



HAL
open science

Evaluation probabiliste de la dangerosité des trajectoires de véhicules en virages

Abdourahmane Koita

► **To cite this version:**

Abdourahmane Koita. Evaluation probabiliste de la dangerosité des trajectoires de véhicules en virages. Sciences de l'ingénieur [physics]. Université Blaise Pascal - Clermont-Ferrand II, 2011. Français. NNT: . tel-00626964v1

HAL Id: tel-00626964

<https://theses.hal.science/tel-00626964v1>

Submitted on 27 Sep 2011 (v1), last revised 24 Jan 2012 (v2)

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.

N° d'ordre : D.U : 2113

E D S P I C : 517

UNIVERSITÉ BLAISE PASCAL - CLERMONT II

ECOLE DOCTORALE

SCIENCES POUR L'INGENIEUR DE CLERMONT-FERRAND

THÈSE

Présentée par :

Abdourahmane KOITA

Ingénieur en robotique

pour obtenir le grade de

DOCTEUR D'UNIVERSITÉ

SPÉCIALITÉ : Mécanique aléatoire

Titre de la thèse :

**Évaluation probabiliste de la dangerosité des trajectoires
de véhicules en virages**

Soutenue publiquement le 23 mars 2011 devant le jury :

M. Michel BASSET,	Professeur, Université Haute Alsace	Rapporteur
M. Fabrice POIRION,	Directeur de recherche-HDR, ONERA	Rapporteur
M. David CLAIR,	Maître de conférences, UBP/C. Ferrand	Examineur
M. Bernard JACOB,	IGPEF, Direction scientifique, IFSTTAR	Examineur
M. Michel FOGLI,	Professeur, UBP/Clermont Ferrand	Directeur de thèse
M. Dimitri DAUCHER,	Chargé de recherche , IFSTTAR	Conseiller d'étude

Les remerciements

Je remercie vivement Madame H el ene Jacquot-GUIMBAL, directrice g en erale de l' tablissement IFSTTAR, de m'avoir donn e l'opportunit e de r ealiser ce travail de th ese au sein de ce laboratoire et surtout pour les moyens mis en oeuvre durant ces ann ees de th ese.

J'adresse mes vifs remerciements   Monsieur Michel BASSET et Monsieur Fabrice POIRION pour avoir accept e d' tre rapporteurs de ce travail de th ese. Je tiens  galement   remercier Monsieur Bernard JACOB et Monsieur David CLAIR de m'avoir fait l'honneur de participer   ce jury de th ese.

Je tiens   exprimer mes remerciements les plus sinc eres   Monsieur Michel FOGLI, mon directeur de th ese, pour m'avoir encadr e, conseill e et soutenu tout au long de la th ese avec rigueur et professionnalisme, et pour la confiance qu'il m'a accord e. Je le remercie  galement de m'avoir ouvert la porte de son savoir immense.

J'adresse  galement un tr es grand merci   Monsieur Dimitri DAUCHER pour son accompagnement tout le long de ce travail de th ese. Sa grande disponibilit e, son aide pr ecieuse et les discussions fructueuses que nous avons eues, ont  t e des facteurs d eterminants pour me permettre de r ealiser ce travail dans de bonnes conditions. Pour tout cela, je lui exprime mes sinc eres remerciements.

Je tiens   adresser mes remerciements   Monsieur Michel BRY, Monsieur Victor DOLCEMASCOLO et Monsieur Bernard JACOB qui sont   l'origine de ce sujet de th ese, pour leurs pr ecieux soutiens, encouragements et conseils avis es.

Le travail pr esent e ici est le r esultat d'une collaboration scientifique, impliquant en particulier des unit es de recherche et des CETE. Ainsi, je tiens   remercier Monsieur Fran ois PEYRET, Monsieur David BETAILLE et Monsieur Miguel ORTIZ de l'unit e MACS, pour m'avoir aid e   r ealiser les exp erimentations qui ont servi   recueillir les  chantillons de mesures sur lesquels j'ai travaill e, ainsi que pour les remarques pertinentes qui m'ont permis de progresser dans mes recherches. Mes remerciements s'adressent aussi   Monsieur Eric VIOLETTE et l'ensemble des partenaires de l'op eration de recherche MTT pour leur contribution   ce travail de th ese.

Je remercie  galement Monsieur Fabrice POIRION pour m'avoir accueilli chaleureusement au sein de son laboratoire   l'ONERA, et pour les discussions constructives que nous avons eues.

Je remercie Monsieur Abderrahmane BOUBEZOUL pour les discussions fructueuses que nous avons eues, à propos des techniques d'apprentissage de type SVM.

Je remercie Monsieur Bernard JACOB pour non seulement sa relecture attentive de ce manuscrit, et pour l'interprétation théorique des résultats expérimentaux présentés.

Ce travail de thèse s'est déroulé au sein de l'unité LEPSIS. J'ai pris beaucoup de plaisir à effectuer ce travail. Outre l'aspect purement scientifique, qui procure parfois des joies indicibles, c'est aussi et surtout l'environnement humain qui m'a permis de mener à bien ma tâche, de faire face aux difficultés, et de passer de bons moments. Je tiens à remercier son chef, Monsieur Didier AUBERT et l'ensemble de mes collègues pour la bonne humeur, les rires, la gentillesse et la disponibilité de tous les instants.

A tous mes amis, qu'ils apparaissent ou non dans la liste, un grand merci !

Ma famille compte aussi beaucoup pour moi, elle a été présente et d'un soutien sans faille. A ma mère, à mon père, mes frères et soeurs, afin de les remercier du fond du coeur pour tout ce qu'ils ont fait pour moi, et pour leur dire toute ma RECONNAISSANCE.

Enfin, mes derniers remerciements, mais non des moindres, seront destinés à ma femme et mes enfants.

Abdourahmane KOITA

Table des matières

Table des matières	iii
Table des figures	vii
Liste des tableaux	xi
Introduction générale	1
I Système Véhicule-Infrastructure-Conducteur (V-I-C)	7
I.1 Introduction	7
I.2 Analyse accidentologique des véhicules légers	8
I.2.1 Accidentologie	8
I.2.2 Statistiques sur les accidents	9
I.2.3 Origine des accidents en virage	10
I.3 Synthèse des travaux réalisés sur le système VIC	14
I.3.1 Véhicule	14
I.3.2 Infrastructure	18
I.3.3 Conducteur	19
I.4 Description et formulations du système V-I-C	21
I.4.1 Description du système	21
I.4.2 Architecture du système	22
I.4.3 Formulations mathématiques du système	23
I.5 Aperçu de la solution envisagée	26
I.5.1 Définition de la trajectoire du véhicule	26
I.5.2 Aléas (incertitudes) du système V-I-C	29

I.5.3	Définition des outils d'analyse fiabiliste du système V-I-C	30
I.5.4	Reconnaissance de la trajectoire par la méthode des SVMs	36
I.5.5	Résumé de l'approche proposée	37
I.6	Conclusion	38
II	Méthodologie de prédiction de trajectoires à risque	39
II.1	Introduction	39
II.2	Classification des trajectoires u	40
II.2.1	Ensemble des trajectoires u observées	40
II.2.2	Méthodes de classification des trajectoires	41
II.2.3	Transformation $3D-2D$ des données de mesures	48
II.2.4	Validation des p classes de trajectoires identifiées	49
II.3	Modélisation stochastique du système V-I-C	51
II.3.1	Processus d'état U du système	51
II.3.2	Caractérisation du processus X	56
II.3.3	Approximation de la densité marginale d'ordre 1 de X	56
II.3.4	Simulation du processus X	62
II.3.5	Validation des modèles Z	69
II.4	Analyse fiabiliste du système V-I-C	70
II.4.1	Variable de contrôle et critères de sécurité	70
II.4.2	Marge de sécurité M	71
II.4.3	Fonction d'état limite G	71
II.4.4	Identification de la loi de la v.a M	72
II.4.5	Estimation de la probabilité de défaillance P_f	73
II.4.6	Variation de P_f en fonction du seuil δ^*	74
II.4.7	Indice de fiabilité conventionnel β_c	77
II.5	Méthode de reconnaissance des trajectoires u	79
II.5.1	Identification de la zone d'apprentissage	79
II.5.2	Construction du modèle de reconnaissance	81
II.5.3	Validation du modèle \mathcal{M} de reconnaissance	83
II.6	Indicateur de risque du système V-I-C	85
II.6.1	Définition du risque	85
II.6.2	Evaluation du risque du système	86
II.6.3	Classes de fiabilité du système	87
II.7	Conclusion	88
III	Application de la méthodologie sur le virage (LCPC/Nantes)	89
III.1	Introduction	89
III.2	Recueil de données expérimentales	92

Table des matières

III.2.1	Choix d'un dispositif métrologique	92
III.2.2	Descriptif de l'environnement du véhicule	92
III.2.3	Protocole expérimental	93
III.2.4	Illustration d'une trajectoire expérimentale	94
III.3	Classification des trajectoires u observées	96
III.3.1	Résultat de la classification	96
III.3.2	Validation des 4 classes de trajectoires identifiées	97
III.3.3	Interprétation physique des classes	98
III.4	Analyse statistique des processus stochastiques	100
III.4.1	Vérification des hypothèses de régularité de U , Z et X	100
III.5	Modélisation stochastique du processus X	104
III.5.1	Loi marginale d'ordre 1 de X	104
III.5.2	Densité spectrale de puissance S_X	109
III.5.3	Validation numérique des modèles probabilistes Z	112
III.6	Analyse fiabiliste du système V-I-C	118
III.6.1	Identification de la loi de Y	118
III.6.2	Estimation de la probabilité de défaillance P_f	122
III.7	Reconnaissance des trajectoires u observées	127
III.7.1	Identification de la zone d'apprentissage \mathcal{S}	127
III.7.2	Résultats de l'affectation des trajectoires	128
III.8	Fiche technique d'utilisation de la méthodologie	132
III.9	Conclusion	133
IV	Conclusion générale	135
V	Annexes	139
V.1	Stationnarité et ergodicité d'un processus X	139
V.1.1	Stationnarité en moyenne quadratique de X	139
V.1.2	Hypothèse d'ergodicité de X	140
V.2	Estimations de la densité de probabilité p_X	140
V.3	Estimation densité spectrale de puissance S_X	142
V.4	Approximations de lois de probabilité	143
V.4.1	Lois usuelles	143
V.4.2	Développement de la densité sur la base des polynômes d'Hermite	144
V.4.3	Coefficients d'aplatissement et d'asymétrie	145
V.5	Généralités et définitions sur les méthodes SVM	146
V.5.1	Classifieur linéaire	146
V.5.2	Classifieur à marge maximale	150
V.6	Liste des publications	151

Table des figures

1	Illustration du système V-I-C	2
2	Concept de l'indicateur de risque	3
I.1	Comparaison des statistiques d'accidents en France pour 2008 et 2009.	10
I.2	Comparaison des statistiques d'accidents en Europe pour 2005 et 2006.	10
I.3	Répartition des facteurs d'accident liés à l'inadaptation de la vitesse	11
I.4	Relation entre l'accélération latérale et la vitesse longitudinale du véhicule	12
I.5	Modèle bicyclette / Véhicule dans son repère	16
I.6	Schéma des entrées/sorties du système V-I-C	22
I.7	Repère galiléen de référence et repère mobile du véhicule	27
I.8	Architecture de la démarche proposée (1 ^{ère} partie)	37
I.9	Architecture de la démarche proposée (2 nd partie)	37
II.1	Transformation de l'espace d'entrée \mathcal{D} en espace de sortie \mathcal{Y}	42
II.2	Illustration de la méthode des nuées dynamiques	44
II.3	Carte topologique avec plusieurs neurones	46
II.4	Transformation des données	48
II.5	Répartition des trajectoires u observées en p familles	49
II.6	Espace roulant du véhicule.	53
II.7	Distance relative $d(t)$	53
II.8	Résumé des grandes étapes de la modélisation probabiliste proposée	68
II.9	Estimation de la probabilité P_f via l'aire sous la courbe.	74
II.10	Variation de la probabilité P_f en fonction du seuil δ^*	74
II.11	Illustration de l'indice de fiabilité β_{HL}	77

