Ref. Ares(2020)3267034 - 23/06/2020

Horizon 2020
European Union Funding
for Research & Innovation

This project has received funding from the Shift2Rail Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement no. 881805 (LOCATE)

Deliverable D 2.3
FMECA Analysis

Project acronym:	LOCATE
Full title	Locomotive bOgie Condition mAinTEnance
Starting date:	01/11/2019
Duration (in months):	24
Call (part) identifier:	S2R-OC-IP5-01-2019
Grant agreement no:	881805
Due date of deliverable:	M5
Actual submission date:	23-06-2020
Responsible/Author:	Manuel Leite / IST
Dissemination level:	PU
Status:	Issued

Reviewed: yes

Document history		
Revision	*Date*	*Description*
1	19-05-2020	First issue
2	05-06-2020	Partners Contributions
2.1	19-06-2020	Final revision

Report contributors		
Name	Beneficiary Short Name	Details of contribution
Manuel Leite	IST	1st DRAFT and reviews
António Andrade	IST	1st DRAFT and reviews
Virgínia Infante	IST	1st DRAFT and reviews
Xiaocheng Ge/Adam Bevan	HUD	Review and corrections
Magno Santos	EVOLEO	Final revision

Table of Contents

1. Executive Summary

The present report conducts a Failure Modes, Effects, and Criticality Analysis (FMECA) analysis of the bogie system of a freight locomotive. The main goal is to identify the potential failure modes of such system, while assessing the risks associated with those failure modes. It ranks failure risks in terms of importance and identifies mitigation actions to address these major risks.

Firstly, a Failure Mode and Effect Analysis (FMEA) is performed, identifying different sources of information regarding failure modes and quantifying severity, occurrence, and detectability for different failure modes for different components of the system. Then, uncertainty quantification of reliability/survival curves is carried out using expert judgment techniques to the responses of a survey. Such information for some failure modes, that resulted from the expert judgment assessment, is then incorporated into the FMEA analysis, and later extended in a FMECA analysis. Finally, common Risk Mitigation strategies (RSM) are proposed, whereas a combination of these is obtained regarding the aim of the project.

It is important to note, that the lack of available and reliable information to quantify the failure rates and associated reliability/survival curves required the exploration of other options to estimate such quantities. Therefore, it was necessary to compile previous studies, to conduct a survey, and to use expert judgment techniques. This comprehensive approach leads to a more robust quantification of the uncertainties associated with failure modes of the bogie system. Consequently, a more generic locomotive bogie FMECA is obtained, which would be applicable to other operators or maintenance firms of similar locomotives.

2. Abbreviations and acronyms

Abbreviation / Acronyms	Description
TD	Technology Demonstrator
WA	Work Action
WP	Work Package
FMEA	Failure Mode and Effect Analysis
FMECA	Failure Modes, Effects and Criticality Analysis

3. Background

The present document constitutes the Deliverable D2.3 “FMECA Analysis” as part of the WP2 – Requirements and Specifications.

It does not contribute any TD/WA.

4. Objective/Aim

The aim of the LOCATE project is to provide the methods and tools by which every Entity in Charge of its Maintenance (ECM) to implement predictive maintenance of bogie, which is one of safety-critical component in a rail vehicle, in order to:

- Ensure safety: The parts concerned are continuously under surveillance.
- Increase availability and reduce cost by avoiding unnecessary controls. Most checks do not result in repair or replacement. The data collected makes a continuous improvement of the maintenance process easier to implement.
- Increase reliability: Interventions are made before any problem in operation.
- Without impact on maintainability: The implementation of surveillance equipment will be done under the control of the people doing the maintenance.

The main objective of LOCATE project is to replace as necessary as possible the preventive conditional or scheduled maintenance of mechanical parts of the bogie by predictive maintenance.

It is expected that a condition –based monitoring maintenance program will:

- Increase availability (concerns only the time to work on the bogie). 30%
- Decrease of the costs (only the maintenance costs of the bogie) 20%
- Increase of the reliability (of the bogies and the components linked) 60% (incidents per unit of route)

The LOCATE will develop tools and methods

- To identify the failures in the bogies, primary and secondary suspensions, wheels, electric traction motor, or transmission. LOCATE development will be able to anticipate these failures from several days to several weeks.
- To do pre-operational and operational planning using the data produced.

5. FMEA Analysis

The FMEA analysis was one of the first engineering techniques introduced in reliability and safety engineering. It was first implemented by the US Military in the 1950s with the *MIL-STD-1629A* [2] guideline, and afterwards, it was developed in industries such as the automotive industry, food processing industry, and electronic equipment industry [3][4]. As one of well-used safety and reliability assessment techniques, FMEA provides a framework that defines, identifies, prioritizes and controls all potential failure modes that may include in the system design, manufacture phases, or functional process of the entire system [1].

In the railway industry, the FMEA analysis was first introduced with the RAMS Guideline [5] and has since been developed. Within LOCATE project, a Failure Mode and Effect Analysis (FMEA) was performed in order to identify and prioritize potential hazards related to the functional failures in subsystems (and components) of the Bogie in terms of risk importance and propose mitigation measures for those risks in subsystems and components, as well as further analysis/modelling.

In the project, we have collected detailed information of select components and conducted a FMEA analysis to identify and asses potential failure modes and risks associated in order to provide reference and guidance for further measures on the maintenance of locomotive bogies. A first section on the FMEA methodology is discussed below and later on it is applied, in order to identify the most critical components and failure modes.

5.1. Risk Priority Number

The procedure for performing a FMEA in the railway industry is recommended in the RAMS Guideline Standard "*BS EN:50126 Railway Applications - The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS) - Part 1: Generic RAMS Process*" [5] and shown in Figure 1:

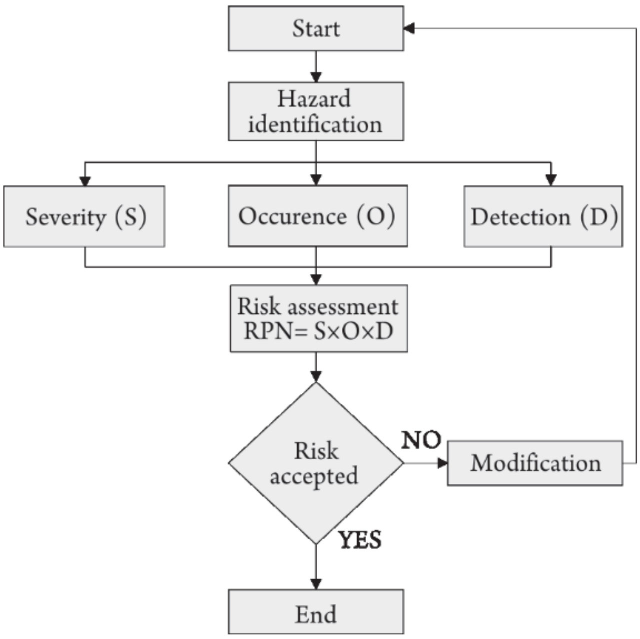


Figure 1: Procedure for implementing a FMEA in the rail industry [5]

As recommended in the RAMS Guideline [5] and the railway wheelset standard BS EN 60812 [6], the importance of risks associated with a hazard in a railway system can be prioritised by using an index, named as the Risk Priority Number (RPN). The Risk Priority Number takes three global indicators into account, namely:

1. <u>Severity, S:</u> a risk indicator corresponding to the consequences of the failure mode

2. Occurrence, O: a risk indicator corresponding to the probability of occurrence of the failure mode
3. Detectability, D: a risk indicator corresponding to the probability that a failure mode is detected (in an early stage)

These indicators can be assessed on a scale from one to ten and the Risk Priority Number (RPN) is obtained as a product of these.

$$RPN = S \times O \times D \qquad (1)$$

Therefore, the Risk Priority Number (RPN) takes values from 1 to 1000.

The criteria to define each of the global indicators goes by the guidelines provided in the *BS EN 60812* [6] standard for the wheelsets. For the Severity (S), the criteria to define the indicator can be verified in Table 1.

Table 1: Definition of the severity in the UIC Guideline (based on EN 60812, Analysis techniques for system reliability – Procedure for failure mode and effect analysis (FMEA)) [11/2006][6]

Rank	Impact	Criteria	Example
1	no impact	No recognisable effect.	
2	very little	Error is noticed by few passengers. Minor changes in structure and dimensions which are below limits.	
3	Little	Error is noticed by few passengers and there are impacts on rolling stock and infrastructure on long term	
4	very low	Error is noticed by many passengers. Due to the failure there is an impact on the quality of rolling stock and on the infrastructure in long term.	
5	Low	Error is noticed by all passengers. Due to the failure there is an impact on the quality of rolling stock and on the infrastructure in mid-term.	less comfort, minor damages on transported goods, eventually higher noise level, increase maintenance cost
6	moderate	Due to the failure there is an impact on the quality of rolling stock and on the infrastructure in short term. Error is noticed by all passengers. Loaded goods can get damaged.	
7	high	Risk of very few light injured people and risk of significant impact on environment and operation. Operation on the line is closed or the line capacity is declined for hours. The loaded goods can get damaged.	small derailment in a shunting yard
8	very high	Risk of few injured people and severe impact on environment. Operation on the line is closed for weeks. Part of the train is destroyed.	major derailment in a shunting yard
9	unsafe with warning	Risk of multiple injured people and few dead people. The impact on environment is very high. Operation on the line is closed for weeks. Large parts of the train are destroyed.	
10	unsafe without warning	Risk of many dead and numerous injured people or the impact on environment is catastrophic. Operation on the line is closed for weeks. The train is destroyed.	derailment ("Viareggio")

As can be seen, the Severity (S) criteria is established concerning financial losses and human fatalities, whereas its lowest score can have no recognizable impact on the functionality of the system and its highest score can bring human losses and a destructive impact in the operation of the system.

For the Occurrence (O), the criteria presented in Table 2 was established by the UIC Guidelines [6]:

Table 2: Definition of the occurrence in the UIC Guideline (based on EN 60812, Analysis techniques for system reliability – Procedure for failure mode and effect analysis (FMEA) [11/2006][6]

Rank	Impact	Probability number of failures per operating-h "vehicle is in use"
1	little - failure is implausible	< 10-9
2	very low: relative very few failures	=< 10-9 till < 3 * 10-8
3	low: relative few failures	=< 3 * 10-8 till < 8 * 10-7
4	moderate: seldom there are failures	=< 8 * 10-7 till < 2 * 10-6
5	moderate: sometimes there are failures	=< 2 * 10-6 till < 5 * 10-6
6	moderate: often there are failures	=< 5 * 10-6 till < 10-5
7	high: repeating failures	=< 10-5 till < 2 * 10-5
8	high: repeating failures in short cycle	=< 2 * 10-5 till < 4 * 10-4
9	very high: Failures in short cycle are nearly not nearly avoidable	=< 4 * 10-4 till < 8 * 10-3
10	very high: Failures in very short cycle which are not avoidable	more than 8 * 10-3 per year

The score criteria of the Occurrence (O) is dependent on the failure rate of the identified hazard. For low failure rates, where the probability of the event to happen is relatively small, lower scores are given. Contrarily, for high failure rates, where the probability of the event to happen is high, the scores are higher. For the Detectability (D), the criteria presented in Table 3 was established by the UIC Guidelines.

Table 3: Definition of the detectability in the UIC Guideline (based on EN 60812, Analysis techniques for system reliability – Procedure for failure mode and effect analysis (FMEA) [11/2006][6]

Rank	Detection	Criteria
1	nearly certain	With a very high probability a failure will be detected in a very early initial stage.
2	very high	With a high probability a failure will be detected in a very early initial stage.
3	high	With a high probability a failure will be detected in an early initial stage.
4	moderate high	With a high probability a failure will be detected after initial stage.
5	moderate	With a moderate probability a failure will be detected while existing a short while before getting critical.
6	low	A failure will be detected while existing a while just before getting critical.
7	very low	A failure will be detected while existing for a long while just before getting critical.
8	little	A failure will be hardly detected in a very late stage.
9	uncertain	The detection of a failure before becoming critical is uncertain.
10	nearly uncertain	The detection of a failure nearly is not possible

The Detectability (D) criteria is mainly based on expert judgment, where low scores mean the failure can be easily detected in early initial stages and high scores mean the failure can be hardly detected.

5.2. Selection of components and identification of their failure modes

The components selected in the LOCATE project are summarised in Deliverable D2.1 and include:

1) Wheelset subsystems;
2) Axle-box;

3) Bogie Frame;
4) Brake System;
5) Suspension system / elements;
6) Electric Traction Motor.

The justification for the selection of the case studies can be found In Deliverable 2.1.
Once these components are selected, FMEA requires the identification of potential component and interface failure modes and their effects to the system and ultimately to the bogie (the system-of-interest). According to MIL-STD-1629A, a failure mode should describe the manner a component fails to fulfil its defined **function**. The key requirements for defining a failure mode can be summarised as follow:

- relates to how the failure is observed;
- describes the manner the failure occurs;
- describes the impact/effect of the failure on the component;
- relates to performance measurement of the component.

Due to limited information on the case studies being analysed, the FMEA analysis was conducted by referring to the findings from the previous EU Project INNOWAG [7], a project focused on lightweight cargo wagons Bogies which shares a number of commonalities with the LOCATE project. For example; in the LOCATE project, many similarities can be found on the functional breakdown of both Bogie systems, as well as typical failure modes and their effects on the system.
The INNOWAG project comprised its study with three different sets of data, all based in three different subsystems of the bogie, namely: the wheelset, the braking system, and the suspension system. Although the different datasets belong to three different operating wagon companies, which have different maintenance policies, a combined FMEA analysis spreadsheet was created to list the most critical components and their failure modes.
Using the methodology described above, an evaluation of the RPN number was performed for all failure modes identified. In accordance with the UIC Guidelines [6], a threshold limit value of $RPN_i = 250$ was set and all the failure modes with a higher value than this acceptable threshold are identified as critical. In addition to this threshold limit value, and to guarantee that all critical failure modes were identified, all failure modes for which its Severity (S) number is $S = 10$ are considered as critical.
The results of the FMEA analysis of the three systems are shown in Table 4.

Table 4: FMEA Analysis

Subsystem	Component	Failure Mode	Failure Rate 1/h	Severity, S		Occurrence, O		Detectability, D		RPN
Wheelset	Axle	Axle crack	1.31E-06	unsafe, without warning	10	low, relative few failures	3	moderate	5	150
	Wheel	Wheel out of round	6.04E-06	very high	8	moderate, often some failures	6	very low	7	336
		Wheel cracks and notches	4.80E-05	very high	8	high, repeating failures in short cycle	8	very low	7	448
		Wheel flat	9.60E-05	very high	8	high, repeating failures in short cycle	8	very low	7	448
		Wheel thermomechanical crack	3.50E-07	very high	8	low, relative few failures	3	very low	7	168
		Wheel build-up material	6.00E-05	very high	8	high, repeating failures in short cycle	8	very low	7	448
		Wheel profile under threshold limit	8.40E-04	unsafe, without warning	10	very high, many failures in short cycles	9	low	6	540
	Axlebox	Absence of the cover box screw	6.00E-05	very high	8	high, repeating failures in short cycle	8	moderate	5	320

		Housing not watertight	1.20E-04	very high	8	high, repeating failures in short cycle	8	moderate	5	320
		Bearing Failure	2.12E-06	unsafe, without warning	10	moderate, sometimes some failure	5	very very low	8	400
Braking System	Brake	Parts of brake rigging hanging	2.01E-05	very high	8	high, repeating failures in short cycle	8	moderate	5	320
		Brake isolating cock	2.01E-05	very high	8	high, repeating failures in short cycle	8	uncertain	9	576
		Cast iron Brake Block	1.08E-04	moderate	6	high, repeating failures in short cycle	8	very low	7	336
		Composite Brake Block	3.12E-05	moderate	6	high, repeating failures in short cycle	8	very low	7	336
		Brake coupling missing	1.20E-04	moderate	6	high, repeating failures in short cycle	8	very high	2	96
	Pneumatic System	Front air valve damaged	6.00E-05	unsafe, without warning	10	high, repeating failures in short cycle	8	moderate	5	400
		Brake cylinder damaged	6.00E-05	moderate	6	high, repeating failures in short cycle	8	very very low	8	384
		Air distributor damaged	3.00E-04	moderate	6	high, repeating failures in short cycle	8	uncertain	9	432
		Slack adjuster damaged	2.40E-04	very high	8	high, repeating failures in short cycle	8	low	6	384
Suspension System	Spring Buckle	Spring Buckle Fracture	6.00E-05	unsafe, without warning	10	high, repeating failures in short cycle	8	very very low	8	640
	Helical Spring	Helical Spring broken	6.00E-05	unsafe, without warning	10	high, repeating failures in short cycle	8	moderate	5	400
	other suspension elements	Bottoming between Axlebox housing and bogie frame	1.44E-06	unsafe, without warning	10	low, relative few failures	3	moderate	5	150

In order to highlight the most critical failure modes that resulted from the FMEA analysis, the background of the cells corresponding to an *RPN* higher than the threshold limit (250) and/or with a Severity (S) indicator equal to 10, were coloured in red and considered critical components.
Following the definition of the subsystems, the critical components, and their failure modes were defined again, with the guidance of the FMEA analysis results and literature review (Table 5).

Table 5: Components and Failure Modes identified as critical

Subsystem ID	Subsystem	Component	Component ID	Failure Mode	Source
1	Wheelset	Axle	1.1	Axle Crack	FMEA
1	Wheelset	Wheels	1.2	Wheel out of round	FMEA
1	Wheelset	Wheels	1.2	Wheel Cracks and notches	FMEA
1	Wheelset	Wheels	1.2	Wheel Build-up Material	FMEA
1	Wheelset	Wheels	1.2	wheel flat	FMEA
1	Wheelset	Wheels	1.2	Profile under the threshold limit	FMEA

1	Wheelset	Bearings	1.3	-	FMEA
2	Axle Box	Axle Box	2.1	Absence of the cover box screw	FMEA
2	Axle Box	Axle Box	2.1	Housing not watertight	FMEA
3	Axle Box	Axle Box	2.1	Bearing Failure	Literature
3	Bogie Frame	Frame	3.1	-	Literature
4	Brake System	Brake	4.1	parts of brake rigging hanging	FMEA
4	Brake System	Brake	4.1	Brake isolating cock	FMEA
4	Brake System	Brake	4.1	Cast Iron Brake Block	FMEA
4	Brake System	Brake	4.1	Composite Brake Block	FMEA
4	Brake System	Pneumatic Braking system	4.2	Front air valve damaged	FMEA
4	Brake System	Pneumatic Braking system	4.2	Brake cylinder damaged	FMEA
4	Brake System	Pneumatic Braking system	4.2	Air distributor damaged	FMEA
4	Brake System	Pneumatic Braking system	4.2	Slack adjuster damaged	FMEA
4	Brake System	Master/Auxiliary Compressor	4.2	-	Literature
4	Brake System	Master/Auxiliary Compressor Driving Motor	4.3	-	Literature
4	Brake System	Servo-motor in the braking system	4.5	-	Literature
4	Brake System	Other Elements of the pneumatic braking system	4.6	-	Literature
4	Brake System	Other Elements of the braking system (pins, sleeves,…,)	4.7	-	Literature
5	Suspension Elements	Spring Buckle	5.1	Spring Buckle Fracture	FMEA
5	Suspension Elements	Helical Spring	5.2	Helical Spring broken	FMEA
5	Suspension Elements	Other Suspension elements	5.4	Bottoming between Axlebox housing and bogie frame	FMEA
6	Electric Traction Module	Power transmission system	6.1	-	Literature
6	Electric Traction Module	Shaft Coupling	6.2	-	Literature
6	Electric Traction Module	Traction Motor	6.3	-	Literature

As shown in Table 5, some components are not disaggregated in their failure modes, mostly due to lack of information or data.

