New Probabilistic Approach to Estimate Vehicle Failure Trajectories in Curve Driving

Abdourahmane Koita, Dimitri Daucher, Michel Fogli

New Probabilistic Approach to Estimate Vehicle Failure Trajectories in Curve Driving

Abdourahmane Koita*
Researcher
University of Paris-Est
IFSTTAR, IM, LEPSIS,
F-75732, Paris, France
abdourahmane.koita@ifsttar.fr

Dimitri Daucher
Researcher
University of Paris-Est
IFSTTAR, IM, LEPSIS,
F-75732, Paris, France
dimitri.daucher@ifsttar.fr

Michel Fogli
Professor
University of Blaise Pascal
Department of LaMI
63175, Clermont Ferrand
michel.fogli@polytech.univ-bpclermont.fr

The vehicle trajectories analysis on dangerous bends is an important task to improve road safety. This paper propose a new methodology to predict failure trajectories of light vehicles in curve driving. It consists to use a stochastic modelling and reliability analysis in order to estimate the failure probability of vehicle trajectories.
Firstly, we build probabilistic models able to describe real trajectories in a given bend. The models are transforms of scalar normalized second order stochastic processes which are stationary, ergodic and non-Gaussian. The process is characterized by its probability density function and its power spectral density estimated starting from the experimental trajectories. The probability density is approximated by using a development on the basis of Hermite polynomials. The second part is devoted to apply a reliability strategy intended to associate a risk level to each class of trajectories. Based on the joint use of probabilistic methods for modelling uncertainties, reliability analysis for assessing risk levels and statistics for classifying the trajectories, this approach provides a realistic answer to the tackled problem. Experiments show the relevance and effectiveness of this method.

1 INTRODUCTION

In spite of road safety improvement in these last years, the light vehicles accidents in bend have very serious consequences in human term for road users [1]. The statistics show that 18% of road accidents and $1/3$ of mortals road accidents in France took place in bend. The main types of these accidents are road departures and vehicle loss control. Currently, the practical solution to reduce these accidents is not only the speed limit indicator via a road sign but also the advanced driver assistance systems such as: ABS, ESP, etc. There exist also several research works based on vehicle dynamic system. These works are undertaken on the topic of the detection and/or control of the vehicle path being able to be dangerous in bend, [2], [3], [4]. Afterwards, we can cite other works based on human factor models like [5], [6], [7], etc. However, the diversity of proposed solutions gives the level of complexity in curve driving.

For most drivers, driving in curves is a skill-based task, but it actually requires combined control of both steering and speed, taking into account the dynamic response and the vehicle environment. Indeed, it is a complex task who implies to consider all the interactions between vehicle and its environment. These interactions can generate discontinuities in the system parameters (or in their evolution). These discontinuities result either by a bad perception of the road for the driver, which then applies unsuited orders, or by large road entries in high quantity to be corrected via the loop of control. In both case, the driving system is disturbed and consequently the risk of accident is high. Among the system interactions, we quote perception, visibility or adherence which are complex physical phenomena, not only, strongly nonlinear but also presenting uncertainties in measurements.

All these reasons make that deterministic approaches are not sufficient. We propose to use probabilistic approaches in order to take into account uncertainties. These approaches will permit to estimate failure probability of trajectory according to each behaviour of control in curve driving. The proposed approach operates on real observed trajectories.

*Address all correspondence to this author.

1.1 Vehicle trajectory definition

The trajectory can be considered in several way according to functional space in which we want to describe vehicle dynamic. Either in **space of parameters**, where it is defined as the graph of function :

$$\begin{cases} \mathcal{T} : T \to \mathbb{R}^6 : \\ t \to \mathcal{T}(t) = (x_1(t), x_2(t), x_3(t), \psi(t), \phi(t), \theta(t))^T \end{cases} \quad (1)$$

Where x_1, x_2 and x_3 are cartesian coordinates of vehicle center of mass and ψ, ϕ and θ the Euler angles.
Or in **space of phasis**, where it is defined in this form:

$$\begin{cases} u : T \to \mathbb{R}^6 : \\ t \to u(t) = (x_1(t), x_2(t), v_1(t), v_2(t), \gamma_1(t), \gamma_2(t))^T \end{cases} \quad (2)$$

Where $T \subset \mathbb{R}_+$ is temporal observation interval of the vehicle movement. The coordinates (x_1, x_2, v_1, v_2, γ_1 and γ_2) describe vehicle movement in the Galilean coordinate system noted $\mathcal{R}_0^A = (O, \vec{e}_1, \vec{e}_2, \vec{e}_3)$. To define this coordinate system, let $\mathcal{B}_0^A = (\vec{e}_1, \vec{e}_2, \vec{e}_3)$ be the orthonormal basis associated to $\mathcal{R}_0^A$. The axis of $\mathcal{R}_0^A$ are noted Ox_1, Ox_2 and Ox_3. $\mathcal{R}_0^A$ is fixed and vehicle movement takes place in the horizontal plane (Ox_1, Ox_2). We make assumption that vehicle moves on a plane roadway without unevenness in which case the pumping effects are reduced. The trajectory u of the vehicle center of mass G is oriented and the point ς is chosen for the origin of the curvilinear abscise s. At any time t, the position of G is defined by this curvilinear abscise $s(t)$, see Fig. 1.

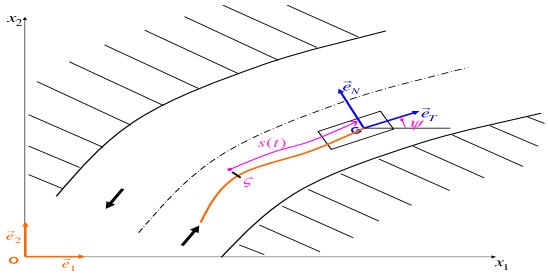

Fig. 1. Galilean coordinate system and Serret-frénet frame

The use of powerful measuring instruments able to acquire real observed trajectories with high-precision. However, we have only experimental measurements of trajectories. This one by nature contain uncertainties because the driving system is an intrinsically random phenomena. It depends on several factors, including the driver and vehicle environment, acting on the trajectories (e.g. effects of random external actions like wind, or unexpected reactions of drivers, or interactions related to a dubious probabilist like an adherence reduction, etc). In fact, in practice the trajectory u cannot be predicted with certainty according to Eqn.(2). Thus, the same driver circulating with the same vehicle on the same road under the same conditions will not reproduce the same trajectory twice. The better way to include such uncertainties is to use stochastic approach.

We propose to consider the vehicle trajectory as a realization of $\mathbb{R}^6$-valued stochastic process $U = (U(t), t \in \mathbb{R})$ defined on $(\Omega, \Im, \mathbb{P})$. We obtain $u(t) = U(t, \omega)$, $\forall t \in T$ where $\omega \in \Omega$:

$$\forall \omega \in \Omega,\ u = U(., \omega) : T \to \mathbb{R}^6 : t \to u(t) = U(t, \omega) \quad (3)$$

Starting from this definition, the main objective is to identify stochastic process by using the real trajectories and to apply a reliability analysis in order to estimate failure trajectory. One calls failure trajectory, any trajectory not respecting the safety conditions defined by legislator.

