



**HAL**  
open science

## Alternative Weibull analysis for road markings: an EM approach

Maxime Redondin, Nadège Faul, Laurent Bouillaut, Allou Same, Dimitri Daucher

### ► To cite this version:

Maxime Redondin, Nadège Faul, Laurent Bouillaut, Allou Same, Dimitri Daucher. Alternative Weibull analysis for road markings: an EM approach. ESREL2018 - European Safety and Reliability Conference, Jun 2018, Trondheim, Norway. pp 587-594, 10.1201/9781351174664-72 . hal-01818448

**HAL Id: hal-01818448**

**<https://hal.science/hal-01818448>**

Submitted on 19 Jun 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Alternative Weibull analysis for road markings: an EM approach

Maxime Redondin & Nadège Faul

*VEDECOR Institute, 77 rue des chantiers, F-78000 Versailles, France*

Laurent Bouillaut & Allou Samé

*Université Paris-Est, Grettia (IFSTTAR), F-77455 Marne-la-Vallée, France*

Dimitri Daucher

*Université Paris-Est, Lepsis (IFSTTAR), F-77455 Marne-la-Vallée, France*

**ABSTRACT:** The quality and reliability of road infrastructure and its equipment play a major role in road safety. This is especially true for autonomous car traffic guided mainly by a GPS system that is, unfortunately, neither precise nor reliable. In order to improve the guidance systems, one option could be to equip the vehicle with a camera reading road markings. Such solution require maintenance strategies guaranteeing markings' perceptibility to the human eye or the autonomous car camera. Currently, the retroreflection luminance of markings is measured for evaluating marking degradation. An important remaining step is a life time analysis depending on the inspection strategy. Since the exact failure time isn't generally observed, feedback database contain many censored data: the left-censure corresponding to a marking failing before the first inspection, the interval-censure that corresponds to markings failing between two inspections, and the right-censure corresponding to a marking that never fails. In the literature, a Weibull analysis was proposed to estimate the markings reliable distributions using the Maximum Likelihood through the Newton-Raphson method. Facing with censored data, this approach couldn't be computed without introducing strong bias in the reliability estimation. For generic interval-censored data, Pradhan and Kundu proposed an alternative, based on the EM algorithm. In our study an extension of the EM algorithm processing left and right censures is proposed. Finally, this algorithm is applicable for all kind of observations, whatever the censure nature. After introducing this EM extension, the paper focuses on the fact that computations are simpler than the Newton-Raphson methods and censored-data are directly estimated. The French National Road 4 markings case is considered to illustrate the proposed approach. Moreover, the proposed algorithm being generic, its application is, of course, not limited to our road marking case study.

## 1 INTRODUCTION

The quality and reliability of road infrastructure and its equipment play a major role in road safety. This is especially true for autonomous car traffic. Currently, autonomous vehicles are guided mainly by a GPS system. Unfortunately, GPS systems are neither precise nor reliable. For example, GPS signals couldn't work in urban canyons or tunnels. In order to improve autonomous vehicle guidance systems, one option would be to equip the vehicle with a camera able to "read" road markings. However, this solution requires a maintenance strategy guaranteeing that road markings remain perceptible to a human eye or an autonomous car camera.

According to both the AFNOR rules (AFNOR 2009) and the available inspection devices, the

retroreflection luminance of markings is the only measure used for evaluating marking degradation. A retroreflective marking reflects light from a vehicle headlight back in the direction of the driver. For waterborne markings, the retroreflective property is guaranteed by glass spheres mixed into the paint during application. The retroreflection luminance is measured in millicandela per square meter and by lux ( $mcd/m^2/lx$ ). A minimum threshold of  $150 mcd/m^2/lx$  is required for a new marking (AFNOR 2009).

Several decay models for retroreflective marking exist in the current literature which mainly calculate retroreflective luminance based on a regression model. For example, Lu (Lu 1995) proposed an exponential regression model function of age of markings, Abboud and Bowman (Abboud & Bowman 2002) de-

veloped exponential models as a function of the Annual Average Daily Traffic (AADT) and the age of markings, Sarasua et al. (Sarasua, Clarke, & Davis 2003) calculated the difference in reflectivity over time, Sitzabee et al. (Sitzabee, Hummer, & Rasdorf 2009) proposed the most complete decay multilinear model as a function of time, the initial retroreflection, the AADT, the lateral locations of markings and marking color. All decay models have a common weakness: they are difficult to apply directly to a given road network. For example, consider a single road. The road is not systematically maintained in its entirety. For safety reasons, road managers maintain only specific areas at a time. This is especially true for road surface maintenance.