Table des figures

II.12 Hyperplan séparateur d'un classifieur linéaire. La quantité $\frac{b}{\ \theta\ }$ exprime la distance entre l'hyperplan séparateur (H) et l'origine.	82
II.13 Diagramme de Farmer : probabilité-gravité.	85
III.1 Représentation simplifiée du système V-I-C.	89
III.2 Observation des trajectoires u pratiquées sur ce virage	90
III.3 Véhicule instrumenté et centrale inertielle.	92
III.4 Vue aérienne du virage (LCPC/Nantes).	93
III.5 Evolution temporelle des coordonnées (x_1 et x_2) d'une trajectoire u	94
III.6 Evolution temporelle des coordonnées (v_1 et v_2) d'une trajectoire u	94
III.7 Evolution temporelle des coordonnées (γ_1 et γ_2) d'une trajectoire u	95
III.8 Projection dans \mathbb{R}^2 du résultat de la classification des trajectoires u	96
III.9 Séparabilité entre les 4 classes de trajectoires.	97
III.10 Schéma de validation des classes \mathcal{C}_p obtenues.	99
III.11 Accélération γ_2 pour la 1 ^{ère} et 2 ^{ème} expérimentation.	99
III.12 Moyenne temporelle des coordonnées (v_1 et v_2) de u	100
III.13 Moyenne temporelle des coordonnées (γ_1 et γ_2) de u	101
III.14 Moyenne temporelle de Z selon le critère K_1 pour les 4 classes.	102
III.15 Moyenne temporelle de Z selon le critère K_2 pour les 4 classes.	102
III.16 Une ω -réalisation de X du processus de contrôle $D(t)$	102
III.17 Une ω -réalisation de X du processus de contrôle $\Gamma_N(t)$	102
III.18 Graphes de R_X selon le processus de contrôle D	103
III.19 Graphes de R_X selon le processus de contrôle Γ_N	103
III.20 Graphes de p_X pour le critère K_1	104
III.21 Graphes de p_X pour le critère K_2	104
III.22 Approximations de p_X pour \mathcal{C}_1 et \mathcal{C}_2 selon le critère K_1	106
III.23 Approximations de p_X pour \mathcal{C}_3 et \mathcal{C}_4 selon le critère K_1	106
III.24 Approximations de p_X pour \mathcal{C}_1 et \mathcal{C}_2 selon le critère K_2	107
III.25 Approximations de p_X pour \mathcal{C}_3 et \mathcal{C}_4 selon le critère K_2	107
III.26 Graphes de S_X pour les 4 classes selon le processus de contrôle $D(t)$	109
III.27 Graphes de S_X pour les 4 classes selon le processus de contrôle $\Gamma_N(t)$	109
III.28 Principe de validation du modèle probabiliste Z	111
III.29 Comparaison des graphes de p_X pour \mathcal{C}_1 et \mathcal{C}_2 selon le critère K_1	113
III.30 Comparaison des graphes de p_X pour \mathcal{C}_3 et \mathcal{C}_4 selon le critère K_1	113
III.31 Comparaison des graphes de p_X pour \mathcal{C}_1 et \mathcal{C}_2 selon le critère K_2	114
III.32 Comparaison des graphes de p_X pour \mathcal{C}_3 et \mathcal{C}_4 selon le critère K_2	114
III.33 Comparaison des graphes de S_X pour \mathcal{C}_1 et \mathcal{C}_2 selon le critère K_1	115
III.34 Comparaison des graphes de S_X pour \mathcal{C}_3 et \mathcal{C}_4 selon le critère K_1	115
III.35 Comparaison des graphes de S_X pour \mathcal{C}_1 et \mathcal{C}_2 selon le critère K_2	116

Table des figures

III.36	Comparaison des graphes de S_X pour \mathcal{C}_3 et \mathcal{C}_4 selon le critère K_2	116
III.37	Approximation de la loi de p_M pour \mathcal{C}_1 et \mathcal{C}_2 selon le critère K_1	120
III.38	Approximation de la loi de p_M pour \mathcal{C}_3 et \mathcal{C}_4 selon le critère K_1	120
III.39	Approximation de la loi de p_M pour \mathcal{C}_1 et \mathcal{C}_2 pour le critère K_2	121
III.40	Approximation de la loi de p_M pour \mathcal{C}_3 et \mathcal{C}_4 selon le critère K_2	121
III.41	Graphes de $\delta^* \rightarrow P_f(\delta^*)$ et $\delta^* \rightarrow \beta_c(\delta^*)$ pour le critère K_1	122
III.42	Graphes de $\delta^* \rightarrow P_f(\delta^*)$ et $\delta^* \rightarrow \beta_c(\delta^*)$ pour le critère K_2	123
III.43	Evolution de P_f en fonction de δ^* selon la classe \mathcal{C}_1	124
III.44	Evolution de P_f en fonction de δ^* selon la classe \mathcal{C}_2	124
III.45	Evolution de P_f en fonction de δ^* selon la classe \mathcal{C}_3	124
III.46	Evolution de P_f en fonction de δ^* selon la classe \mathcal{C}_4	124
III.47	Probabilité de défaillance avec critère du Sup pour \mathcal{C}_1 et \mathcal{C}_2	126
III.48	Probabilité de défaillance avec critère du Sup pour \mathcal{C}_3 et \mathcal{C}_4	126
III.49	Zone d'apprentissage du virage.	127
III.50	Choix de la base d'apprentissage optimale du virage.	127
III.51	Résumé de la méthodologie pour cette configuration de virage.	131
V.1	Approximations de p_X et F_X	140
V.2	Hyperplan séparateur d'un classifieur linéaire. La quantité $\frac{b}{\ \theta\ }$ exprime la distance entre l'hyperplan séparateur (H) et l'origine.	147
V.3	Impact de la maximisation de la marge sur la capacité de généralisation.	148
V.4	Hyperplans de séparation.	149
V.5	Hyperplans canoniques.	149

Liste des tableaux

II.1	Probabilité de défaillance en fonction de l'indice de fiabilité β_c pour EC0.	78
II.2	Matrice de confusion SVM Multi-classes.	83
II.3	Risque associé à quelques activités.	86
II.4	Classes de conséquences du système V-I-C.	87
III.1	Répartition en nombre de trajectoires	98
III.2	Coefficients d'aplatissement et d'asymétrie de X	105
III.3	Comparaison des méthodes d'approximation de la loi marginale de p_X	108
III.4	Critères de validité des simulations à partir des modèles Z	112
III.5	Comparaison des moments μ^p entre mesures et simulations pour K_1	117
III.6	Comparaison des moments μ^p entre mesures et simulations pour K_2	117
III.7	Validation des ajustements de la loi p_M des maxima de Z	119
III.8	Critères de validité des ajustements de la loi des max.	119
III.9	Validation du modèle Multi-class (SVM) par une matrice de confusion.	128
III.10	Probabilité de défaillance par classe de trajectoires.	130

Introduction générale

Malgré une amélioration de la sécurité routière ces dernières années, les accidents de la route préoccupent toujours les pouvoirs publics et l'industrie automobile. Les accidents de véhicules légers en virages ont des conséquences très graves au plan humain pour les usagers de la route. Les statistiques montrent que 18% des accidents de la route et 1/3 des accidents mortels en France ont eu lieu en virages (BAAC, 2006). L'accidentologie des véhicules légers en virages reste un enjeu majeur pour la sécurité routière. Les principaux types de ces accidents sont les *pertes de contrôle* et les *sorties de route*.

La solution pratique qui existe actuellement pour réduire ces accidents est la limitation de la vitesse par des panneaux de signalisation. Il existe également des travaux de recherche. En effet, ces travaux sont menés sur le thème de la détection et/ou de la correction de trajectoire pouvant être dangereuse en virage par le biais de l'analyse du triptyque Véhicule-Infrastructure-Conducteur. Cela consiste par exemple à définir une vitesse critique à respecter ou un profil de vitesse en fonction des caractéristiques de la route et de l'état dynamique du véhicule. Cet état du véhicule est obtenu par des modèles cinématiques ou dynamiques. On peut faire plusieurs reproches à ce type de travaux. Non seulement, ils sont difficilement validables actuellement sur le terrain, mais surtout ne considèrent pas toutes les interactions entre le véhicule et son environnement.

D'autres approches existent avec et sans modèles qui intègrent les systèmes d'aide à la

conduite tels que : ABS, ESP, ASR, etc. Ces approches ne tiennent pas compte de toutes les interactions du système de la conduite. Alors que la conduite est une habileté complexe qui implique de prendre en compte l'ensemble des interactions du triptyque Véhicule-Infrastructure-Conducteur. Dans cette thèse, ce triptyque est appelé système V-I-C. La figure (1) représente une illustration de ce système.

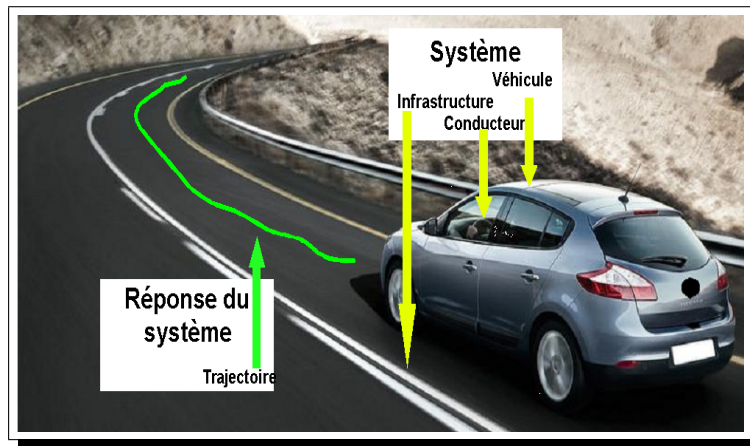


Figure 1 – Illustration du système V-I-C

La modélisation de ce système fait intervenir des problèmes non linéaires et des incertitudes très difficilement modélisables par les approches existantes. Il se révèle fort difficile de concevoir un modèle permettant de décrire ce système dans toute sa complexité. En conséquence, nous proposons d'élaborer plusieurs modèles partiels *vis-à-vis* de critères de défaillance liés à des comportements de conduite. Chaque modèle décrit et prédit une situation particulière présentant un intérêt du point de vue sécurité routière.

Ces modèles seront basés sur l'observation du système V-I-C. Dans le cadre de l'opération de recherche MTT (Météorologie des Trajectoires et du Trafic), le LCPC s'est doté d'observatoires de trajectoires en virage. En utilisant ces observatoires, nous pouvons observer le système V-I-C *via* des trajectoires mesurées. Ces trajectoires caractérisent toute la complexité de ce système puisqu'elles sont la résultante de ce système.

A partir de ces observations de trajectoires pratiquées dans une configuration de virage donnée, l'objectif principal de ce travail de thèse consiste à construire un indicateur de risque de défaillance de la trajectoire. Cet indicateur permet, à partir d'une trajectoire en cours de réalisation à l'abord d'un virage, de détecter si le véhicule auquel elle est associée présentera ou non un risque de violer un critère de sécurité donné.

Pour atteindre cet objectif, nous proposons d'abord de construire des modèles probabilistes simples et robustes à partir d'observations expérimentales issues des observatoires de trajectoires. Chaque modèle est spécifique à un comportement de conduite donné. Ce choix de modélisation permet de prendre en compte la grande variabilité du triplet V-I-C. Il permet également de s'affranchir d'éventuelles difficultés dans l'alimentation des paramètres gouvernant le modèle. Les études récentes montrent une tendance à analyser le système de conduite en milieu naturel. Car il est souvent judicieux d'observer un phénomène pour mieux le modéliser « naturalistic driving ». Cette approche de modélisation semble d'un grand intérêt du fait qu'elle intègre l'automobiliste comme un acteur de la conduite, et pas seulement comme une boîte noire.

Ensuite, cette procédure de modélisation couplée à une analyse fiabiliste permet d'associer une probabilité de défaillance à chaque comportement de conduite. Enfin, nous proposons de développer des méthodes de reconnaissance de trajectoires (à partir d'observations partielles, par exemple des données de mesures en entrée de virage) afin de pouvoir affecter chaque trajectoire à un comportement de conduite identifié dont on a calculé préalablement ses probabilités de défaillance.