1.2 Suggested methodology

Initially, we propose to acquire experimental trajectories in real site and to use statistical analysis in order to identify several classes of trajectories. A Class of trajectories is characterized by a subset of trajectories having the sufficiently homogeneous statistical properties.
Starting from the trajectories of each class, we build a probabilistic model based on stochastic process identification. Each process will permit to simulate the dynamic of driving system according to failure criteria. These criteria relate to functional projection of vectorial stochastic process in $\mathbb{R}$. This approach of modeling is a major contribution in road safety because it makes abstraction to all the nonlinear interactions which are not modeled preciously.
Then, we will use simulations of stochastic processes in the reliability analysis application in order to evaluate the probability of failure according to each class of trajectories.
Lastly, a recognition model of trajectory was developed in earlier study, see [8]. It uses the data observations in bend's entry to assign each trajectory u to a membership class. This result associate to the knowledge of the failure probability is an indication to predict failure trajectory in curve driving.

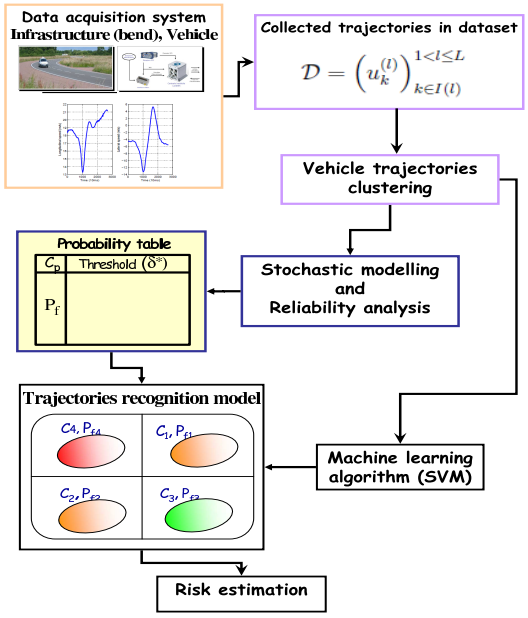

Fig. 2. Concept of failure trajectory prediction in curve driving

2 STATISTICAL ANALYSIS OF TRAJECTORIES

The collected discrete observations constitute a sample of trajectories. It is a L-family in this form:

$$\mathcal{D} = \left(u_k^{(l)}\right)_{k\in I(l)}^{1<l\leq L} \tag{4}$$

Where $u_k^{(l)} = u^{(l)}(t_k)$, L is the number of observed trajectories, $I(l) = \{1_l, 2_l, \ldots, K_l\}$ and $K_l \in \mathbb{N}^*$ is the number of discretizations. For each $l \in \mathcal{L} = \{1, \ldots, L\}$ is associated a single course of the driver during the bend:

$$u_k^{(l)} = \left(x_{1,k_l}^{(l)}, x_{2,k_l}^{(l)}, v_{1,k_l}^{(l)}, v_{2,k_l}^{(l)}, \gamma_{1,k_l}^{(l)}, \gamma_{2,k_l}^{(l)}\right)^T \in \mathbb{R}^6 \; ; \tag{5}$$

For any $k_l \in I(l)$, the vector $u_{k_l}^{(l)}$ is an observation of vehicle trajectory at instant t_k. This vector is constituted of the positions, velocity and accelerations of center of mass G in $\mathcal{R}_0^A$.

It is not obvious, in advance that we can attach all the L trajectories of $\mathcal{D}$ to the same process taking into account for example the habituation effects or the drivers heterogeneity. However, it is very difficult to create experimental observations conditions to obtain the same stochastic process U. For that, it is judicious to seek similar classes of trajectories on the totality of the course in bend.

Each class noted $\mathcal{C}_p$ is made up of L_p experimental trajectories derived from the same stochastic process. This $\mathbb{R}^6$-valued process noted U_p is defined on probability space $(\Omega, \Im, \mathbb{P})$. With $p \in \mathcal{P} = \{1, \ldots, P\}$, where P is the number of identified classes. To distribute the trajectories of $\mathcal{D}$ in P classes having homogeneous statistical properties, we used robust clustering algorithms.

2.1 Trajectories clustering methods

Clustering is based on the measure of proximity or dissimilarity between the trajectories in dataset $\mathcal{D}$. It consists to find subsets of $\mathcal{D}$ wich are homogeneous and/or well-separated. Homogeneity means that trajectories in the same cluster (or class) must be similar. The separation between trajectories in different clusters must differ one from the other. In general, clustering methods use machine learning techniques. We consider it as a function f_θ from $\mathcal{D}$ to the output space $\mathcal{Y}$:

$$f_\theta : \mathcal{D} \to \mathcal{Y} \tag{6}$$

Where $\mathcal{Y} \subset \mathbb{N}$ is classes $\mathcal{C}_p$ index of the partition $\mathcal{D}$, and f_θ is a function which depend on the chosen clustering algorithm. Several algorithms exist such as k-means: [9], mean shift [10] and Self Organizing Maps [11], etc. In this study, we choose the k-means algorithm because it gives better results than any other methods for our observations data.

2.1.1 k-means clustering algorithm

The k-means [9] is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. It uses P centroids noted N_p where $p \in \mathcal{P}$. Each class $\mathcal{C}_p$ is characterized by its centroid. It corresponds to the barycenter or the average of the elements which compose it. This algorithm aims to find clusters centers. These centers should be placed them as much as possible far away from each other, see [12]. In this study, the particularity of this algorithm is to introduce a specific distance between trajectories.

Then, it is necessary to calculate the distance $D(u, N_p)$ between each trajectory $u^{(l)}$ and cluster center N_p:

$$D(u, N_p) = \frac{1}{L_p} \sum_{u^{(2)}\in N_p} d^2(u^{(1)}, u^{(2)}) \tag{7}$$

Where d is an application of $\mathcal{D}\times\mathcal{D}$ in $\mathbb{R}_+$ verifying the properties of symmetry, separation and triangular inequality. The distance d between an object $u_p^{(1)}$ to another object $u_p^{(2)}$ must be selected. Several distances are used for similarity measurement. The concept of distance between trajectories u becomes complicated compared to Euclidean distance because, on the one hand, the dimension of phasis space, on the other hand, the dependence of the coordinates of u. In this study, we chose **Mahalanobis distance** (cf. Koita et al. [13]) in order to take into account the correlation between the parameters of u. This distance defined by:

$$d_M(u^{(1)}, u^{(2)}) = \sqrt{(u_k^{(1)} - u_k^{(2)})^T \tilde{\Sigma}^{-1} (u_k^{(1)} - u_k^{(2)})} \tag{8}$$

Where $\tilde{\Sigma}$ is the estimate of covariance matrix of random vector:

$$(U_p(t_1), \ldots, U_p(t_K))^T = (U_{p,1}, \ldots, U_{p,K})^T \tag{9}$$

The use of this distance will permit to assign each trajectory u to the cluster center N_p whose distance $D(u, N_p)$ from the center N_p is minimum of all the N_p with $p \in \mathcal{P}$. An object $u \in \mathcal{C}_p$ if:

$$D(u, N_p) = min\{D(u, N_r) \; ; \; r = 1, \ldots, p\} \tag{10}$$

The class $\mathcal{C}_p$ is then defined as follow:

$$\mathcal{C}_p = \{u \in \mathcal{D} \; ; \; D(u, N_p) \leq D(u, N_r) \; , \; \forall p \neq r\} \tag{11}$$

For each $\mathcal{C}_p$ generated in the previous step, its centroid N_p is recalculated. After we have these P new centroids, a new binding has to be done between the same subset and the nearest new center. A loop has been generated. As a result

of this loop, we may notice that the N_p centers change their location, step by step until no more changes are done or in other words centers do not move any more.