In a previous study, a clustering approach able to segment a road network according to past inspections was proposed (Redondin, Bouillaut, Daucher, & Faul 2017). Each cluster was interpreted with respect to a specific area of the road network and admits its own retroreflective luminance evolution over time. This fact leads to each cluster having its own optimum maintenance strategy. An important remaining step is to do a life time analysis. This work could confirm the necessity of one maintenance strategy by cluster and is the first step to develop any maintenance model.

Currently, road markings are monthly or yearly inspected by a retroreflectometer. Thus, such a periodical approach isn't able to determine the exact failure time for any marking. Moreover, in feedback database, three kinds of censure are observed: the left-censure corresponding to a marking failing before the first inspection, the interval-censure that corresponds to markings failing between two inspections, and the right-censure corresponding to a marking that never fails.

A Weibull analysis was proposed by Sathyanarayanan et al (Sathyanarayanan, Shankar, & Donnell 2008), which consists of estimating the markings reliable function using a Weibull distribution. Its parameters are estimated according to the Maximum Likelihood Estimator (MLE). In attendance of censored-data, the likelihood function depends on the probability density, the cumulative distribution and the reliable function of a Weibull distribution. For interval-censored data, Pradhan and Kundu (Pradhan & Kundu 2014) found instances where the Newton-Raphson does not compute. An alternative, based on the EM algorithm (McLachlan & Krishnan 2008) has been proposed.

Section 2 proposed an extension of the EM algorithm processing left and right censures. The proposed algorithm is applicable for all observation vectors, independently of the nature of the censure. After introducing this EM extension, this paper will focus on a simpler computations than the Newton-Raphson methods (Bain & Englehardt 1975). Moreover, censored-data are directly estimated. Then, the

proposed approach could be applied for other cases studies and are not restricted to road markings considered in this paper.

As in, the French National Road 4 inspection database is considered to illustrate the proposed approach and its use for analysing maintenance cycles. Two applications are proposed : on the one hand, the first take into account the whole maintenance cycle and the EM Algorithm produces one global Weibull analysis. On the other hand, the second segments first the cycle according to a clustering approach (Redondin, Bouillaut, Daucher, & Faul 2017). Each cluster admits its own Weibull analysis taking into account different decay profiles.

## 2 IMPROVED EM ALGORITHM

### 2.1 Censored Weibull distribution

In this paper, a Weibull distribution  $\mathcal{W}(\alpha, \beta)$  is defined by its associated probability density function (1) where  $\alpha > 0$  and  $\beta > 0$  are respectively the associated scale and shape parameters.

$$f(t) = \begin{cases} \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1} e^{-\left(\frac{t}{\alpha}\right)^\beta} & \text{if } t > 0 \\ 0 & \text{if } t \leq 0 \end{cases} \quad (1)$$

The observed data  $(t_1, \dots, t_n)$  is assumed independent identically distributed. In this paper, an observed failure is a failure established according to a given inspection strategy.  $\forall i = 1, \dots, n$ ,  $t_i$  is assumed to be the first observed failure moment of  $i$ . Three censorship cases are assumed.

- If a failure has occurred before the first inspection, then it is called left-censored.
- If a failure has occurred between two inspections, then it is called interval-censored.
- If a failure hasn't been observed, then it is called right-censored.

If  $t_i$  is interval-censored, then its associated interval is  $[l_i, r_i]$ . The interval is clearly defined according to the current inspection strategy. A censor detector is associated for each observation as follows  $\forall i = 1, \dots, n$ .

$$\delta_i = \begin{cases} 0 & \text{if } t_i \text{ is uncensored} \\ 1 & \text{if } t_i \text{ is left-censored} \\ 2 & \text{if } t_i \text{ is interval-censored} \\ 3 & \text{if } t_i \text{ is right-censored} \end{cases} \quad (2)$$

From now, the observed data is subdivided according to the censor detector.

$$\mathcal{T} = \left\{ t \in ((t_1, \delta_1), \dots, (t_n, \delta_n)) \mid \delta_i = 0 \right\} \quad (3)$$

$$\mathcal{X} = \left\{ x \in ((t_1, \delta_1), \dots, (t_n, \delta_n)) \mid \delta_i = 1 \right\} \quad (4)$$

$$\mathcal{Y} = \left\{ y \in ((t_1, \delta_1), \dots, (t_n, \delta_n)) \mid \delta_i = 2 \right\} \quad (5)$$

$$\mathcal{Z} = \left\{ z \in ((t_1, \delta_1), \dots, (t_n, \delta_n)) \mid \delta_i = 3 \right\} \quad (6)$$