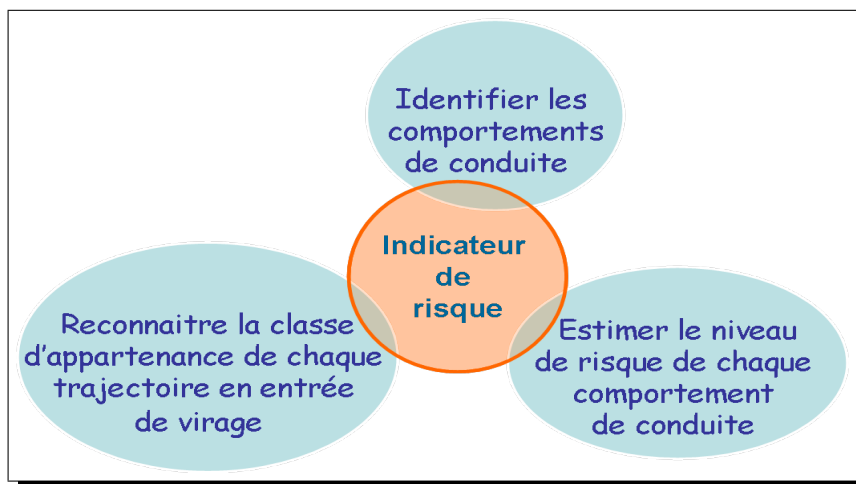


Figure 2 – Concept de l'indicateur de risque

L'originalité de la méthodologie réside à plusieurs niveaux : l'emploi de méthodes statistiques pour le traitement et la classification des trajectoires observées, l'utilisation de méthodes probabilistes pour la modélisation des incertitudes (aléas) du système et le recours à l'analyse fiabiliste pour l'estimation des niveaux de risque associés aux trajectoires.

Cette méthodologie utilise la connaissance du passé du système V-I-C pour prédire son futur comme le dit le neurophysiologiste Alain Berthoz : « La mémoire du passé n'est pas faite pour se souvenir du passé, elle est faite pour prévenir le futur. La mémoire est un instrument de prédiction ».

Du point de vue de sa structuration, outre la présente introduction générale, le manuscrit comprend trois chapitres, une conclusion générale et des annexes.

Le **premier chapitre** est composé de quatre parties. La première consiste en une analyse accidentologique portant sur les véhicules légers. Cette analyse montre non seulement le besoin d'améliorer la sécurité routière en France mais décrit également quelques facteurs accidentogènes des véhicules en virage.

La deuxième partie fait une synthèse des travaux réalisés par les chercheurs et industriels dans le domaine de l'accidentologie des véhicules légers. Cet état de l'art, nous permet de montrer les limites des méthodes déterministes pour traiter la défaillance du système. En effet, ces approches, non seulement ne considèrent que des sous systèmes tels que le véhicule ou l'infrastructure ou encore le conducteur, mais en outre ne prennent pas en compte les incertitudes des paramètres des modèles ou leurs interactions.

Nous donnons ensuite dans la troisième partie une description complète du système V-I-C ainsi que les diverses formulations mathématiques possibles de ce système en faisant apparaître les difficultés de mise en oeuvre. Nous montrerons effectivement pourquoi ce système n'est pas modélisable dans sa globalité de façon réaliste.

Enfin, la quatrième partie donne un aperçu de la nouvelle méthode que nous proposons pour prédire les trajectoires potentiellement dangereuses en entrée de virage. Dans cette partie, nous introduisons également quelques définitions nécessaires pour la compréhension de la méthodologie proposée.

Le **deuxième chapitre**, centré sur la méthodologie proposée dans cette thèse, comprend également quatre parties. Dans la première, après une description du principe de la construction de la base de données des trajectoires expérimentales, nous décrivons les méthodes de classification les mieux adaptées à cette étude. Le but de la classification est d'identifier les différents comportements de conduite dans la base de données de trajectoires observées.

La deuxième partie décrit les différentes étapes pour construire les modèles probabilistes à partir des trajectoires observées. Pour construire ces modèles, nous avons préalablement proposé plusieurs critères de défaillances *via* des fonctions d'états limites associées. A chacun de ces critères est associé un unique processus stochastique scalaire normalisé, faiblement stationnaire, ergodique et non gaussien. Ce processus est partiellement caractérisé par sa loi marginale d'ordre 1 et sa densité spectrale de puissance. Ces caractéristiques statistiques sont estimées à partir des observations expérimentales disponibles. La loi marginale d'ordre 1 de chaque processus est approximée par une technique basée sur le développement de la densité de probabilité sur la base hilbertienne des polynômes d'Hermite normalisés. Chacun de ces processus est caractéristique d'un comportement de conduite spécifique. Leur simulation requiert un algorithme approprié dont nous donnons une description détaillée en fin de partie.

Dans la troisième partie, pour des critères de sécurité donnés, nous estimons les probabilités de non-respect de ces critères. Le but de cette analyse fiabiliste est d'associer un niveau de risque de défaillance de trajectoires à chaque comportement de conduite.

Enfin dans la quatrième partie, nous décrivons la mise en oeuvre d'un modèle de reconnaissance de trajectoires basé sur les algorithmes SVM. Ce modèle permettra d'affecter une trajectoire à un comportement de conduite en entrée de virage. Connaissant préalablement le niveau de risque de chaque comportement de conduite, il est possible de prédire la dangerosité de chaque trajectoire pour une configuration de virage donnée.

Dans le **troisième chapitre**, nous montrons la faisabilité de la méthodologie décrite dans le deuxième chapitre en traitant une application sur le virage du LCPC/Nantes. Ce travail est complété par plusieurs validations expérimentales. A travers cette application, nous prouvons que la méthode proposée est facile à mettre en oeuvre.

Enfin, nous donnons quelques éléments généraux de conclusion et perspectives.

Systeme Véhicule-Infrastructure-Conducteur (V-I-C)

I.1 Introduction

L'objet de ce chapitre I est dans un premier temps de mettre en exergue les enjeux de l'accidentologie des véhicules légers en virages et de comprendre certains facteurs accidentogènes. Cette compréhension nécessite dans un premier temps la mise en place d'une analyse accidentologique du système V-I-C.

Ensuite, nous ferons une synthèse des travaux existants dans la littérature autour de la modélisation de ce système en insistant évidemment sur les forces et faiblesses de chaque méthode.

Puis, nous décrirons la complexité de ce système et les formulations mathématiques possibles. Dans cette description, nous montrerons pourquoi la modélisation du système dans sa globalité est une tâche très difficile voire impossible.

Enfin, sur la base de ces éléments, nous proposerons une démarche basée sur une approche probabiliste et fiabiliste des trajectoires expérimentales de véhicules légers en virages. Dans cette démarche, nous présenterons des définitions et des outils nécessaires à la compréhension de la méthodologie proposée au chapitre II.

I.2 Analyse accidentologique des véhicules légers

I.2.1 Accidentologie

L'accidentologie est une science relativement récente dont le but est avant tout de comprendre les mécanismes accidentels, et d'identifier les besoins qui permettraient de réduire le nombre et la gravité des accidents de la route. Les accidents de la route sont des événements pouvant reposer sur des mécanismes complexes et difficilement prévisibles. En effet, ces phénomènes font intervenir des paramètres aussi variés que la configuration géométrique de l'infrastructure, les conditions météorologiques, les caractéristiques mécaniques et dynamiques de chaque véhicule, et bien entendu le comportement des différents conducteurs impliqués. Tous ces éléments, dont beaucoup sont difficilement quantifiables, voire mesurables, vont faire en sorte que pour des mises en situation *a priori* très similaires, certaines vont aboutir à un accident tandis que d'autres ne présenteront pas de danger visible. Dans la plupart des cas, les études d'accidents prennent la forme de relevés statistiques à grande ou moyenne échelle. Les informations qui en découlent peuvent être exploitées dans certaines des situations, notamment l'annonce d'arrivée sur points noirs¹. Bien entendu, l'immense majorité des véhicules traversent un point noir en toute sécurité. Inversement de nombreux accidents peuvent avoir lieu sur des emplacements non identifiés comme point noir.

Pour réduire le nombre d'accidents, il est indispensable de bien en connaître les causes et les mécanismes. C'est la raison pour laquelle on a recours aux connaissances accumulées par l'accidentologie. La compréhension de ces mécanismes nécessite, entre autres, des études approfondies sur les accidents réels et la construction de modèles permettant à l'accidentologue de structurer le déroulement de l'accident et de déterminer les interactions entre les facteurs et les causes l'ayant provoqué.

Les études accidentologiques approfondies orientées vers la sécurité primaire (les mécanismes accidentels, la prévention et l'évitement de l'accident) datent des années 1950 et se sont développées dans les années 1970 et 1980, notamment en France à l'Observatoire National Interministériel de Sécurité Routière (ONISR) puis au département Mécanismes d'Accident de l'Institut National de Recherche sur les Transports et leur Sécurité (INRETS). Les constructeurs automobiles, par l'intermédiaire de leur Laboratoire d'Accidentologie et de Biomécanique commun (LAB), ont développé ce type d'études dans

¹lieux où se concentrent les accidents

I.2. Analyse accidentologique des véhicules légers

les années 1990, dans le cadre du programme de recherche VSR (Véhicule et Sécurité Routière) en partenariat avec l'INRETS. Ces études sont destinées aux constructeurs automobiles, au CESAAR pour les poids lourds et au LAB pour les véhicules légers. Elles apportent les connaissances accidentologiques pour comprendre les situations accidentogènes et éventuellement permettent de spécifier des dispositifs d'évitement de l'accident.

Comprendre les mécanismes accidentels implique notamment de recueillir de l'information sur les accidents de la route, puis d'analyser cette information. Au début du programme VSR, le LAB a choisi de développer une méthode d'investigation prospective et systématique qui consiste à établir un catalogue évolutif de l'information à collecter sur les accidents. L'analyse de l'ensemble de ces éléments, accident par accident, permet ensuite de procéder à une reconstruction cinématique et cognitive de l'accident. Puis, ces analyses agrégées pour tous les accidents aident à mieux comprendre les mécanismes accidentels, voire, dans certains cas, à fournir sous forme de *statistiques* des résultats épidémiologiques sur l'incidence ou les risques relatifs des facteurs de risque identifiés.

I.2.2 Statistiques sur les accidents

En 2009, le bilan de l'insécurité routière pour la France métropolitaine selon la direction de la sécurité routière [29] s'élevait à 73390 accidents corporels, 4274 tués et 91669 blessés dont 32930 blessés hospitalisés, ce qui représente par rapport à 2008 (figure I.1), des baisses de 9.7% du nombre d'accidents, de 7.5% des tués, de 11.2% des blessés et de 14.7% des blessés hospitalisés. Cette forte accélération de la baisse annuelle du nombre de tués sur la route (7.5% après 2% en 2008) est toutefois légèrement en dessous des 8.2% nécessaires pour atteindre l'objectif fixé par les pouvoirs publics de moins de 3000 tués en 2012. Les progrès de l'année 2009 sont principalement dus à la baisse des vitesses moyennes (-1.1%), probablement avec l'effet des radars.

Cependant, IRTAD (International Road Traffic and Road Accident) [45] a fait une comparaison entre la France et les autres pays européens sur le nombre de tués par accidents de la route en 2006 rapporté à la population (par million d'habitants). Les résultats de la figure (I.2) montrent que la France compte 40 à 50% de tués par habitant plus que la Suède ou le Royaume-Uni. Il reste des marges pour améliorer la sécurité routière en France et atteindre le niveau de sécurité des meilleurs pays européens.

Des études du SETRA [31] et de l'INRETS [40] ont mis en évidence que le risque d'accident

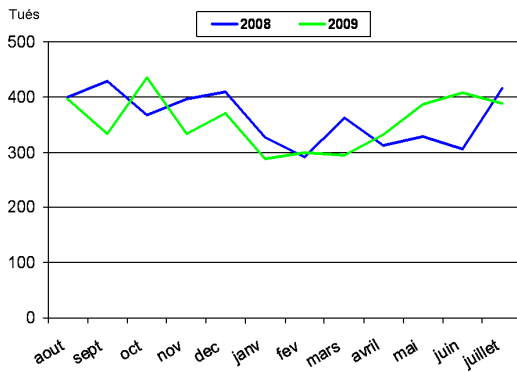


Figure I.1 – Comparaison des statistiques d'accidents en France pour 2008 et 2009.