The P identified classes of trajectories are stable against the number of iterations. Each $\mathcal{C}_p$ is regarded as a different behavior of control. It is necessary to identify all the classes, and for each $\mathcal{C}_p$ must be representative. The use of the validity criteria is an effective way to ensure the algorithms convergence and classes stability. Because the unstable classes can generate homogeneity problems and consequently a bad identification of stochastic processes U_p. We recall that the elements of each $\mathcal{C}_p$ are regarded as discrete realizations of the same U_p.

3 STOCHASTIC MODELING OF V-I-D SYSTEM

The objective of this section is to develop stochastic models able to simulate the vehicle dynamic according to criteria of failure. These models are specific to the classes $\mathcal{C}_p$ and consist to identify stochastic processes U_p [14]. However, observations data shown that, it is not judicious technically to model the vectorial stochastic process U_p.

3.1 Vectorial stochastic process U_p

For each $\mathcal{C}_p$, $p \in \mathcal{P}$, previously defined, we associate a unique $\mathbb{R}^6$-valued stochastic process $U_p = (U_p(t), t \in \mathbb{R})$ defined on $(\Omega, \Im, \mathbb{P})$ such as $\forall t \in \mathbb{R}$:

$$Up(t) = (X_1(t), X_2(t), V_1(t), V_2(t), \Gamma_1(t), \Gamma_2(t))^T \qquad (12)$$

The process U_p describes the random behavior of the vehicle dynamic system. It is partially known through the trajectories of $\mathcal{C}_p$ to which it is associated. The components X_1, X_2, V_1, V_2, Γ_1 and Γ_2 are indexed processes on $\mathbb{R}$ in values of $\mathbb{R}$. They describe respectively the vehicle positions, velocities and acceleration in the basis $\mathcal{B}_0^A$. For reasons of notations simplification, we consider :

$$U = U_p \qquad (13)$$

However, the objective of this work is to predict the failure trajectory by using reliability analysis. This analysis requires to choose criteria of failure. These criteria relate on the processes of control resulting from the translation of studied problem in reliability analysis language. It consists to make a functional projection of U on $\mathbb{R}$ in order to obtain Z.

3.2 Formulation of the process Z

A process of control associated with the stochastic process U is a $\mathbb{R}$-valued process $Z = (Z(t), t \in \mathbb{R})$ defined on $(\Omega, \Im, \mathbb{P})$ such as:

$$Z(t) = F(U(t)) \ , \ \forall t \in \mathbb{R} \qquad (14)$$

Where F is a functional of $\mathbb{R}^6$ in $\mathbb{R}$. We suppose subsequently that Z is a second order stochastic process. F is relating to the criterion chosen to characterize the failure trajectory. In this paper, we choose the vehicle **lateral acceleration** as a failure criterion because several studies show that it plays an important role in the vehicle lane crossing accidents.

3.2.1 Failure criterion definition

Lateral acceleration and its variation (Jerk) are criteria often used in the literature to estimate the **dangerousness of trajectory**. For example, Rasmussen [15] proposed that the speed choice strategy of drivers in curves is based on dynamically adjusting a safety margin of lateral acceleration. When entering a given curve, the driver reduces the initial speed to avoid reaching some maximum value in lateral acceleration inside the curve. This maximum lateral acceleration is estimated subjectively by individual drivers, depending on their own driving experience, the road handling performance of their car, road and weather conditions, and personal level of acceptable risk.

Then, under normal driving conditions, a vehicle is able to turn without skidding, provided that its lateral acceleration (usually referred to as "'centrifugal force"' by drivers) is kept below the saturation threshold of its tires (i.e., the maximum grip force defined by the road adherence conditions). This threshold value depends very little on vehicle speed [16]. If drivers actually tried to maximize the performance of vehicle, they would indeed adapt the speed in curves so as to systematically reach this maximum cornering acceleration.

This parameter Γ_N is used as failure criterion because the going beyond a safety threshold (δ^*) can generate vehicle loss of control, see Revue [17]. This function must be limited by δ^* to ensure vehicle stability on the road. This criterion is defined by:

$$sup_{t \in T} |\Gamma_N(t)| > \delta^* \qquad (15)$$

This criterion relates on the $\mathbb{R}$-valued process of control $\Gamma_N = (\Gamma_N(t), t \in \mathbb{R})$ defined on $(\Omega, \Im, \mathbb{P})$. Note that $\Gamma_N(t)$ is easy to determine because it is a coordinate of the trajectory $u(t)$ in the Serret-Frenet frame. We have a relation between the Galilean coordinate system $\mathcal{R}_0^A$ and Serret-Frenet frame $\mathcal{R}_G^{SF}$ in this form:

$$\Gamma_N = -\gamma_1 sin(\psi) + \gamma_1 cos(\psi) \qquad (16)$$

Where ψ is yew angle, γ_1 and γ_2 are vehicle acceleration in Galilean coordinates system. We suppose that Γ_N is a relevant criterion to analyse the failure of driving system. This criterion is represented by Z for following steps in this paper.

$$Z = \Gamma_N \qquad (17)$$

Now, we need to simulate Z but experimental observations of Γ_N show that Z is not a stationary process. We will build a representation allowing to express Z like an affine function of stationary process by using normalization techniques to obtain a process X. We recall that the process X is obtained via different steps in this section. The logical sequence of these steps is illustrated in Fig. 3.

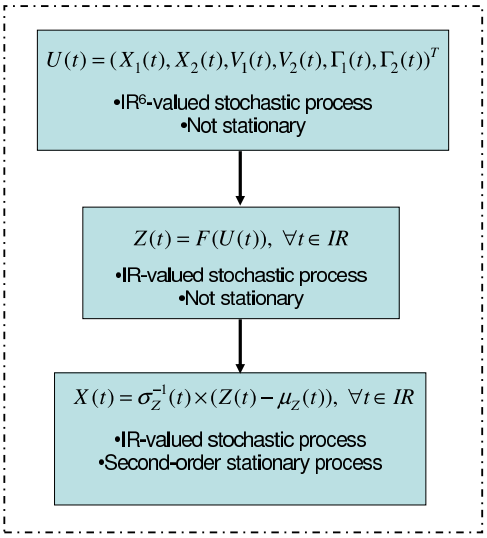

Fig. 3. Approach of stochastic modelling of driving system.