In the case of an uncensored Weibull distribution, the MLE is the couple  $(\alpha, \beta)$  which maximizes the associated likelihood function (7). This couple annuls also simultaneously the two partial derivatives of the log-likelihood function. For the specific case of a Weibull distribution, the solution of this non-linear equations system is currently done by a Newton-Raphson approach (Bain & Englehardt 1975).

$$L(\alpha, \beta) = \prod_{i=1}^n f(t_i) \quad (7)$$

The likelihood function associated to a censored Weibull distribution (8) depends on the reliable function  $S$  for interval and right censorship cases and it also depends on the cumulative function  $F = 1 - S$  for the left censorship case.

$$L(\alpha, \beta) = \prod_{t \in \mathcal{T}} f(t) \prod_{x \in \mathcal{X}} F(x) \prod_{y \in \mathcal{Y}} R(r) - R(l) \prod_{z \in \mathcal{Z}} R(z) \quad (8)$$

Reminder : let's take  $W$  a Weibull distribution  $\mathcal{W}(\alpha, \beta)$ , the reliable function (9) is the probability that the time of failure is later than some specified time  $t > 0$ .

$$R(t) = P(W > t) = e^{-\left(\frac{t}{\alpha}\right)^\beta} \quad (9)$$

For the road markings study, Sathyanarayanan et al (Sathyanarayanan, Shankar, & Donnell 2008) chose this classic approach for a Weibull analysis. In a breast cancer case study and restricted only to interval-censored data, Pradhan and Kundu (Pradhan & Kundu 2014) showed different examples where this approach does not compute. Specifically, the Newton-Raphson approach isn't able to estimate  $(\alpha, \beta)$ . An alternative based on a EM Algorithm is proposed. This paper proposed an extension of the algorithm, processing left and right censures.

The EM algorithm interprets censored data like missing data to estimate and computes iteratively into two steps :

1. The Expected Step estimates censored data according to a given Weibull distribution  $\mathcal{W}(\alpha, \beta)$ .

2. The Maximization Step estimates a Weibull distribution  $\mathcal{W}(\alpha, \beta)$  by the MLE according to both uncensored and completed data.

The algorithm computes until one distribution converges. This point is detailed in section 2.4. First, sections 2.2 and 2.3 present the extended algorithm through one given iteration.

## 2.2 Improved Expected Step

Let's take  $W$  a Weibull distribution  $\mathcal{W}(\alpha, \beta)$ . To simplify the EM formalism, the substitution  $\alpha \leftarrow 1/\alpha^\beta$  is suggested by Pradhan and Kundu (Pradhan & Kundu 2014). According to that, the density function is done by (10) in this section.

$$f(t) = \begin{cases} \alpha\beta t^{\beta-1} e^{-\alpha t^\beta} & \text{if } t > 0 \\ 0 & \text{if } t \leq 0 \end{cases} \quad (10)$$

The Expected Step calculates the estimated likelihood function defined as the maximized expectation of the likelihood function (7) conditioned to uncensored data (11).

$$L_c(\alpha, \beta) = E[L(\alpha, \beta) \mid \mathcal{T}] \quad (11)$$

This conditional expectation leads to produce three estimators  $\forall x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}$ . Each one is a conditional expectation adapted for one specific censure : (12) and (14) estimates respectively the left and right censures and (13) is proposed by Pradhan and Kundu (Pradhan & Kundu 2014) for the interval-censures case.

$$\hat{x} = E[W \mid W < x] = \frac{\alpha\beta \int_0^x t^\beta e^{-\alpha t^\beta} dt}{1 - e^{-\alpha x^\beta}} \quad (12)$$

$$\hat{y} = E[W \mid l < W < r] = \frac{\alpha\beta \int_l^r t^\beta e^{-\alpha t^\beta} dt}{e^{-\alpha l^\beta} - e^{-\alpha r^\beta}} \quad (13)$$

$$\hat{z} = E[W \mid W > z] = \frac{\alpha\beta \int_z^{+\infty} t^\beta e^{-\alpha t^\beta} dt}{e^{-\alpha z^\beta}} \quad (14)$$

Finally, the completed likelihood function associated to a censored Weibull distribution is defined by (15).