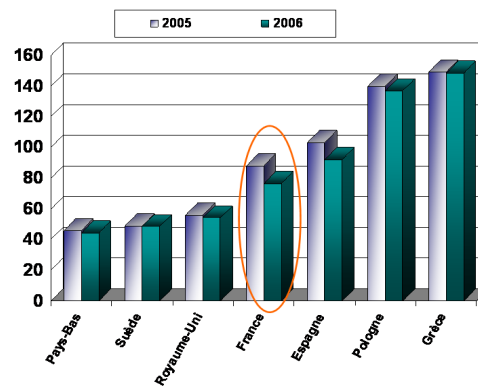


Figure I.2 – Comparaison des statistiques d'accidents en Europe pour 2005 et 2006.

en courbe est 5 à 10 fois plus élevé qu'en alignement droit ; 40% des accidents mortels se produisent en virage. En outre, le LAB a montré qu'une meilleure gestion des difficultés en virage pourrait permettre d'éviter 15% des accidents mortels. D'où l'intérêt, dans ce travail de thèse de se focaliser, exclusivement à l'accidentologie des Véhicules légers en virages. A partir de cette étude accidentologique, plusieurs *facteurs accidentogènes* sont déterminés dans le paragraphe suivant.

I.2.3 Origine des accidents en virage

Dans le but d'accéder à une meilleure connaissance des conditions et mécanismes des accidents de perte de contrôle en virage, le département mécanisme d'accident de l'INRETS s'est appuyé sur l'analyse approfondie de 84 cas d'accidents en virage afin de disposer d'une base d'étude adaptée à la problématique [14]. L'analyse concerne les aspects cinématiques et le rôle de quelques caractéristiques d'aménagement des virages où les accidents se produisent. Les accidents concernés sont issus des EDA² [40] et survenus sur un secteur expérimental de l'INRETS autour de Salon de Provence. L'étude [34] a distingué deux familles de pertes de contrôle en courbe, qui se produisent dans des conditions différentes :

- Les pertes de contrôle liées à une erreur de direction (guidage) : elles représentent 35% des cas et se produisent le plus souvent dans des virages à gauche de grand rayon. L'état du conducteur (endormissement, malaise, accomplissement d'une tâche annexe et alcoolémie) est le premier responsable de ce genre de perte de contrôle.

²Etudes Détaillées d'Accidents

I.2. Analyse accidentologique des véhicules légers

- Les pertes de contrôle liées à un problème dynamique : elles représentent 65% des cas. Elles résultent d'une vitesse inadaptée compte tenu des capacités du conducteur, du véhicule et des caractéristiques de l'infrastructure.

Nous illustrons nos propos par la figure (I.3) en représentant la répartition des facteurs d'accident liés à l'inadaptation de la vitesse du véhicule en virage. On remarque que la vitesse inadaptée au tracé de la route est le cas le plus fréquent (38%), suivi (24%) de la vitesse dépassant le maximum autorisé. Les autres cas représentent des taux inférieurs à 15%.

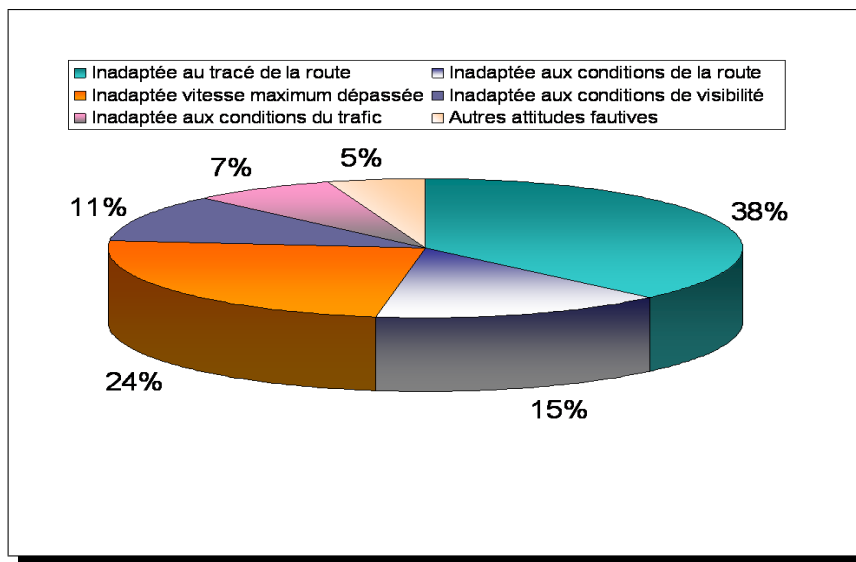


Figure I.3 – Répartition des facteurs d'accident liés à l'inadaptation de la vitesse

Les trois causes principales sont donc : (i) vitesse inadaptée au tracé de la route, (ii) vitesse inadaptée aux capacités du conducteur et (iii) vitesse inadaptée au trafic. En virage, il n'est pas facile pour le conducteur de déterminer la vitesse appropriée. Par contre, les conducteurs sous-estiment souvent la vitesse de franchissement ou la dangerosité du virage ; leur vitesse est donc souvent inadaptée. Il est donc nécessaire de les aider à adapter leur vitesse à la configuration du virage.

Bien que la vitesse soit un facteur déterminant pour l'accidentologie des véhicules en virage, les études ont mis en évidence d'autres facteurs tels que l'inadaptation de l'accélération latérale du véhicule.

Une étude de Dakhlallah et al [24] a montré que l'accélération latérale augmente quand la vitesse longitudinale du véhicule augmente dans le virage. Cette étude a été réalisée pour un virage de rayon 150m. On distingue sur la figure (I.4) quatre plages de variation de l'accélération latérale en fonction de la vitesse longitudinale.

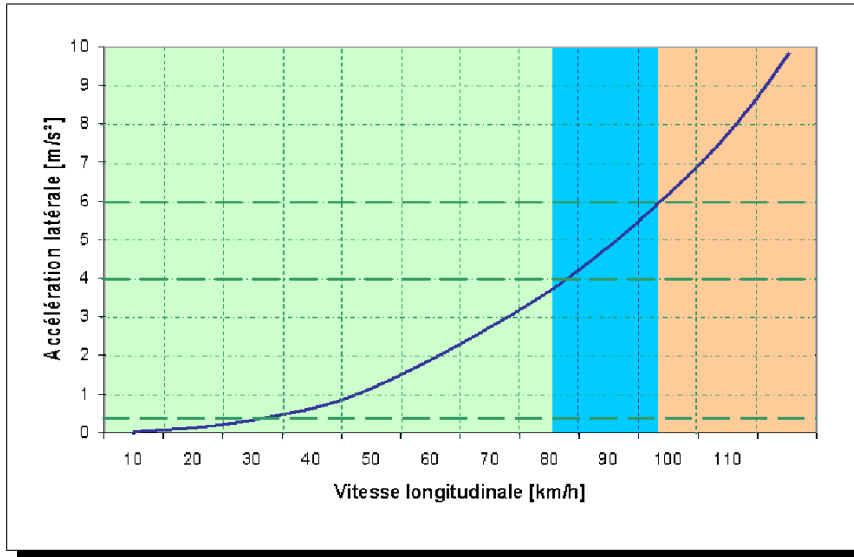


Figure I.4 – Relation entre l'accélération latérale et la vitesse longitudinale du véhicule

La **première** plage de 0 à $0.5 m/s^2$ est appelée plage de petits signaux. Elle concerne le comportement de réaction en ligne droite à des sollicitations de type ornières ou vent latéral ou en virage à faible vitesse. Les forces centrifuges sont négligeables et les pneus ne doivent pas développer de forces latérales.

La **deuxième** de 0.5 à $4m/s^2$ est appelée plage ou zone de fonctionnement linéaire du pneumatique. Les phénomènes étudiés sont entre autres les réactions transitoires de lacet, les changements de voie et les combinaisons de mouvements dynamiques transversaux et longitudinaux tels que les réactions aux transferts de charge dans les virages. La plage de petits signaux et la plage linéaire sont donc déterminantes pour l'appréciation subjective du comportement dynamique des véhicules par les conducteurs. Ces deux premières plages définissent la zone de conduite où le véhicule est contrôlé par le conducteur.

La **troisième** de 4 à $6m/s^2$ est appelée plage transitoire. Le véhicule se comporte soit de façon linéaire, soit de façon non linéaire selon sa conception. La zone de conduite est critique (possibilité de perte de contrôle ou de sortie de route).

I.2. Analyse accidentologique des véhicules légers

La **quatrième** située au-delà de $6m/s^2$ est appelée plage limite de danger. Elle n'est atteinte que dans des situations extrêmes par rapport aux compétences des conducteurs. Dans cette plage, le comportement des véhicules est fortement non linéaire et instable du point de vue contrôle pour un conducteur moyen. L'accident est presque inévitable car le conducteur n'ayant pas l'habitude de s'y retrouver.

L'accélération latérale est donc un facteur accidentogène déterminant en virage comme la vitesse longitudinale. En conduite sécurisée, elle ne dépasse généralement pas $4m/s^2$ selon avis d'experts. Nous verrons dans la partie application de la méthodologie au chapitre (III) si cette hypothèse est toujours vérifiée pour une conduite apaisée.

Cette *analyse accidentologique* a décrit d'une part, l'influence de quelques facteurs accidentogènes du système V-I-C. D'autre part, la France dispose d'une marge de manoeuvre pour réduire l'accidentologie des véhicules en virages, notamment les sorties de route et les pertes de contrôle auxquelles nous nous intéressons dans ce travail de thèse.

Nous avons énoncé dans l'*introduction générale* que la seule solution pratique utilisée actuellement pour réduire ces accidents est la limitation de la vitesse par des panneaux de signalisation. Il existe également des *travaux de recherche* qui ont permis de comprendre et de réduire ces accidents. Certains de ces travaux sont décrits dans la section suivante.

I.3 Synthèse des travaux réalisés sur le système VIC

L'objet de cette section est de faire un état de l'art autour de la modélisation du système V-I-C en virages. La plupart de ces travaux sont basés sur trois familles de modèles : *modèles de véhicule*, *modèles liés à la géométrie de la route* et *modèles conducteurs*.

I.3.1 Véhicule

Le véhicule léger est un solide en mouvement, généralement modélisé par un modèle **cinématique** ou un modèle **dynamique**. Les modèles cinématiques de trajectoires décrivent uniquement la trace et la loi horaire (vitesses, accélérations et jerks) du centre de gravité du véhicule. Les modèles dynamiques (avec masses, inerties et forces) décrivent l'évolution des variables d'état du système obtenu à partir des équations de mouvement du véhicule.

I.3.1.1 Modèles cinématiques

Ces modèles ont pour but de représenter les trajectoires observées et se raccordant à une description linéaire de la géométrie des routes (segments de droite, arcs de cercle, arcs de clothoïde pour le cas plan). Les trajectoires peuvent suivre des chemins à *courbure polynômiale*, ou à *coordonnées polynômiales*. Souvent, les trajectoires respectent des critères de minimisation de l'accélération latérale (elle est mathématiquement liée à la courbure). Dans les deux cas, il s'agit de trouver une alternative à l'utilisation d'arcs de cercle pour assurer la jonction entre deux lignes droites avec une continuité de la courbure aux points de jonction.

Chemins à courbure polynômiale : Les clothoïdes sont des courbes cinématiques dont la courbure est une fonction linéaire de l'abscisse curviligne.

$$R(t) = \begin{pmatrix} X_0 + \int_{l_0}^{l_1} \cos(\tau_0 + \kappa_0 l(t) + \frac{1}{2}cl(t)^2)dl \\ Y_0 + \int_{l_0}^{l_1} \sin(\tau_0 + \kappa_0 l(t) + \frac{1}{2}cl(t)^2)dl \end{pmatrix} \quad (\text{I.1})$$

où τ_0 est l'angle de la tangente, κ_0 la courbure initiale, c le coefficient de variation de la courbure avec l'abscisse curviligne l et (X_0, Y_0) le point initial. Les travaux de Kanayama et Miyake [48] illustrent cette approche. D'autres fonctions particulières existent telles que les spirales cubiques, à courbure quadratique de l'abscisse curviligne, Kanayama et Hartman [49]. L'*avantage* de ces chemins est leur formulation simple *vis-à-vis* des paramètres de configuration. Un de leurs inconvénients est que leur projection n'est pas analytique, c'est-à-dire que leurs coordonnées ne sont pas calculables mais peuvent seulement être

I.3. Synthèse des travaux réalisés sur le système VIC

obtenues de manière approchée. Un autre inconvénient est qu'ils contiennent une infinité de morceaux, à partir du moment où ils contiennent un segment de droite. Intuitivement, cela se produit dès qu'il s'agit de relier deux configurations suffisamment éloignées.