3.3 Scalar process X

We consider the $\mathbb{R}$-valued stochastic process $X = (X(t), t \in \mathbb{R})$ defined on $(\Omega, \Im, \mathbb{P})$ such as:

$$X(t) = \sigma_Z^{-1}(t)\,(Z(t) - \mu_Z(t)) \tag{18}$$

Where $\mu_Z(t) = \mathbb{E}[Z(t)]$ and $\sigma_Z(t) = \mathbb{E}[(Z(t) - \mathbb{E}[Z(t)])^2]$ are respectively the temporal mean and standard deviation of Z. A process is second order stationary if three properties are checked. The first two properties imply that the mean function $\mu_X(t)$ and standard deviation function $\sigma_X(t)$ must be constant. The third property means that the autocorrelation function R_X depends only on the difference between t_1 and t_2 and only needs to be indexed by one variable rather than two variables. This also implies that R_X depends only on $\tau = (t_1 - t_2)$:

$$\begin{cases} \mathbb{E}[X(t)] = \mu_X \ \ (constant)\ ;\ \forall t \in T \\ \mathbb{E}[(X(t) - \mathbb{E}[X(t)])^2] = \sigma_X \ \ (constant)\ ;\ \forall t \in T \\ Cov(X(t_1), X(t_2)) = C_{XX}(\tau);\ \forall t_1 \forall t_2; \end{cases} \tag{19}$$

By constructing, the stochastic process X is such as $\mu_X(t) = 0$ and $\sigma_X(t) = 1$. A statistical study of X, for the considered example shown that these assumptions are approximately checked and therefore X can be reasonably considered as a second-order stationary process. Now, we are going to describe the characterization of X. For that, it is necessary to calculate the first order marginal density function p_X and power spectral density function S_X by using real observations data. The marginal density was estimated by standard estimator and gaussian kernel estimator. The power spectral density was estimated by Welch estimator with Hamming window, [18]. For autocorrelation function, we used estimator based on the periodogram [19].

Then, we approximate the *estimate* of p_X, F_X and S_X by theoretical laws. These approximations are necessary to simulate the stochastic process $X = (X(t), t \in \mathbb{R})$. For the first order margin density approximation of X, two great ways exist. Firstly, we use an approximation in the class of usual laws [20]. The experimental data are shown that this technique is not sufficient for laws approximation. Secondly, we build an hilbertian approximation on a functional basis. We choose the basis of Hermite polynomials, see [19].

3.3.1 Approximation of X on Hermite polynomial basis

To define Hermite polynomials, we consider the Hilbert space $L^2(\mathbb{R}, \nu)$. It is the real function space defined $\nu-$almost everywhere on $\mathbb{R}$ and square-integrable by report to the canonical spectral measure $\nu(dx) = f(x)dx$, bearing of scalar product:

$$((\varphi_1, \varphi_2)) = \int_{\mathbb{R}} \varphi_1(x)\varphi_2(x)f(x)dx \tag{20}$$

Where $f(x) = \frac{1}{\sqrt{2\pi}} exp\left(-x^2/2\right)$, $x \in \mathbb{R}$ is probability density function of the standard gaussian law.

The Hermite polynomial $H_\alpha(x)$ in $\mathbb{R}$, index $\alpha \in \mathbb{N}$ are defined by the recurrence relation:

$$\begin{cases} H_0(x) & = 1 \\ H_{\alpha+1}(x) & = xH_\alpha - \frac{d}{dx}H_\alpha(x) \end{cases} \tag{21}$$

The derivate of $H_\alpha(x)$ by report to x can be written :

$$\frac{d}{dx}H_\alpha(x) = \alpha H_{\alpha-1}(x) \tag{22}$$

The Hermite polynomials are also defined by:

$$H_m(x) = \int_{\mathbb{R}} (x + iy)^m f_1(x)dx \tag{23}$$

Let $p_X : \mathbb{R} \to \mathbb{R}_+$ be the probability density of target. We seek p_X in this form :

$$p_X(x) = f(x) \sum_{\alpha=0}^{+\infty} q_\alpha h_\alpha(x) \tag{24}$$

Where $q_\alpha \in \mathbb{R}$ and $(h_\alpha)_\alpha$ represent the orthonormal basis of normalized Hermite polynomials, with for each $\alpha \in \mathbb{N}$:

$$h_\alpha(x) = (\alpha!)^{-1/2} H_\alpha(x) \tag{25}$$

They form a orthonormal basis of $L^2(\mathbb{R}, \nu)$ and $(\alpha!)^{-1/2}$ is a normalization factor such that $((h_j, h_k)) = \delta_{jk}$.

From Eqn.(24), we can carry out an approximation at order N in this form :

$$p_X^N(x) = f(x) \sum_{\alpha=0}^{N} q_\alpha h_\alpha(x) \tag{26}$$

Where N is an integer ≥ 0 to determine.

To estimate the coefficients $(q_\alpha)_{1<\alpha<N}$, we used the *projection* of p_X on the basis of Hermite polynomials.

3.3.2 Estimation of coefficients q_α by margin law integration.

The marginal law P_X can be written:

$$p_X(x)dx = q(x)\nu(dx) \tag{27}$$

Where $q(x) = p_X f^{-1}(x)$. In the particular case where $q \in L^2(\mathbb{R}, \nu)$, the development of q can be written on hilbertian basis $(h_\alpha)_\alpha$.

$$q(x) = \sum_{\alpha=0}^{+\infty} q_\alpha h_\alpha(x) \tag{28}$$

The series of right member of the Eqn. (24) is convergent in $L^2(\mathbb{R}, \nu)$. For each index $\alpha \in \mathbb{N}$, the real q_α is obtain by projection on the basis $(h_\alpha)_\alpha$ in the form:

$$\begin{aligned} q_\alpha &= ((q, h_\alpha)) \\ &= \textstyle\int_{\mathbb{R}^N} q(x) h_\alpha(x) f(x) dx \end{aligned} \tag{29}$$

With $h_\alpha(x) = 1$ for $\alpha = 0$, the normalization condition $\int_{\mathbb{R}} p_X(x)dx = 1$ implies $q_\alpha = 1$ for $\alpha = 0$.
Using a formula (29), the coefficients $(q_\alpha)_{1<\alpha<N}$ can be expressed as follow :

$$\bar{q}_\alpha = \int_{\mathbb{R}} \hat{p}_X(x) h_\alpha(x) dx \tag{30}$$

Where $\hat{p}_X$ is an estimate of the density p_X, obtained from the experimental statistical sample of process X marginal law. To get the numerical values of $(q_\alpha)_{1<\alpha<N}$, we used numerical scheme for integrating such as Simpson technique.

Then, after having truncated the development of probability density on Hermite polynomials basis with a reasonable order N, we obtained an approximation p_X^N of the first order marginal density p_X of X. In the large majority of treated cases the truncation order of the Hermite development did not exceed $N = 7$.

The approximation result of p_X could be better or less better which is due to the regularity and/or the smoothness of the estimated density of p_X. If it presents too many fluctuations, the Hermite method of approximation can not give good performance in spite of a high order α of truncation. If the development on the basis of Hermite polynomials would not be sufficient, to improve the quality of approximations, we will use the method of Edgeworth basis. It consists to gather the terms $(q_\alpha)_{0<\alpha<N}$ having the same order of magnitude to approximate the target probability density p_X as well as possible, [21].

We note $\tilde{p}_X$, $\tilde{F}_X$, $\tilde{S}_X$, $\tilde{R}_X$ the estimate functions obtained from approximation of the functions p_X, F_X and S_X. From these quantities previously, we can simulate the process X by using the stochastic process simulation methods, (as in [22] and [23]).

3.4 Simulation of process X

Let us recall that our aim is to simulate a good approximation of X. We use a simulation method for non-Gaussian stochastic process based on the use of Hermite polynomials. This method requires only the knowledge of F_X and R_X. This step is fundamental and the main steps of this method is described:

Let $F_{\mathcal{N}}$ be the cumulative distribution function of a $\mathbb{R}-$valued standard Gaussian random variable. We suppose that $F_N^{-1} \circ F_{\mathcal{N}}$ belongs to the Hilbert space:

$$L^2(\mathbb{R}) = f : \mathbb{R} \to \mathbb{R} | \int_{\mathbb{R}} |f(x)|^2 dx < +\infty \tag{31}$$

Equipped with the scalar product:

$$< f, g >= \int_{\mathbb{R}} f(x) g(x) \frac{exp(-x^2/2)}{\sqrt{2\pi}} dx \tag{32}$$

$\forall f \in L^2(\mathbb{R}), \forall g \in L^2(\mathbb{R})$

$F_X^{-1} \circ F_{\mathcal{N}}$ can be projected on the basis $(h_n)_{n\in\mathbb{N}}$ of L^2.

This basis is constituted by the normalized polynomials of Hermite. Thus we have:

$$F_X^{-1} \circ F_{\mathcal{N}} = \sum_{n\in\mathbb{N}} f_n h_n \tag{33}$$

With $\forall n \in \mathbb{N}$:

$$f_n = \int_{\mathbb{R}} (F_X^{-1} \circ F_{\mathcal{N}})(x) h_n(x) \frac{exp(-x^2/2)}{\sqrt{2\pi n!}} dx \qquad (34)$$

Let G be a $\mathbb{R}-$valued zero mean stationary standard Gaussian process on $(\Omega, \Im, \mathbb{P})$. Let F_G be the cumulative distribution function associated with G. Let R_G be the autocorrelation function associated with G. Let $L = (L(t), t \in \mathbb{R})$ be a $\mathbb{R}-$valued second order stationary stochastic process defined on $(\Omega, \Im, \mathbb{P})$ such that:

$$L(t) = (F_X^{-1} \circ F_{\mathcal{N}})(G(t)), \forall t \in \mathbb{R} \qquad (35)$$

This process possesses the following properties:

1. The probability distribution of L has F_X for cumulative distribution function
2. L has the first and second same moment than X
3. $L(t) = \sum_{n \in \mathbb{N}} f_n h_n(G(t)), \forall t \in \mathbb{R}$
4. $R_L(t) = \sum_{n \in \mathbb{N}} f_n^2 (n!)(R_G(t))^n, \forall t \in \mathbb{R}$

Consequently, L is an approximation of X.

In practice we approach X by a process $L^M = (L^M(t), t \in \mathbb{R})$ whose autocorrelation function is $\tilde{R}_X$. This process is defined on $(\Omega, \Im, \mathbb{P})$ and:

$$L^M(t) = \sum_{n=0}^{M} f_n h_n(G(t)), \forall t \in \mathbb{R} \qquad (36)$$

With $M \in \mathbb{N}^*$ fixed, $(f_n)_{n=0,\ldots,M}$ given by Eqn. (34) and G is previous Gaussian process.

To make an approximation of X we shall simulate L^M by using Eqn. (36). The problem becomes the simulation of process G. For it, we use a classical method [18] based on the use of the PSD S_G of G. We chose to take $S_G = \tilde{S}_X$. Now, we call back the steps of the method to simulate S_G:

In practice, the PSD S_G is only known at the points $\lambda_j = -\lambda_L + (j + 1/2)\Delta\lambda, j \in 0, 1, \ldots, N-1$ of a regular $N-$partition of $[-\lambda_L, \lambda_L]$. We have:

$$S_G \equiv (S_G(\lambda_j), j = 0, 1, \ldots, N-1) \qquad (37)$$

Where λ_j and $\Delta\lambda$ (time increment) are imposed by the Shannon's rule.

The proposed algorithm of simulation is based on the spectral representation of G. Let $(\phi_j)_{j=0,1,\ldots,N-1}$ be a family of $\mathbb{R}-$valued independent uniform on $[0, 2\pi]$ random varaibles on $(\Omega, \Im, \mathbb{P})$. Then, we have the following result:

The $\mathbb{R}-$valued process $G_N = (G_N(t), t \in \mathbb{R})$ defined on $(\Omega, \Im, \mathbb{P})$ such as:

$$G_N(t) = \sqrt{2\Delta\lambda} Re \left[\sum_{j=0}^{N-1} e^{i\lambda_j t} e^{i\phi_j} \sqrt{S_G(\lambda_j)} \right] \qquad (38)$$

Converge in distribution to G. The algorithm of simulation is obtained by sampling the approximation G_N of G at the points $t_j = j\Delta t, j = 0, \ldots, N-1$ of the domain T of simulation. It appeals to the notion of FFT (Fast Fourier Transform). With this result, we obtain the simulation results of L^M. We point out that L^M is an approximation of the process X.

Now the use of the simulations of X and Eqn. (39) allowed to obtain the simulations of Z. We point out that the estimate of $\mu_Z(t)$ and $\sigma_Z(t)$ are available by using observations data of Γ_N. The Eqn. (39) is obtained by taking into account the relation (18).

$$Z(t) = \mu_Z(t) + \sigma_Z(t) X(t) \qquad (39)$$

The simulations of Z are necessary in reliability analysis section.

4 RELIABILITY ANALYSIS OF V-I-D SYSTEM

The objective of this secion is to evaluate the risk of failure trajectory according to each class $\mathcal{C}_p$. For that, the vehicle lateral acceleration Γ_N is chosen as a criterion of failure. This criterion relates to the function $\Gamma_N = (\Gamma_N(t), t \in \mathbb{R})$. This function represented by Z is a $\mathbb{R}$-valued process of control defined on $(\Omega, \Im, \mathbb{P})$. The process Z was partially identified in previous section. From Z, we define the variable of control, the safety margin and the associated limit state. Then, we identify the law for maxima according to each Z. Finally, we will estimate the probability of failure noted P_f relating to each $\mathcal{C}_p$.

4.1 Variable of control

It is a variable such as the values it takes with beyond an acceptable threshold δ^* characterizes a failure trajectory of driving system. Let Y be a variable of control defined on $(\Omega, \Im, \mathbb{P})$ in values of $\mathbb{R}$ such as:

$$Y = sup_{t \in T} |Z(t)| \qquad (40)$$

We remember that $Z(t) = F(U(t)), \forall t \in \mathbb{R}_+$, where $F : \mathbb{R}^6 \to \mathbb{R}_+$ is a functional who operates on the stochastic process U. The simulations resulting from Z enable to obtain the realizations of Y through the relation (40). From Y, we associate a safety margin.