6. Reducing Uncertainty

There are some challenges when applying an adaptation of a classic method to the real-world scenarios. In the LOCATE project, one of the main challenges is the uncertainty embedded in the decision-making process, due to the availability of reliable information from FGC of the real-life behaviour of some subsystems and components. To mitigate the impact of the uncertainty and to gain more confidence in our analysis in order to accomplish a more realistic assessment aligned with the real case study of FGC freight locomotives, a survey was conducted to experts and expert judgment techniques were used to quantify survival curves and failure curves of the most critical subsystem of the bogie: the wheelset subsystem.

6.1. Expert Judgement Techniques

Expert Judgement (also known as expert elicitation) is an effective tool to explore the sources of uncertainty and to answer questions where data is scarce or expensive to obtain [8]. According to Taylor et al. [9], elicitation is the procedure of developing the expertise of a person about one or more uncertain quantities into a probability distribution. Consequently, the success of such elicitations depends not only on the type and format of the questions but also on the personality, experience, and technical background of each expert [10]. Therefore, the definition of an expert is not only based on great knowledge of the subject matter but, according to Wood and Ford [11], someone who represents problems in terms of formal principles and solves problems with its acknowledged strategies.

Once the elicitation process is completed, and assuming the Decision Maker (DM) has access to more than one expert, the weighting of each expert can follow a mathematical or a behavioural approach, in order to produce a single aggregated distribution. Mathematical aggregation methods create single evaluations per variable by applying analytical models to each assessment, such as the Bayesian methods, Opinion pooling, or the Cooke's Method. Behavioural aggregation methods, on the other hand, comprise a synergy of the experts to accomplish a homogeneity on the assessment of the variable of interest [12]. A typical behavioural method is the Delphi method, which implicates various rounds of experts providing their assessments, sharing that information with all the other experts, and then allowing them to review their assessment to move towards a general opinion. This is commonly known as a group elicitation [13].

From the mathematical models mentioned, the Opinion Pool method is the most widely used technique due to its simplicity. The simplest decision-making process is seen as the linear opinion pool, where the aggregated distribution is obtained through an equal-weighted average of the individual distributions. Nevertheless, this method has its inconsistencies, since the weighting process does not contemplate the experts who are recognized to be better and to have more expertise in the required field. Consequently, a more refined method of the Linear Opinion Pool was developed by Roger M. Cooke, a mathematician from the US, called the Structured Expert Judgement, also known as the Cooke Method, which has been validated over the years as a more accurate and informative assessment as the equal weighting of experts [14].

For the present study, the mathematical model to estimate the weight of each expert has been used and the associated Structured Expert Judgement was selected as the elicitation method.

6.1.1. Structured Expert Judgement (SEJ) – The Cooke Method

The Cooke Method is an approach for eliciting and mathematically aggregate expert judgment based on the principle of objective calibration scoring and hypothesis testing in classical statistics [15]. The method consists of two types of questions: target questions and calibration questions. Target questions are, as Cooke describes, the variables of interest, i.e. those variables that one wants to quantify, and that cannot be assessed with other methods. Calibration questions are questions that are either known to the expert at the time of the elicitation, or will be known during the analysis period and provide the experts' know-how on the specific topic. Experts are then scored based on their performance on the calibration questions, and their assessments are weighted (according to their scores) and combined.

6.1.2. Calibration Questions – Theoretical Background

In the Cooke Method, each expert quantifies his/her uncertainty for each calibration question, whereas his/her score is based on two variables: i) *the calibration score*, which measures the statistical accuracy of the expert, and ii) the *information score*, which measures the informativeness of the experts' assessments. The uncertainty quantification can take many forms, nevertheless, it tends to take a common structure due to application purposes: the experts commonly specify their fifth (5%), fiftieth (50%) or median and ninety-fifth (95%) percentiles for the estimate of each uncertainty, and thus each expert provide a 90% range of possible values. With the 5% quantile, the expert is assessing the lower bound, meaning the expert believes that the true value (also known as realization) has a 5% chance of being below that value and a 95% chance of being above. Similarly, with the 95% quantile, the expert is assessing the upper bound, and thus he/she believes that there is a 95% chance that the true value lies below that value and a 5% chance of being above. The experts' best guess, the 50% quantile, also known as the median value, specifies that there is an equal chance that the true value is lower or higher than the value given. Consequently, there is a 90% confidence from the expert that the true value lies between the lower and the upper bound, as we can verify in the following formula:

$$p(x) = \begin{cases} v_1, \ x \in [0, 0.05[\\ v_2, \ x \in [0.05, 0.5[\\ v_3, \ x \in [0.5, 0.95[\\ v_4, \ x \in [0.95, 1] \end{cases} \quad (2)$$

These three quantile assessments (5%, 50% and 95%) define four intervals or inter-quantile ranges: i) one from 0% up to the 5%, ii) one from the 5% up to the 50%, iii) one from the 50% up to the 95% and finally iv) one from the 95% up to 100%. This leads to the theoretical probability vector:

$$p = (p_1, p_2, p_3, p_4) = (0.05, 0.45, 0.45, 0.05) \quad (3)$$

Which gives the expected proportion of realizations in each interval. In practice, the inter-quantile ranges of the expert do not usually capture the true realizations at the expected frequency. If *N* quantities are assessed, each expert may be regarded as a statistical hypothesis, namely, each realization falls in one of the four inter-quantile intervals with probability vector p.

Assuming one has $x_1, \dots, x_N$ realizations of these quantities. One may then form the sample distribution of the expert's inter-quantile intervals as:

$$\begin{aligned} &s_1(e) = \#\{\, i \mid x_i \leq q_{5\%} \,\}/N \\ &s_2(e) = \#\{\, i \mid q_{5\%} < x_i \leq \ q_{50\%}\}/N \\ &s_3(e) = \#\{\, i \mid q_{50\%} < x_i \leq \ q_{95\%}\}/N \\ &s_4(e) = \#\{\, i \mid q_{95\%} \leq x_i \,\}/N \\ &s(e) = (s_1, \dots, s_4) \end{aligned} \quad (4)$$

Vector $s(e)$ is the empirical probability vector of an expert e.

In order to measure how different the vector $s(e)$ is from vector p, one can apply the relative information of vector $s(e)$ with respect to vector p, also known as the Kullback-Leibler (K-L) divergence or distance, which measures the difference between two distributions. This divergence is given by:

$$I(s, p) = \ \sum_{i=1}^{n} s_i \ln\left(\frac{s_i}{p_i}\right) \quad (5)$$

where n is the number of inter-quantile ranges. The divergence is equal to 0, if $s_i = p_i$, otherwise it is positive. If the realizations are indeed drawn independently from a distribution with quantiles as stated by the expert, then:

$$T = 2 \times m \times I(s, p) \sim \chi^2_{(3)} \quad (6)$$

is asymptotically following a chi-square distribution variable with 3 degrees of freedom, where m is the number of seeding variables (calibration questions). Based on this result, the calibration score is given by:

$$C(e) = 1 - F_{\chi^2_3}(t) \quad (7)$$

where F is the cumulative distribution function of the chi-square probability distribution. The calibration score can vary from 0 to 1. The greater the calibration score, the more statistically accurate is the expert.
Unlike the calibration score, the information score is calculated for each calibration question separately. Generally, the information in a distribution is the degree to which the distribution is concentrated. To determine the information score for each expert in each question, one first needs to determine the intrinsic range. That is, one needs to obtain bounds that are determined by expert assessments and the realizations. For this, one first takes the minimum between all the 5% quantiles of each expert and the realization itself ($\min(5\%_1, 5\%_2, \ldots, 5\%_e, realization) = X$), which is considered as the lower bound L. Likewise, the maximum between all the 95% quantiles of each expert and the realization, which is assumed as the upper bound U. The intrinsic range in each seed variable (calibration question) i is given by:

$$[L_i, U_i] = [L, U] \quad (8)$$

The intrinsic range is then determined by extending the interval by an overshoot k. The extended intrinsic range is then given by:

$$[L_i^*, U_i^*] = [L - k(U - L), U + k(U - L)] \quad (9)$$

Typically, the overshoot k is 10%. Hence, the information score including all assessments for each expert is calculated by the following formula:

$$Inf(e) = (1/N) \sum_{i=1,\ldots,N} I(f_{e,i}|g_i) \quad (10)$$

Where $I(f_{e,i}|g_i)$ for 3 quantiles is given by:

$$I(f_{e,i}|g_i) = I(e_i) = 0.05 \times ln\frac{0.05}{q_{5\%} - L_i^*} + 0.45 \times ln\frac{0.45}{q_{50\%} - q_{5\%}} + 0.45 \times ln\frac{0.45}{q_{95\%} - q_{50\%}} + 0.05 \times ln\frac{0.05}{U_i^* - q_{95\%}} + \ln(U_i^* - L_i^*) \quad (11)$$

Where g_i is the background density for each seeding variable and $f_{e,i}$ is expert e's density for seeding variable *i*.

The combined score of an expert e will serve as an (unnormalized) weight for each expert. It is based on both calibration score and information score, and is obtained by the following formula:

$$w'_{e,\alpha} = C(e) \times Inf(e) \times 1_\alpha(C(e) \geq \alpha) \quad (12)$$

Where $1_\alpha(C(e) \geq \alpha) = 1$ if $C(e) \geq \alpha$ and 0 otherwise. The use of a cut off threshold α is imposed by the requirement that the weights w'_e should be an *asymptotically strictly proper scoring rule,* meaning the long-run expected weights should correspond to the expert's true beliefs. Nevertheless, for the current project, the α threshold was considered zero, to take into account all experts' opinions, despite ones being considered as having better know-how.

Finally, the weights are then normalized across all experts:

$$w_e = \frac{w'_e}{\sum w'_e} \quad (13)$$

6.1.3. Illustrative Example

In order to have a better perception of how the calibration and information scores are obtained, an illustrative example is presented.

<u>Calibration Score:</u>
Taking an example of two expert assessments for 10 calibration questions about the failure of different Bogie components, like the primary suspension or the wheelset, before an inspection of the bogie is performed. The three quantile assessments of the experts are provided and after the inspection period, the realizations of the number of failures of each component are also taken into account. When observing the assessments of the first expert, one can verify that in 3 out of the 10 calibration questions, the experts overestimated, meaning that in 3 questions, the realization of the assessment is below his 5% quantile. In addition, in 1 calibration question, the expert underestimated the realization, meaning the realization was above its 95% quantile. For the remaining 5 questions, 4 questions had the realizations between the 5% and the 50% quantiles, while 2 questions had the realization between 50% and 95%. For the second expert, 1 question

had the realization under 5%, 6 questions had the realization between 5% and 50%, 2 questions between 50% and 95%, and 1 question was underestimated.
The following assessments lead to the vector of observed proportions of realizations for each expert:

$$s_{e_1} = (s_1, s_2, s_3, s_4) = \left(\frac{3}{10}, \frac{4}{10}, \frac{2}{10}, \frac{1}{10}\right) = (0.3, 0.4, 0.2, 0.1)$$
$$s_{e_2} = (s_1, s_2, s_3, s_4) = \left(\frac{1}{10}, \frac{6}{10}, \frac{2}{10}, \frac{0}{10}\right) = (0.1, 0.6, 0.2, 0.1)$$

The Kullback-Leibler (K-L) divergence (4) for both exerts is then obtained as:

$$I_{e_1}(s,p) = 0.3\ln\left(\frac{0.3}{0.05}\right) + 0.4\ln\left(\frac{0.4}{0.45}\right) + 0.2\ln\left(\frac{0.2}{0.45}\right) + 0.1\ln\left(\frac{0.1}{0.05}\right) = 0.3975$$
$$I_{e_2}(s,p) = 0.1\ln\left(\frac{0.1}{0.05}\right) + 0.6\ln\left(\frac{0.6}{0.45}\right) + 0.2\ln\left(\frac{0.2}{0.45}\right) + 0.1\ln\left(\frac{0.1}{0.05}\right) = 0.1491$$

With the (K-L) divergence, it is possible to calculate the statistical accuracy, namely the calibration score for both experts. The calibration score is then obtained with the formula (6):

$$C(e_1) = 1 - F_{\chi^2_{(3)}}(2 \times 10 \times 0.3975) = 1 - F_{\chi^2_{(3)}}(7.95) = 0.047$$
$$C(e_2) = 1 - F_{\chi^2_{(3)}}(2 \times 10 \times 0.1491) = 1 - F_{\chi^2_{(3)}}(2.982) = 0.394$$

As one can verify, the calibration score of expert 2 was higher than the calibration score of expert 1. Therefore, the assessment of expert 2 was statistically more accurate than the assessment of expert 1.
The calibration score is then obtained for all the experts' assessment. The higher the calibration score, the more accurate is the expert.

<u>Information Score:</u>
For the same two experts and the respective assessments, let one consider that for one specific calibration question, namely the question where it is asked about the number of failures of the wheels due to cavities, the realization is 16 failures and the experts' assessments were the following:

Table 6: Experts assessment as an illustrative example

	5%	50%	95%
Expert 1	7	10	15
Expert 2	6	7	10

To calculate the information score of each expert for this calibration question, one first needs to determine the intrinsic range. For this specific case, the intrinsic range is given by:

$$[L_i, U_i] = [\min(5\%_1, 5\%_2, \mathit{realization}), \max(95\%_1, 95\%_2, \mathit{realization})] = [6,16]$$

By extending the interval with a 10% overshoot, the intrinsic range for the following seed variable is:

$$[L_i^*, U_i^*] = [6 - 0.1(16-6), 16 + 0.1(16-6)] = [5,17]$$

Consequently, for each expert the information score is the following:

i. $I(e_1) = 0.05 \times ln\frac{0.05}{7-5} + 0.45 \times ln\frac{0.45}{10-5} + 0.45 \times ln\frac{0.45}{15-10} + 0.05 \times ln\frac{0.05}{17-15} + \ln(17-5) = 0.179$

ii. $I(e_2) = 0.05 \times ln\frac{0.05}{6-5} + 0.45 \times ln\frac{0.45}{7-6} + 0.45 \times ln\frac{0.45}{10-7} + 0.05 \times ln\frac{0.05}{17-10} + \ln(17-5) = 0.875$

As we can verify intuitively, the distribution of expert 2 is more concentrated than the distribution of expert 1. Hence, expert 2 is more informative than expert 1 and, therefore, expert 2 has a higher information score. For all the calibration questions, one information score is obtained for each expert, whereas the final information score is obtained through an average of all information scores of each expert in each question, as we can verify with the formula (10).

<u>Combined Score:</u>
After all assessments, the unnormalized weights of each expert are obtained with the formula (12) and

divided by the sum of each expert's unnormalized weight to obtain each final weight of each expert (13).

6.2. Case Study - LOCATE

As already mentioned, the purpose of using expert judgment techniques is to estimate failure rates and survival curves of the most critical subsystem: the wheelset subsystem.
Therefore, in order to define the target variables of the expert judgment, one needs to disaggregate the wheelset subsystem in its components, to try to understand which of these might be the most desirable for the project and select those for further analysis. As a result, and based on what FGC considered critical for the analysis, the following wheelset components were identified as target variables:

i. Wheels
ii. Axles

In order to obtain a robust result, one had to construct the target questions, calibration questions, and a list of experts that would benefit the project.

6.2.1. Target Questions

Concerning the goal of this study, the target questions were formulated to try to obtain the failure rates of both components by requesting the number of failures of one batch of 1000 identical components of both types. For this, FGC provided the typical mean distances between inspections for both components, to have a better reference of the real case scenario. The defined intervals were the following:

i. <u>Wheels</u>: seven equally distant intervals with a range of 15000km each. Starting at zero, the first interval was [0-15000km], the second [15000km-30000km], and so on until the seventh interval, which was defined from [90000kms - infinite].
ii. <u>Axles</u>: seven equally distant intervals with a range of 300000km each. Starting at zero, the first interval was [0-300000km], the second [300000km-600000km], and so on until the seventh interval, which was defined from [1800000kms - infinite].

For each interval, the expert is requested to estimate the number of components (of each type) that would fail in each specific interval. The sum of all failures in all intervals is the amount of the batch, which was set equal to 1000 components. In order to adapt these failure rates with the FMEA analysis presented in the first chapter, the failure modes were aggregated in each component and considered as a whole, meaning the failure rate of the wheels is assumed to be equal to all failure modes assigned as critical, like “Wheel Cracks and notches” or “Wheel flats”. For the axle, the existent failure mode was not aggregated since it is the most known failure mode of this component.

6.2.2. Calibration Questions

The calibration questions were formulated with the guidance of two European Railway Agency (ERA) annual reports from the past 3 years, namely the *ERA report on Railway Safety and Interoperability in the EU* from 2018 [16] and 2019 [17], and also from the European Standards available for the wheelset component. Therefore, a total of 13 questions were formulated, whereas 4 questions out of 13 were used for the survey. All questions were specific to the components topic and all questions were based on the reliable data published, and thus all questions had their actual realizations.

6.2.3. List of Experts

In order to obtain reasonable results, which would reflect the real case scenario of FGC, a robust list of experts in wheelset components was established. This list was constructed with the support of the UIC Experts list, namely, the *List of recognized UIC experts to elaborate expertise on braking components 2019*

[18] and the *List of experts recognized by UIC and relative expertise of wheels 2012* [19]. In addition to UIC lists, the list of experts had also the support of the vast network of the project. It should be noted that the expert's list is composed of wheelset experts not only of freight locomotives but also passenger locomotives. The survey was established using the public platform *Google Forms.* The survey contained a brief introduction to the project, a brief illustrative example to explain the reader on the format of the elicitation, the calibration questions, and, finally, the target questions. The survey was held anonymously and from the entire list of experts, 6 experts completed the survey.

6.3. Results

This subsection comprises the results of the expert judgment performed to get the failure rates of the wheels and axle component. All in all, 6 experts completed the survey and the results are presented below. The calibration and target questions can be seen in the Appendix section.

2.3.1 Expert Weighting - Calibration

Table 7 summarizes the expert's performance in the four calibration questions.