Chemins à coordonnées polynômiales : Les clothoïdes sont des courbes simples permettant de constituer des trajectoires admissibles en respectant les critères de continuité sur la courbure. Il existe d'autres solutions pour parvenir au même résultat, notamment en utilisant différents formalismes mathématiques pour représenter les coordonnées des chemins sous forme polynômiales. On peut utiliser une représentation paramétrique de la trajectoire pour définir des chemins dont certains points de passage sont imposés (en général 2 ou 3). Le degré du polynôme est fonction du nombre de points de passage imposés. Par exemple Komoriya et Tanie [53] ont utilisé des fonctions B-splines d'ordre 3 afin de satisfaire la continuité de la courbure. En 2003, ces travaux sont améliorés par Lauffenburger [57] en utilisant des B-splines avec une contrainte sur le secteur angulaire.

Le principal *avantage* de ces chemins est la simplicité de l'expression analytique des coordonnées de leurs points. Cette expression analytique garantit la simplicité et surtout l'exactitude de leur détermination, contrairement aux clothoïdes où les coordonnées sont obtenues par intégration et peuvent souffrir d'imprécision. Néanmoins, l'*inconvénient* majeur de ces chemins est que, si la contrainte d'orientation est implicitement prise en compte, la vérification de la contrainte sur la courbure demande des calculs coûteux. D'autres travaux récents utilisent une technique de filtrage pour reconstruire la trajectoire du véhicule dans un plan (cf. Zamora[106], Joly [47]). Nous citons également les travaux de Basset[1].

Les modèles cinématiques sont souvent simples à construire. Ils sont intéressants pour identifier les trajectoires qui s'écartent de façon significative de la trajectoire sécurisée. Leur faiblesse réside dans la non prise en compte de la dynamique du véhicule (donc manque de réalisme). Ils peuvent générer un nombre important de trajectoires irrégulières ou physiquement non réalisables. Cette famille de modèles ne convient donc pas pour représenter le système V-I-C dans le but de faire une analyse fiabiliste.

I.3.1.2 Modèles dynamiques de véhicule

Bien qu'à l'heure actuelle, il existe beaucoup de modèles dynamiques reproduisant de manière relativement fidèle le comportement d'un véhicule léger, ce dernier demeure un système complexe dont la modélisation est rendue délicate par la variation de ses pa-

ramètres (vitesse, adhérence, ...). Les modèles dynamiques peuvent être plus ou moins simples (quart de véhicule, bicyclette) ou complexe (modèle 3D à quatre roues). La figure (I.5) montre le modèle bicyclette et le véhicule avec les axes de translations (X, Y, Z) et de rotations (ϕ, ψ, θ).

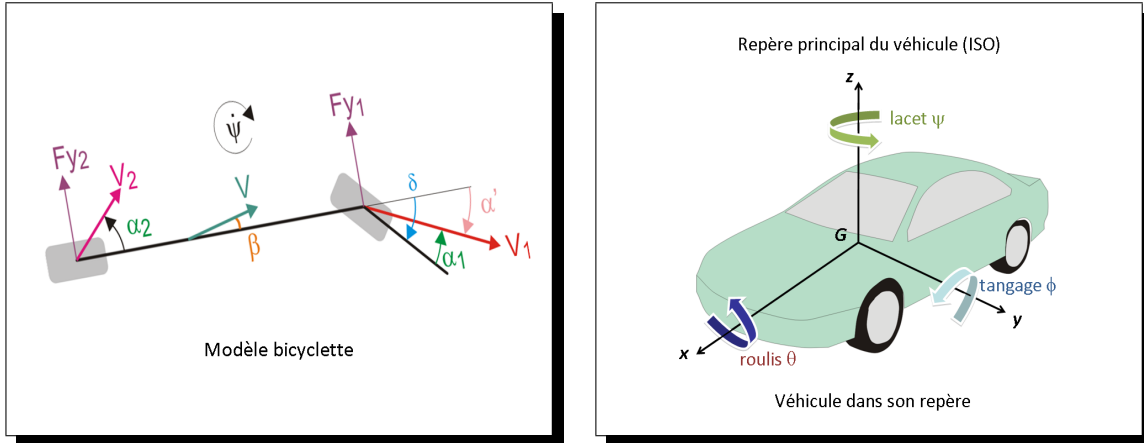


Figure I.5 – Modèle bicyclette / Véhicule dans son repère

où V est la vitesse longitudinale du véhicule, F_{y1} et F_{y2} les forces latérales avant/arrière, δ l'angle de braquage des roues, β l'angle de dérive, α' l'angle de dérive du train avant, α_1 l'angle de dérive du pneumatique et $\dot{\psi}$ le taux de lacet.

Les modèles dynamiques servent à la conception de nouveaux véhicules ou à la compréhension des phénomènes régissant leur comportement. Ils décrivent le comportement du système à partir des lois de la physique. Par exemple, en faisant le choix d'un vecteur d'état $\mathbf{q}=[X, Y, Z, \phi, \psi, \theta]$, le modèle dynamique s'écrit sous la forme matricielle suivante :

$$M(q)\ddot{q} + f(q, \dot{q}) = \Gamma \quad (\text{I.2})$$

où $M(q)$ est la matrice d'inertie généralisée, f la matrice des forces extérieures et Γ le vecteur des forces généralisées. Nous citons ici quelques travaux relatifs à la modélisation dynamique. D'abord, le modèle bicyclette à deux roues en ligne obtenu en supposant que les deux roues de chaque essieu sont confondues en une roue virtuelle située au milieu de l'essieu dans l'axe longitudinal du véhicule. Ce modèle plan présente beaucoup d'avantages, en facilite la mise en oeuvre de certaines lois de contrôle et de commande. Il est très utilisé dans les études de la dynamique du véhicule telles que Canudas [17], Mammar [61], Mendoza [63], Bouteldja [13], Imine [44], Stephant [95], etc. Cependant, il ne reproduit pas

I.3. Synthèse des travaux réalisés sur le système VIC

les mouvements de roulis. Le modèle « lacet-dérive » très simple, permet une utilisation embarquée dans des applications temps-réel telles que dans Riedel [83], Glaser [41], etc.

Pour étendre le domaine de validité de ces modèles, il est possible d'y ajouter un ou plusieurs degrés de liberté (ddl), tels que le roulis ou le tangage. Les modèles les plus complexes peuvent intégrer jusqu'à 30 ddl comme le modèle de Chao [19]. Ils ont pour objectif d'étudier des phénomènes complexes comme les régimes transitoires du pneumatique, l'un des organes du véhicule les plus délicats à étudier et à modéliser (cf. Pacejka [73] et Basset [2]). Ce pneumatique assure l'interface entre la route et le véhicule et a pour rôle de transmettre les efforts. Il a été décrit dans les travaux de Sammier [87]. La plupart des modèles à 4 roues sont issus des travaux de Peng [74], de Tomizuka [98]. Les travaux de Brossard [15], Gillespie [39] donnent une synthèse des modèles les plus courants.

En complément des modèles dynamiques de véhicule, il existe des simulateurs pour générer des trajectoires (trace et loi horaire), de manière plus précise que les modèles de véhicules. ARCSIM (Vehicle Dynamics Simulation Software) du centre américain de recherche militaire en automobile (Michigan, USA), ASM (Automotive Simulation Model) du fournisseur mondial d'outil d'ingénierie dspace, CarMaker de la compagnie IPG CarMaker, Callas/Prosper de la société SERA-CD en France, ainsi que d'autres, figurent parmi les simulateurs professionnels de la dynamique du véhicule.

Cette famille de modèles régissant des équations de mouvement du véhicule rendent compte de la dynamique du système mieux que les modèles cinématiques. Ils sont donc utiles pour analyser le comportement dynamique du véhicule. La principale *difficulté* est leur validation expérimentale, avec une identification difficile des paramètres tels que la hauteur de centre de gravité, la masse, etc. En outre, ils ne prennent pas en compte des facteurs accidentogènes (la visibilité, la perception, etc.) nécessaires pour l'analyse de la défaillance du système V-I-C. Nous verrons plus tard dans ce chapitre que ces modèles sont incomplets car ils sont basés sur des hypothèses simplificatrices des équations qui gouvernent le système V-I-C dans le but d'obtenir des modèles utilisables en pratique.

Le paragraphe suivant décrit comment le système de conduite a été étudié en tenant compte de la modélisation de l'*infrastructure*.

I.3.2 Infrastructure

Le passage d'un virage est une tâche délicate qui fait appel aux capacités de guidage et de contrôle du conducteur. Certains virages peuvent être considérés comme plus difficiles que d'autres. En général, les accidents surviennent dans des virages présentant un défaut de conception tel qu'une irrégularité de la courbure. Les irrégularités de courbure sont quantifiées par la notion de « rayon mini », qui prend en compte la courbure locale de la route, et la compare à la courbure générale du virage.

Pour représenter ces phénomènes, il existe des modèles reliant le système de conduite à la géométrie de la route. Par exemple, Revue [81] explique l'adéquation de l'infrastructure aux contraintes dynamiques du véhicule pour définir le critère d'état limite sur la courbure. La conception cinématique de la route doit être telle qu'elle n'entraîne pas une rupture de l'équilibre transversal en courbe, des accélérations résiduelles ou des variations d'accélération brutales préjudiciables à la sécurité et au confort des usagers. L'équilibre transversal du véhicule se définit à partir du dévers δ , de la déformation, de l'élasticité et de la dérive du pneumatique α , des conditions d'équilibre dans les courbes, du rayon R de la courbe, de l'évolution de l'accélération transversale. Le dévers vers l'intérieur de la courbe a pour effet de faire jouer un rôle actif à la pesanteur, la composante du poids du véhicule parallèle au plan de la chaussée compense partiellement la force centrifuge. La loi de Coulomb nous permet d'écrire :

$$F_y = C_f \alpha F_z = \mu F_z \quad (\text{I.3})$$

où C_f est le coefficient d'envirage, F_z est la charge instantanée de la roue, F_y la sollicitation transversale et μ le coefficient d'adhérence. A partir de la condition d'équilibre d'un essieu, on peut alors écrire :

$$\frac{F_{y1}}{F_{z1}} = \frac{F_{y2}}{F_{z2}} = \frac{F_{y1} + F_{y2}}{F_{z1} + F_{z2}} = \frac{F - \delta P}{P + \delta F} = C_f \alpha \quad (\text{I.4})$$

où $F = \frac{MV^2}{R}$ est la force centrifuge et $P = Mg$ le poids exercé sur l'essieu. L'équation (I.4) permet d'établir la relation liant les caractéristiques de l'infrastructure de la forme :

$$R_{min} = \frac{V^2}{g} \left(\frac{1 - \mu \delta}{\delta + \mu} \right). \quad (\text{I.5})$$

En complément à ce modèle lié à la géométrie de la route, nous citons trois modèles de

I.3. Synthèse des travaux réalisés sur le système VIC

contact pneumatique/chaussée : le modèle de LuGre, le modèle semi-empirique de Guo et le modèle de Pacejka. Ce dernier, bien qu'ayant un domaine de validité limité, est encore largement utilisé comme référence dans le monde industriel. Une des difficultés est l'estimation de ses coefficients à partir d'essais expérimentaux. Cette opération, réalisée couramment sur bancs d'essais par les manufacturiers de pneumatiques, demande souvent à être complétée de manière à tenir compte des variations de conditions (revêtement, protocole d'essais, etc.). Ces modèles sont décrits dans Nadji [69], Delanne [26], Basset [2].

L'infrastructure est une composante importante dans l'étude dynamique du système V-I-C mais elle est souvent représentée d'une manière simplifiée, et s'appuie sur des représentations conventionnelles. Ce type d'approche ne considère pas également les interactions du système telles que la visibilité ou la perception donc ne répond pas à nos objectifs. Le paragraphe suivant décrit la représentation du système par des *modèles conducteurs*.

I.3.3 Conducteur

L'accidentologie a mis en évidence la responsabilité directe du conducteur dans la majorité des accidents de la route ; il est le seul acteur contrôlant le véhicule et son environnement. En effet, l'infrastructure peut être non coopérative ou dégradée, le véhicule a aussi ses limites liées à ses caractéristiques cinématiques et dynamiques. En revanche, le conducteur a des défaillances dues à ses états psychiques et physiques (fatigue, inconscience, maladie, ivresse, etc.). Les conséquences de ces défaillances sont l'hypovigilance et la mauvaise prise de décision due à un mauvais jugement de la situation accidentogène. La plupart des accidents sont liés, directement ou non, à une erreur humaine [35].