$$L_c(\alpha, \beta) = \prod_{t \in \mathcal{T}} f(t) \prod_{x \in \mathcal{X}} f(\hat{x}) \prod_{y \in \mathcal{Y}} f(\hat{y}) \prod_{z \in \mathcal{Z}} f(\hat{z}) \quad (15)$$

## 2.3 Improved Maximization Step

According to the MLE formalism, the couple  $(\alpha, \beta)$  annuls simultaneously the two partial derivatives of the completed log-likelihood function.

$$\partial_\alpha L_c = \frac{n}{\alpha} - \sum_{t \in \mathcal{T}} t^\beta - \sum_{x \in \mathcal{X}} \hat{x}^\beta - \sum_{y \in \mathcal{Y}} \hat{y}^\beta - \sum_{z \in \mathcal{Z}} \hat{z}^\beta \quad (16)$$

$$\begin{aligned} \partial_\beta L_c &= \frac{n}{\beta} + \sum_{t \in \mathcal{T}} \ln t + \sum_{x \in \mathcal{X}} \ln \hat{x} + \sum_{y \in \mathcal{Y}} \ln \hat{y} + \sum_{z \in \mathcal{Z}} \ln \hat{z} \\ &- \alpha \left[ \sum_{t \in \mathcal{T}} t^\beta \ln t + \sum_{x \in \mathcal{X}} \hat{x}^\beta \ln \hat{x} + \sum_{y \in \mathcal{Y}} \hat{y}^\beta \ln \hat{y} + \sum_{z \in \mathcal{Z}} \hat{z}^\beta \ln \hat{z} \right] \end{aligned} \quad (17)$$

This non-linear equations system doesn't admit an obvious solution. Restricted to interval-censored data, Pradhan and Kundu (Pradhan & Kundu 2014) solved this non-linear equations system by a fixed point approach. This choice is due to an estimation of  $\alpha$  done by the equation  $\partial_\alpha L_c = 0$ . Processing left and right censures, the extending estimator is :

$$\alpha = \frac{n}{\sum_{t \in \mathcal{T}} t^\beta + \sum_{x \in \mathcal{X}} \hat{x}^\beta + \sum_{y \in \mathcal{Y}} \hat{y}^\beta + \sum_{z \in \mathcal{Z}} \hat{z}^\beta} \quad (18)$$

This estimator (18) is completely dependent on  $\beta$  : an estimation of  $\beta$  is enough. Furthermore replacing  $\alpha$  by (18) in (17), a function  $g(\beta)$  as  $\partial_\beta L_c = g(\beta) - \beta$  is extracted (19). To simplify, let's take  $\mathcal{S} = \mathcal{X} \cup \mathcal{Y} \cup \mathcal{Z}$ .

$$g(\beta) = \frac{1}{\frac{\sum_{t \in \mathcal{T}} t^\beta \ln t + \sum_{s \in \mathcal{S}} \hat{s}^\beta \ln \hat{s}}{\sum_{t \in \mathcal{T}} t^\beta + \sum_{s \in \mathcal{S}} \hat{s}^\beta} + \frac{\sum_{t \in \mathcal{T}} \ln t + \sum_{s \in \mathcal{S}} \ln \hat{s}}{n}} \quad (19)$$

The fixed point of  $g$  is the point  $\beta$  as  $g(\beta) = \beta$  also defined as the convergence point of the sequence (20). The initial value is arbitrary. This point is detailed at section 2.4

$$(u_n)_{n \geq 0} = \begin{cases} u_0 > 0 & \text{arbitrary} \\ u_n = g(u_{n-1}) & n > 0 \end{cases} \quad (20)$$

To conclude,  $\alpha$  is finally deduced from (18).

## 2.4 Iteration process and convergence

The Weibull distribution estimated at the algorithm iteration  $k > 0$  is denoted  $\mathcal{W}(\alpha_k, \beta_k)$ . The initial distribution  $\mathcal{W}(\alpha_0, \beta_0)$  is arbitrary. In this paper,  $\alpha_0$  and  $\beta_0$  are the MLE where the censorship is not taken into account.