Table 7: Experts assessment on the four Calibration Questions (CQ1- CQ4)

	CQ1			CQ2			CQ3			CQ4		
	5%	50%	95%	5%	50%	95%	5%	50%	95%	5%	50%	95%
Expert 1	40	50	80	10	50	90	50	60	80	40	50	60
Expert 2	1800	2074	2400	5	20	35	200	500	800	50	100	150
Expert 3	1500	2000	2500	20	30	40	300	600	900	50	100	150
Expert 4	270	346	422	15	30	45	10	25	40	800	1700	2600
Expert 5	1700	1900	2100	2	8	10	50	75	100	100	200	300
Expert 6	1500	1750	2000	60	70	80	40	90	200	30	50	200
Realization	**1789**			**58**			**104**			**18**		

For each expert, 3 quantiles (5%, 50% and 95%) were given in order to define the distribution range of each expert in each question. Table 7 also compares each expert's assessment in each question with the true realization, which is provided in the last row. Of course, this realization was not provided to the experts when completing the survey.

After assessing the expert's calibration questions results, the expert weights were obtained with the support of free software *EXCALIBUR.* This software enables to calculate the expert's weight according to the theoretical mathematical aggregation presented in the first subsection of this chapter. By introducing the expert's assessments, as well as the realizations of each calibration question in the software, a summarized table with all the relevant scores is obtained.

Each expert's calibration score, information score, unnormalized, and final weight can be observed in Table 8.

Table 8: Experts calibration score, information score, unnormalized and normalized weight

	Calibration Score	Information Score	Unnormalized weight	Normalized weight
Expert 1	0.01043	2.76600	0.02885	**0.28400**
Expert 2	0.00022	1.28500	0.00028	**0.00277**
Expert 3	0.01043	1.26300	0.01318	**0.12970**
Expert 4	0.00022	1.61300	0.00035	**0.00348**
Expert 5	0.01043	2.09000	0.02180	**0.21460**
Expert 6	0.02197	1.69100	0.03714	**0.36550**

As can be seen from Table 8, Expert 6 has the largest weight, due to his/her good performance in accurately replying to the calibration questions. In relation to the other experts, Expert 6 was statistically more accurate. Despite not being the most informative, since its information is not the narrowest one (Check experts assessments in Table 7), his/her accuracy stands out in comparison with the remaining experts. On the other hand, Expert 2 is the least accurate as well as one of the least informative experts in the poll. Therefore, his opinion on the target variable will have a relatively poor impact on the result of the analysis.

6.3.2. Target Variables

Having computed the weights of each expert, after analysing their performance on the calibration questions, it is now time to assess to assess their responses to the target variables. In this case, the target questions are related with the failure rates of each component being analysed. For each target variable, the weight of the expert is taken into account, being the following list a descendant ranking of the most impactful expert in each assessment on the target variables: *Expert 6, Expert 1, Expert 5, Expert 3, Expert 4* and *Expert 2*.

6.3.2.1. Wheels

For this component, Table 9 summarizes the expert's assessment in each interval mentioned in subsection 2.2.1.

Table 9: Experts assessments on wheels for each interval

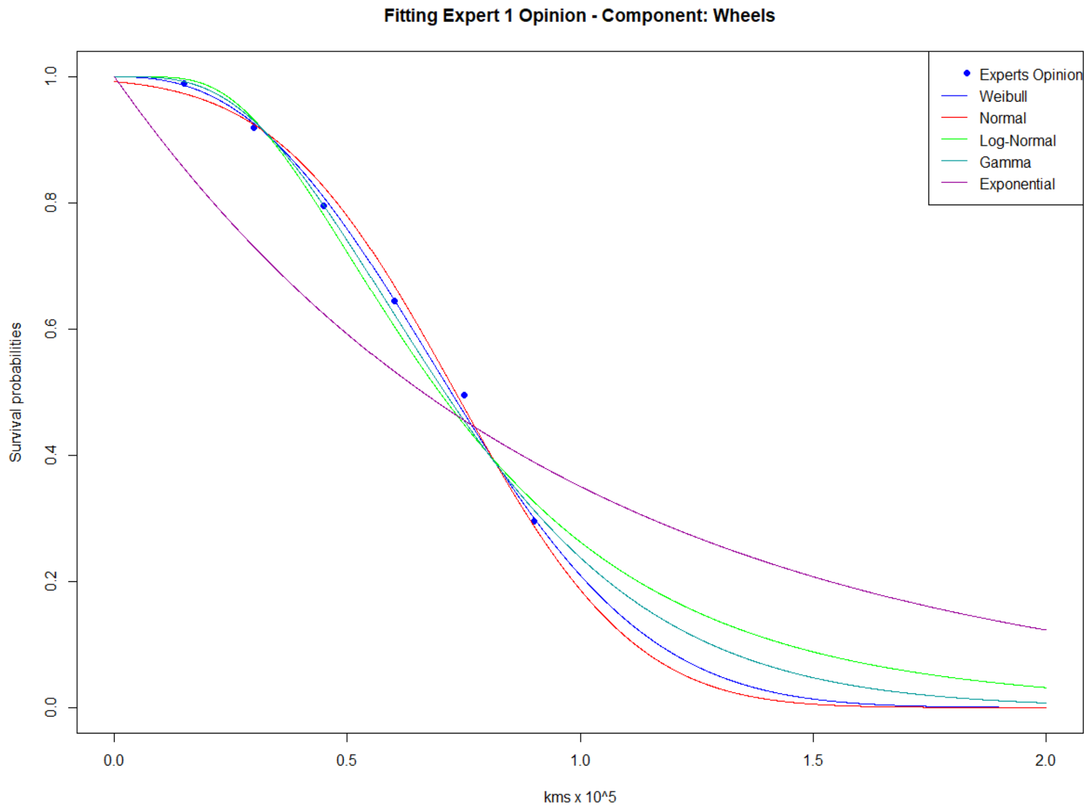
	[0 , 15] 10^3 km	**[15 , 30] 10^3 km**	**[30 , 45] 10^3 km**	**[45 , 60] 10^3 km**	**[60 , 75] 10^3 km**	**[75 , 90] 10^3 km**	**[90 , +∞] 10^3 km**	**Sum**
Expert 1	10	70	125	150	150	200	295	1000
Expert 2	10	40	50	100	100	100	600	1000
Expert 3	10	50	100	150	150	200	340	1000
Expert 4	2	8	20	40	70	90	770	1000
Expert 5	10	30	70	100	300	400	90	1000
Expert 6	10	20	70	100	150	200	450	1000

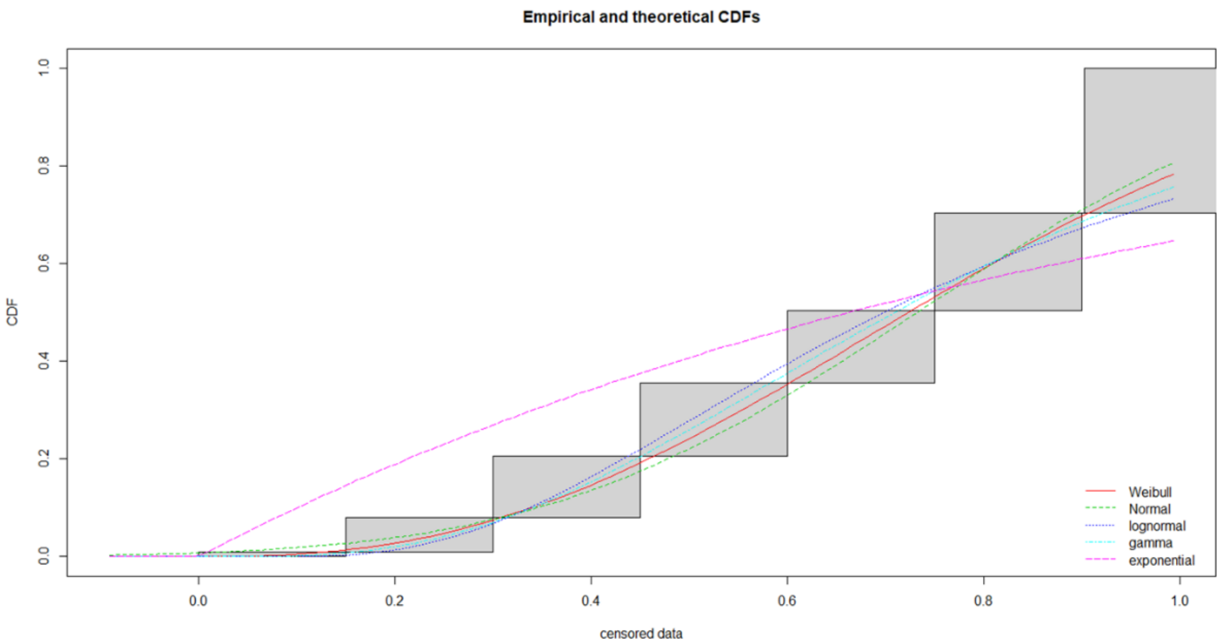
As already mentioned, for each given interval, each expert assessed its opinion on how many components will fail of the batch of 1000 components. With this assessment, and regarding each expert's weight, one can perform a survival analysis for each component, in order to define a survival/reliability curve and, consequently, obtain the failure rate of the desired component.
For this analysis, two approaches were considered:

i) <u>Approach 1</u>: First performing a survival analysis for each expert's opinion and then combining them, by taking into account the expert's weight to obtain a final survival curve with its respective failure rate.
ii) <u>Approach 2</u>: First combine expert assessments, by using a weighted expert opinion from all the experts, and then performing the survival analysis.

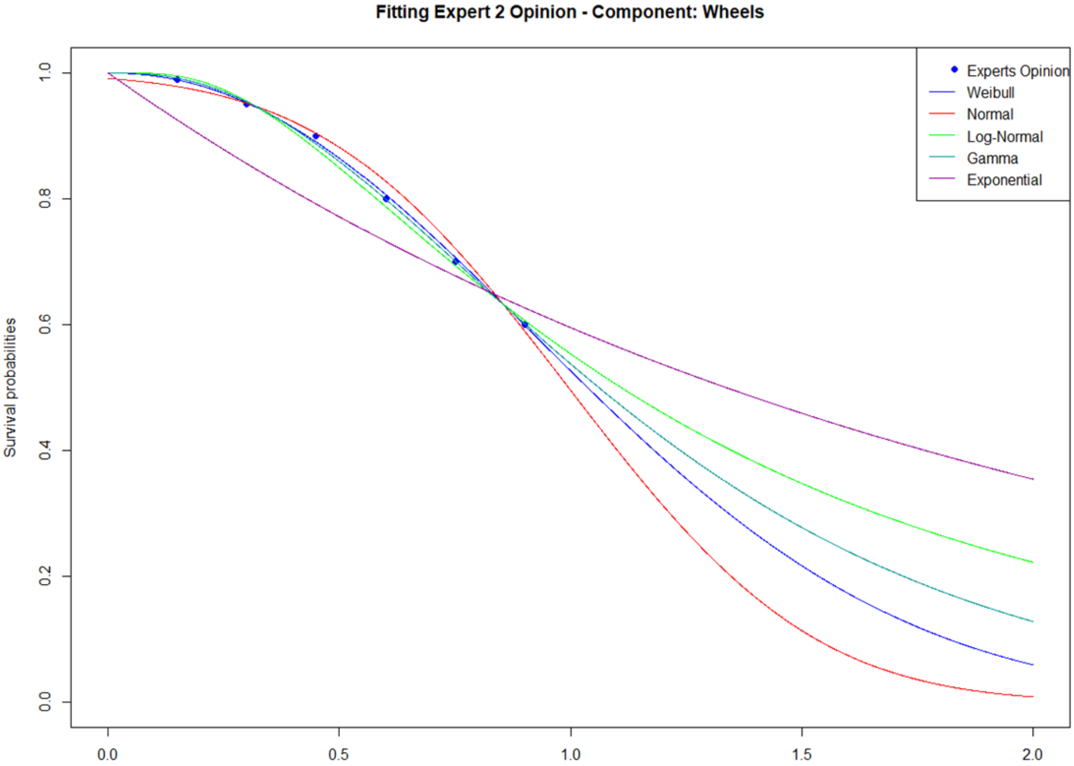
For both survival analysis, the function *Fitdistcens* from the package *Fitdistrplus* in *RStudio* was used. This function allows fitting the data of each expert to a distribution, considering the data as interval censored.
<u>Approach 1:</u>
For this first approach, a survival analysis on the opinion of each expert is performed. For each expert, 5 probability distributions were considered, namely: Weibull, Normal, Lognormal, Gamma and Exponential. These statistical distributions are commonly used in survival analysis and reliability analysis. With the support of the *Fitdistcens* function, each opinion was fitted to each distribution and afterwards compared.
In the following figures, for each expert, the five statistical distributions are fitted to the expert's opinion, as well as the opinion itself (represented with points). In addition, a figure comparing the empirical and the theoretical cumulative distribution functions is shown to emphasize which distribution best fits the expert's opinion. In this case, the best distribution is the one that follows the intervals of the *cdf* the best way.

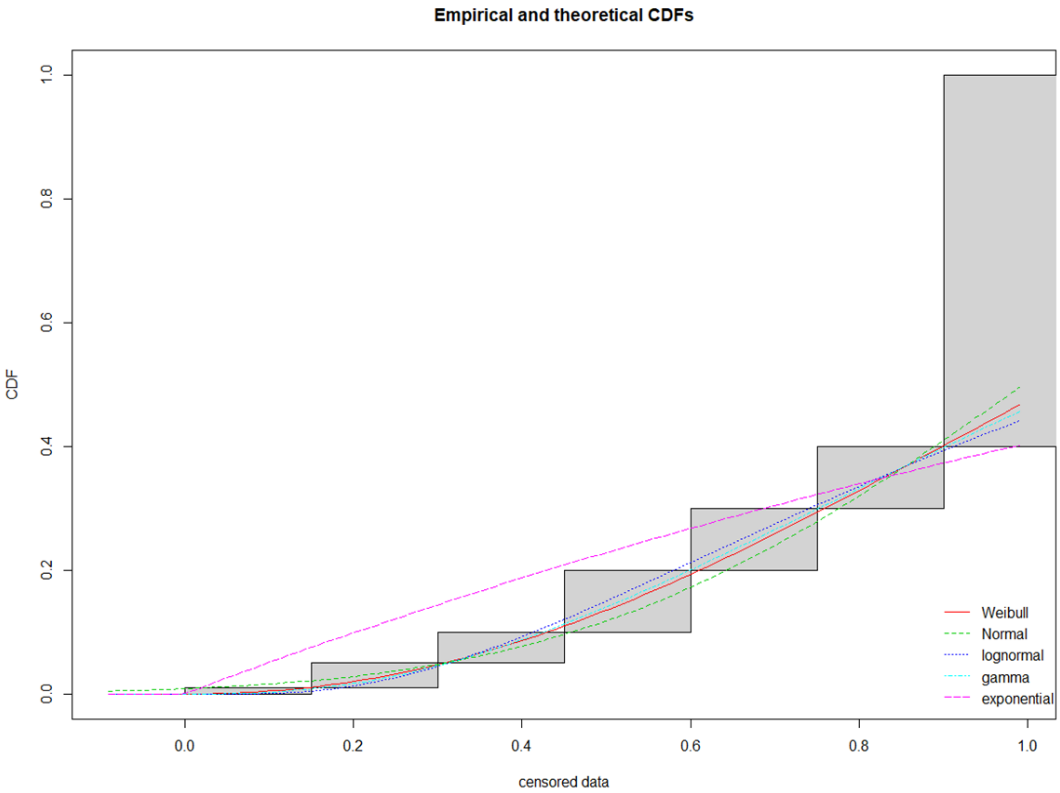
Figures 2 and 3: Fitting Expert 1 opinion – wheels



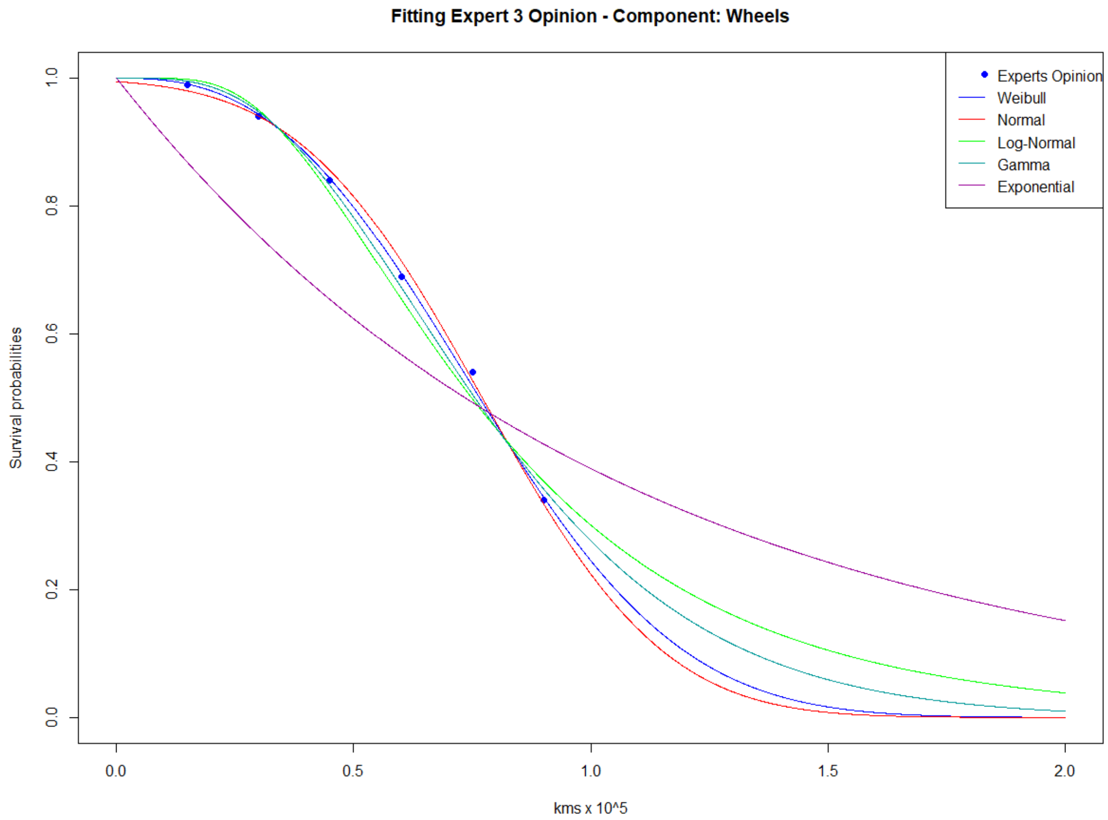


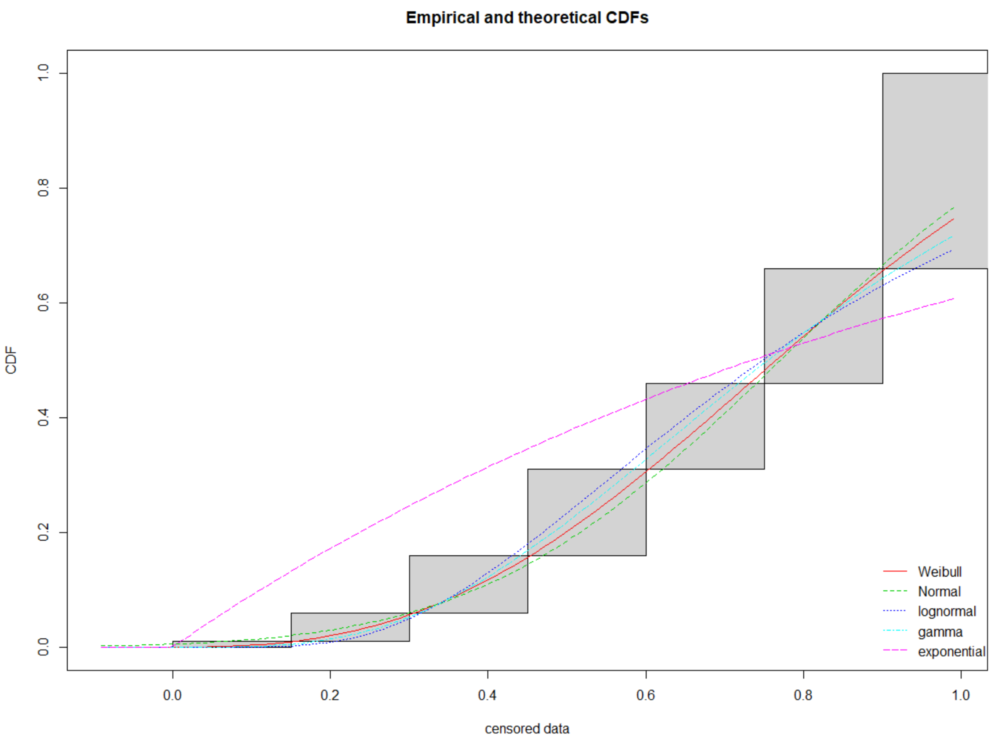
Figures 4 and 5: Fitting Expert 2 opinion – wheels



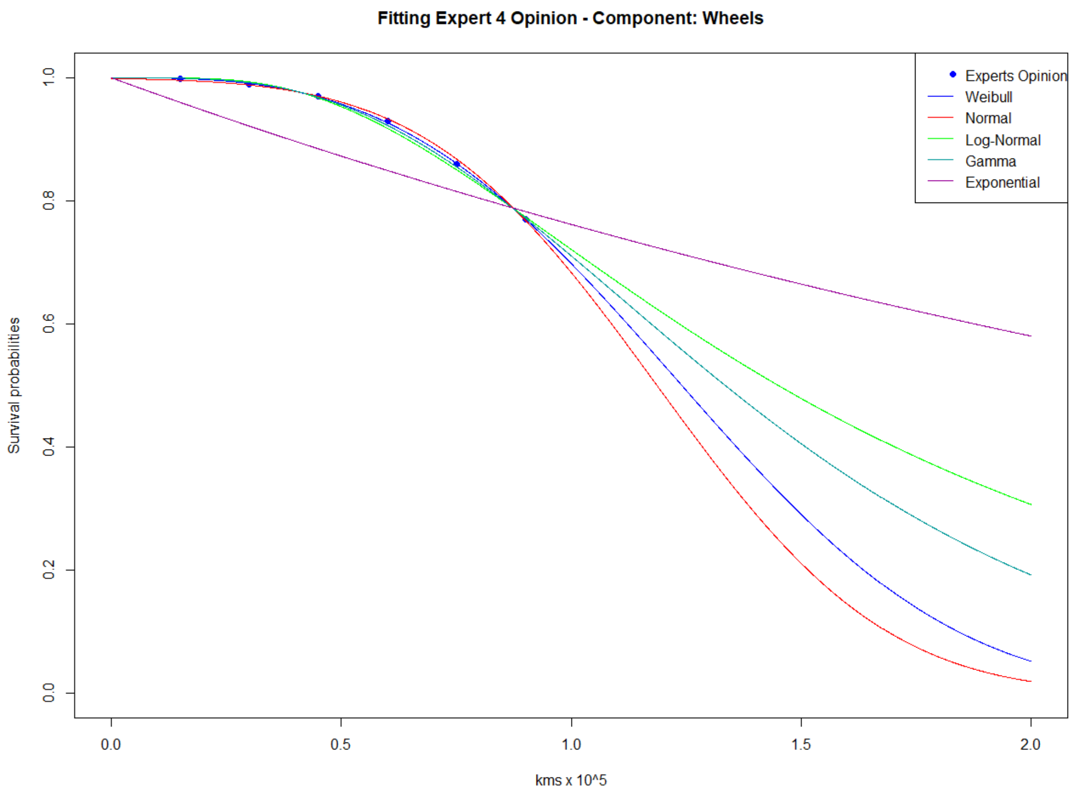


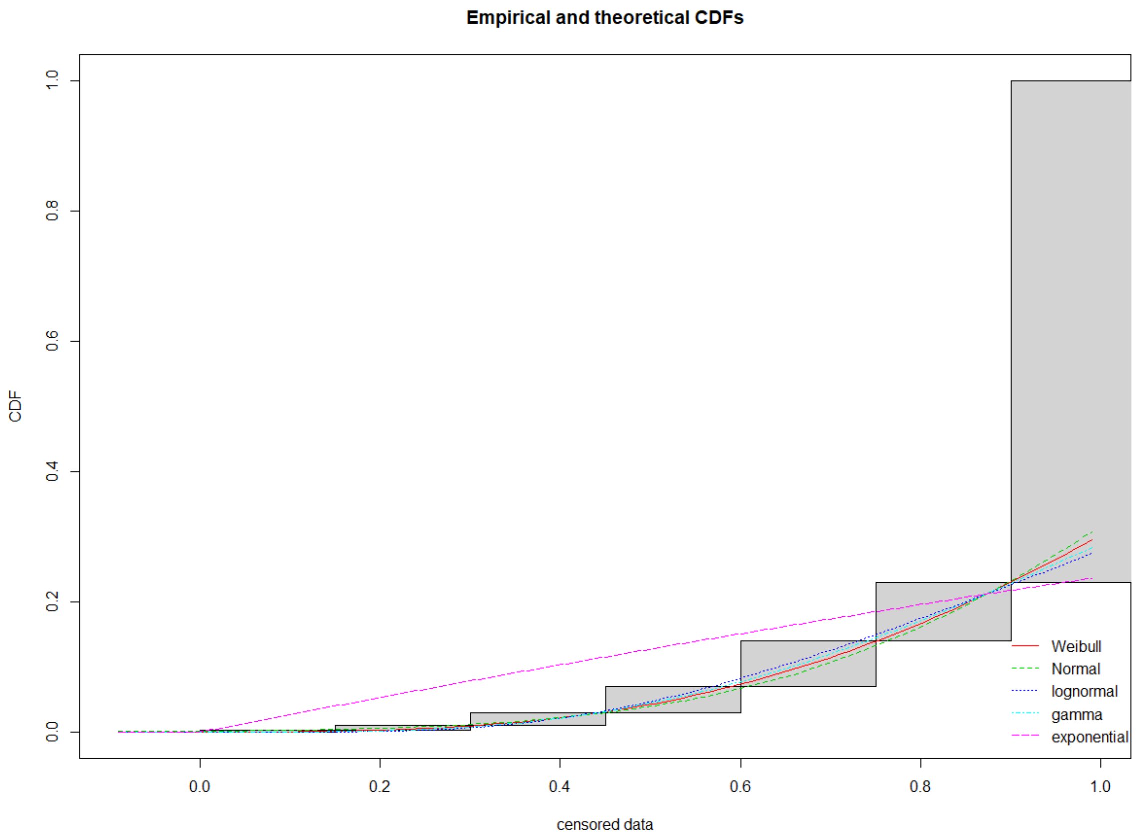
Figures 6 and 7: Fitting Expert 3 opinion - wheels



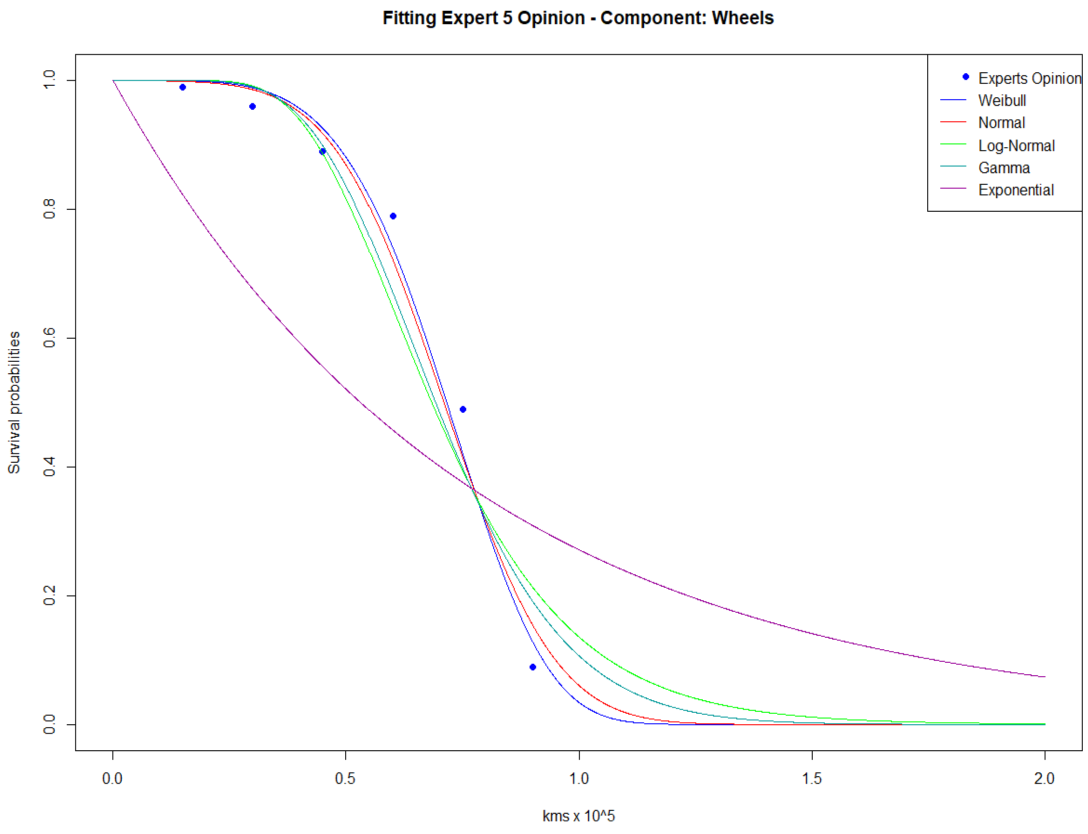


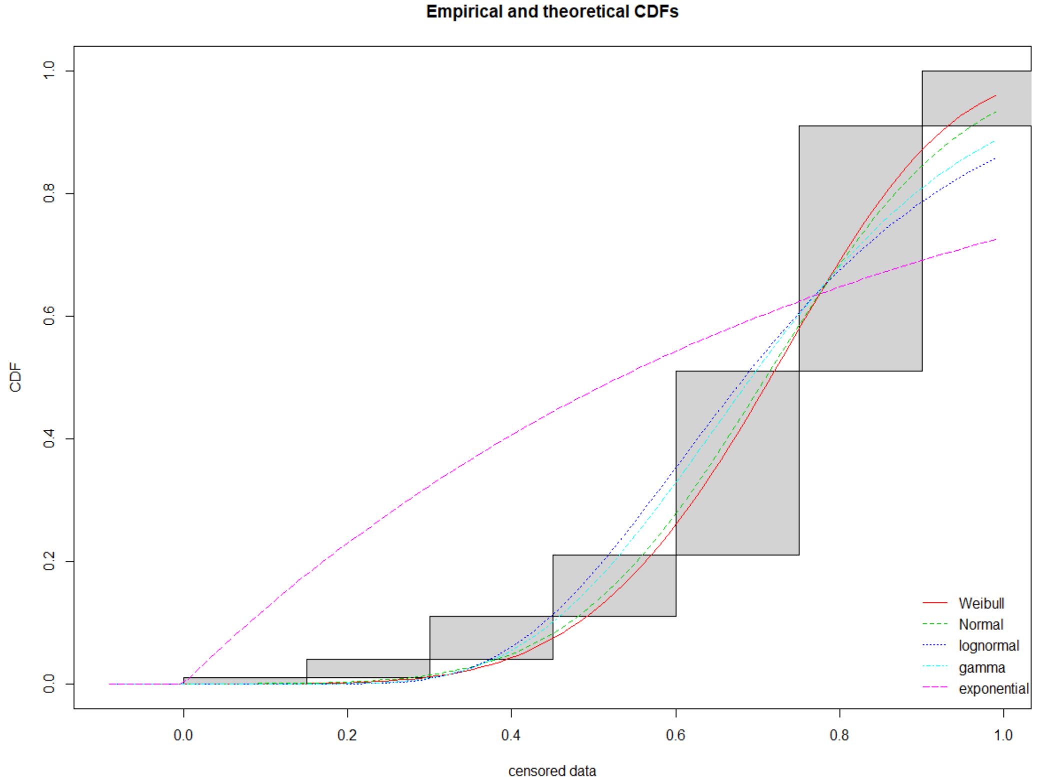
Figures 8 and 9: Fitting Expert 4 opinion – wheels



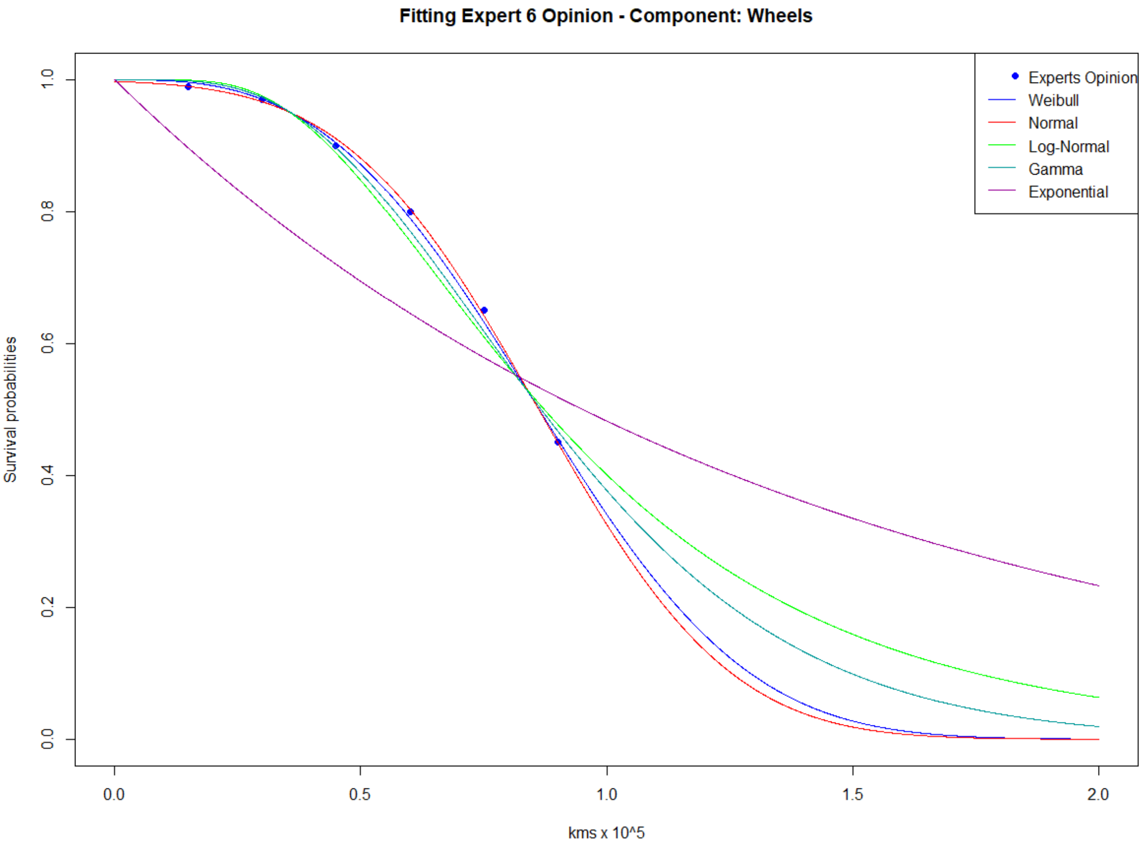


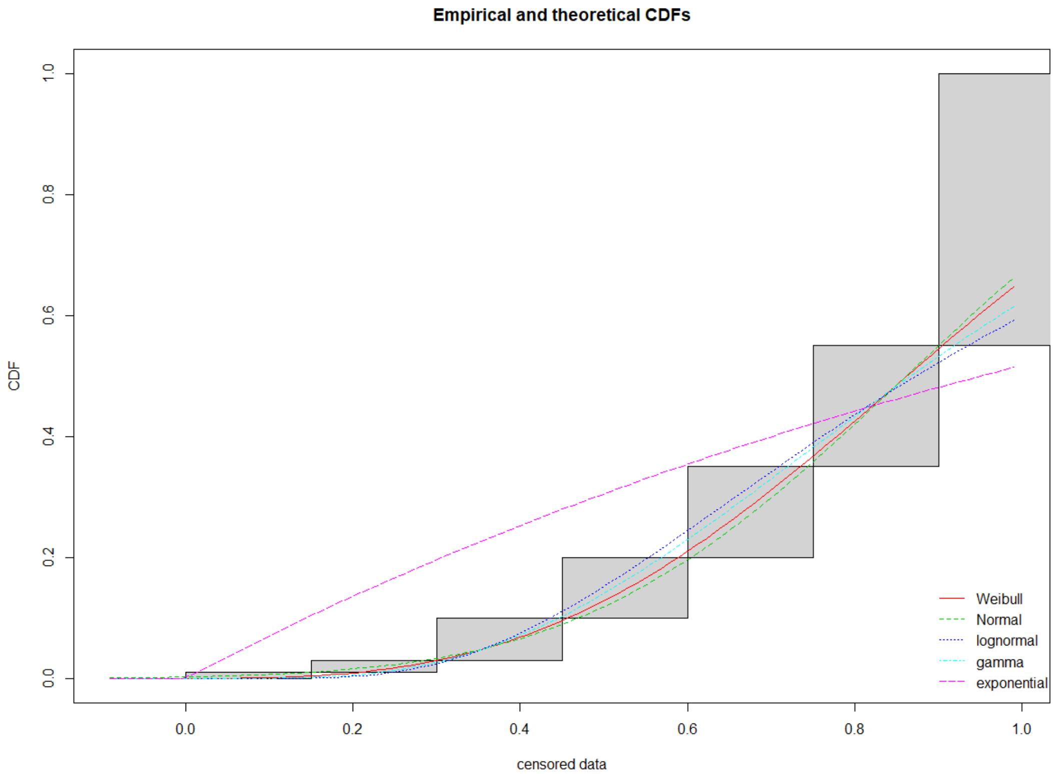
Figures 10 and 11: Fitting Expert 5 opinion – wheels





Figures 12 and 13: Fitting Expert 6 opinion – wheels





In order to compare and decide which distribution best fits the opinion, a goodness-of-fit test with the likelihood function was performed. The likelihood function, commonly known in statistics, measures the goodness-of-fit of a statistical model, in this case of a distribution to a data sample, which in this case is the expert's opinion. The higher the value of the log-likelihood function, the best the distribution fits the data sample. In addition, a second criterion to compare the goodness-of-fit was used, namely the Akaike information criterion (AIC). The AIC estimates the quality of each model, given a collection of models for the given data. The quality of each model is estimated relative to each of the other models. A good model is the one that has minimum AIC among all the other models. Consequently, a lower AIC value indicates a better fit.

In Tables 10 and 11, one can verify the log-likelihood values as well as the AIC values for each distribution in each expert's opinion fitting.

Table 10 and 11: Log-likelihood and AIC value for each expert's opinion fitting

	Log-likelihood					
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6
Weibull	-1752.54	-1325.09	-1688.09	-862.71	-1594.49	-1513.30
Normal	-1768.13	-1341.58	-1696.93	-864.38	-1609.03	-1511.90
Gamma	-1759.01	-1325.40	-1697.03	-864.32	-1722.43	-1527.04
Lognormal	-1775.26	-1330.69	-1715.27	-868.33	-1799.34	-1547.84
Exponential	-2010.37	-1421.24	-1950.86	-965.60	-2395.86	-1767.56

	AIC					
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6
Weibull	3509.07	2654.18	3380.18	1729.42	3192.99	3030.61
Normal	3540.25	2687.15	3397.87	1732.75	3222.07	3027.80
Gamma	3522.01	2654.79	3398.06	1732.65	3448.87	3058.07
Lognormal	3554.52	2665.38	3434.54	1740.66	3602.68	3099.68
Exponential	4022.73	2844.47	3903.71	1933.19	4793.71	3537.13

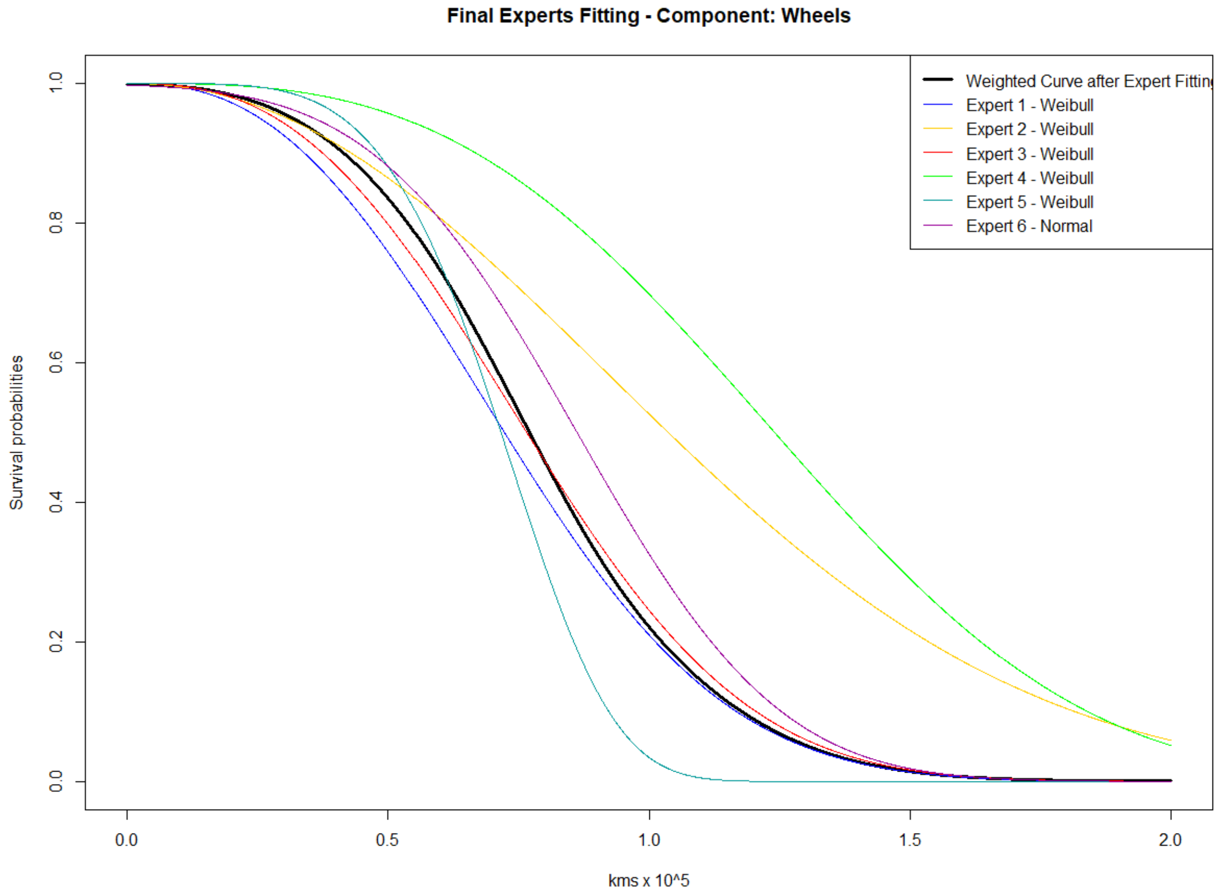
Overall, for Experts 1 to 5, the distribution that best fits the data is the Weibull distribution, due to higher log-likelihood values on one hand, and to low AIC values on the other hand. For Expert 6, the best distribution is the normal distribution.
For each of the chosen distributions, the parameters were determined (Table 12).

Table 12: Distribution parameters for each expert opinion fit - Wheels

Expert 1		Expert 2		Expert 3		Expert 4		Expert 5		Expert 6	
Weibull		Weibull		Weibull		Weibull		Weibull		Normal	
Shape	Scale	Shape	Scale	Shape	Scale	Shape	Scale	Shape	Scale	Mean	SD
2.49365	0.83719	2.14123	1.22972	2.64162	0.8788	3.04383	1.399	4.73445	0.77342	0.86147	0.30596

All the scale variables, as well as the mean variables, are in 10^5 km.
Having a distribution fitted to each expert's opinion, there is a need to develop a final curve which represents the know-how of each expert combined. Therefore, in order to create a final curve, each weight of each expert is multiplied by the corresponding distribution function, as one can see in Figure 14.

Figure 14: Weighted curve after fitting expert opinion - wheels



One can identify that the experts with the highest weights have the most impact in the final curve, like Expert 6 and Expert 5, where the values and slopes of the final curve take similarities from these experts' opinions.

<u>Approach 2:</u>

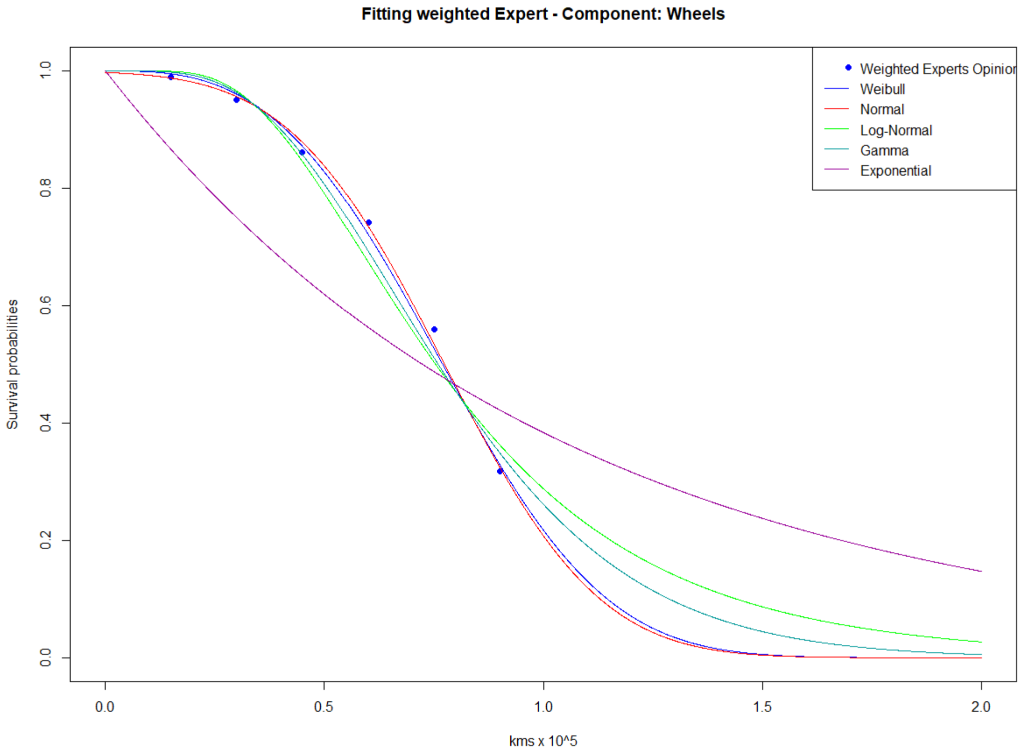
For the second approach, a combined weighted opinion was first obtained, following a survival analysis of the combined opinion. The combined weighted opinion is obtained by multiplying each expert's weight to each opinion in every interval. The combined expert opinion can be seen in Table 13.

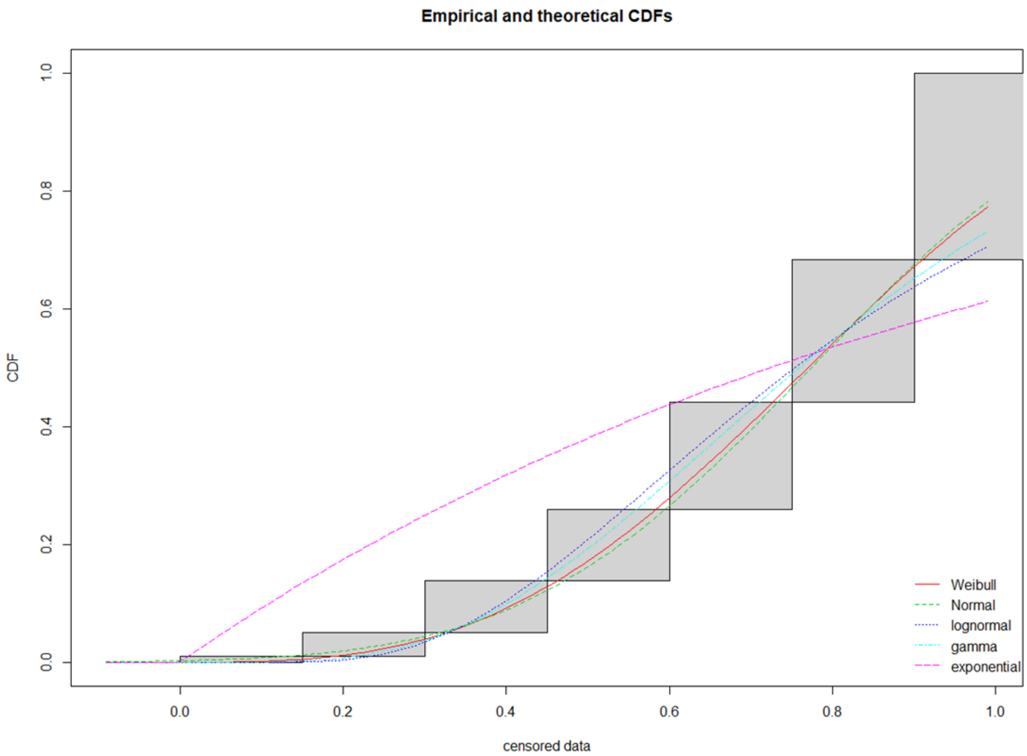
Table 13: Combined weighted expert opinion - wheels

	[0 , 15] 10^3 km	[15 , 30] 10^3 km	[30 , 45] 10^3 km	[45 , 60] 10^3 km	[60 , 75] 10^3 km	[75 , 90] 10^3 km	[90 , +∞] 10^3 km	Sum
Combined Expert	10	40	89	121	182	242	316	1000

Like on the first approach, a survival analysis on the combined weighted opinion is performed and 5 probability distributions are considered. In Figures 15 and 16, one can verify the 5 fitted distributions as well as the combined expert opinion.

Figures 15 and 16: fitting the combined weighted expert opinion





Likewise, a log-likelihood test is performed, and both the log-likelihood and the AIC values are obtained, in order to get the 'best' model to the sample data. Table 14 describes the values obtained in the goodness-of-fit tests.

Table 14: Log-likelihood and AIC value for the combined weighted expert opinion fitting - wheels

Log-likelihood	
	Combined Expert
Weibull	-1673.156
Normal	-1672.078
Gamma	-1696.812
Lognormal	-1726.253
Exponential	-2008.391

AIC	
	Combined expert
Weibull	3350.312
Normal	3348.156
Gamma	3397.623
Lognormal	3456.506
Exponential	4018.783

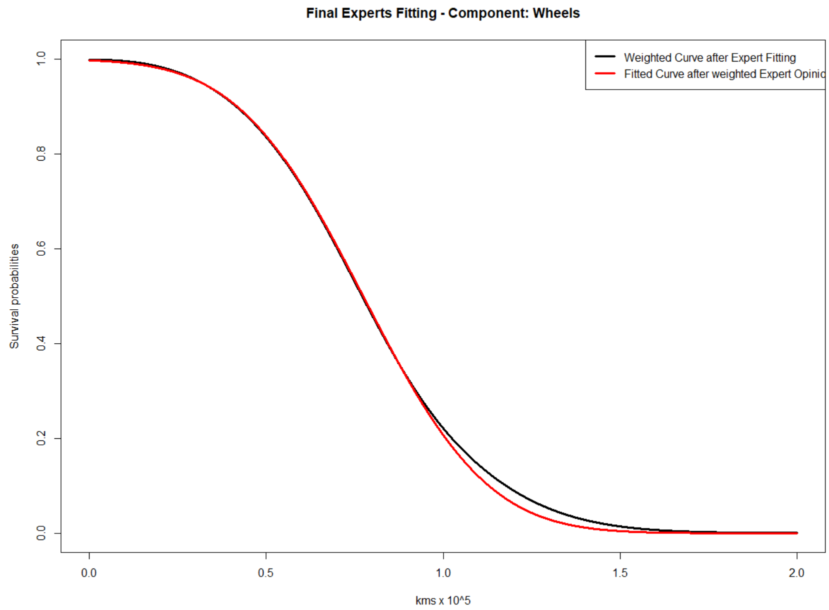
From the Log-likelihood and AIC values, the most suitable distribution to fit the data is the Normal distribution. The distribution parameters are given in Table 15.

Table 15: distribution parameters for the combined weighted expert opinion - wheels

Normal	
Mean	SD
0.774	0.277

Once the best distribution is chosen, a comparison is drawn between Approach 1 and Approach 2, concerning the distribution that best describes the wheel failure in real case scenarios.
Therefore, both distributions are represented (Figure 17), to identify and compare the similarities or divergences.

Figure 17: Comparison between Approach 1 (black) and Approach 2 (red) results



Although the curves are very similar, the red curve (Approach 2) shows lower survival probabilities for high values of kilometres. This means that the probability of the component not to fail for a high value of kilometres is lower than the probability of the black curve. Therefore, the red curve is more conservative, since it is safer in terms of maintenance to think a component is going to fail earlier. Therefore, failures rates will be estimated

based on the red curve.
Finally, by choosing the red curve, one has to calculate with the given distribution its mean value in order to find the mean distance between failure (MDBF). Since the red curve is a normal distribution, its mean is given by:

$$\hat{\mu} = 0.77375x10^5\ km = MDBF$$

With the MDBF, the failure rate is easily obtained with:

$$f_1 = \frac{1}{MDBF} = 1.2924\ x\ 10^{-5}\frac{1}{km} \quad (14)$$

6.3.2.2. Axle

For the second component, Table 16 summarizes the expert's assessment in each interval mentioned in the subsection 2.2.1.

Table 16: Experts assessments on the axle for each interval

	[0, 300] 10^3 km	[300, 600] 10^3 km	[600, 900] 10^3 km	[900, 1200] 10^3 km	[1200, 1500] 10^3 km	[1500, 1800] 10^3 km	[1800, +∞] 10^3 km	Sum
Expert 1	0	0	5	5	5	10	25	1000
Expert 2	0	2	4	6	8	10	970	1000
Expert 3	0	2	5	10	15	20	948	1000
Expert 4	20	80	150	250	300	150	50	1000
Expert 5	10	50	150	250	400	100	40	1000
Expert 6	2	18	40	100	120	150	570	1000

Like for the first component, for each given interval, each expert assessed its opinion on how many components will fail of the batch of 1000 components. Similarly, two approaches were performed to analyse which of these would best fit the expert's assessments and weights.

<u>Approach 1:</u>

Once again, a survival analysis was performed on the opinion of the expert. For each of the 6 experts, one can verify in Annex A the distribution fitting obtained.

Likewise, the distribution comparison was done with a goodness-to-fit test and Tables 17 and 18 present the log-likelihood and the AIC values obtained.

Table 17 and 18: Log-likelihood and AIC value for each expert's opinion fitting - axle

	Log-likelihood					
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6
Weibull	-152.72	-179.91	-277.19	-1721.19	-1586.19	-1305.05
Normal	-154.12	-181.66	-278.94	-1718.28	-1584.86	-1310.45
Gamma	-152.61	-179.83	-277.06	-1774.63	-1640.92	-1306.16
Lognormal	-152.34	-179.67	-277.00	-1830.28	-1690.66	-1311.95
Exponential	-161.84	-188.67	-298.18	-2278.22	-2308.23	-1506.89

	AIC					
	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6
Weibull	309.44	363.81	558.39	3446.38	3176.37	2614.10

Normal	312.25	367.32	561.88	3440.57	3173.73	2624.91
Gamma	309.21	363.66	558.13	3553.26	3285.85	2616.32
Lognormal	308.69	363.34	558.01	3664.55	3385.31	2627.89
Exponential	325.68	379.35	598.35	4558.43	4618.45	3015.78

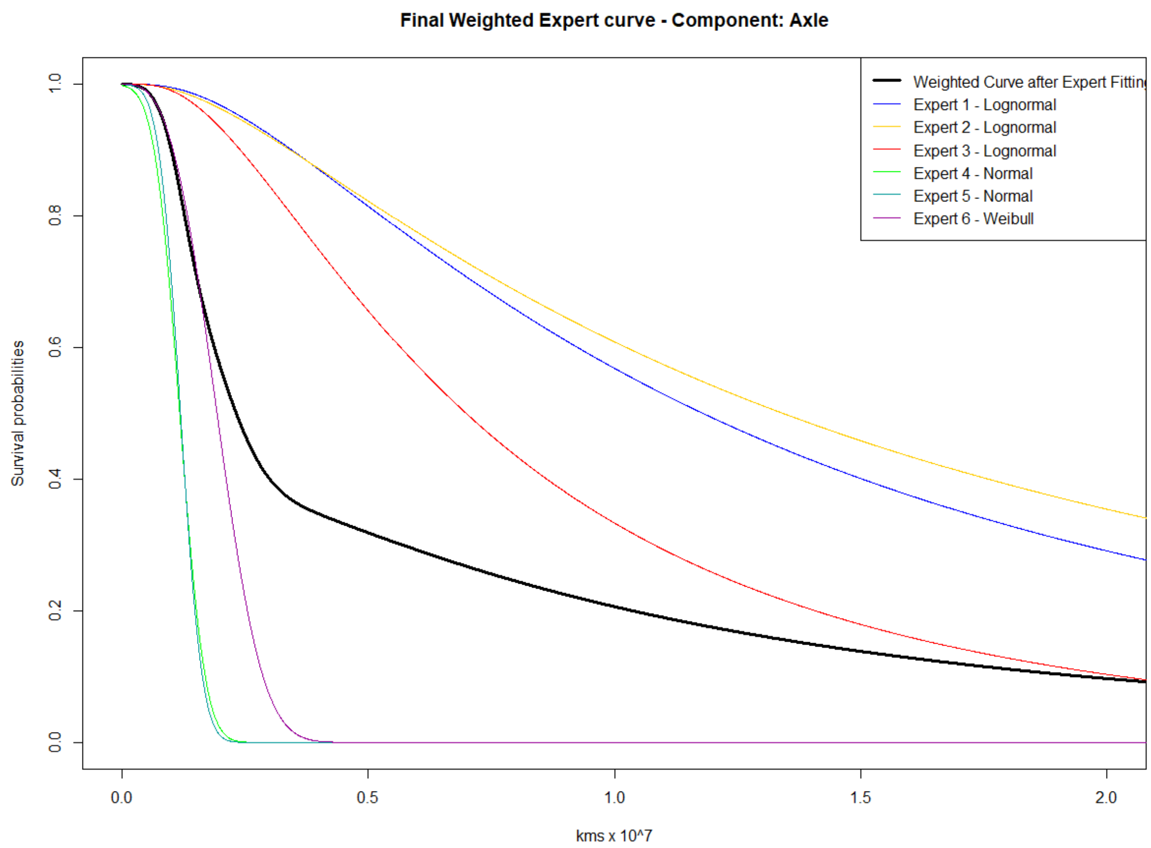
As shown in Tables 17 and 18, for Experts 1, 2 and 3 the Lognormal distribution is the distribution that best fits each expert's opinion. For Experts 4 and 5, the Normal distribution is the most suitable. Finally, for Expert 6 the best distribution is the Weibull distribution.
For each of the chosen distributions, the parameters are determined and presented in Table 19.

Table 19: Distribution parameters for each expert opinion fit - Axle

Expert 1		Expert 2		Expert 3		Expert 4		Expert 5		Expert 6	
Lognormal		Lognormal		Lognormal		Normal		Normal		Weibull	
Meanlog	SDlog	Meanlog	SDlog	Meanlog	SDlog	Mean	SD	Mean	SD	Shape	Scale
0.16467	0.95948	0.2939	1.06807	-0.35963	0.83351	0.11669	0.040669	0.11669	0.04067	3.02929	0.21748

All the scale variables, as well as the mean variables, are in 10^7 km. After the fitting process, a final curve was created with regard to each of the expert's performance in the calibration questions. The final curve, with the expert's weights, is represented in Figure 18.

Figure 18: Weighted curve after fitting expert opinion – axle



Since the expert's opinions diverge a lot, the final weighted curve is formed by a combination of expert 6, expert 1, and expert 5 slopes. Here, we can confirm the impact of a high weighted expert.

Approach 2:

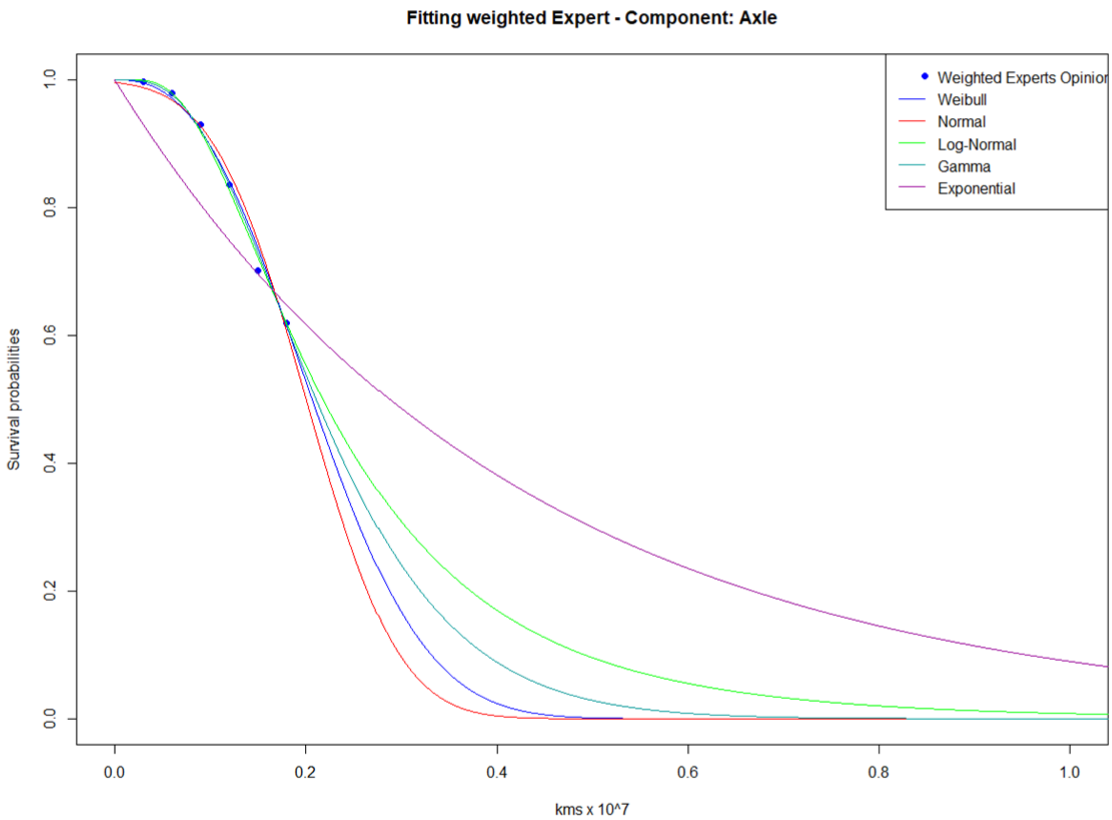
In the same way as the first component, for this second approach, a combined weighted opinion was first obtained, followed by a survival analysis of the combined opinion. For the axle, the combined expert opinion can be observed in Table 20.

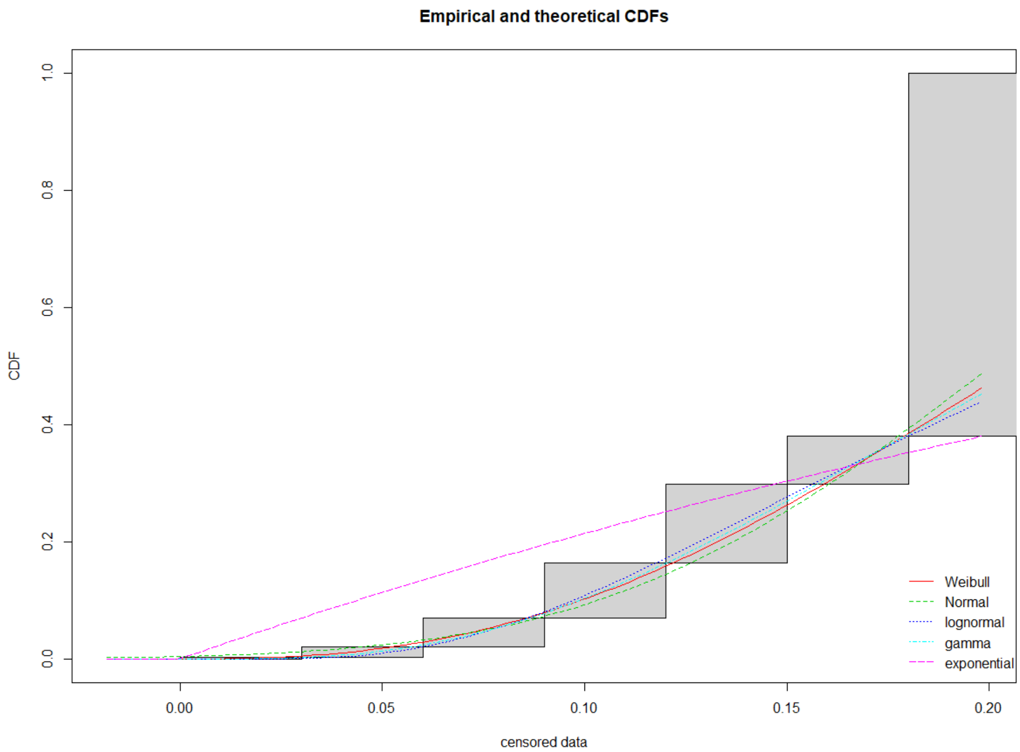
Table 20: combined weighted expert opinion - axle

	[0, 300] 10^3 km	[300, 600] 10^3 km	[600, 900] 10^3 km	[900, 1200] 10^3 km	[1200, 1500] 10^3 km	[1500, 1800] 10^3 km	[1800, +∞] 10^3 km	Sum
Combined Expert	3	18	49	94	134	82	620	1000

Again, survival analysis is performed to the combined experts' opinions and 5 statistical distributions were obtained. In Figure 19 and 20, the plot of the 5 fitted statistical distributions, as well as the combined expert opinion, are shown.

Figures 19 and 20: Fitting the combined weighted experts' opinions – axle





Like on the first component, a log-likelihood test is performed, and both the log-likelihood and the AIC values are taken. The goodness-of-fit values are presented in Table 21.

Table 21: Log-likelihood and AIC value for the combined weighted expert opinion fitting - axle

Log-likelihood	
	Combined Expert
Weibull	-1244.474
Normal	-1258.268
Gamma	-1240.023
Lognormal	-1239.683
Exponential	-1378.015

AIC	
	Combined Expert
Weibull	2492.948
Normal	2520.536
Gamma	2484.045
Lognormal	2483.366
Exponential	2758.03

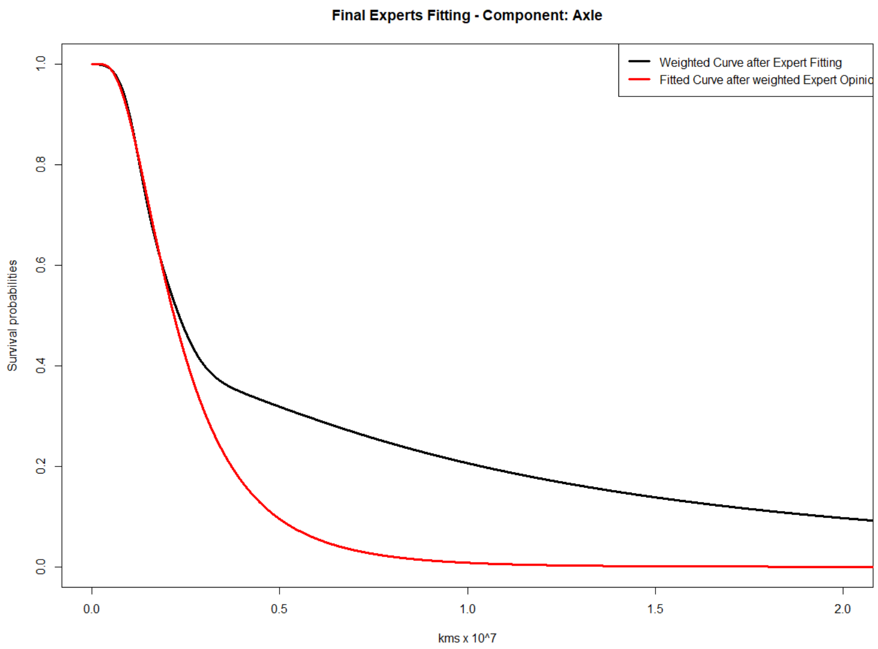
From the Log-likelihood and AIC values, the most suitable distribution to fit the data is the lognormal distribution. The estimated parameters are given in Table 22.

Table 22: Distribution parameters for the combined weighted expert opinion – axle

Lognormal	
Meanlog	SDlog
-1.5219	0.63316

Having assessed both statistical distributions for both the approaches, a comparison is made to determine which of the distributions is more conservative. Therefore, both distributions are represented, to identify and compare the similarities or divergences (Figure 21).

Figure 21: Comparison between approach 1 (black) and approach 2 (red) result



Once again, the red curve (approach 2) is more conservative, i.e. for higher values of kilometres, the axle has lower survival probabilities than the black curve. Consequently, the red curve is assumed to be the final distribution function chosen to best describe the case study under analysis.

Lastly, and after selecting the red curve, one has to estimate the mean value to find the MDBF and therefore to find the failure rate. For a lognormal distribution, the mean value is given by the following formula:

$$mean = \exp\,(\mu + \frac{\sigma^2}{2})$$

Where μ is the meanlog and σ the sdlog. The mean value is:

$$mean = \exp\left(-1.5219 + \frac{0.63316^2}{2}\right) = 0.266748\, x10^7\ km = MDBF$$

Finally, the failure rate is obtained with the MDBF:

$$f_2 = \frac{1}{MDBF} = 3.749\ x\ 10^{-7}\ km^{-1} \tag{15}$$

With both failure rates estimated, one can consolidate the FMEA analysis with a more realistic and therefore conservative case-scenario.

7. FMECA Analysis (consolidating FMEA with Expert Judgment)

After performing a FMEA analysis in the first chapter in order to identify and prioritise the most relevant subsystems and components of the Bogie system, the second chapter conducted an expert judgment to obtain failure rates of two components. This third chapter combines these two analyses to have a better assessment of a real case scenario and performs a criticality analysis, to rank the most critical components and subsystems.

Nevertheless, to combine these two analyses, one must first modify the failure rate units that resulted from the expert judgment. The average cargo locomotive speed was assumed to be equal to 40km/h. Therefore, with a given MDBF the units conversion is obtained with the following formula:

i. $$\mathrm{MTBF_1} = \frac{\mathrm{MDBF_1}}{\frac{40\ \mathrm{km}}{\mathrm{h}}} = 1933.93h \rightarrow \lambda_1 = \frac{1}{\mathrm{MTBF_1}} = 5.171x10^{-4}\frac{1}{h}$$

ii. $$\mathrm{MTBF_2} = \frac{\mathrm{MDBF_2}}{\frac{40\ \mathrm{km}}{\mathrm{h}}} = 66687h \rightarrow \lambda_2 = \frac{1}{\mathrm{MTBF_2}} = 1.5x10^{-5}\frac{1}{h}$$

Considering the occurrence of the FMEA analysis in Table 2, one can verify that for component 1 (wheels), the impact would be ranked with number 8, meaning the occurrence is high with repeating failures in a short cycle. Therefore, and having in mind that every wheel failure mode is ranked with an 8 or higher in terms of severity and detectability (Tables 3 and 4, respectively), one can assume that the component is critical. For the axle, the occurrence with the new failure rate would be ranked with a 7, also high. Bearing in mind that the axle has a severity of 10, this component should always be considered critical, since a failure could bring possible fatalities.

In order to perform a criticality analysis, one has to calculate first the modal criticality and then the item's criticality. In the latter, the items considered, are the assumed critical subsystems analysed. For this purpose, the following formulas were considered:

i. $$C_m = \lambda \times S \times \beta \times O \qquad (16)$$

ii. $$C_i = \sum_i^n C_m \qquad (17)$$

Where: λ is the failure rate of each failure mode, S is the severity number associated with each failure mode, β the failure rate of the effect of each failure mode, and O the operating hours of each failure mode. Considering that in this project the failure rate of the effect was not assumed due to lack of information on the failure modes effects and that the operating hours are the same for each component and therefore for each failure mode, β and O are assumed to be 1.

Following this and considering the critical subsystems, with its critical components and the associated failure modes that resulted from the FMEA analysis, one can adapt Table 5 and insert the new failure rates estimated using the expert judgment techniques and the severity numbers (Table 23). It is important to emphasize that the severity numbers linked to the components and failure modes not mentioned in the FMEA analysis, were obtained from the literature.

Table 23: Critical components based on the consolidated FMEA with Expert Judgment

Subsystem ID	Subsystem	Component	Component ID	Failure Mode	Severity	Failure rate (1/h))	Source
1	Wheelset	Axle	1.1	Axle Crack	10	1.5E-05	Expert Judgment
		Wheels	1.2	(Wheel out of round, Wheel cracks and notches, wheel build up material, wheel flat, profile under threshold)	8	5.171E-04	Expert Judgment
		Bearings	1.3	-	9	2.12E-06	FMEA
2	Axle Box	Axle Box	2.1	Absence of the cover box screw	8	6.00E-05	FMEA
		Axle Box	2.1	Housing not watertight	8	1.20E-04	FMEA
		Axle Box	2.1	Bearing Failure	10	2.12E-06	Literature
3	Bogie Frame	Frame	3.1	-	9	1.18E-05	Literature
4	Brake System	Brake	4.1	parts of brake rigging hanging	8	2.01E-05	FMEA
		Brake	4.1	Brake isolating cock	8	2.01E-05	FMEA
		Brake	4.1	Cast Iron Brake Block	6	1.08E-04	FMEA
		Brake	4.1	Composite Brake Block	6	3.12E-05	FMEA
		Pneumatic Braking system	4.2	Front air valve damaged	10	6.00E-05	FMEA
		Pneumatic Braking system	4.2	Brake cylinder damaged	6	6.00E-05	FMEA
		Pneumatic Braking system	4.2	Air distributor damaged	6	3.00E-04	FMEA
		Pneumatic Braking system	4.2	Slack adjuster damaged	8	2.40E-04	FMEA
		Master/Auxiliary Compressor	4.2	-	9	1.09E-04	Literature
		Master/Auxiliary Compressor Driving Motor	4.3	-	9	2.60E-05	Literature
		Servo-motor in braking system	4.5	-	9	8.76E-06	Literature
		Other Elements of the pneumatic braking system	4.6	-	9	1.92E-04	Literature
		Other Elements of the braking system (pins, sleeves,…,)	4.7	-	9	1.28E-04	Literature
5	Suspension Elements	Spring Buckle	5.1	Spring Buckle Fracture	10	6.00E-05	FMEA
		Helical Spring	5.2	Helical Spring broken	10	6.00E-05	FMEA
		Other Suspension elements	5.4	Bottoming between Axle-box housing and bogie frame	10	1.44E-06	FMEA
6	Electric Traction Module	Power transmission system	6.1	-	9	3.99E-04	Literature
	Electric Traction Module	Shaft Coupling	6.2	-	9	6.98E-05	Literature
	Electric Traction Module	Traction Motor	6.3	-	9	7.82E-06	Literature

As can be verified, all failure modes from the wheels were aggregated to a general failure mode. This general failure mode has a higher combined failure rate than the failure modes themselves. Therefore, a more conservative and realistic scenario was analysed, which leads to a better criticality assessment of the Bogie. After performing the criticality calculations, the Table 24 was obtained.

Table 24: Criticality Analysis

Subsystem ID	Subsystem	Component	Component ID	C_m	C_i	Ranking
1	Wheelset	Axle	1.1	1.50E-04	**4.3059E-03**	**2**
		Wheels	1.2	4.14E-03		
		Bearings	1.3	1.91E-05		
2	Axle Box	Axle Box	2.1	4.80E-04	**1.4612E-03**	**4**
		Axle Box	2.1	9.60E-04		
		Axle Box	2.1	2.12E-05		
3	Bogie Frame	Frame	3.1	1.06E-04	**1.0620E-04**	**6**
4	Brake System	Brake	4.1	1.61E-04	**1.0011E-02**	**1**
		Brake	4.1	1.61E-04		
		Brake	4.1	6.48E-04		
		Brake	4.1	1.87E-04		
		Pneumatic Braking system	4.2	6.00E-04		
		Pneumatic Braking system	4.2	3.60E-04		
		Pneumatic Braking system	4.2	1.80E-03		
		Pneumatic Braking system	4.2	1.92E-03		
		Master/Auxiliary Compressor	4.2	9.81E-04		
		Master/Auxiliary Compressor Driving Motor	4.3	2.34E-04		
		Servo-motor in braking system	4.5	7.88E-05		
		Other Elements of the pneumatic braking system	4.6	1.73E-03		
		Other Elements of the braking system (pins, sleeves,...,)	4.7	1.15E-03		
5	Suspension Elements	Spring Buckle	5.1	6.00E-04	**1.2144E-03**	**5**
		Helical Spring	5.2	6.00E-04		
		Other Suspension elements	5.4	1.44E-05		
6	Electric Traction Module	Power transmission system	6.1	3.59E-03	**4.2896E-03**	**3**
		Shaft Coupling	6.2	6.28E-04		
		Traction Motor	6.3	7.04E-05		

Based on the criticality analysis, a consolidated ranking of the most critical subsystems is obtained by ordering the subsystem with the highest combined C_i score. The following list ranks the most critical subsystems:

1. Brake System
2. Wheelset components
3. Electric Traction Module
4. Axle Box
5. Suspension System
6. Bogie Frame

Intuitively, it can be verified that one of the main reasons the braking system is considered to be the most critical subsystem is due to excessive failure modes linked to its components and the information obtained for this subsystem.

By identifying the most critical subsystems, it is possible to implement risk mitigation strategies, to optimize the operation and the cost associated with the maintenance of the bogie. The common risk mitigation strategies to decrease severity and occurrence, and increase detectability are the following:

1. Implement redundancy to reduce the risk of losing the function (decrease Occurrence);
2. Apply specific test in simulated operating conditions to check the reliability of a component (decrease Occurrence and increase detectability);
    i. Creation of a functional simulation model that simulates the real-time condition of the components and subsystems combined, according to the operation of FGC;
3. Increase the frequency of inspections (decrease Occurrence and increase detectability);
4. Change the maintenance type to predictive maintenance, monitoring the condition of the components (decrease Occurrence and increase detectability) – by implementing sensors:
    i. Monitoring of bogie stability through the implementation of sensors (e.g. accelerometers) that are able to monitor the movement of each bogie and identify situations/conditions which might increase the risk of derailment;
    ii. Axle-box monitoring using vibration and temperature sensors to detect any unusual behaviour;
    iii. Vibrations and temperature sensors for monitoring any unusual behaviour of the electric engines.
5. Apply specific test to ensure maintainability of components that require a long time to repair (decrease Severity);
    i. Control with sensors the real-time of failure of the most critical components;
6. Prepare specific training and procedures to allow falling back to a safe degraded mode in an emergency (decrease Severity)
    i. providing intermediate system repair to the most critical subsystems;
7. Keep spares on-site so that time to repair is shortened (decrease Severity).

A summary of all the previous strategies and their impact on the three indexes mentioned in section 1.1 are presented in Table 25.

Table 25: Risk Mitigation Strategies and their impact on the Severity, Occurrence, and Detectability

	Decrease Severity	Decrease Occurrence	Increase Detectability
Strategy 1		X	
Strategy 2		X	X
Strategy 3		X	X
Strategy 4		X	X
Strategy 5	X		
Strategy 6	X		
Strategy 7	X		

After extensive analysis to decide which strategy one needs to implement in the project, a combination of some of these was obtained in order to follow the path of the project of providing a continuous monitoring system.

By starting at strategy 1, one can assume that this strategy is not appropriate in an already operating cargo locomotive since this strategy is usually implemented in a design phase of a product. Strategy 2, which mentions the implementation of specific tests in order to monitor the reliability of the system in real-time, provides a good continuous monitoring system. Therefore, this strategy is aligned with the project and taken into consideration. Strategies 3 and 6, which focus on an increasing frequency of inspections and intermediate repairs, are exactly one of the key tasks to eradicate in the project, since FGC wants to reduce the number of (potentially unnecessary) inspections and intermediate system repairs for each cargo locomotive. In fact, such strategies are considered to be an output of a continuous monitoring system. By implementing a predictive maintenance type, as it is specified in strategy 4 and taken into consideration for the project, the inspection frequency is increased by introducing sensors, which will trigger unusual behaviours on real-time conditions of the most critical components. Concisely, a continuous monitoring system provides a remote real-time inspection frequency and therefore one can predict when is suitable to provide a system repair regarding the condition of the component or subsystem. In addition to strategy 4, strategy 5 ensures the maintainability of the components with high severity numbers by employing sensors. Once again, this strategy is aligned with the goal of having a continuous monitoring system. Finally, strategy 7 is associated with strategies 3 and 6 since this strategy is an output of a continuous monitoring system. With a continuous monitoring system, it is possible to predict the failure of a component and therefore plan beforehand the number of spare parts to have on-site.

To conclude, a continuous monitoring system, which is the goal of the LOCATE project, is obtained with a combination of strategies 2 to 5. This enables to reduce the occurrence of several failure modes, since these are being monitored and one can predict the most advantageous time to replace or repair the component, increase the detectability of critical failure modes, by increasing the probability of detecting the failure mode before it turns critical with abnormal behaviours, and to decrease the severity, by providing condition-based repairs to the most critical components.

8. Conclusions

This deliverable presents the activities performed in Task 2.3 of the LOCATE project. After performing a FMEA analysis, based on references and literature in Section 1, an expert judgment assessment was carried out in Section 2 to obtain information about the failure rates and survival probabilities for two components identified as critical in the FMEA Analysis. In Section 3, a criticality analysis was completed, in order to consolidate the new failure rates with the information of the FMEA analysis and therefore assess the most critical subsystems. Finally, risk mitigation strategies regarding the most critical subsystems are proposed, whereas a combination of these is considered in order to implement a continuous monitoring system, which is the goal of the LOCATE project.

With this, a generic FMECA methodology is obtained, and one can relate these outputs with the previous deliverables and highlight the potential benefits of the FMECA analysis for supporting the definition of Risk Mitigation Strategies. Such risk mitigation strategies lead to the definition of a predictive maintenance system for the locomotive bogies with the installation of sensors (to be developed in WP3), modelling of dynamic theoretical behaviour (to be developed in WP4) and operational behaviour and maintenance scheduling models (to be developed in WP5).

References

[1] M. G. Park, “RAMS management of railway systems,” no. August, p. 260, 2014.

[2] M. Szkoda and G. Kaczor, “Application of FMEA Analysis to Assess the Safety of Rail Vehicles,” *23rd Int. Symp.*, no. June, pp. 1–9, 2015.

[3] M. Szkoda and G. Kaczor, “Analiza rams w cyklu istnienia kolejowych środków transportu,” *J. Konbin*, vol. 41, no. 1, pp. 83–106, 2017.

[4] M. Szkoda and M. Satora, “The application of failure mode and effects analysis (FMEA) for the risk assessment of changes in the maintenance system of railway vehicles,” *Czas. Tech.*, no. 8, pp. 159–172, 2019.

[5] BS EN:50126, “Railway Applications - The Specification and Demonstration and Reliability, Availability, Maintainability and Safety (RAMS) - Part 1: Generic RAMS Process,” *Bs En 50126-1 2017*, p. 106, 2017.

[6] BS EN 60812, “EN 60812: Analysis techniques for system reliability – Procedure for failure mode and effects analysis (FMEA),” p. 98, 2006.

[7] C. Ulianov and P. Hyde, “Innovative Monitoring and Predictive Maintenance Solutions on Lightweight Wagon,” p. 114ff, 2017.

[8] A. J. Butler, M. K. Thomas, and K. D. M. Pintar, “Systematic review of expert elicitation methods as a tool for source attribution of enteric illness,” *Foodborne Pathog. Dis.*, vol. 12, no. 5, pp. 367–382, 2015.

[9] P. Taylor *et al.*, “Journal of the American Statistical Association Statistical Methods for Eliciting Probability Distributions Statistical Methods for Eliciting,” no. August 2013, pp. 37–41.

[10] T. R. Anthony O’Hagan, Caitlin E. Buck, Alireza Daneshkhah, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley, *Uncertain Judgements: Eliciting Experts Probabilities*, vol. 170, no. 4. 2007.

[11] L. E. Wood and J. M. Ford, “Structuring interviews with experts during knowledge elicitation,” *Int. J. Intell. Syst.*, vol. 8, no. 1, pp. 71–90, 1993.

[12] R. M. Cooke and L. H. J. Goossens, “Expert judgement elicitation for risk assessments of critical infrastructures,” *J. Risk Res.*, vol. 7, no. 6, pp. 643–656, 2004.

[13] G. Rowe and G. Wright, “The delphi technique as a forecasting tool,” *Int. J. Forecast.*, vol. 2070, no. 99, pp. 353–375, 1999.

[14] A. R. Colson and R. M. Cooke, “Expert elicitation: Using the classical model to validate experts’ judgments,” *Rev. Environ. Econ. Policy*, vol. 12, no. 1, pp. 113–132, 2018.

[15] R. M. Cooke, *Experts in Uncertainty: Opinion and Subjective Probability in Science (Environmental Ethics and Science Policy)*, vol. 44, no. 2. 1991.

[16] European Railway Agency, *Report on Railway Safety and Interoperability in the EU 2018*. 2018.

[17] E. R. Agency, “Safety Overview 2019: Main figures based on CSI Data (up to 2017),” vol. 33, no. 0, p. 368, 1390.

[18] J. Rey, F.- Paris, and A. Bouvresse, “Liste der von der UIC anerkannten Experten zur Gutachtenerstellung von Bremskomponenten Liste des experts reconnus par l ’ UIC pour l ’ élaboration d ’ expertises sur les composants de freinage , List of recognised UIC experts to elaborate expertises on braking components geordnet nach Schwerpunkten – Tätigkeitsfeldern listed according their main area of knowledge Der genaue Umfang der Tätigkeitsfelder und die Adresse des entsprechenden Experten ist der anhängenden Liste zu entnehmen : Le domaine d ’ activité et l ’ adresse des experts se trouvent dans la liste suivante : Their activity areas as well as address and telephone number are indicated in the appended list : Rückfragen richten Sie bitte an / Questions à poser à / Questions to be addressed : UIC Gleitschutzeinrichtungen / Anti-enrayeurs / Wheel-slide protection systems (UIC 541-05),” vol. 33, no. August, pp. 1–26, 2019.

[19] T. Base and T. Base, “RSF Sector Rolling Stock UIC - Rad Fachleute / Specialistes Roue / Wheel experts Liste der Experten die seitens der UIC berechtigt sind die Bewertung von Eisenbahnrädern Mitglieder Räder / Membres Roues / Members Wheels,” vol. 5, 2012.

Appendix

For the experts assessment, in order to obtain the failure rates of the wheels and the axles, the following calibration questions and target questions were formulated:

Calibration Questions:

1. Knowing that the average number per year of Railway Significant Accidents (i.e. accidents which include resulting fatalities and serious injuries) in Europe between 2010-2015 was 2074, how many significant accidents in Europe were there in 2016?

   5%___________ 50%___________ 95%_____________

2. From the significant accidents in 2016, what was the percentage of fatalities and weighted serious injuries (FWSI) per significant accident? [%]

   5%___________ 50%___________ 95%_____________

3. In 2017, there were 1908 significant accidents in Europe. How many of these accidents were caused by derailments of trains?

   5%___________ 50%___________ 95%_____________

4. In 2017, there were in the 28 EU Countries 10026 total precursors. From these total precursors, how many belonged to the *"Broken Wheels and Broken Axles"* type?
   5%___________ 50%___________ 95%_____________

Target Questions:

Wheels:

Based on your experience and knowledge, from a sample of n=1000 locomotive wheels, how many would fail in each interval?

1) 0 - 15.000 kms ___________
2) 15.000kms - 30.000kms ___________
3) 30.000kms - 45.000kms ___________
4) 45.000kms - 60.000kms ___________
5) 60.000kms - 75.000km ___________
6) 75.000kms - 90.000km ___________
7) 90.000kms – infinite ___________

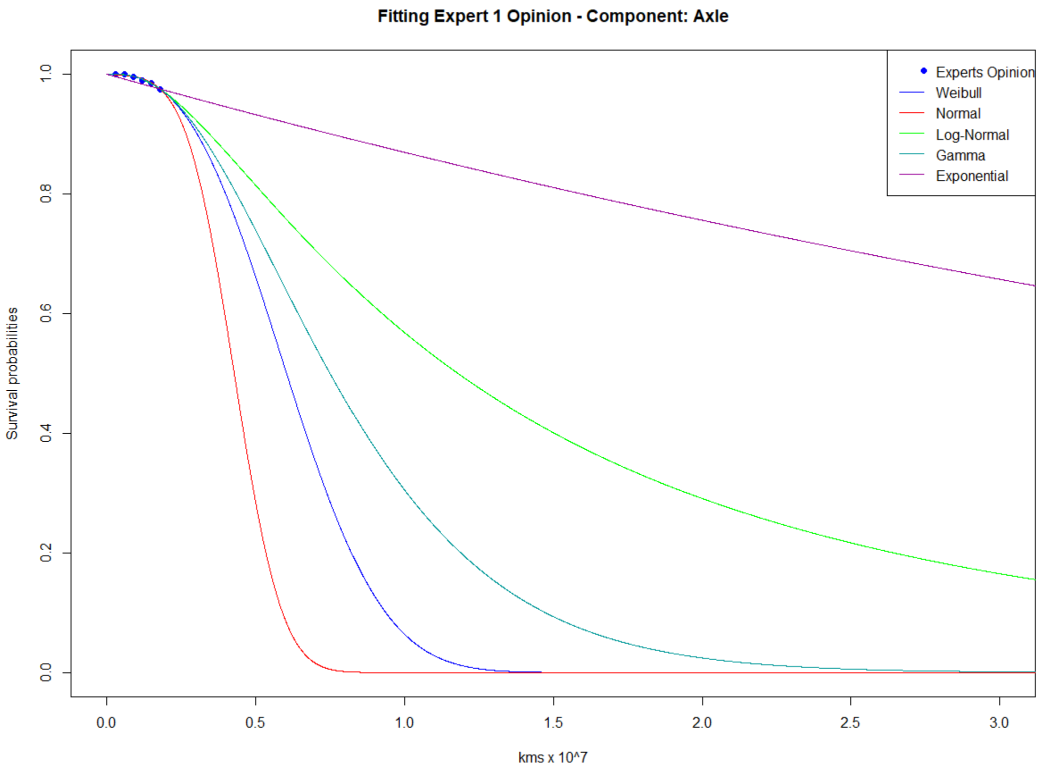
Axle:

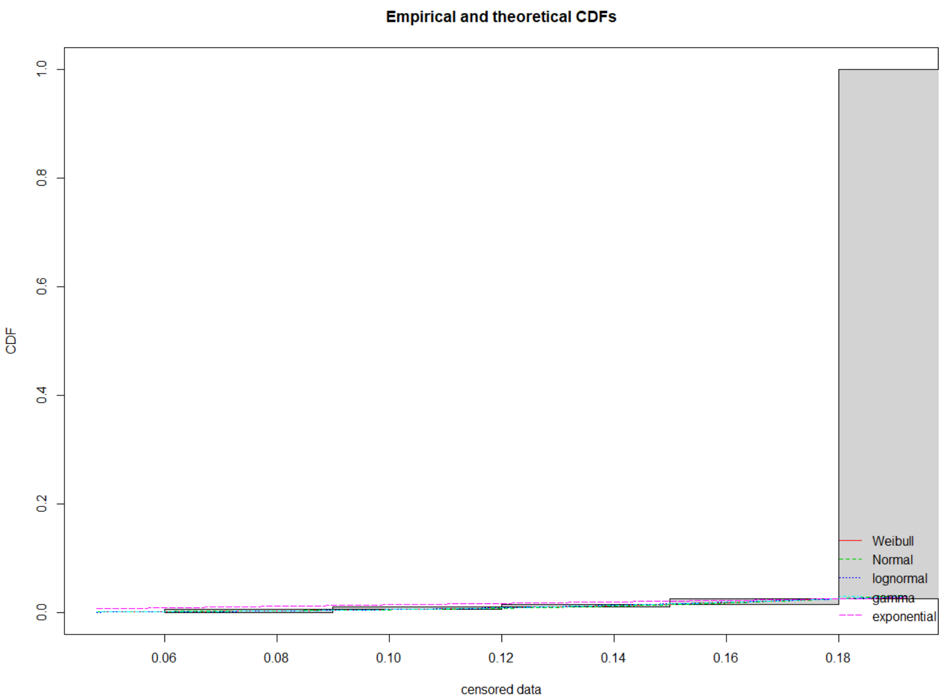
Based on your experience and knowledge, from a sample of n=1000 locomotive axles, how many would fail in each interval?

1) 0 - 300.000 kms ___________
2) 300.000kms – 600.000kms ___________

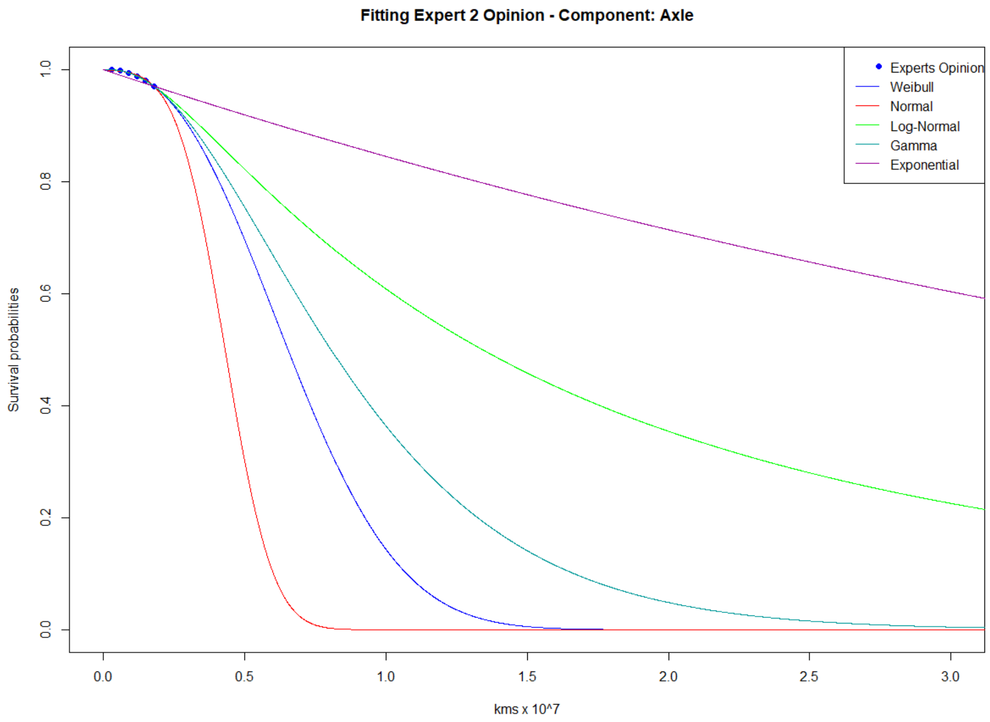
3) 600.000kms - 900.000kms ___________
4) 900.000kms - 1200.000kms ___________
5) 1200.000kms - 1500.000km ___________
6) 1500.000kms - 1800.000km ___________
7) 1800.000kms – infinite ___________

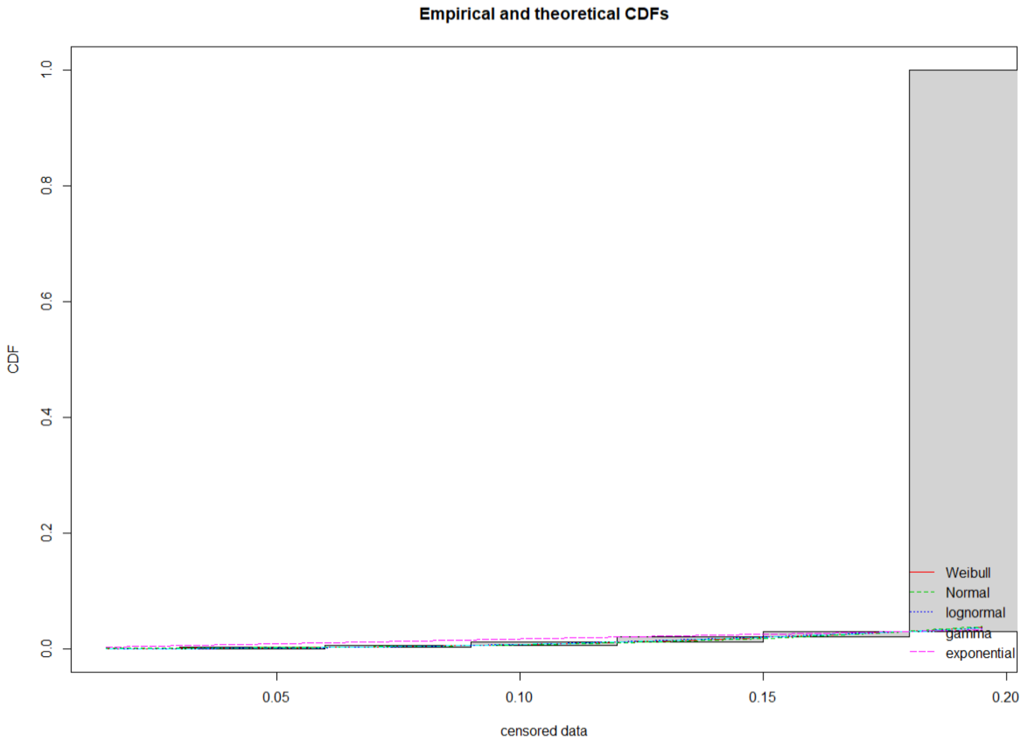
Figures A.1 and A.2: Fitting Expert 1 opinion - axle



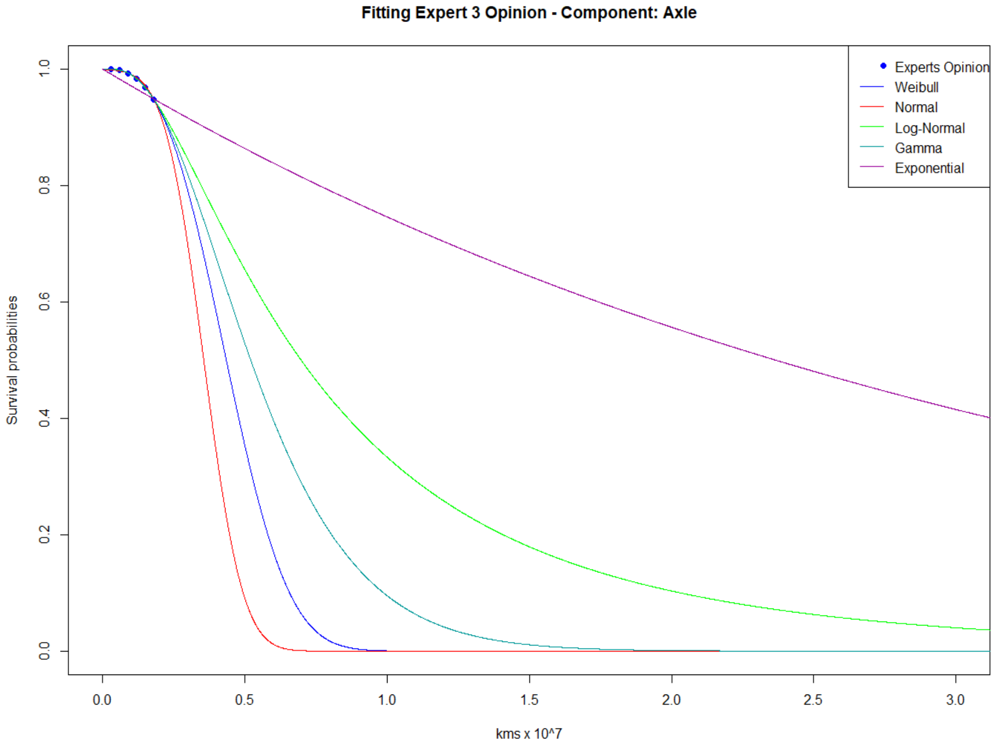


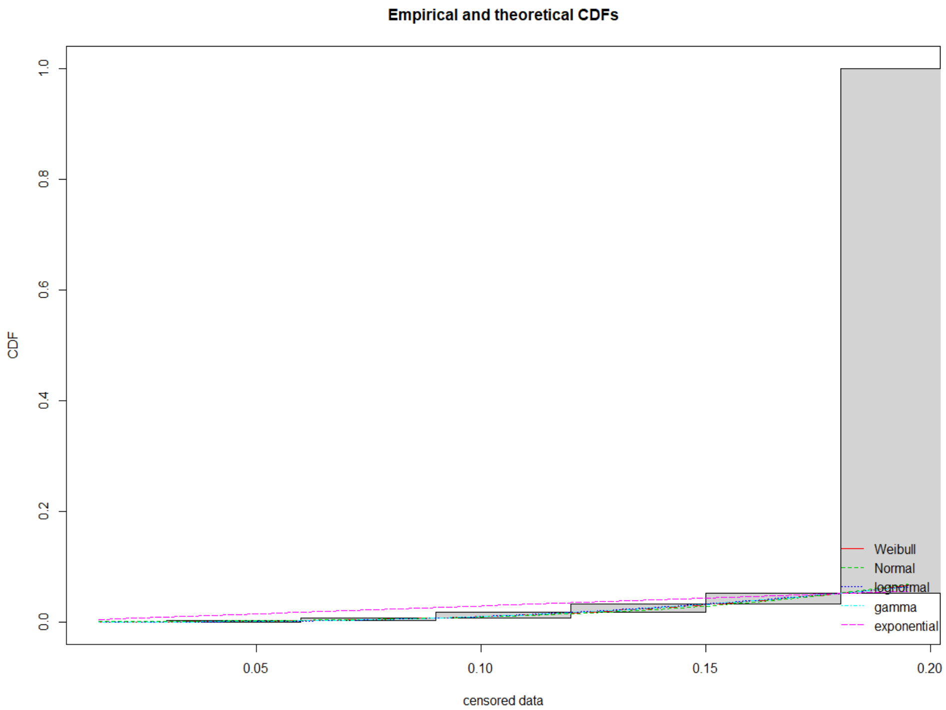
Figures A.3 and A.4: Fitting Expert 2 opinion – axle



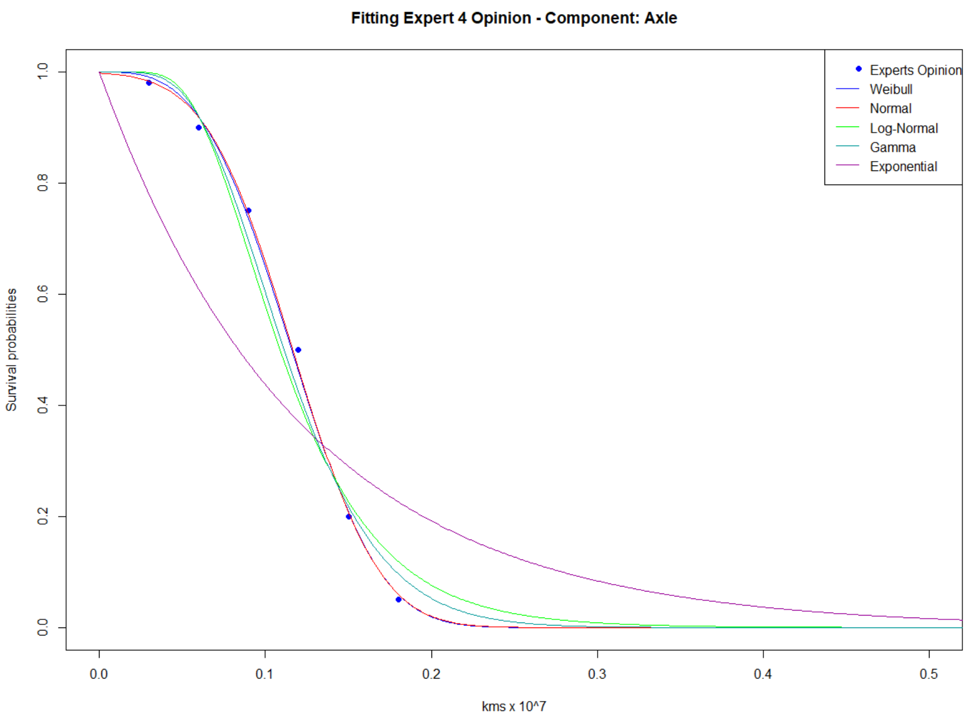


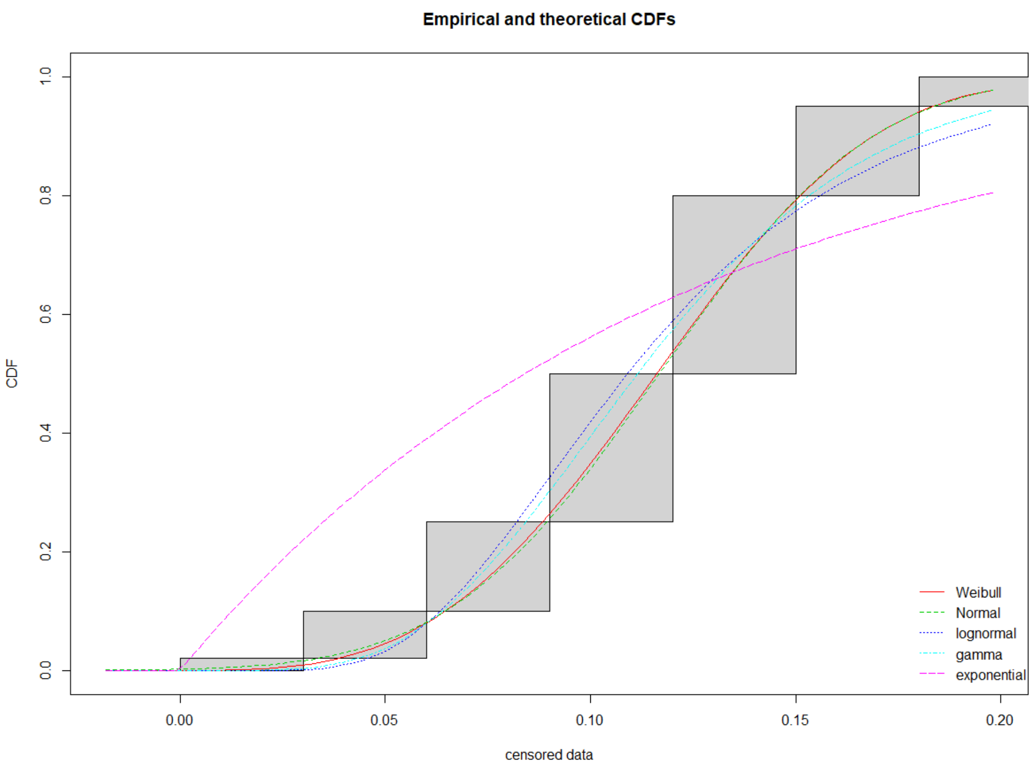
Figures A.5 and A.6: Fitting Expert 3 opinion – axle



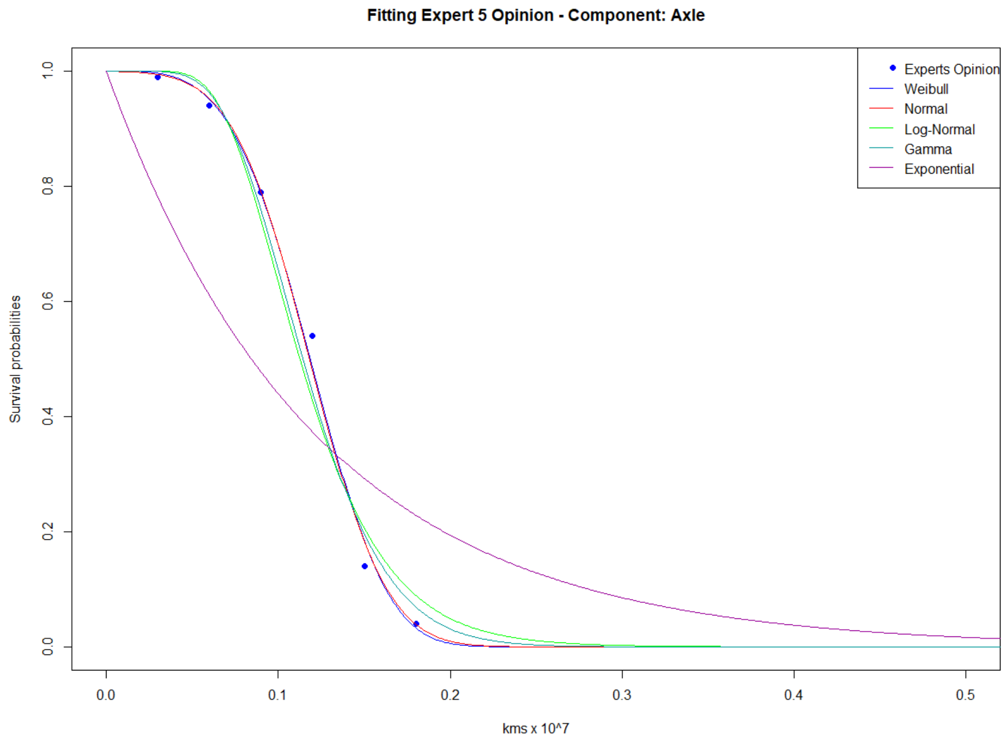


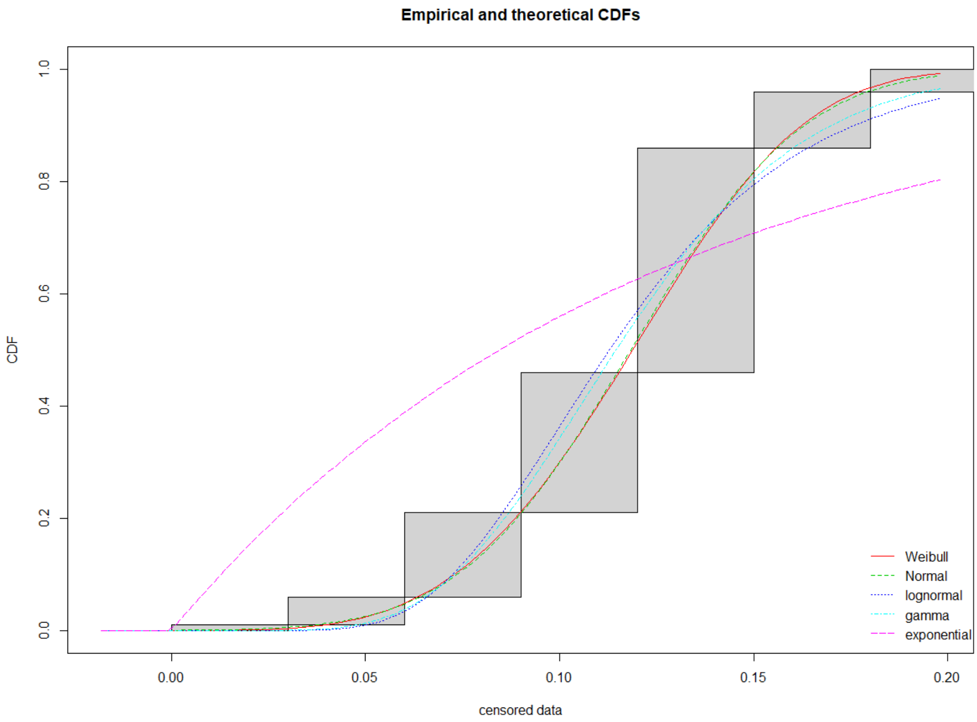
Figures A.7 and A.8: Fitting Expert 4 opinion – axle





Figures A.9 and A.10: Fitting Expert 5 opinion – axle





Figures A.11 and A.12: Fitting Expert 6 opinion – axle