Les études sur le conducteur dans son environnement ont permis le développement de modèles comportementaux (basés sur la perception) et une classification des conducteurs. Parmi ces modèles, citons les modèles continus, discrets, hybrides et cognitifs tels que Lauffenburger et al.[56]. A partir de cette perception visuelle, des études complémentaires ont donné des modèles basés sur la classification des conducteurs. Cette classification résulte d'études statistiques pour définir le profil du conducteur (inexpérimenté/expérimenté) et élaborer un modèle cognitif.

I.3.3.1 Classification des conducteurs

Les automobilistes sont regroupés selon deux classes définies par Rothengatter et al. [85] : conducteurs « **inexpérimentés** » et « **très expérimentés** ». Le principe consiste tout

d'abord à définir les caractéristiques, principalement en terme de trajectoire, des différentes classes de conducteurs. Ces caractéristiques dépendent de l'âge, de l'expérience, des caractéristiques physiques, des mesures proprioceptives et extéroceptives. Les critères ainsi définis pourront ensuite être utilisés pour l'identification des paramètres du modèle en vue du développement d'un générateur de trajectoires. Le conducteur « inexpérimenté » adopte une conduite que l'on pourrait qualifier de « conduite en régime établi », avec des sollicitations longitudinale et transversale relativement découplées : le freinage s'effectue presque entièrement avant l'entrée dans la courbe alors que la reaccélération est limitée par les efforts transversaux auxquels est soumis le conducteur. Un conducteur *très expérimenté* adopte une stratégie qui consiste à *attendre pour voir*, au lieu de débiter sa manoeuvre sans connaissance du profil du virage et de sa sortie. Ces traits de caractère correspondent d'ailleurs aux deux stratégies de conduite présentées par Savkoor et al[89].

La méthode décrite dans ces travaux est insuffisante car il est nécessaire de prendre en compte l'information de secteur angulaire dans la détermination de la position latérale du véhicule. Messaoudene [64] propose de considérer le couple (rayon de courbure, secteur angulaire) pour définir l'évolution des positions latérales des points de plongée et de corde par l'intermédiaire d'un modèle de représentation identifié en utilisant le critère des moindres carrés. Enfin, pour définir un modèle de trajectoires qui convienne aux deux classes de conducteurs citées ci-dessus, il est impératif que les points de plongée et de corde, ainsi que leurs caractéristiques respectives (position, pente et courbure), puissent être imposés dans la détermination de l'expression analytique du spline polaire.

L'hypothèse de réduire le système V-I-C à un *modèle conducteur* ne permet pas de tenir compte des interactions nécessaires pour étudier la défaillance du système.

Cette synthèse de *l'état de l'art* sur la modélisation du système V-I-C que nous venons de décrire montre que beaucoup de travaux de recherche ont contribué de manière significative à la compréhension et à l'amélioration de la sécurité routière. Malheureusement, chacun de ces travaux a considéré le système comme la modélisation d'un sous système. Certes, ces approches ont simplifié le problème pour obtenir des solutions utilisables, mais elles n'ont pas tenu compte de l'ensemble des facteurs accidentogènes qui peuvent conduire à la défaillance. Par exemple, les phénomènes tels que l'adhérence, la visibilité, la perception,... sont difficilement modélisables par les approches déterministes citées.

La section suivante *décrit la complexité* de ce système afin d'apporter des solutions.

I.4 Description et formulations du système V-I-C

A travers la section précédente, nous remarquons que ce système n'a jamais été modélisé dans sa globalité en tenant compte de l'ensemble des interactions. Car la modélisation de ces interactions nécessite des hypothèses simplificatrices par les approches déterministes. Alors qu'il est dit dans l'introduction que la conduite est une habileté complexe qui implique de prendre en compte toutes les interactions entre le *véhicule*, l'*infrastructure* et le *conducteur*.

L'objet de cette section est de décrire de façon plus ou moins détaillée la complexité de système. Cette description consiste à mettre en exergue les interactions du système et les discontinuités sur ces interactions. A partir de cette description, nous proposerons des formalismes mathématiques adéquats du système. Nous montrerons également les *difficultés* techniques et scientifiques de la mise en oeuvre de ces formalismes.

I.4.1 Description du système

L'activité de la conduite est composée de plusieurs interactions entre les trois sous-systèmes :

- Le **véhicule** est une structure mécanique, liée à la chaussée par des pneumatiques, met directement en oeuvre les décisions prises par le conducteur. Il est soumis à des contraintes cinématiques et dynamiques.
- L'**infrastructure** routière caractérise le tracé et la surface de la route. Elle fournit une très grande quantité d'informations au conducteur, et interagit avec le véhicule. Particulièrement, les virages constituent une configuration pouvant générer des dangers, de par la contrainte qu'ils imposent au conducteur de suivre une trajectoire complexe et de la corriger durant tout le temps nécessaire au virage.
- Le **conducteur**, à partir de la perception de son environnement, analyse et décide ses actions sur le véhicule à travers des opérateurs de commande : le volant, la pédale d'accélération et la pédale de frein.

La modélisation de ce système complexe dans sa globalité nécessite de comprendre les différentes interactions entre ces trois sous systèmes. Pour une meilleure compréhension de ce système, nous le représentons sous la forme d'une architecture.

I.4.2 Architecture du système

La figure (I.6) montre les différentes interactions du système V-I-C. Ces interactions peuvent engendrer des discontinuités dans les paramètres (ou dans leur évolution). Ces discontinuités se traduisent soit par une mauvaise perception de la route par le conducteur, qui applique alors des commandes inadaptées, soit par des entrées route en quantité trop importantes pour être corrigées par la boucle de conduite. Dans les deux cas, le système est perturbé par conséquent le risque d'accident est élevé.

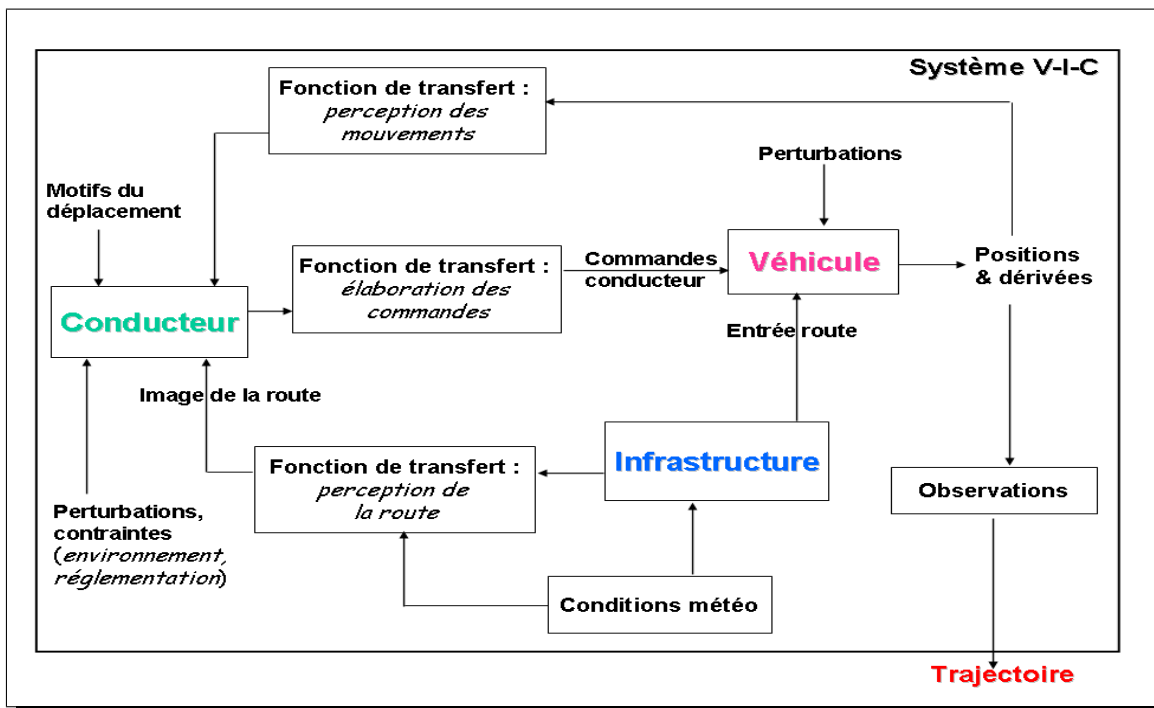


Figure I.6 – Schéma des entrées/sorties du système V-I-C

Parmi les interactions de ce système (cf. figure I.6), nous citons la perception, la visibilité ou encore l'adhérence qui sont des phénomènes physiques complexes non seulement fortement *non linéaires* mais présentant également des *incertitudes* sur leurs estimations. Une modélisation par les approches déterministes citées précédemment ne permet pas de représenter ce système en tenant compte de ce caractère aléatoire.

Nous allons maintenant proposer des *formulations mathématiques* possibles de ce système et la difficulté de leurs mises en oeuvre.

I.4. Description et formulations du système V-I-C

I.4.3 Formulations mathématiques du système

Une modélisation réaliste du système V-I-C en tenant compte des incertitudes consiste à décrire son évolution par une équation différentielle stochastique (EDS) vectorielle de la forme :

$$\begin{cases} \dot{\mathbb{X}}(t) &= F(t, \mathbb{X}(t)) + \xi(t, \mathbb{X}(t))\dot{W}(t) , \quad t > 0 \\ \mathbb{X}(0) &= \mathbb{X}_0 \text{ p.s.} \end{cases} \quad (\text{I.6})$$

où $\mathbb{X} = (\mathbb{X}(t), t \in \mathbb{R}_+)$ est un processus défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ indexé sur \mathbb{R}_+ à valeurs dans \mathbb{R}^m ; $F : \mathbb{R}_+ \times \mathbb{R}^m \mapsto \mathbb{R}^m : (t, \omega) \mapsto F(t, \omega)$; $\xi : \mathbb{R}_+ \times \mathbb{R}^m \mapsto \text{Mat}_{\mathbb{R}}(m, d) : t, x \mapsto \xi(t, x)$; \dot{W} est un bruit blanc gaussien standard à valeurs dans \mathbb{R}^d ; \mathbb{X}_0 est une v.a donnée, définie sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R}^m , et où $(\Omega, \mathfrak{F}, \mathbb{P})$ est l'espace de probabilité de base sur lequel seront supposées définies toutes les grandeurs aléatoires (processus stochastiques et variables aléatoires) considérées dans ce travail.

Dans l'équation différentielle stochastique (I.6), qui, en toute rigueur, doit être lue au sens des distributions aléatoires, \mathbb{X} représente l'état du système, F son évolution générale, ξ l'intensité des fluctuations aléatoires autour de cette évolution générale et \mathbb{X}_0 la condition initiale. F est encore appelée la dérive du système et ξ son opérateur de diffusion.

Le vecteur \mathbb{X} est défini par l'état du triplet Véhicule-Infrastructure-Conducteur. Il contient les éléments suivants :

- les caractéristiques cinématiques et dynamiques du véhicule que sont : position, vitesse, accélération, jerk, forces, etc.
- les caractéristiques de l'infrastructure telles que le dévers, la pente, la courbure, l'adhérence, etc.
- les actions du conducteur que sont : les commandes exercées sur véhicule, la distance de visibilité, la perception, etc.

L'équation (I.6) admet une écriture au sens des équations différentielles stochastiques de Itô de la forme :

$$\begin{cases} d\mathbb{X}(t) &= F(t, \mathbb{X}(t))dt + \xi(t, \mathbb{X}(t))dW(t) \\ \mathbb{X}(0) &= \mathbb{X}^0 \text{ p.s.} \end{cases} \quad (\text{I.7})$$

La résolution de (I.7) peut s'effectuer de différentes manières mais cela suppose la connaissance de F et de ξ . Si tel est le cas (I.7) peut alors être résolue soit au sens des *solutions fortes*, soit au sens des *solutions faibles*. Dans le premier cas, il est nécessaire de connaître

des trajectoires de \dot{W} pour obtenir des trajectoires du processus solution \mathbb{X} . Dans le second, il est nécessaire en toute rigueur de connaître la loi du processus solution \mathbb{X} mais en pratique on ne peut atteindre que certaines caractéristiques statistiques de cette loi.