4.2 Margin of safety

For each criterion of safety is associated a margin of safety noted M. It's a random variable define on $(\Omega, \Im, \mathbb{P})$ in values of $\mathbb{R}$. This variable is a scalar which measures in a point of the mechanical system, the difference between Y and an acceptable value δ^* of Y. When δ^* is fixed, we obtain the observations of M by the relation :

$$M = \delta^* - Y \tag{41}$$

These observations are known because, one had previously the realizations of Y. The safety margin is also used to define the events of safety noted E_S and of failure E_F as we will define.

4.3 Limit state function G

The limit state function G is defined as linear or non-linear separation between safe and failure domains. It defines the events E_s and E_f associated with the states **reliable** and **failing** such as:

$$G(y) = \delta^* - y \tag{42}$$

For each $y \in \mathbb{R}, G(y) > 0$ characterise reliable state and $G(y) < 0$ failing state. Consequently, E_s and E_f are defined by :

$$\begin{cases} E_s = \{\omega \in \Omega : G(Y(\omega)) > 0\} \\ E_f = \{\omega \in \Omega : G(Y(\omega)) < 0\} \end{cases} \tag{43}$$

and check :

$$E_s \cup E_f = \Omega \;\; ; \;\; E_s \cap E_f = \oslash \tag{44}$$

With these two events are associated the domains of safety D_S and failure D_F of the model, such as :

$$\begin{cases} D_s = \{y \in \mathbb{R} : G(y) > 0\} \\ D_f = \{y \in \mathbb{R} : G(y) < 0\} \end{cases} \tag{45}$$

and who check :

$$D_s \cup D_f = \mathbb{R} \;\; ; \;\; D_s \cap D_f = \oslash \tag{46}$$

We make assumption that limit state $G(Y) = 0$ is not included in the failure domain and $G(Y) = M$. The margin M associated to a limit state function represents an indicator of safety. The next step will consist to identify the law of M, necessary to estimate the probability of failure P_f.

4.4 Identification of the law of M

The random variable M is defined by the law P_M on $(\mathbb{R}, \beta)$, which will be supposed to admit a density p_M by report to measure of Lebesgue dy on $(\mathbb{R}, \beta)$:

$$P_M(dy) = p_M(y)dy \tag{47}$$

That is

$$\forall B \in \mathcal{B} \; , \; P_M(B) = \int_B p_M(y)dy \tag{48}$$

For obvious reasons to do with (48), this density must verify:

$$Supp(p_M) \subset D_G \tag{49}$$

Where $Supp(p_M)$ indicates the support of p_M, D_G the domain of definition. We estimate the probability density function p_M of M by using the realizations of M. It is necessary to find a good approximation of p_M. It can be approximated by extreme values laws such as: Gumbel, Weibull or Frechet. These extreme values laws are stable. However, these 3 types of extreme values law are not exhaustive. Other laws not checking the stability principle can exist in some cases (rare in practice). Their interest would be limited because of instability of their form. If extreme values laws would not be adapted, one can make an approximation by a development on the Hermite polynomial basis. From a good approximation of p_M, we estimate the failure probability $P_f = \mathbb{P}(E_F)$.

4.5 Estimation of the failure probability P_f

The **probability of failure** requires the simulation of M and consequently of model Z. Note that the simulation of Z was made previously. By fixing $\delta^* \in \mathbb{R}_+^*$, we can estimate P_f by:

$$P_f = \mathbb{P}(M < 0) = \int_{-\infty}^{0} p_M(x)dx \tag{50}$$

The estimation of P_f may be made by using *Monte Carlo methods*. We can also obtain Eqn.50 by using *distribution function* of M, see [24]. Note that one of the **difficulties** for application part is to choose the threshold δ^*. We propose to vary δ^* in an acceptable values interval and to estimate P_f. It is the function $\pi_p(\delta^*)$ corresponding with the probability to accept a class $\mathcal{C}_p$ containing a number of trajectories exceeding the safety threshold δ^*. This variation allows to identify *geometrical rupture* (linear and nonlinear zone of variation) for each $\mathcal{C}_p$. From $\pi_p(\delta^*)$, we will compare the failure probability P_f for the classes $\mathcal{C}_p$, with $p \in \mathcal{P}$. This comparison will permit to show that risk level of failure trajectory is different according to classes. We will use real observed trajectories to build abacuses $\pi_p(\delta^*)$ of failure probability per class $\mathcal{C}_p$.

By using recognition model necessary to assign each trajectory u to its membership class $\mathcal{C}_p$, we obtain the risk level of failure trajectory in bend's entry.

5 APPLICATION ON IFSTTAR/Nantes BEND

The objective of this section is to apply the proposed methodology. After experiment conditions description, we will present clustering results, stochastic models validation, reliability analysis result and discussion.

5.1 Experiment conditions

Experiment took place on the IFSTTAR test track, located in Nantes (France). We used a test car to record vehicle kinematics and dynamic parameters at discrete time (frequency $100Hz$). On board the test car, a set of high-precision accelerometers, gyrometers, GPS sensors were used. In order to avoid obstructing the road traffic circulation, the test car was used on a closed road.

We acquire a set of trajectories physically carried out with various instructions and various drivers. We mobilized 35 voluntary drivers with 10 trajectories per driver, which makes a set of 350 trajectories to be analyzed. The participants are selected by the following criteria (age, gender, driving experience). To carry out these tests, 2 instructions were given to drivers: a **fast driving** in order to minimize run time and a **normal driving** in order to maximize passenger comfort. For each instruction, the driver is supposed to be on an open road and must adhere to the highway code. To avoid habituation effects with a given instruction, the two instructions are alternated.

The reader will find in Koita [25] more information on experimental data acquisition. The collected data is represented by the dataset $\mathcal{D}$.

5.2 Trajectories clustering result

From the real trajectories in $\mathcal{D}$, a statistical analysis (clustering) was used to identify 4 classes of trajectories with an optimal separability between classes. These classes are stable with regard to the number of iterations. The Fig. 4 illustrates the projection of clustering result on two components (or axis) of u.

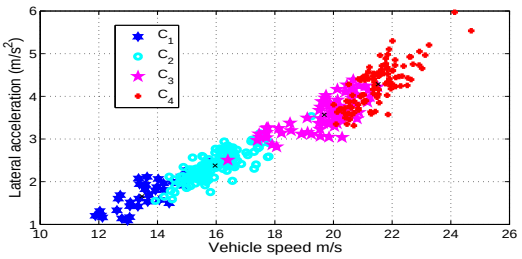

Fig. 4. The four identified classes obtained by k-means algorithm.

The result in Tab. 1 shows that the dynamic of classes is very different. Consequently, the behaviour of control is not the same according to classes.

Table 1. Identified classes description.

Classes	*Number*	ΔC_{max}	Γ_{max}	V_{max}
$\mathcal{C}_1$	35	$7\ km^{-1}$	$2.9m/s^2$	$18m/s$
$\mathcal{C}_2$	125	$5.1\ km^{-1}$	$3.80m/s^2$	$20.5m/s$
$\mathcal{C}_3$	72	$3.4\ km^{-1}$	$5.1m/s^2$	$24m/s$
$\mathcal{C}_4$	107	$2.9\ km^{-1}$	$6.5m/s^2$	$26m/s$

Starting from the real trajectories, we built stochastic models X specific to each $\mathcal{C}_p$ by respecting the procedure describe in section 3.

5.3 Validation of stochastic models X

The processes U and Z are not stationary, hence the need to use the process X. In this study, the *lateral acceleration* is chosen to characterize failure criterion. The Fig. 5 represents evolution of ω-realization of X. It is specific to the class $\mathcal{C}_2$.

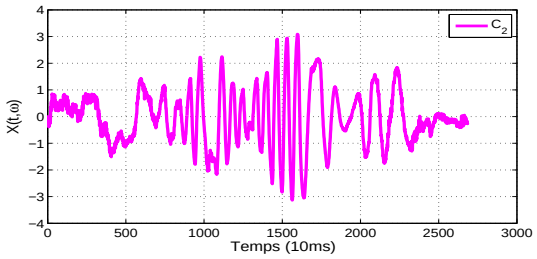

Fig. 5. Evolution of ω-realization of X

We are going to check stationarity assumptions of X. The Fig. 6 shows a fast decrease of the autocorrelation function $R_X(\tau)$. The variation of R_X depends on $\tau = (t_1 - t_2)$. In addition, stationarity tests are used. We conclude that X is Second-order stationary process.

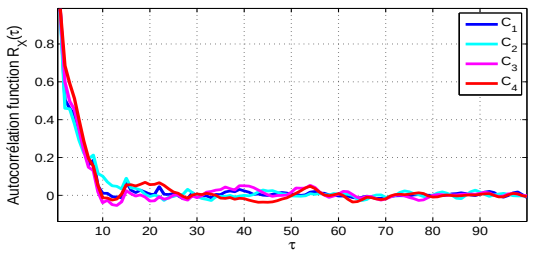

Fig. 6. Autocorrelation function R_X for each class $\mathcal{C}_p$.

Afterwards, we estimate the probability density function p_X. After this statistical estimation, we made an approximation by usual laws and the development of p_X in the basis

of Hermite polynomials. Whereas, the usual laws give less precise approximations in tail of distribution than Hermite method of approximation. The precision of the models $X = (X(t), t \in \mathbb{R})$ depends as well on the central tendency than tails of distribution. In this study, the Hermite method is used but one does not guarantee a positive approximation of p_X to a high-order of truncation α due to the multiplicity of the Hermite polynomials $h_\alpha(x)$.

To validate the models X, we compared the statistical characteristics (p_X, S_X) of real measurements and simulations data. The Fig. 7 shows the superposition of the graphs of p_X between data measurements and simulations.

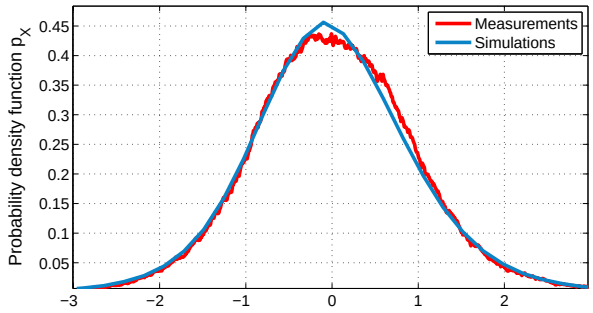

Fig. 7. Comparison of the p_X graph for class $\mathcal{C}_2$.

In addition to the graphic comparison in Fig. 7, statistical tests of law conformity are used. The null hypothesis $\mathcal{H}_0$ is not rejected for a $p-$value of 5%, [26].

In the Fig. 8, we note again a superposition of power spectral density S_X between measurements and simulations. This superposition is still more precise for $\mathcal{C}_2$ and $\mathcal{C}_4$ than $C1$ and $C3$. This is still with the representativeness of these classes in real trajectories.

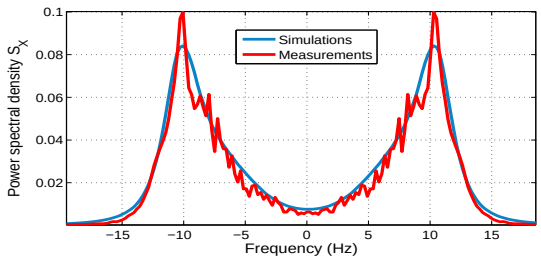

Fig. 8. Comparison of the S_X graph for $\mathcal{C}_2$.

These results made it possible to validate the identification of the process X and consequently the validation of the model Z. We can consider that the data simulated from Z have the similar statistical properties like the real trajectories in each $\mathcal{C}_p$. This result checks also an assumption posed at the beginning of this study. It consists to suppose that all trajectories in each $\mathcal{C}_p$ are resulting from the same stochastic process U_p. Then, the simulations of Z are used to identify the law of p_M in reliability analysis.

5.4 Reliability analysis result

The objective is to evaluate the failure probability of trajectories. Starting from an estimate of p_M, we can approximated p_M by the Gumbel law for $\mathcal{C}_2$, see Fig. 9. By using

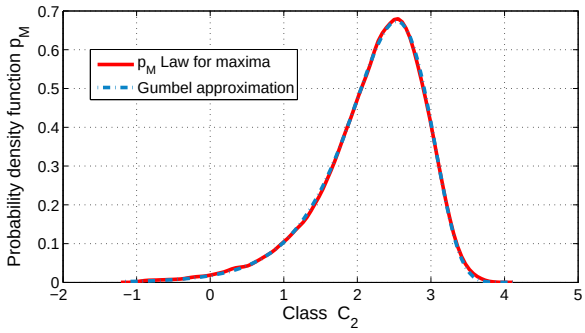

Fig. 9. Target law approximation p_M by Gumbel law.

the Kolmogorov-Smirnov test, the null hypothesis $\mathcal{H}_0$ is not rejected for a $p-$value of 5%, see Tab. 2.

Table 2. Statistical test results for extreme law approximation.

	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$
$\mathcal{H}_0$	0	0	0	0
$p-value$	0.10	0.52	0.08	0.166

The Gumbel distribution function is used to calculate the failure probability P_f. The Fig. 10 represents the evolution of P_f as function of the safety threshold δ^* according to 4 identified classes.

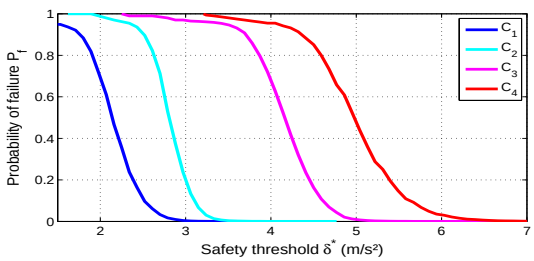

Fig. 10. Evolution of $P_f(\delta^*)$ as function of δ^*.

5.5 Discussion

The classes of trajectories have different levels of risk. We note that P_f decrease quickly with an increase in the value of δ^* for $\mathcal{C}_1$ and $\mathcal{C}_2$ corresponding to a **normal driving**. Whereas it decrease less quickly with an increase in the value of δ^* for $\mathcal{C}_3$ and $\mathcal{C}_4$ corresponding to a **fast driving**. This result makes it possible to conclude that according to the failure criterion (*lateral acceleration for this application*), $\mathcal{C}_1$ is protected of all the identified classes, $\mathcal{C}_4$ is riskiest and $\mathcal{C}_3$ has a risk more high than $\mathcal{C}_2$. An order relation is noted between identified classes, in this form :

$$\mathcal{C}_1 \prec \mathcal{C}_2 \prec \mathcal{C}_3 \prec \mathcal{C}_4 \tag{51}$$

We note that the graphs in Fig. 10 are the abacuses functions $\pi(\delta)$. From these functions, when the membership class $\mathcal{C}_p$ of each trajectory u is known, one cans obtain its probability of failure P_f.

In the earlier study, we describe the implementation of a robust recognition model. This model makes it possible to predict the membership class $\mathcal{C}_p$ of each trajectory u in the bend's entry. In this paper, a relation was established between $\mathcal{C}_p$ and P_f according to δ^*. The classes are homogeneous in probability of failure. The prediction of the membership class will make it possible to know P_f for each trajectory u. The results showed contrary to the vehicle dynamic modeling where parameters identification is difficult that it is possible to model the vehicle trajectory parameters with high-precision by using some real observed trajectories. This new approach based on observations data makes a better use of reliability analysis of failure trajectory. The readers can find more description of this work in Koita [27].

6 CONCLUSION

The objective of this study was to predict failure trajectory, starting from a given bend configuration and representative observations of real trajectories.

For this application, we have identified 4 classes of trajectories by using clustering methods on experimental data. The trajectories of each class $\mathcal{C}_p$ are regarded as realizations of stochastic process U_p. This process offers a realistic description for the observed random variability of driving system. Through the observations data, we checked that the processes U and Z are not stationary. However, stationary assumptions on second-order of the process $X = (X(t), t \in \mathbb{R})$ are checked.

Then, adequate approximations were made for X. From simulations of X, the models Z are validated. And the simulations of Z permitted also to estimate the failure probability P_f for each $\mathcal{C}_p$. The comparison of P_f function allowed to associate a risk level to each $\mathcal{C}_p$.

Lastly, by using the trajectories recognition model, each trajectory u is assigned to its membership class $\mathcal{C}_p$ of which one calculated its probability P_f beforehand. With this methodology and in this bend configuration, we are able to predict a dangerous trajectory according to its lateral acceleration. This information is an indication for driver to control his vehicle or at least to be vigilant on its control. It cans also concern an infrastructure manager to identify dangerous drivers in the traffic. So the present results of this new approach are promising for road safety. This proposed methodology is related to the data acquisition system quality and not to the studied mobile. Consequently, it can be extended almost to all kinds of mobile objects (motorcycles, pedestrians,...). At the same time, it is relatively flexible and thus able to adapt to an evolving environment (bend, intersection,...).

Achnowledgements

This work was supported by the French Ministry of Transport (MEDDTL). The authors would like to thank F. Peyret, M. Ortiz and D. Betaille (IFSTTAR/MACS) for their contribution during experiment in Nantes.

References

[1] IRTAD. "International trafic safety data and analysis: Roads accident statistics". *Journal Name*.

[2] Basset, M., Baujon, J., and Gissinger, G.-L. "Intelligent driver aid system based on trajectory reconstruction for active safety". *IFAC World Congress*.

[3] Mammar, S. "Lateral control assistance and automation of vehicles: approaches by robust command". *University of Evry, Complex Systems laboratory*.

[4] Lauffenburger, J., Basset, M., Coffin, F., and Gissinger, G. "Driver-aid system using path-planning for lateral vehicle control". *Journal of Control Engineering Practice*.

[5] Gibson, J., and Crooks, L. "A theoretical field-analysis of automobile-driving". *American Journal of Psychology,* **51**.

[6] Godthelp, H. "Vehicle control during curve driving". *Human Factors,* **28**.

[7] Rasmussen, J. "Information processing and human-machine interaction". *An approach to cognitive engineering*.

[8] Boubezoul, A., Koita, A., and Daucher, D. "Vehicle trajectories classification using support vectors machines for failure trajectory prediction". *IEEE/ACTEA'09*.

[9] Diday, E. "Nouvelle mthode en classification automatique et reconnaissance des formes. k-mean method". *Journal of Applied Statistics*.

[10] Comaniciu, D., and Meer, P. "Mean shift analysis and applications".

[11] Kohonen, T., 1986. Learning vector quantization for pattern recognition. Technical report tkk-f-a601, University.

[12] Duda, R. O., Hart, P. E., and Stork, D. G., 2001. *Pattern Classification*. John Wiley &Sons.

[13] Koita, A., and Daucher, D. "Stochastic analysis of vehicle trajectories in bend: toward a risk indicator construction".

[14] Bouleau, N., 1988. *Stochastic processes and applications*. Hermann.

[15] Reymond, G., Kemeny, A., Droulez, J., and Berthoz, A. "Role of lateral acceleration in curve driving: Driver model and experiments on a real vehicle and a driving simulator". *Human Factors and Ergonomics Society,* **43**(3), p. 483495.

[16] Gillespie, T. "Fundamentals of vehicle dynamics". *Warrendale, PA: Society of Automotive Engineers*.

[17] Revue, A., 2003. "Contribution of geographic information systems to road safety: multisensors approach to evaluate risks". PhD Thesis, University of Mines Paris.

[18] Fogli, M., 2000. Statistical mechanics. Dea course, University of Blaise Pascal in Clermont-Ferrand.

[19] Soize, C., 1993. *Mathematical methods for signal analysis*. MASSON.

[20] Bouleau, N., 2002. *Engineering probabilities: random variables and simulation*. Hermann.

[21] Pugachev, V., 1982. *Probability Theory and Mathematical Statistics*.

[22] Poirion, F., Puig, B., and Soize, C. "Non-gaussian simulation using hermite polynomial expansion". *Computational Stochastic mechanics,* **1**.

[23] Puig, B., 2003. "Modelling and Simulation of non-Gaussian processes". PhD Thesis, University of Paris 6.

[24] Lemaire, M., 2005. *Structural reliability, couplage mecano-fiabiliste statique*. Hermes.

[25] Koita, A., and Daucher, D., 2008. Mise en oeuvre d'un protocole exprimental. Technical report, LCPC.

[26] Saporta, G., 1978. *Statistical theory and methods*. Technip, Paris.

[27] Koita, A., 2011. "Probabilistic evaluation of the vehicles trajectories dangerousness in bend". PhD Thesis, University of Blaise Pascal, Clermont-Ferrand.