Sections 2.2 and 2.3 show that one algorithm iteration could be reduced to an estimation of  $\beta$  done by the convergence point of (20) depending on the function  $g$  (19) conditioned by estimated data (12-14). As  $\alpha_k$  is directly defined both by  $\beta_k$  and (18), the next sequence  $(\beta_k)_{k \geq 0}$  is clear-defined as :

$$(\beta_k)_{k \geq 0} = \begin{cases} \beta_0 > 0 \\ \beta_k = \lim_{n \rightarrow +\infty} \begin{cases} u_0 = \beta_{k-1} \\ u_n = g(u_{n-1}) \end{cases} \end{cases} \quad (21)$$

$|\beta_k - \beta_{k-1}| < 10^{-4}$  is the selected stopping criterion. Finally, the EM Algorithm convergence is defined both by the convergence of the sequence  $(\beta_k)_{k \geq 0}$  and (18) :

$$\begin{cases} \beta = \lim_{k \rightarrow +\infty} \beta_k \\ \alpha = \frac{n}{\sum_{t \in \mathcal{T}} t^\beta + \sum_{x \in \mathcal{X}} \hat{x}^\beta + \sum_{y \in \mathcal{Y}} \hat{y}^\beta + \sum_{z \in \mathcal{Z}} \hat{z}^\beta} \end{cases} \quad (22)$$

Again, the stopping criterion is  $|\beta_k - \beta_{k-1}| < 10^{-4}$ . To conclude, the substitution  $\alpha \leftarrow 1/\sqrt[\beta]{\alpha}$  returns the current Weibull distribution (1).

## 3 MAINTENANCE CYCLE 2008-2012

### 3.1 Presentation of the French National Road 4

The broken centerline of the French National Road 4 (NR4) is considered to illustrate the proposed approach. The NR4 runs between Paris and Strasbourg. Since 2007, the section of this road between Courgi-vaux and Vaclerc ( $\sim 102$  km) has been managed by the DIR Est, which inspects this section of the road once a year. Inspections are organized in September in collaboration with CEREMA Est. The selected retroreflector is an Ecodyn.

The maintenance cycle selected is composed of 73 measures annotated with a PR. To simplify, each measure is interpreted as one marking. Markings laid in March 2008 and replaced in March 2012. The marking material chosen is supposed to be the same. The cycle is localized around three cities : Courgi-vaux, Sommesous and Vitry-le-François. The direction heading toward Vaclerc is chosen.

Four inspections are available : 6, 18, 30 and 42 months (after March 2008). The retroreflection luminance of a given marking  $i$  at the inspection point  $t$  is denoted  $RL_t(i) \in \mathbb{N}^*$ . If  $RL_t(i) \leq 150 \text{ mcd/m}^2/\text{lx}$  then the marking  $i$  is failed at  $t$ . Let's take  $\tau$  be the first time when a given marking is observed failing.

$$\tau = \min \left\{ t \in \{6, 18, 30, 42\} \mid RL_t(i) \leq 150 \right\} \quad (23)$$

According to  $\tau$ ,  $t_i$  is the first time when the marking  $i = 1, \dots, 73$  is observed failing. If a marking  $i$  hasn't been observed, then the marking is right-censored and  $\tau = \emptyset$ . In this situation,  $t_i = 42+$  tentatively.

$$t_i = \begin{cases} \min \tau & \text{if } \tau \neq \emptyset \\ 42+ & \text{else} \end{cases} \quad (24)$$

The adapted censor detector is deduced (25). From now on, if  $t_i = 42+$  then  $t_i = 42$ . This formalism assumes three censorship intervals ( $[6, 18]$ ,  $[18, 30]$  and  $[30, 42]$ ) and  $t_i$  isn't an uncensored observation. Finally, The observed data is finally defined as  $((t_1, \delta_1), \dots, (t_{73}, \delta_{73}))$ .

$$\delta_i = \begin{cases} 1 & \text{if } t_i = 6. \\ 2 & \text{if } t_i = 18, 30, 42. \\ 3 & \text{if } t_i = 42+. \end{cases} \quad (25)$$

According to inspection campaigns, the monitoring of the retroreflective luminance is presented by the figure 1. All censor case are presented : two failures are observed during the 6th month (September 2008), a group of markings which never failed until the 42nd month (September 2011) exists and the majority of failure is interval-censored. At least two decay models are also presented.

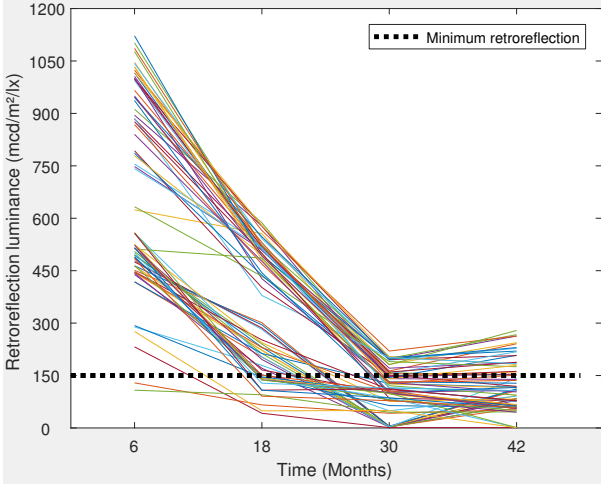


Figure 1: Maintenance cycle on the NR4 broken centerlines between September 2008 and September 2011.

Two approaches could be produced : a global and a clustering Weibull analysis. The global analysis consists in selecting the whole monitoring and estimating one global Weibull distribution. This approach is interesting if the observed data is small. This situation could happen for two reasons. Firstly, current retroreflectometer like the Ecodyn produce generally one average measure every 100m. This fact reduces considerably the data set size. Secondly, current maintenance strategy could concern only one specific area. The clustering analysis consists in segmenting first the whole monitoring and estimating one Weibull distribution by cluster. This approach is interesting to produce one specific maintenance strategy by cluster over time, but the observed data must be large and must present a diversified censorship case.

### 3.2 Global Weibull analysis

The table 1 presents the Weibull analysis. The first column indicates the observed data associated to the

Table 1: Weibull analysis of the 2008-2012 maintenance cycle

Observed failure	Numbers	Estimated failure $\mathcal{W}(31.65, 2.20)$
(6, 1)	2 (3%)	4.12
([6, 18], 2)	13 (18%)	12.94
([18, 30], 2)	39 (53%)	24.01
([30, 42], 2)	2 (3%)	35.44
(42, 3)	17 (23%)	50.41

censor indicator. For example, (6, 1) corresponds to failures observed during the 6th month (September 2008) and the left-censored case. The second column indicates the number of observation : for example, there are 2 left-censored data and they correspond to 3% of the observed data. The last column indicates the failure time (in month) estimated by the EM Algorithm. For example, according to the Weibull distribution  $\mathcal{W}(31.65, 2.20)$ , left-censored data are estimated at 4.12 months (July 2008).

Table 1 confirms observations made on the monitoring. An important case of interval-censored data (74%) is presented. Particularly the main failure is emerged between 18 and 30 months (53%). Right-censored data are the second most important case (23%). Only two markings are left-censored.

The EM Algorithm converges after 12 iterations to the Weibull distribution  $\mathcal{W}(31.65, 2.20)$ . According to this distribution, several estimated failures are proposed. Interval-censored failures are estimated 12.94, 24.01, 35.44 months (March 2009-2010, February 2011). This period corresponds to the average of each interval. Left and right censored failures are respectively estimated to 4.12 months (July 2008) and 50.41 months (May 2012). The maintenance campaign in March 2012 is finally warranted.

Finally, the EM Algorithm is able to produce one Weibull analysis adapted to the whole monitoring.

### 3.3 Clustering Weibull analysis

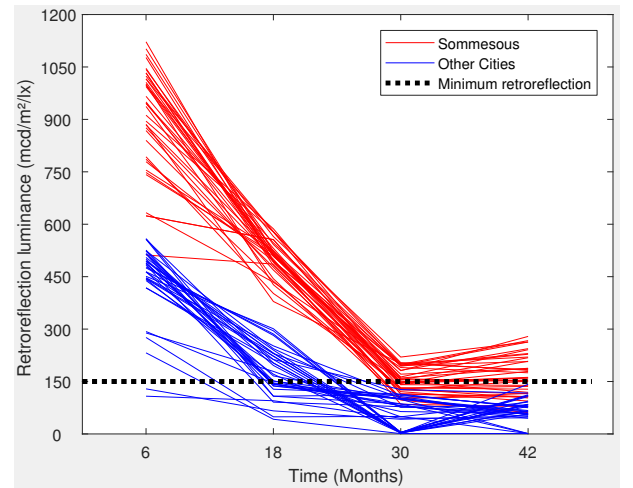


Figure 2: Maintenance cycle 2008-2012 by clusters.

Based on an Agglomerative Hierarchical Clustering, the clustering process proposed in the previous

ESREL conference could be restricted to the maintenance cycle (Redondin, Bouillaut, Daucher, & Faul 2017). Two clusters are proposed : Sommesous and other cities. Figure 2 distinguishes clusters on the monitoring. Sommesous (red) represents markings which admit a strong retroreflection luminance, a fast decay between 6 and 30 months and finally a stationary decay until the 42 Months (September 2011). Other cities (blue) present a fast decay between 6 and 18 months and a slow decay until the 42 Months.

Table 2: Weibull analysis of the maintenance cycle 2008-2012 by clusters.

Cluster	Observed failure	Numbers	Estimated failure
Sommesous $W(41.35, 3.5)$	$([18, 30], 2)$	18 (25%)	24.95
	$([30, 42], 2)$	2 (3%)	35.11
	$(42, 3)$	17 (23%)	49.84
Other Cities $W(18.83, 6.61)$	$(6, 1)$	2 (3%)	5.21
	$([6, 18], 2)$	13 (18%)	15.23
	$([18, 30], 2)$	21 (29%)	20.14

The table 2 presents the Weibull analysis by clusters. The structure is similar to table 1 excepted the added first column to indicate both the cluster and the Weibull distribution estimated. The EM Algorithm proposed the Sommesous Weibull distribution  $W(41.35, 3.5)$  and the Other Weibull distribution  $W(18.83, 6.61)$  respectively after 19 and 38 iterations.

Sommesous markings failed mainly between 18 and 30 months and focused all right-censored data. Their failure are respectively estimated at 24.95 months (March 2010) and 49.84 months (April 2012). Other markings failed mainly between 6 and 18 months or 18 and 30 months and focused all left-censored data. Their failures are respectively estimated at 15.23 months (June 2009), 20.14 months (November 2009) and 5.21 months (August 2008).

The clustering process clearly separated the monitoring into two decay profiles. Sommesous markings had a strong retroreflectivity and failed either within 30th month or after the 42nd month while Others markings had a weaker retroreflectivity and failed before the 30th month. Finally, the EM Algorithm is also able to produce one specific Weibull analysis by cluster.

### 3.4 Comparison

Estimated failures between the global and the clustering models are equivalent to one month. Markings failed between 18 and 30 months in Other Cities are the main difference. The global model estimated the failure in March 2010 and the clustering model estimated the failure in November 2009. The difference is due both to the EM Algorithm computation and the current inspection strategy. First, the September 2010 inspection observed that 53% of markings failed

since September 2009 and the associated interval-censored data is  $([18, 30], 2)$  for all markings. Second, the global approach ignores different decay profiles. Therefore in this case, the EM-Algorithm produced one global estimation.

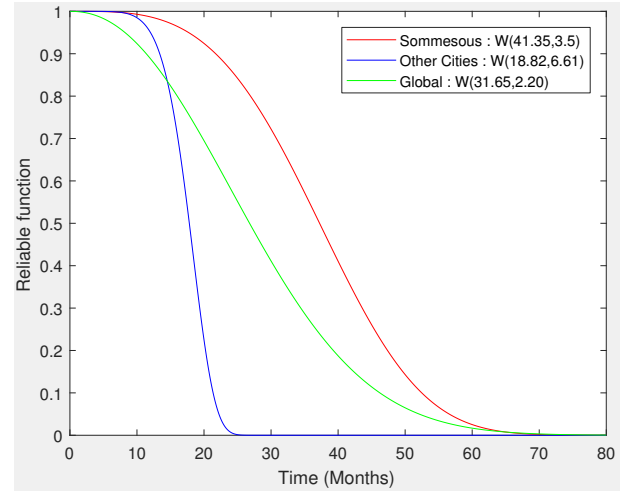


Figure 3: Associated reliable functions according to the maintenance 2008-2012.

The figure 3 compares reliable functions according to the global model and the two clusters. The global reliable function (green) underestimates Sommesous over time (-0.12 in average). Other Cities is slightly underestimated the 14th first months (-0.03 in average) and overestimated next months (+0.12 in average) in particular after the 23rd months. Finally, the global reliable function is an average estimation.

The global model is a compromise estimation according to different observations : two left-censored observations, important failure between 18 and 30 months, the quarter of observations is right-censored... The clustering model suggests first several clusters according to the decay profile and each one admit its own Weibull distribution. Therefore the Sommesous Weibull distribution is estimated according to two facts : without failures before the 18th month and all right-censored observations.

The clustering model admits two main disadvantages. First, the clustering process tends to isolate markings which admit one specific censure. For example, all right-censored observation could be gathered into a third cluster extracted from Sommesous. In this situation, the MLE is based on the only observation  $(42, 3)$  and doesn't compute. The second problem is each cluster admits its own Weibull analysis. The model admits a total of 57 EM Algorithm iterations, whereas the global model admits only 12.

However, The global reliable function concludes to a poor compromise : Sommesous and Other markings reliable are respectively underestimated and overestimated.

Finally, the clustering model has done a more reliable Weibull analysis but the clustering process and the number of EM Algorithm iteration should be monitored.

## 4 CONCLUSIONS

The current literature presents a first Weibull analysis for road markings (Sathyanarayanan, Shankar, & Donnell 2008). Censored Weibull distributions are estimated by the current approach based on the MLE upgraded by a Newton-Raphson approach. Restricted to interval-censored data, Pradhan and Kundu (Pradhan & Kundu 2014) discovered several limits to this method and proposed a first alternative based on a EM Algorithm.

The introduced extended EM Algorithm is a credible alternative to the MLE of a censored Weibull distribution. This is specially true on the censored data management. Indeed, the EM Algorithm replaces censored data by a failure iteratively estimated at once the Weibull distribution. Furthermore, the EM Algorithm is upgraded by a fixed point approach. This method is simpler than the Newton-Raphson. Indeed, the fixed point depends only to uncensored data and estimated data and doesn't need any upstream verification.

The introduced algorithm isn't limited to our case study. Pradhan and Kundu themselves presented a breast cancer case study for example. The extension accepts a greater variety of situations : uncensored and right-censored data, no-uncensored data, no-interval-censored data... From the moment the decay monitoring over the time and the maximum decay level are both clear, the EM Algorithm is able to produce and rank all censored data.

The EM Algorithm is able to estimate several Weibull distribution in our road markings study. Two approach is proposed. The global model is interesting in the case where the observed data is small. However, the estimated Weibull distribution is an average compromise between different decay profiles. In the NR4 case, the global reliable function overestimates Other markings and underestimates Sommesous markings. The clustering analysis segments first the whole monitoring and estimates one Weibull distribution by cluster. This situation is interesting if the observed data is large and could distinguishes different decay profiles. In the NR4 case, reliable functions are more reliable.

This paper shows that the cluterer approach is more reliable that the global model. However, this approach needs a monitoring of the clustering process. Indeed, the current process tends to isolate one specific censure by cluster. Consequently, the MLE could be based on only one observed data and it formalism cannot compute. An alternative based on a mixture model also produces by an EM Algorithm is currently investigated. This approach could estimate directly the optimum segmentation and associated it in one mixture Weibull distribution.

Finally, the Weibull analysis completed by an EM approach is adapted in a road markings study. Furthermore, for given sections of the road network, reliable functions are able to indicate directly each section ad-

mits a premature aging for example. These facts lead to the development of a probabilistic opportunistic maintenance model adapted to a whole road network or a segmented road network.

## 5 ACKNOWLEDGMENT

VEDECOM and IFSTTAR thank the DIR Est and the CEREMA for inspection data on the National Road 4.

## REFERENCES

- Abboud, N. & B. Bowman (2002). Cost and longevity-based scheduling of paint and thermoplastic striping. *Transportation Research Record* 1794.
- AFNOR (2009). Nf en 1436+a1 - road marking materials - road marking performance for road users. *AFNOR Editions*.
- Bain, L. & M. Englehardt (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston, Inc.
- Lu, J. (1995). Performance of traffic markings in cold regions. *University of Alaska Fairbanks - Transportation Research Center (Report No. INE/TRC 95.03)*.
- McLachlan, G. & T. Krishnan (2008). *The EM Algorithm and Extensions*. John Wiley and Sons,.
- Pradhan, B. & D. Kundu (2014). Analysis of interval-censored data with weibull lifetime distribution. *Sankhya B : The Indian Journal of Statistics* 76, 120–139.
- Redondin, M., L. Bouillaut, D. Daucher, & N. Faul (2017). Temporal clustering for retroreflective markings. *European Safety and Reliability Conference 2017 (Portoriz, Slovenia)* (76), 7p.
- Sarasua, W. A., D. B. Clarke, & W. J. Davis (2003). Evaluation of interstate pavement marking retroreflectivity. *South Carolina Department of Transportation (Report No. FHWA-SC-03-01)*.
- Sathyanarayanan, S., V. Shankar, & E. T. Donnell (2008). Pavement marking retroreflectivity inspection data : A weibull analysis. *Journal of the Transportation Research Board*. 2055, 63–70.
- Sitzabee, W., J. Hummer, & W. Rasdorf (2009). Pavement marking degradation modeling and analysis. *Journal of infrastructure systems* 15(3), 190–199.