Dans un cas comme dans l'autre, il est nécessaire en pratique de simuler le processus solution \mathbb{X} . Pour simuler la solution des EDS du type (I.7), il est nécessaire d'utiliser des schémas de discrétisation. On en trouve plusieurs dans la littérature [3], possédant chacun ses caractéristiques propres.

Par exemple, le schéma d'Euler possède de mauvaises propriétés de stabilité, comme cela est déjà connu pour les équations différentielles ordinaires. Il impose le choix d'un pas de temps très petit. Il est donc nécessaire d'utiliser d'autres schémas, mais nous ne pouvons pas en faire une étude complète ici, voir [50]. Notons que l'ordre fort $\frac{1}{2}$ du schéma d'Euler-Maruyama peut être insuffisant, et il est intéressant de rechercher des schémas fortement convergents d'un ordre 1 ou supérieur. Un schéma très utilisé par les EDS du type (I.7) est celui de Milstein [50] qui s'écrit en dimension 1, en supposant ξ dérivable et $\xi\xi'$ borné :

$$\begin{cases} \mathbb{X}_{i+1} &= \mathbb{X}_i + hF(\mathbb{X}_i) + \xi(\mathbb{X}_i)\Delta W_i + \frac{1}{2}\xi(\mathbb{X}_i)\xi'(\mathbb{X}_i)((\Delta W_i)^2 - h) , \quad 0 \leq i \leq N \\ \mathbb{X}(0) &= \mathbb{X}_0 \end{cases} \quad (\text{I.8})$$

où h est le pas de discrétisation et $\Delta W_i = W_{i+1} - W_i$ l'accroissement du brownien.

On peut également résoudre l'équation (I.7) par une autre méthode. En effet, le processus solution de (I.7) est un processus de diffusion. Il s'agit donc d'un processus Markovien. Un tel processus est bien entendu entièrement caractérisé par la seule connaissance de sa loi initiale (qui est donnée ici) et de sa famille de probabilités de transitions. Pour déterminer cette famille, il suffit de résoudre l'équation aux dérivées partielles de Fokker-Planck associée à ce processus de Markov (cf. Soize[94]).

Néanmoins, compte tenu de la complexité de la description de l'évolution du véhicule dans son contexte, il est très difficile voire impossible d'obtenir les expressions de F et de ξ . On peut construire des approximations simplificatrices telles que dans les travaux cités dans le paragraphe (I.3.1.2). Le problème est d'autant plus complexe que les paramètres les plus importants de ces modèles sont difficilement identifiables en pratique.

L'expérience montre que le système non linéaire (I.7) n'admet pas de solution stationnaire pour le problème considéré car certains de ses paramètres varient de manière croissante

I.4. Description et formulations du système V-I-C

en fonction du temps. Par contre, il est possible de disposer d'observations d'une partie du vecteur d'état \mathbb{X} . Ce vecteur d'observation associé au vecteur d'état est noté U . Il est obtenu par la relation suivante :

$$U(t) = \eta(\mathbb{X}(t)) \tag{I.9}$$

où $\eta : \mathbb{R}_+ \times \mathbb{R}^m \rightarrow \mathbb{R}^q : t, x \rightarrow \eta(t, x)$ est une fonction généralement très complexe également. Le processus $U = (U(t), t \in \mathbb{R}_+)$ est défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ indexé sur \mathbb{R}_+ à valeurs dans \mathbb{R}^q tel que $q < m$.

Nous venons de mettre en évidence que la modélisation du système V-I-C dans sa globalité est une tâche extrêmement difficile à réaliser. Ceci est dû à la multiplicité de ses paramètres qui, en interagissant, donnent lieu à des problèmes non linéaires et des incertitudes difficilement modélisables.

Pour contourner ces difficultés techniques, nous proposons d'une part, de trouver une représentation pertinente simplifiée du système V-I-C. D'autre part, au lieu de modéliser ce système directement, nous proposons de procéder à une analyse fiabiliste sur cette représentation afin de pouvoir estimer la défaillance de ce système.

Plusieurs travaux tels que Revue [81], Lauffenberger [56], ont montré non seulement que la trajectoire est la résultante du système V-I-C mais également que les pertes de contrôle ou les sorties de route sont dues à une mauvaise estimation de la *trajectoire* du véhicule. La défaillance du système est donc due à une défaillance de la trajectoire. D'où le choix dans ce travail de thèse, d'une méthodologie qui est basée sur l'*observation des trajectoires* du système V-I-C.

Cette méthodologie consiste à identifier plusieurs familles de trajectoires pour une configuration de virage jugée accidentogène. Ensuite, il faut associer un niveau de risque à chaque famille de trajectoires identifiée. Enfin, il suffit de reconnaître la trajectoire en entrée de virage et de l'affecter à sa famille d'appartenance dont on connaît son niveau de risque.

Nous allons présenter dans la section suivante, un *aperçu de cette méthodologie*.

I.5 Aperçu de la solution envisagée

L'objectif de cette thèse est de prédire les trajectoires dangereuses en entrée de virage. La méthode proposée s'appuie sur des approches probabiliste (construction de modèle), fiabiliste (estimation de la probabilité de défaillance) et statistique (reconnaissance de trajectoire). Ces approches opèrent sur les trajectoires observées. Dans cette section, nous présentons quelques définitions et outils indispensables à la compréhension de la démarche.

I.5.1 Définition de la trajectoire du véhicule

La trajectoire peut être considérée de plusieurs façons selon l'espace fonctionnel dans lequel on veut décrire le mouvement du véhicule :

- soit dans l'**espace des paramètres**, où elle est définie comme le graphe de la fonction :

$$\mathcal{T} : T \rightarrow \mathbb{R}^6 : t \rightarrow \mathcal{T}(t) = (x_1(t), x_2(t), x_3(t), \psi(t), \phi(t), \theta(t))^T \quad (\text{I.10})$$

où x_1 , x_2 et x_3 sont les coordonnées cartésiennes du centre de masse du véhicule et ψ , ϕ et θ les angles d'Euler.

- soit dans l'**espace des phases**, où elle est définie comme le graphe de la fonction :

$$u : T \rightarrow \mathbb{R}^6 : t \rightarrow u(t) = (x_1(t), x_2(t), v_1(t), v_2(t), \gamma_1(t), \gamma_2(t))^T \quad (\text{I.11})$$

où $T \subset \mathbb{R}_+$ est l'intervalle temporel d'observation du mouvement du véhicule. Les coordonnées $(x_1, x_2, v_1, v_2, \gamma_1$ et $\gamma_2)$ décrivent le mouvement du véhicule dans la base du repère Galiléen de référence $\mathcal{R}_0^A = (O, \vec{e}_1, \vec{e}_2, \vec{e}_3)$. Pour définir ce repère, soit $\mathcal{B}_0^A = (\vec{e}_1, \vec{e}_2, \vec{e}_3)$ la base orthonormée associée à \mathcal{R}_0^A . Les axes de \mathcal{R}_0^A sont notés Ox_1 , Ox_2 et Ox_3 . Ce repère est supposé être fixe dans le temps. Le mouvement du véhicule a lieu dans le plan horizontal (Ox_1, Ox_2) .

Avec une telle définition de la trajectoire, on fait l'hypothèse que le véhicule se déplace sur une chaussée *plane* sans dénivellation auquel cas les effets de pompage sont réduits. La trajectoire u du centre de masse G du véhicule est orientée et on choisit sur elle un point ς pris pour origine de l'abscisse curviligne s . A chaque instant t , la position de G est définie par cette abscisse curviligne $s(t)$, cf. figure (I.7).

I.5. Aperçu de la solution envisagée

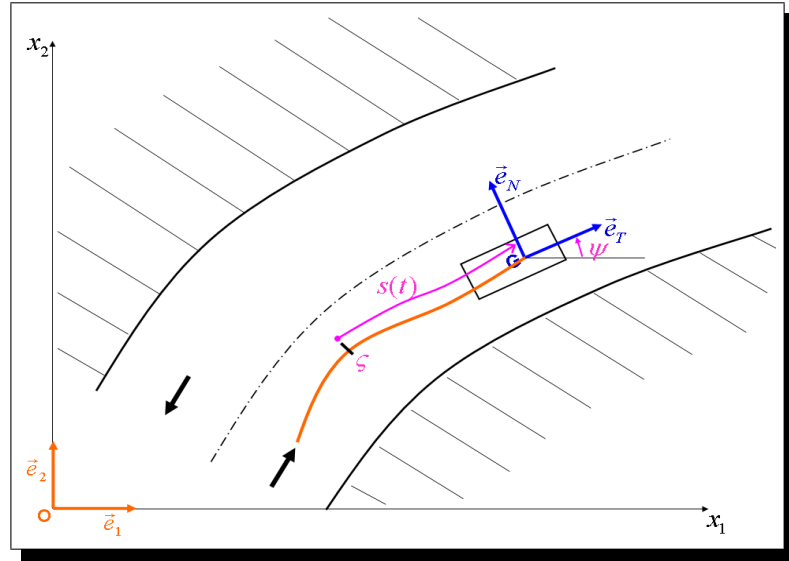


Figure I.7 – Repère galiléen de référence et repère mobile du véhicule

Changement de repère

On peut observer également la trajectoire de G dans le repère mobile ou de Serret-Frénet $\mathcal{R}_G^{SF} = (G, \vec{e}_T, \vec{e}_N, \vec{e}_b)$. Pour définir ce repère, soit $\mathcal{B}_G^{SF} = (\vec{e}_T, \vec{e}_N, \vec{e}_b)$ la base orthonormée associée à \mathcal{R}_G^{SF} .

Rappelons que l'abscisse curviligne s de G est une fonction de t . Le repère \mathcal{R}_G^{SF} et la base \mathcal{B}_G^{SF} sont des fonctions de s et donc de t par l'intermédiaire de s . Le bipoint position \vec{OG} est une fonction de s et donc aussi de t par l'intermédiaire de s .

Les vecteurs de la base \mathcal{B}_G^{SF} sont définis par :

$$\begin{cases} \vec{e}_T &= \frac{d\vec{OG}}{ds} \\ \vec{e}_N &= \rho \frac{d\vec{e}_T}{ds} \\ \vec{e}_b &= \vec{e}_T \wedge \vec{e}_N \end{cases} \quad (\text{I.12})$$

où ρ est le rayon de courbure de la trajectoire en G ; c'est une fonction de s et donc aussi de t par l'intermédiaire de s .

\vec{e}_T , \vec{e}_N et \vec{e}_b sont des fonctions de s de norme 1 et donc de t par l'intermédiaire de s . Le vecteur \vec{e}_T est tel que son support est tangent à la trajectoire en G et contenu dans le plan de coordonnées $(x_1 O x_2)$ du repère fixe \mathcal{R}_0^A . Son sens positif est celui défini par l'orientation positive de la trace de u .

Chapitre I. Système Véhicule-Infrastructure-Conducteur (V-I-C)

Les coordonnées de la trajectoire u sont obtenues dans le repère de Serret-Frénet de la forme :

– Position du centre de masse G du véhicule

Relativement au repère fixe \mathcal{R}_0^A , le point G est défini par ses coordonnées (x_1, x_2, x_3) , avec $x_3 = 0$: $G = (x_1, x_2, 0)$.

– Vitesse du centre de masse G du véhicule

Relativement à la base fixe $\mathcal{B}_0^A = (\vec{e}_1, \vec{e}_2, \vec{e}_3)$

$$\vec{V} = \frac{d\vec{OG}}{dt}, \text{ d'où } , \vec{V} = v_1\vec{e}_1 + v_2\vec{e}_2$$

Relativement à la base de Serret-Frénet $\mathcal{B}_G^{SF} = (\vec{e}_T, \vec{e}_N, \vec{e}_b)$

$$\vec{V} = \frac{d\vec{OG}}{dt} = \vec{e}_T \frac{ds}{dt}$$

D'où : $\vec{V} = V\vec{e}_T$ où $V = \frac{ds}{dt}$ est la vitesse longitudinale au point G .

– Accélération du centre de masse G du véhicule

Relativement à la base fixe $\mathcal{B}_0^A = (\vec{e}_1, \vec{e}_2, \vec{e}_3)$

$$\vec{\Gamma} = \frac{d\vec{V}}{dt}, \text{ d'où } , \vec{\Gamma} = \gamma_1\vec{e}_1 + \gamma_2\vec{e}_2$$

Relativement à la base de Serret-Frénet $\mathcal{B}_G^{SF} = (\vec{e}_T, \vec{e}_N, \vec{e}_b)$

$$\vec{\Gamma} = \frac{d\vec{V}}{dt} = \frac{d(V\vec{e}_T)}{dt} = \frac{dV}{dt}\vec{e}_T + V\frac{d\vec{e}_T}{dt} = \frac{dV}{dt}\vec{e}_T + \frac{V^2}{\rho}\vec{e}_N$$

D'où : $\vec{\Gamma} = \Gamma_T\vec{e}_T + \Gamma_N\vec{e}_N$

où $\Gamma_T = \frac{dV}{dt}$ est l'accélération longitudinale et $\Gamma_N = \frac{V^2}{\rho}$ est l'accélération transversale du véhicule.

Pour passer de la base \mathcal{B}_0^A à \mathcal{B}_G^{SF} , nous utilisons la matrice de passage suivante :

$$P_{\mathcal{B}_0^A \rightarrow \mathcal{B}_G^{SF}} = \begin{bmatrix} \vec{e}_T & \vec{e}_N & \vec{e}_b \\ \cos(\psi) & -\sin(\psi) & 0 \\ \sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{matrix} \vec{e}_1 \\ \vec{e}_2 \\ \vec{e}_3 \end{matrix} \quad (\text{I.13})$$

I.5. Aperçu de la solution envisagée

Liens entre les composantes de \vec{V} dans \mathcal{B}_0^A et \mathcal{B}_G^{SF}

$$\begin{pmatrix} v_1 \\ v_2 \\ 0 \end{pmatrix}_{\mathcal{B}_0^A} = P_{\mathcal{B}_0^A \rightarrow \mathcal{B}_G^{SF}} \begin{pmatrix} V \\ 0 \\ 0 \end{pmatrix}_{\mathcal{B}_G^{SF}} \quad (\text{I.14})$$

Liens entre les composantes de $\vec{\Gamma}$ dans \mathcal{B}_0^A et \mathcal{B}_G^{SF}

$$\begin{pmatrix} \Gamma_T \\ \Gamma_N \\ 0 \end{pmatrix}_{\mathcal{B}_G^{SF}} = P_{\mathcal{B}_0^A \rightarrow \mathcal{B}_G^{SF}}^{-1} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ 0 \end{pmatrix}_{\mathcal{B}_0^A} \quad (\text{I.15})$$

$$\Rightarrow \begin{cases} \Gamma_T = \gamma_1 \cos(\psi) + \gamma_2 \sin(\psi) \\ \Gamma_N = -\gamma_1 \sin(\psi) + \gamma_2 \cos(\psi) \end{cases} \quad (\text{I.16})$$

Nous rappelons que ψ est un paramètre non seulement mesuré par les capteurs mais il est également connu à travers la relation suivante :

$$\text{tg}(\psi) = \frac{v_2}{v_1} \quad (\text{I.17})$$

Nous venons de montrer le passage du repère Galiléen de référence au repère de Serret-Frénet. Cependant, nous ne disposons que d'observations expérimentales de trajectoires en virage. Celles-ci sont par nature entâchées d'incertitudes. C'est pourquoi les méthodes de modélisation déterministes citées dans la section (I.3) sont mal adaptées. Nous verrons dans le paragraphe suivant comment tenir compte de ces *incertitudes* (aléas).

I.5.2 Aléas (incertitudes) du système V-I-C

L'aléa porte sur les phénomènes intrinsèquement aléatoires, liés ou non à l'*environnement du véhicule* ou au *conducteur*, agissant sur les trajectoires (par ex. effets d'actions extérieures comme le vent, ou de réactions inattendues du conducteurs, ou encore d'interactions liées à un incertain probabiliste comme une réduction d'adhérence sur une plaque de verglas, une flaque d'huile, etc.). En pratique, la trajectoire ne peut être prédite avec certitude et il est nécessaire de prendre en compte les incertitudes qui l'affectent. Ainsi, le même conducteur circulant avec le même véhicule sur la même route dans les mêmes conditions ne reproduira pas deux fois la même trajectoire au sens de la définition (I.11).

Pour tenir compte de ces incertitudes, nous assimilons toute trajectoire $t \rightarrow u(t)$ à une réalisation d'un processus stochastique $U = (U(t), t \in T)$ défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R}^6 . On aura alors $u(t) = U(t, \omega)$, $\forall t \in T$, avec $\omega \in \Omega$:

$$\forall \omega \in \Omega, u = U(., \omega) : T \rightarrow \mathbb{R}^6 : t \rightarrow u(t) = U(t, \omega) \quad (\text{I.18})$$

A partir de cette définition de la trajectoire, nous proposons une approche fiabiliste pour évaluer le niveau de risque de défaillance des trajectoires.

I.5.3 Définition des outils d'analyse fiabiliste du système V-I-C

L'analyse fiabiliste classique consiste à se donner des critères de défaillance, à leur associer des états limites et à estimer les probabilités que ces critères ne soient pas respectés. Nous allons rappeler quelques concepts fiabilistes tels : défaillance, types et critères de défaillance, états limites, probabilité de défaillance, etc. Ces concepts sont nécessaires et utiles pour la compréhension de l'approche proposée.

I.5.3.1 Défaillance du système V-I-C

Le *système* est réputé défaillant s'il produit une trajectoire défaillante, c'est-à-dire, une trajectoire qui n'appartient pas au sous-ensemble des trajectoires admissibles pour lequel des conditions de sécurité spécifiées ont été définies. Ces conditions portent sur les coordonnées de la trajectoire (positions, vitesses et accélérations). A partir de cette définition, pour étudier la défaillance du système, nous déterminons différents types de défaillance de ce système.

I.5.3.2 Types de défaillance

Différents types de défaillance de trajectoires (en virage) peuvent être trouvés dans la littérature, qui se traduisent par une perte de contrôle ou une sortie de voie du véhicule. Parmi ceux-ci, on peut citer :

- la sortie du véhicule de l'espace roulant,
- une perte d'adhérence du véhicule rendant celui-ci incontrôlable par le conducteur,

Par rapport à ces types de défaillance, plusieurs critères (ou marqueurs) construits à partir de variables de base engendrent (seul ou de manière combinée) la défaillance de la

I.5. Aperçu de la solution envisagée

trajectoire. La question à résoudre est la suivante : comment une défaillance de la trajectoire est-elle révélatrice de la défaillance du système V-I-C ? Pour cela, il faut traduire en langage fiabiliste chaque type de défaillance en critère de défaillance.

I.5.3.3 Critères de défaillance

La défaillance de la trajectoire peut être caractérisée de plusieurs façons selon le critère de défaillance considéré.

Soit on utilise une **distance** par rapport à une trajectoire de référence. Celle-ci est sûre par construction. La probabilité de sortie de route est alors une fonction croissante de cette distance. La notion de distance entre trajectoires se complique par rapport à la notion de distance usuelle du fait, d'une part, de la dimension de l'espace des phases, et d'autre part, de la dépendance des variables d'accélération, vitesses et positions. La notion naturelle de distance est associée à la notion d'écart en accélération et/ou en vitesse et/ou en position. L'*inconvenient* de cette notion, est qu'elle conduit à sommer des écarts sur des variables de dimensions physiques différentes. Dans cette étude, nous chercherons à définir une distance plus cohérente. Nous ferons également un choix de la trajectoire de référence nécessaire pour le calcul de la distance entre trajectoires.

Soit on utilise une **fonctionnelle des coordonnées** de la trajectoire u dans l'espace des phases. Cette fonctionnelle peut-être de la plus simple (une seule coordonnée) jusqu'à la plus complexe (toutes les coordonnées). Nous choisirons les fonctionnelles les plus pertinentes pour étudier la défaillance des trajectoires.

Les deux types de caractérisation du critère de défaillance du système sont intéressants. Le *premier* génère la notion de risque conditionnel, i.e. la probabilité d'accident sachant un niveau de risque de référence donné. Le *second* est anticipatif en ce sens qu'il donne la possibilité au conducteur de choisir entre avoir un accident ou effectuer telle manoeuvre pour continuer le roulage.

A partir des critères de défaillance à déterminer au chapitre (II), nous essayerons d'associer un modèle à chaque critère afin de simuler la dynamique du système *vis-à-vis* de ces critères de défaillance. A cet effet, nous proposons d'utiliser les méthodes probabilistes car elles permettent de construire des modèles simples, robustes et plus réalistes (prise en compte des aléas) que tous les modèles présentés dans la section (I.3).

I.5.3.4 Simulation de l'état dynamique du système V-I-C

L'intérêt de la simulation est de permettre d'envisager toutes les situations critiques, et de mener des études quantitatives dans ces situations. Les modèles qui permettent ces simulations sont issus de la phase d'identification et sont appelés modèles probabilistes « boîte noire ». Ils permettent de prédire le comportement d'un système dynamique en fonction de sollicitations spécifiques. Ils sont utilisés dans plusieurs domaines. La plupart de ces applications sont traitées dans l'ouvrage de Bouleau [11].

Ces modèles probabilistes consistent à caractériser et identifier des processus stochastiques. Une modélisation utilisant des processus stochastiques est adaptée pour traiter le problème avec la théorie des valeurs extrêmes. Cette théorie est une branche des probabilités et statistiques dont l'objectif est de caractériser le comportement stochastique des queues de distribution de processus aléatoires (c.f. Jacob[46]). Les résultats, obtenus dans un cadre asymptotique, permettent d'établir des modèles mathématiques qui peuvent alors servir de guides dans des problématiques d'extrapolation.

Notons que la modélisation probabiliste des trajectoires discrètes de véhicules en virages est originale. Pour mener à bien le problème posé, plusieurs questions restent ouvertes techniquement et théoriquement. Entre autres, faut-il construire un modèle unique ou un modèle par comportement de conduite ? Quels sont les algorithmes les mieux adaptés pour identifier chaque comportement de conduite ? Est-ce que les trajectoires considérées comme des réalisations du processus U sont indépendantes ? Est-ce que les processus U modélisant les observations u sont stationnaires ou non ? Si non peut-on se ramener à une situation stationnaire ? Est-ce que les caractéristiques statistiques (les densités de probabilité, les densités spectrales de puissance, les fonctions d'autocorrélation, etc.) nécessaires à la simulation des processus stochastiques concernés pourront être correctement estimées ? Quelles seront les meilleures méthodes d'estimation de ces caractéristiques ? Quelle sera la meilleure méthode de simulation de ces processus ?

Dans ce manuscrit, nous apporterons des réponses précises à ces questions en nous appuyant, d'une part sur des travaux de Fogli[36] et Soize[93] pour l'estimation et l'identification des processus et d'autre part sur des travaux de Poirion[76], Puig[79], Krée[55], Bernard[5] et Fogli[37] pour la simulation des processus stochastiques. En utilisant ces techniques de simulation, on pourra alors évaluer le risque une fois définies les fonctions d'états limites.

I.5. Aperçu de la solution envisagée

I.5.3.5 Etats limites

Un état limite est un ensemble de trajectoires constitué de la frontière ou la réunion des frontières d'un *domaine de sécurité*, dans l'espace des trajectoires muni de sa topologie. Les états limites ne sont pas uniques, mais pour un même contexte on peut considérer plusieurs états limites, liés à des domaines de sécurité inclus (emboîtés) les uns dans les autres, ou pas, en fonction des situations que l'on veut éviter et de la modélisation adoptée. Le lecteur peut se référer au glossaire de l'opération de recherche (MTT [67]).

A titre d'exemple, pour une route à double sens sans séparateur physique central, un état limite pourra inclure le franchissement de la ligne de séparation de voies par sens (risque de collision frontale ou violation d'une interdiction de dépasser), tandis qu'un autre état limite pourra l'ignorer (dépassement autorisé).

On retiendra l'existence de deux types classiques d'états limites, rencontrés en fiabilité des structures et des systèmes. Ces deux types se transposent à la sécurité routière :

- l'Etat Limite Ultime (ELU), ainsi appelé car son franchissement conduit à une défaillance irréversible du système ou à la perte de l'intégrité du système. On peut citer par exemple les pertes de contrôles de véhicules non rattrapables, des chocs avec des atteintes à l'intégrité des personnes, ou des dégâts matériels significatifs, etc.
- l'Etat Limite de Survie (ELS), ainsi dénommé car son franchissement conduit à une défaillance temporaire et réversible du système. