additional features are not used in the principal components computation and are shown in black in the various plots. Fig. 8 shows the most significant correlations between damage intensities and the first three principal components.



(a) PC1: all damage intensities are positively correlated with the first principal component



(b) The PC2 principal component differentiates between riding with high damage intensity for a zone with a Basquin coefficient of 8 on B2 or H2 roads, and those with high damage intensity for a zone with a Basquin coefficient of 4 on C or H1 roads. For C and H1, damages are correlated with high velocities S90



(c) The PC3 principal component contrasts rides with high damage intensities measured on the front axle with rides with high damage intensities measured on the rear axle. Note that these damage intensities on the rear axle are correlated with vertical damage on the front axle.

Fig. 8: Correlations with the first three principal components of the AFM

First principal component (PC1): it can be seen that all damage intensities are positively correlated with the first principal component. This component, which is a linear combination of the damage intensities, can be considered as a ”severity score” to compare drivings in terms of damage caused to a part: a high value for this first principal component indicates driving that causes significant damage in all local contexts, regardless of road type;

Second principal component PC2): this principal component allows us to distinguish between driving with high damage intensity for a zone with a Basquin coefficient of 8 (e.g. sheet metal edges) when the vehicle is on B2 or H2 roads, and those with high damage intensity for a zone with a Basquin coefficient of 4 (e.g. welds) when the vehicle is on C or H1 roads. For C and H1 roads, damage intensities are correlated with high velocities (SQ90, the 90% Speed Percentile). Recall that the Basquin coefficient is a measure that is physically related to the way a material responds to fatigue:

- a zone with a Basquin exponent between 3 and 5 (e.g., a weld bead) will always show initial microcracks due to material shrinkage after fabrication (see [36]). Consequently, even loads slightly above the

fatigue limit will always result in significant damage;

- this is not the case for a zone with Basquin exponent between 6 and 20 (e.g., a sheet metal edge), where much higher loads, compared to the fatigue limit, are required to cause significant damage (see [20]).

Sudden acceleration and deceleration cause significant load transfers, resulting in increased vertical and longitudinal loads. This results in high stresses on chassis parts. Hard accelerating and braking are characteristic of aggressive driving. The PC2 component differentiates between damages caused by aggressive behavior when acceleration and braking, and damages caused by a smoother driving style.

Third principal component (PC3): this principal component allows us to distinguish between driving with significant damage measured on the front axle and driving with significant damage measured on the rear axle. Note that these damage intensities on the rear axle are negatively correlated with vertical damage intensities on the front axle. These situations are once again characterized by aggressive behavior, but this time when cornering and overtaking (see [37]):

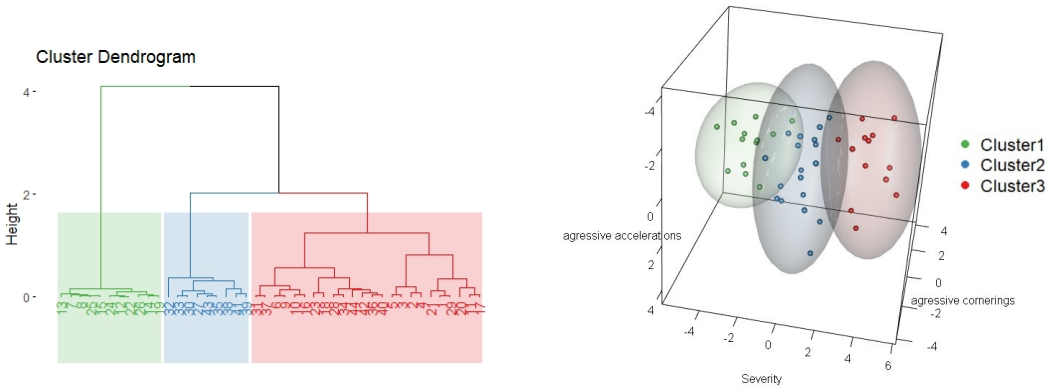
- if the driver is too fast when entering the corner, the load on the front axle increases significantly, which can lead to understeer;
- if the driver suddenly depresses the accelerator when the vehicle is already in the corner, the load on the rear axle increases significantly, which can lead to oversteer.

Again, oversteer and understeer are characteristic of aggressive driving, especially when cornering and overtaking. Component PC3 differentiates damage caused by poor speed management when cornering and overtaking.

Clustering. Two drivers are considered similar if they show proximity in the fatigue damage space. Clustering involves constructing a partition of the set of drivers into groups, or clusters, such that drivers within the same cluster show similarity, while drivers in different clusters do not show similarity to each other. By characterizing each cluster, we can discover the underlying factors associated with it. In our fatigue design context, this makes it possible to identify the driving behaviors that, for certain road types, will cause damage in certain local contexts.

The 44 rows of the data set $Y[44\times 6\cdot 26]$ correspond to the drivers and the columns correspond to the 26 damage intensities for each of the 6 road types. To tackle the "curse of dimension" and to avoid overlearning, we perform a classification based on $PC[44\times 23]$, the projection of the scatterplot $Y[44\times 6\cdot 26]$ into the space of the first 23 principal components. We then combined two clustering methods. First, we use Hierarchical Ascending Classification (HAC): this allows us to obtain an optimal number of classes using the Ward jump criterion [38]. We then use the HAC solution to initialize the k-means algorithm, which allows us to obtain

more homogeneous classes in terms of damage intensity (see Fig. 9 (a)). By projecting the data cloud onto the first three principal components, we can characterize the three clusters identified (see Fig. 9 (b)). Using PC1, the three clusters can be ranked according to their damage severity, from least severe (Cluster



(a) Clusters with CAH

(b) Clusters with the k-means algorithm

Fig. 9: Fig.(a): HAC was performed on the first 23 principal components ot the MFA. Using Ward jump criterion, three clusters are retained. This solution is then improved using the k-means algorithm. Fig.(b): using PC1, the three clusters can be ranked according to their damage severity, from least severe (Cluster 1) to most severe (Cluster 3).

1) to most severe (Cluster 3). Usually, using the pseudo-damage for a given local context, it is possible to compare drivers in terms of damage, taking the ”target severe customer” to be at the 99th percentile of the distribution. Using PC1, it becomes possible to generalize the notion of damage severity for a set of local contexts. Within each cluster there are drivers who are more or less aggressive when accelerating and braking, or more or less aggressive when cornering or overtaking, especially in urban traffic.

6. Reference Itinerary Analysis

During the useful lifetime of the vehicle, the conditions of use may vary, as may the owners or drivers. Therefore, sizing specifications are defined based on a total distance covered (e.g. 240,000 km) and a certain combination of road types. By combining the different road types, it becomes possible to assess the contribution of fatigue resulting from each driving condition associated with a specific road type, on the total distance covered during the vehicle lifetime. Here we look at the total itinerary created by combining all six road types. It is an illustrative example of the reference itineraries that can be used to represent typical customer usage. The dataset $Y[44 times 26]$ has 44 rows corresponding to the drivers and 26 columns corresponding to the calculated damage magnitudes from the load histories along the entire itinerary for the 26 local contexts (Table 1).

6.1. Correlations Structure Analysis

Using an EFA, we obtain a two-factor model to explain the relationships between the 26 damage values (see Fig. 10). The second factor is strongly correlated with longitudinal and lateral loads on the front axle and can be associated with maneuvers that primarily involve the front steering of the vehicle. The first factor is related to other maneuvers: acceleration/braking, high speed overtaking, etc. The two factors are correlated.

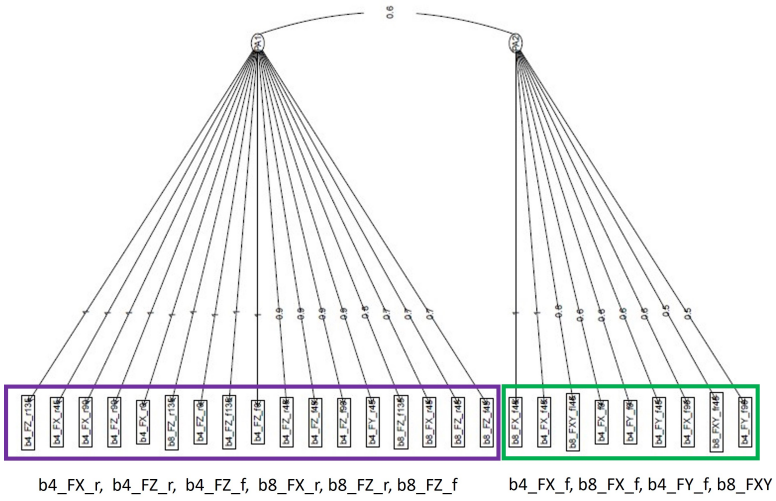


Fig. 10: Total itinerary: Exploratory Factor Analysis of damage magnitudes for the 26 local contexts. The second factor is associated with forward steering maneuvers, while the first factor is associated with various maneuvers (acceleration/braking, high-speed overtaking). The two factors are correlated.

6.2. Damage Magnitude Multi-dimensional Distribution

Just as the load spectrum represents the distribution of load amplitudes in load histories, here we are looking for the distribution of damage magnitudes in load histories. The damage magnitude is proportional to an ”equivalent stress” (see [3] Chapter 3) and then can be used to estimate the fatigue strength of a part zone for the corresponding local context. We then seek to model the multi-dimensional probability density function of the damage magnitudes for the 26 local contexts. First, we perform a PCA to reduce the sampling noise of the data set $Y[44 \times 26]$.

6.3. Principal Components Analysis

The first seven components of the PCA explain more than 95% of the total variance in the data (see Fig. 11).

Fig. 12 shows that the first principal component is positively correlated with all damage magnitudes: it allows to distinguish globally severe drivings. The second component is positively correlated with lateral and longitudinal damage magnitudes: it distinguishes aggressive drivings during cornering.

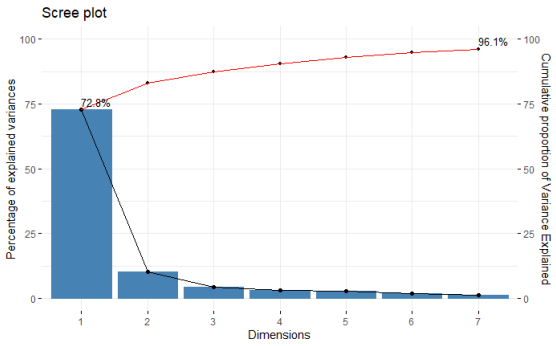


Fig. 11: PCA Scree plot: The first seven PCA components explain over 95% of the total variation in the data, while the remaining components can be assimilated to sampling noise.

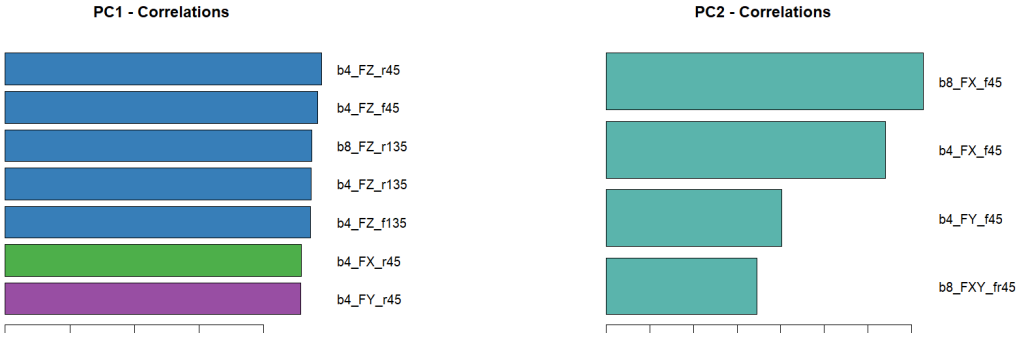


Fig. 12: Correlations with the first two principal components: the first principal component is positively correlated with all damage measures, making it an overall indicator of fatigue severity. The second component is positively correlated with the damage magnitudes of lateral and longitudinal loads: It helps to distinguish aggressive driving when cornering.

Using the data collected on the sample of 44 drivers, we obtain an empirical distribution of damage magnitudes. Our aim is then to estimate the probability density function so that we can extrapolate beyond the observed ranges of values.

6.4. Probabilistic PCA

As shown earlier, the first seven PCA components account for more than 95% of the total data variance (see Fig. 11). We then restrict our analysis to data reconstructed from the first seven principal components:

$$\tilde{Y} = PC[44 \times 7]\, A[7 \times 26]. \tag{9}$$

We propose to fit this data set $PC[44 \times 7]$ with a probabilistic model.

In order to account for the heterogeneity of driving profiles along the entire itinerary, it is common to use mixture models [16, 39, 40]. We have chosen here a multi-dimensional Gaussian mixture in which the mixture components correspond to the driving behaviors.

6.4.1. Multi-dimensional Gaussian Mixture distributions

A mixture of κ multi-dimensional Gaussian distributions (MMG) is defined as follows [41, 42]: for an integer $d \in \mathbb{N}^*$, and $\mathbf{x} \in \mathbb{R}^d$, the probability density function is

$$f_{MMG}(\mathbf{x} \,|\, \boldsymbol{\theta}) = \sum_{k=1}^{\kappa} \alpha_k \, \phi(\mathbf{x} \,|\, \boldsymbol{\mu}_k \,,\, \boldsymbol{\Sigma}_k), \tag{10}$$

where for $k = 1, 2, \ldots, \kappa$:

- $\phi(\mathbf{x} \,|\, \boldsymbol{\mu} \,;\, \boldsymbol{\Sigma})$ is a multi-dimensional Gaussian probability density function with mean $\boldsymbol{\mu}$ and covariances matrix $\boldsymbol{\Sigma}$:

$$\phi(\mathbf{x} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{d}{2}} \, |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \, \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \, \boldsymbol{\Sigma}^{-1} \, (\mathbf{x} - \boldsymbol{\mu})\right); \tag{11}$$

  where $|\boldsymbol{\Sigma}|$ is the determinant of the matrix $\boldsymbol{\Sigma}$.

- $0 < \alpha_k < 1$ is the mixing proportion of the k-th component, with $\sum_{k=1}^{\kappa} \alpha_k = 1$;

Let us denote $\boldsymbol{\theta} = (\kappa, (\boldsymbol{\mu}_k), (\boldsymbol{\Sigma}_k))$, the set of model parameters. In our problem, κ represents the number of different driving styles. For a fixed k, the parameters $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are generally fitted by maximum likelihood, using the Expectation-Maximization (EM) algorithm [43].

6.4.2. Expectation-Maximization (EM) algorithm

Let us denote $\mathbf{z} = (z_1, \ldots, z_n)^\top$ the component labels: for $k = 1, \ldots, \kappa$,

$$z_i = k \text{ if the } i\text{-th driver adopts the } k\text{-th driving style.}$$

Unfortunately, the values of $\mathbf{z}$ are not known *a priori*: it is a latent variable. The EM algorithm is designed to maximise likelihood despite the presence of a latent variable [44]:

1. Initialize the means $(\boldsymbol{\mu}_k^{(0)})$, covariance matrix $(\boldsymbol{\Sigma}_k^{(0)})$ and proportion $(\alpha_k^{(0)})$ of each Gaussian component.
2. For $\ell \geq 1$, repeat until convergence:
   - Expectation-step: Calculate the posterior probabilities:

$$p(z = k \,|\, \mathbf{x}_i \,;\, \boldsymbol{\mu}_k^{(\ell)} \,;\, \boldsymbol{\Sigma}_k^{(\ell)}) = \frac{\alpha_k^{(\ell)} \cdot f(\mathbf{x}_i \,|\, \boldsymbol{\mu}_k^{\ell} \,;\, \boldsymbol{\Sigma}_k^{b})}{\sum_{j=1}^{\kappa} \alpha_j^{(b)} \cdot f(\mathbf{x}_i \,|\, \boldsymbol{\mu}_j^{\ell} \,;\, \boldsymbol{\Sigma}_j^{\ell})} = w_{ik}^{(\ell)}.$$

     The posterior probabilities $(w_{ik}^{(\ell)})$, provide the ability to classify a driver to a cluster (i.e a driving behaviour), according to its driving characteristics $\mathbf{x}_i$. Each driver is assigned to the component with the highest posterior probability.

- Maximization-step: Update the parameters of each component by maximizing the likelihood on the data assigned to that component. In the Gaussian case, this means calculating

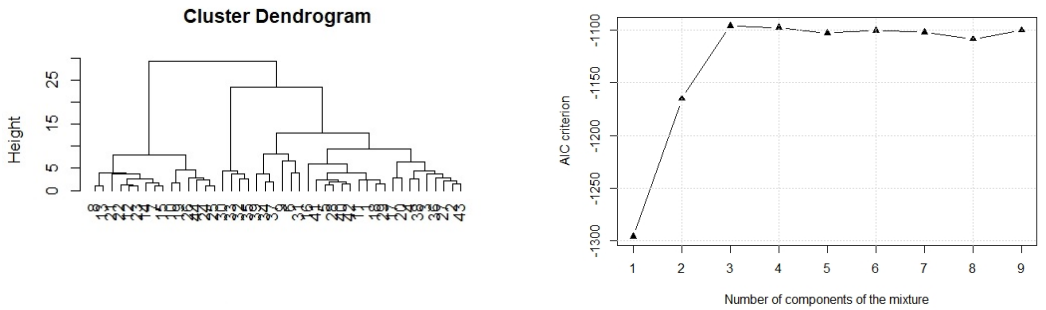
$$\boldsymbol{\mu}_k^{(\ell+1)} = \frac{\sum_{i=1}^{n} w_{ik}^{(\ell)} \mathbf{x}_i}{\sum_j w_{jk}^{(\ell)}} \qquad \boldsymbol{\Sigma}_k^{(\ell+1)} = \frac{\sum_{i=1}^{n} w_{ik}^{(\ell)} \left(\mathbf{x}_i - \boldsymbol{\mu}_k^{(\ell+1)}\right)\left(\mathbf{x}_i - \boldsymbol{\mu}_k^{(\ell+1)}\right)^t}{\sum_j w_{jk}^{(\ell)}} \qquad \alpha_k^{(\ell+1)} = \frac{\sum_{i=1}^{n} w_{ik}^{(\ell)}}{n}$$

3. Convergence criterion: a commonly used indicator for evaluating the convergence of the EM algorithm is the stability of the log-likelihood between two iterations.

It's worth noting that the EM algorithm is an iterative process and its performance depends heavily on the initial conditions. Sub-optimal initialisation can result in slow convergence or a sub-optimal solution [45]. To initialize the EM algorithm, we use a solution obtained from the hierarchical ascending classification dendrogram (see Fig. 13a).

6.4.3. Model Selection

Optimal selection of the number of components to use, is achieved by minimizing a bias-variance criterion that strikes a balance between modeling bias (i.e., the error introduced by overly simplistic modeling) and estimator variance (i.e., the error introduced by overly complex modelling). The Akaike Information Criterion (AIC) is one of the most widely used [46]. The three-components mixture model is thus selected (see Fig. 13b).



(a) Initialization of the EM algorithm using the solutions proposed by CAH.

(b) Model selection: AIC criterion maximized by 3-Component Mixture Model

Fig. 13: Multi-dimensional Gaussian Mixture distributions with the EM algorithm

Fig. 15 show the resulting three-components multi-dimensional Gaussian mixture model. Fig. 14 shows the projections of the mixture of multi-dimensional Gaussian distributions onto the seven principal components.

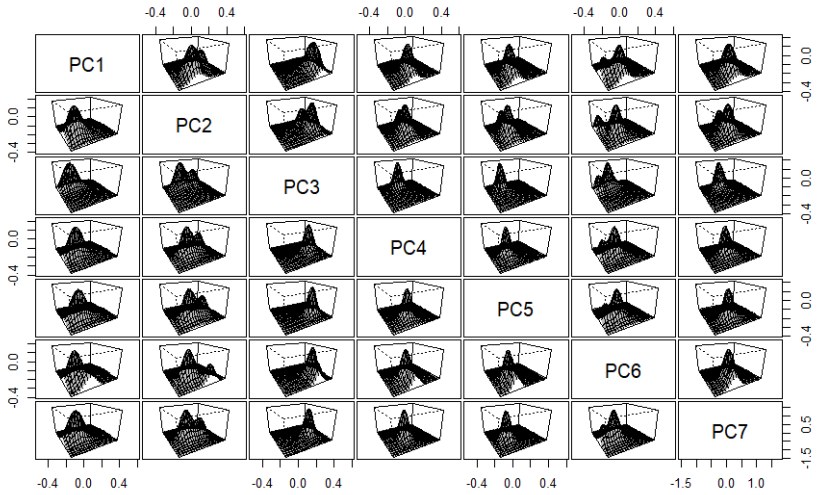


Fig. 14: Visualization of the mixture of three multi-dimensional Gaussian distributions obtained by the EM algorithm on data restricted to the first seven principal components: projections by dimension pairs

Fig. 15 shows the projections of the mixture of multi-dimensional Gaussian distributions onto the first two principal components, together with the marginal distributions: two components of the mixture that have high density values on the large coordinates of PC1, indicating frequent severe behaviours in all local contexts (see Fig. 12). These two components are distinguished by more or less high frequencies on the second principal component (PC2), i.e., more or less frequent aggressive behaviour when cornering. Finally, the last component corresponds to cases of less severe behaviours.

6.5. Damage Magnitudes Distribution

We fitted a multi-dimensional Gaussian distributions mixture model to $\mathbf{PC}[7]$, the vector of the first seven principal components. To obtain the damage magnitudes distribution, we use the reconstruction procedure with the PCA transition matrix (see Eq.8):

$$\mathbf{Y} = \mathbf{PC}[7]\, A[7 \times 26]. \tag{12}$$

Thus, the distribution of $\mathbf{Y}$, the damage magnitudes, is also a mixture of multi-dimensional Gaussian distributions. In this way, we can analyze and reconstruct a damage magnitude distribution for any desired local context or driver profile. Fig. 16 shows two reconstruction examples of the joint distribution and marginal distributions of damage magnitudes of two local contexts. Correlations between the different contexts are clearly visible. For example, when a suspension spring is exposed to both longitudinal and lateral loads during a sharp cornering maneuver, it becomes clear that the risk of generating high stress concentrations is notable, especially for cluster 3 drivers (represented by red triangles). Similarly, simultaneous longitudinal and vertical loads during braking or acceleration can also present similar risks.

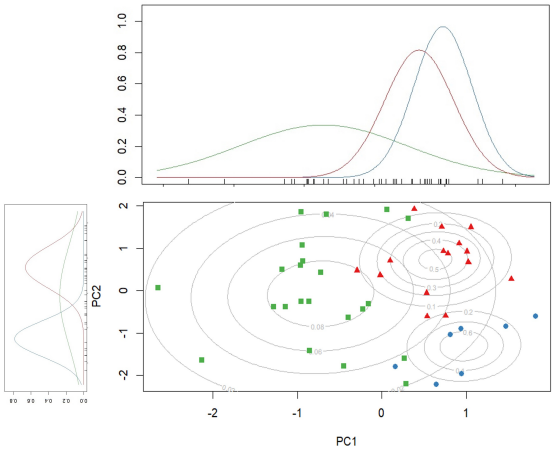
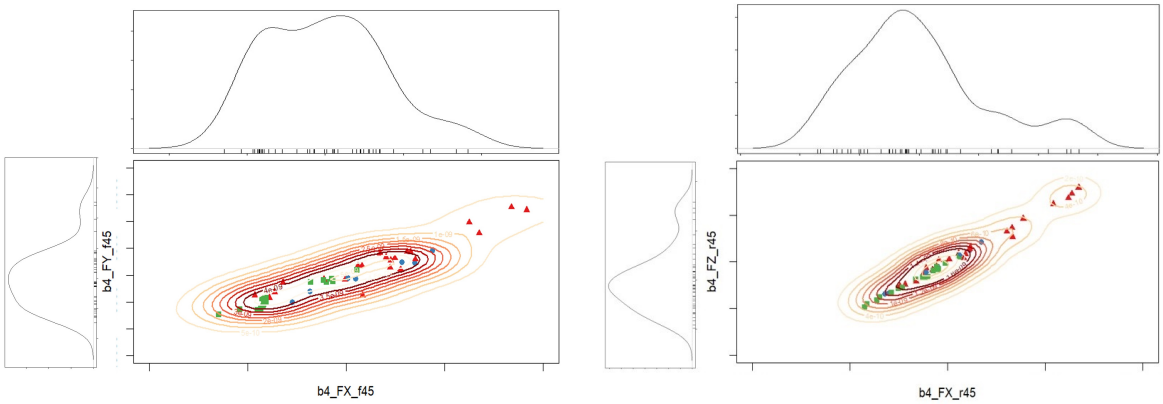


Fig. 15: Projections of the mixture of three Gaussian distributions onto the first two principal components and the marginal distributions. Classification is done with posterior probabilities. Two components of the mixture have high density values on the large PC1 coordinates, indicating severe behavior common to all local contexts. These two components stand out with higher or lower frequencies on the second principal component (PC2), i.e., more frequent aggressive behavior during cornering (cluster 3 with red triangles) and less frequent aggressive behavior (cluster 2 with blue dots). The last component corresponds to less severe behaviors (cluster 1 with green squares).



longitudinal in phase and lateral in phase on the front axle: $b4\_FX\_f45$ and $b4\_FY\_f45$

longitudinal in phase and lateral in phase on the front axle: $b4\_FX\_r45$ and $b4\_FZ\_r45$

Fig. 16: Joint and marginal distributions of local contexts on the front axle. Data have been centered

Note that the magnitude distributions are multi-modal: this can be explained by the different driving styles identified by the mixture model. For example, for the $b4\_FX\_f45$ local context, Fig. 17 shows the 99% quantile of the damage magnitude distributions conditional on the clusters. A cluster of severe behaviors is clearly identified for this local context.

Note that the probabilistic PCA allows us to overcome the challenge of high-dimensional estimation while avoiding over-learning [47].

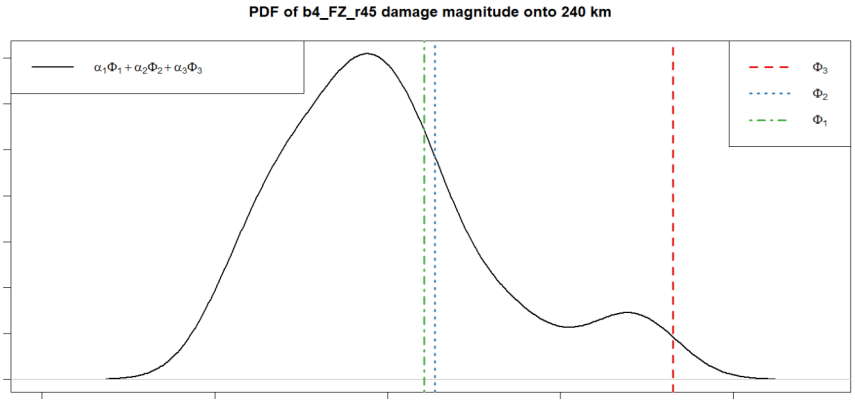


Fig. 17: Local context $b4\_FZ\_r45$ damage magnitude distribution. Vertical lines represent the 99% quantile of the damage magnitude distributions conditional on the clusters (upper bound of a right-sided 95% confidence interval cf. Eq. 13). Values have been suppressed to protect data confidentiality.

It is also important to note that there is a relationship between the load spectra and these damage magnitudes distributions: the load spectrum corresponds to the probability density of the load amplitudes observed in the load histories of a local context [48, 39]. Here, we obtain a multi-dimensional probability density of equivalent damage amplitudes observed in the load histories of all the local contexts studied.

6.6. Defining Validation Test Conditions

This probabilistic model can also be used for the definition of test environments that should have effects similar to those observed in real-world conditions [28, 3]:

- numerical simulation tests on a finite element model of a component: they are used to validate resistance values by exposing the component to high levels of severity;
- rig tests on a prototype of a multi-component part: designed to ensure part reliability and compliance with specifications, these tests are performed at a moderate level of severity;
- proving ground tests with a vehicle: their aim is to validate the performance, safety and handling of the vehicle by exposing it to severe driving conditions.

To implement these various tests, it is necessary to define load scenarios and load levels. The EFA helps to select the relevant loads to define scenarios, and the quantiles of the magnitude distribution are used to define load levels.

Proposition: Conditional to a driving profile, let $DM_p(F, b)$ the p-quantile of the damage magnitude of a local context (F, b). Let $(\bar{y}, s_y)$ be the average and standard deviation of the sample of the damage magnitude in the local context (F, b), for n drivers of the considered driving profile, then the $100(1-\alpha)\%$ confidence

interval of $DM_p(F, b)$ is:

$$\left[\bar{y} - t(1-\alpha/2\,;\, n-1, -\sqrt{n}\,z_p)\frac{s_y}{\sqrt{n}}\,;\, \bar{y} - t(\alpha/2\,;\, n-1, -\sqrt{n}\,z_p)\frac{s_y}{\sqrt{n}}\right], \tag{13}$$

where $t(\beta\,;\, k, \delta)$ is the β-quantile of the Student distribution of order k and non-centrality parameter δ, and z_p is the p-quantile of the gaussian distribution $N(0; 1)$.

Proof: The distribution of $\mathbf{Y}$, the damage magnitudes, is a mixture of multi-dimensional Gaussian distributions (see Eq. 12). For each Gaussian component of the mixture, we can use a pivotal method to construct confidence intervals for the parameters and, in particular, for the p-quantiles (see [49] Appendix E).

6.7. Extrapolation

Durability certification ensures that the vehicle performance and structural integrity will be maintained for 240,000 km, although the tests are limited to a 240 km itinerary. It is therefore necessary to extrapolate the results of the 240 km tests to the full 240,000 km distance. Obviously, the cumulative damage over 240,000 km is not simply 1,000 times the cumulative damage over 240 km: a driver will never take a curve the same way every time, or drive over a pothole the same way. This variability inherent in driving conditions is used as a basis for introducing probabilistic models into the analysis of stress spectra or to account for the statistical variability of rainfall count matrices. Our proposal is to extrapolate from damage magnitude distributions.

Proposition: Let DM be the damage magnitude on the reference itinerary for a local context, and DM_c it's extrapolation to an itinerary c times as long. If the driving conditions remain stationary, then for large c (i.e., $c \geq 30$), DM_c probability density function can be approximated by:

$$f(x) = \frac{(1/b)\, x^{1/b-1}}{\sqrt{2\pi \cdot c\sigma^2}} \exp\left(-\frac{(x^{1/b} - c\mu)^2}{2 \cdot c\sigma^2}\right), \tag{14}$$

where μ and σ are the mean and standard-deviation of DM^b: $\mu = \mathbb{E}\left[DM^b\right]$ and $\sigma = \sigma(DM^b)$.

Proof: If a driver travels the reference route with his vehicle, the number of occurrences of cycles with amplitude ΔF_i has an intrinsic random nature, we denote it as the random variable ν_i (see [50]). Therefore Eq. 2 gives us the pseudo-damage, which is also random:

$$DM^b = \sum_i \nu_i \cdot (\Delta F_i)^b.$$

If the driver repeats this operation c times, D_c, the total pseudo-damage is equal to the sum of the successive n random pseudo-damages $DM_1^b, DM_2^b, \ldots, DM_c^b$:

$$D_c = \sum_{\ell=1}^{c} DM_\ell^b = \sum_{\ell=1}^{c} \sum_i \nu_{\ell i} (\Delta F_i)^b.$$

The stationarity of driving conditions allows us to have $(\nu_{\ell i})$ independent and identically distributed. By invoking the central limit theorem [51], D_c is then distributed according to a Gaussian distribution of mean $c\,\mu$ and standard-deviation $\sqrt{c}\sigma$. This allows us to calculate the density of the damage magnitude $DM_c = (D_c)^{1/b}$ □.

In the previous proposal, we assume that driving conditions, including driving style, do not change from one trip to the next within the same identified cluster. Thus, calculations are performed for each cluster identified above. Fig. 18 shows the distribution of pseudo-damage extrapolation for a local context as well as the 99% percentile for each cluster. It can be seen that with extrapolation, the distributions conditional on the clusters become well separated.

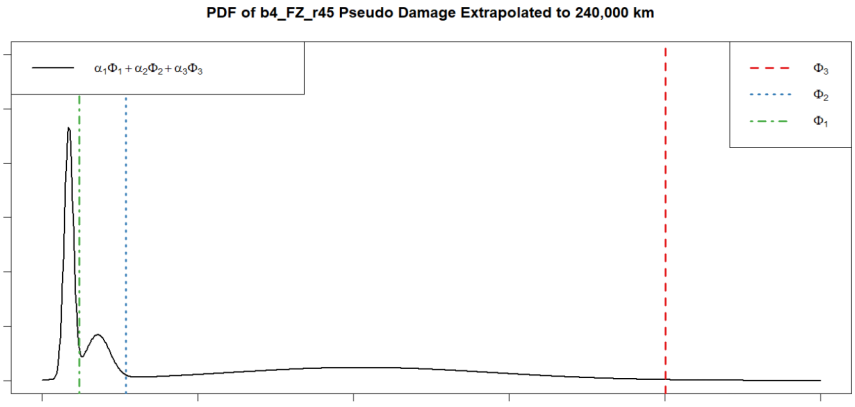


Fig. 18: Local context $b4\_FZ\_r45$ pseudo-damage extrapolation. Vertical lines represent the 99% quantile of the distributions conditional on the clusters. Values have been suppressed to protect data confidentiality.

In addition, it is now possible to simulate pseudo-damage on different road mixes and with driver profile changes.

7. Conclusion

The vehicle chassis plays a critical role in safety by supporting the weight of the vehicle while providing stability, maneuverability and handling. Fatigue analysis of chassis parts is then a critical step in the development of any new vehicle. This analysis considers multiple input loads (i.e., loads acting in the three longitudinal, lateral, and vertical directions, measured at each wheel and the various combinations between the left and right wheels of the front and rear axles). The complexity of the analysis lies in the need to also study the correlations between these loads, as their interactions play a decisive role in the generation of stress concentrations. What's more, loads must be collected on a variety of road types to accurately reflect real-world conditions experienced by customers, resulting in a wide variety of loads to analyze.

In this article, we have presented a methodology that meets the challenge of the complexity of multidimen

-sional load analysis and provides a more complete and in-depth perspective on the development of a personal vehicle chassis.

1. As a first step, we introduce the concept of ”local context”, which takes into account both the load direction and the material behavior through the use of a Basquin coefficient.
2. A field measurement campaign was conducted with a sensor-equipped vehicle: 26 damage magnitudes were computed for six different road types, involving a total of 44 drivers.
3. We then use unsupervised statistical analyses on ”damage intensities”, which correspond b-th root of the pseudo-damage per kilometer. An exploratory factor analysis (EFA) of damage intensities allows us to identify two main sources of variation for each road type. For most road types, the first factor is related to vehicle steering (lateral and longitudinal loads) and the second to acceleration and braking (longitudinal and vertical loads). These two factors are correlated with each other. With 44 drivers and 6x26 variables, we face the risk of overlearning in our statistical analyses. To address this risk, we perform a multiple factor analysis (MFA) to reduce sampling noise. Then, using a clustering method, we obtain clusters of pairs (driver ; road type). The first principal component of the MFA can be seen as a ”severity score” for comparing driving in terms of damage to a part: a high value of this first principal component indicates driving that causes significant damage both longitudinally, laterally, and vertically, regardless of the road type. The second and third components help us identify severe driving profiles in terms of damage associated with poor speed management. This may include acceleration/braking in fast lanes or high-speed cornering in urban areas.
4. During the life of the vehicle, the conditions of use and the owners or drivers may vary. For this reason, durability specifications are based on the total mileage traveled by the vehicle (e.g., 240,000 km) and a specific combination of road types. Therefore we will look at fatigue design using a combination of all six road types over a distance of 240 km. An EFA allows us to identify two main sources of variations of the damage magnitudes. The second factor is associated with forward steering maneuvers, while the first factor is associated with various maneuvers (acceleration/braking, high-speed overtaking). The two factors are correlated. Just as the load spectrum represents the distribution of load amplitudes in load histories, here we’re looking for the distribution of damage amplitudes in load histories. To reduce sampling noise and avoid overlearning, we use probabilistic PCA: to account for the heterogeneity of driving styles, we use a mixture model of multi-dimensional Gaussian distributions on the first principal components. The parameters are estimated by maximum likelihood using the EM algorithm. Next, the probability density function of the extent of damage is constructed using the PCA transition matrix and then extrapolated to cover a distance several times longer.

Significant advantages of this approach include:

- The use of local contexts makes it possible to analyze damage caused by multi-input loads without having to refer to a specific chassis part to which those loads are applied;
- Unsupervised statistical analyses allow us to study the latent structure of correlations between damage intensities and the latent structure of driver behavior:
  - Exploratory Factor Analysis (EFA) enables the design office to select the local context(s) that best describe fatigue on the reference route. This then proves invaluable in guiding the design of chassis parts;
  - Multiple Factor Analysis (MFA) helps to deal with the complexity of the hierarchical structure of multivariate data resulting from the consideration of different road types. First, it solves the problem of overlearning due to the limited number of drivers and the large number of local contexts to be considered. In addition, the first principal component of the MFA is a linear combination of the damage intensities of all local contexts, which allows us to compare the different pairs (driver; road type) in terms of induced damage. In other words, we obtain a multidimensional measure of severity. We then use it to establish profiles of more or less severe drivers.
- The mixing model for damage magnitude probability distribution function, offers the possibility of a more detailed analysis of the damage distribution according to different driving profiles, whether more or less severe, and allow extrapolation and simulations for different road mixes and driver profile changes.

All of this is made possible by field data collection, which adds a more realistic dimension to reference loads for design, simulation and test rig validation.

The main objective of this work is to propose a methodology for multidimensional load analysis. Two axes of perspectives are outlined: on the one hand, the statistical analysis of more complete data sets, including in particular the consideration of the payload effect, and on the other hand, the applications of this methodology in the design process, in particular to define validation test levels for chassis components.

Acknowledgements

This work was carried out within the framework of the partnership between Stellantis and the CNRS with the financial support of the ANRT for the CIFRE contract n° 2020/0182.

Authors contributions

Emilien Baroux: performed Conceptualization, Writing - original draft, Writing review. **Patrick Pamphile:** performed Conceptualization, Writing - original draft, Writing review & Editing, Supervision.

Benoit Delattre: performed Conceptualization, Supervision. **Andrei Constantinescu:** performed Supervision. **Laurent Rota:** performed Supervision.

References

[1] J. Robertson, Chassis design and analysis, in: J. Happian-Smith (Ed.), An introduction to modern vehicle design, Butterworth-Heinemann, 2002, Ch. 6, pp. 125–155.

[2] G. Fischer, M. Streicher, V. V. Grubisic, Durability approval of leaf springs under operational loading, Tech. rep., SAE Technical Paper (1998). doi:10.4271/982839.

[3] P. Johannesson, M. Speckert (Eds.), Guide to Load Analysis for Durability in Vehicle Engineering, John Wiley & Sons, Ltd, 2014. doi:10.1002/9781118700518.

[4] D. Barton, J. D. Fieldhouse, Automotive Chassis Engineering, Elsevier, 2018. doi:10.1007/978-3-319-72437-9.

[5] B. Heißing, M. Ersoy, Chassis handbook: fundamentals, driving dynamics, components, mechatronics, perspectives, Springer Science, 2010. doi:10.1007/978-3-8348-9789-3.

[6] J. Schijve (Ed.), Fatigue of Structures and Materials, Springer, 2009. doi:10.1007/978-1-4020-6808-9.

[7] A. Morel, A. Bignonnet, G. Germain, F. Morel, Teaching durability in automotive Applications using a reliability approach, International Journal On Interactive Design And Manufacturing (ijidem) 4 (4) (2010) 281–287. doi:10.1007/s12008-010-0110-8.

[8] T. Svensson, P. Johannesson, Reliable fatigue design, by rigid rules, by magic or by enlightened engineering, Procedia Engineering 66 (2013) 12–25. doi:10.1016/j.proeng.2013.12.058.

[9] M. Köhler, S. Jenne, K. Pötter, H. Zenner, Load Assumption for Fatigue Design of Structures and Components, Springer Berlin Heidelberg, 2017. doi:10.1007/978-3-642-55248-9.

[10] P. Heuler, H. Klätschke, Generation and use of standardised load spectra and load–time histories, International Journal of Fatigue 27 (8) (2005) 974–990. doi:10.1016/j.ijfatigue.2004.09.012.

[11] V. V. Grubisic, G. Fischer, Methodology for effective design evaluation and durability approval of car suspension components, in: SAE Technical Paper Series, SAE International, 1997, pp. 21–33. doi:10.4271/970094.

[12] E. Baroux, Reliability fatigue design under complex loadings : from specification to validation, Ph.D. thesis, Université Paris-Saclay et Institut Polytechnique de Paris (2023).
URL http://www.theses.fr/2023IPPAX042

[13] S. M. Tipton, A review of the development and use of Neuber's rule for fatigue analysis, SAE technical paper seriesdoi:10.4271/910165.

[14] P. Heuler, T. Bruder, H. Klätschke, Standardised load-time histories - a contribution to durability issues under spectrum loading, Materialwissenschaft Und Werkstofftechnik 36 (11) (2005) 669–677. doi:10.1002/mawe.200500936.

[15] E. Bellec, M. L. Facchinetti, C. Doudard, S. Calloch, S. Moyne, M. P. Silvestri, Modelling and identification of fatigue load spectra: Application in the automotive industry, International Journal of Fatigue 149 (2021) 106222. doi:https://doi.org/10.1016/j.ijfatigue.2021.106222.

[16] J. Klemenc, M. Fajdiga, Improved modelling of the loading spectra using a mixture model approach, International Journal of Fatigue 30 (7) (2008) 1298–1313. doi:10.1016/j.ijfatigue.2007.08.024.

[17] M. Nagode, J. Klemenc, M. Fajdiga, Parametric modelling and scatter prediction of rainflow matrices, International Journal of Fatigue 23 (6) (2001) 525–532. doi:10.1016/s0142-1123(01)00007-xISTEX.

[18] E. Castillo, A. Fernández-Canteli, A unified statistical methodology for modeling fatigue damage, Springer Science & Business Media, 2009. doi:10.1007/978-1-4020-9182-7.

[19] S. Woo, Reliability design of mechanical systems, Springer, 2017. doi:10.1007/978-3-319-50829-0.

[20] V. Chmelko, M. Margetin, The performance of selected multiaxial criteria under tension/torsion loading conditions, International Journal of Fatigue 135 (2020) 105532. doi:10.1016/j.ijfatigue.2020.105532.

[21] M. Nihei, P. Heuler, C. Boller, T. Seeger, Evaluation of mean stress effect on fatigue life by use of damage parameters, International Journal of Fatigue 8 (3) (1986) 119–126. doi:10.1016/0142-1123(86)90002-2.

[22] L. Susmel, P. Lazzarin, A bi-parametric wöhler curve for high cycle multiaxial fatigue assessment, Fatigue & Fracture of Engineering Materials & Structures 25 (1) (2002) 63–78. doi:10.1046/j.1460-2695.2002.00462.x.

[23] G. Marsh, C. Wignall, P. R. Thies, N. Barltrop, A. Incecik, V. Venugopal, L. Johanning, Review and application of rainflow residue processing techniques for accurate fatigue damage estimation, International Journal of Fatigue 82 (2016) 757–765. doi:10.1016/j.ijfatigue.2015.10.007.

[24] M. A. Miner, Cumulative damage in fatigue, Journal of Applied Mechanics 12 (3) (1945) A159–A164. doi:10.1115/1.4009458.

[25] I. Raoult, B. Delattre, Equivalent fatigue load approach for fatigue design of uncertain structures, International Journal of Fatigue 135 (2020) 105516. doi:10.1016/j.ijfatigue.2020.105516.

[26] C. Lipson, N. J. Sheth, R. L. Disney, Reliability prediction - mechanical stress/strength interference, Tech. rep., University of Michigan (1967).

[27] J. J. Thomas, G. Perroud, A. Bignonnet, D. Monnet, Fatigue design and reliability in the automotive industry, in: G. Marquis, J. Solin (Eds.), Fatigue Design and Reliability, Vol. 23 of European Structural Integrity Society, Elsevier, 1999, pp. 1–11. doi:10.1016/s1566-1369(99)80025-9.

[28] A. Bignonnet, J. J. Thomas, Fatigue assessment and reliability in automotive design, Tech. rep., SAE Technical Paper (2001). doi:10.4271/2001-01-4061.

[29] J. J. Thomas, A. Bignonnet, G. Perroud, Fatigue design and experimentations with variable amplitude loadings in the automotive industry, in: P. C. McKeighan, N. Ranganathan (Eds.), Fatigue Testing and Analysis Under Variable Amplitude Loadings Conditions, ASTM International, West Conshohocken, PA, 2005, pp. 381–394. doi:10.1520/STP11317S.

[30] D. Chojnacki, B. Delattre, Towards a better understanding of mechanical stress applied by passenger vehicle customers with optimized instrumentation and relevant data post-processing methodologies, Procedia Structural Integrity 38 (2022) 362–371. doi:10.1016/j.prostr.2022.03.037.

[31] S. A. Mulaik, Foundations of Factor Analysis, Chapman and Hall/CRC, 2009. doi:10.1201/b15851.

[32] M. Decker, G. Savaidis, Measurement and analysis of wheel loads for design and fatigue evaluation of vehicle chassis components, Fatigue & Fracture of Engineering Materials & Structures 25 (12) (2002) 1103–1119. doi:10.1046/j.1460-2695.2002.00593.x.

[33] S. Ayesha, M. K. Hanif, R. Talib, Overview and comparative study of dimensionality reduction techniques for high dimensional data, Information Fusion 59 (2020) 44–58. doi:10.1016/j.inffus.2020.01.005.

[34] C. Eckart, G. Young, The approximation of one matrix by another of lower rank, Psychometrika 1 (3) (1936) 211–218. doi:10.1007/bf02288367.

[35] J. Pagès, Multiple Factor Analysis by Example Using R, CRC Press, 2014. doi:10.1201/b17700.

[36] H. Remes, P. Gallo, J. Jelovica, J. Romanoff, P. Lehto, Fatigue strength modelling of high-performing welded joints, International Journal of Fatigue 135 (2020) 105555. doi:10.1016/j.ijfatigue.2020.105555.

[37] T. D. Gillespie, Fundamentals of Vehicle Dynamics, SAE International, 1992. doi:10.4271/r-114.

[38] F. Husson, J. Josse, J. Pagès, Principal component methods-hierarchical clustering-partitional clustering: why would we need to choose for visualizing data, Tech. rep., Applied Mathematics Department (2010).

[39] M. Nagode, J. Klemenc, Modelling of load spectra containing clusters of less probable load cycles, International Journal of Fatigue 143 (2021) 106006. doi:10.1016/j.ijfatigue.2020.106006.

[40] M. Burger, K. Dreßler, M. Speckert, Load assumption process for durability design using new data sources and data analytics,

729 International Journal of Fatigue 145 (2021) 106116. doi:10.1016/j.ijfatigue.2020.106116.
730 [41] G. J. McLachlan, D. Peel, Finite mixture models, Annual Reviews, 2000. doi:10.1002/0471721182.
731 URL https://doi.org/10.1002/0471721182
732 [42] S. Frühwirth-Schnatter, G. Celeux, C. P. Robert, Handbook of Mixture Analysis, CRC press, 2019. doi:10.1201/
733 9780429055911.
734 [43] G. J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, Taylor & Francis, 2008. doi:10.1002/9780470191613.
735 [44] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal
736 Statistical Society: Series B (Methodological) 39 (1) (1977) 1–22. doi:10.1111/j.2517-6161.1977.tb01600.x.
737 [45] C. Biernacki, G. Celeux, G. Govaert, Choosing starting values for the EM algorithm for getting the highest likelihood in
738 multivariate Gaussian mixture models, Computational Statistics & Data Analysis 41 (3-4) (2003) 561–575. doi:10.1016/
739 s0167-9473(02)00163-9ISTEX.
740 [46] C. Maugis, G. Celeux, M. Martin-Magniette, Variable selection for clustering with Gaussian mixture models, Biometrics 65 (3)
741 (2009) 701–709. doi:10.1111/j.1541-0420.2008.01160.x.
742 [47] K. Murphy, Machine Learning : A Probabilistic Perspective, MIT press, 2012.
743 [48] P. Johannesson, J.-J. Thomas, Extrapolation of rainflow matrices, Extremes 4 (2001) 241–262. doi:10.1023/a:
744 1015277305308.
745 [49] W. Q. Meeker, G. J. Hahn, L. A. Escobar, Statistical intervals: a guide for practitioners and researchers, Vol. 541, John Wiley,
746 2017.
747 [50] D. Socie, Modelling expected service usage from short-term loading measurements, International Journal of Materials and
748 Product Technology 16 (4/5) (2001) 295. doi:10.1504/ijmpt.2001.001272.
749 [51] R. Jiang, Introduction to quality and reliability engineering, Springer, 2015. doi:10.1007/978-3-662-47215-6.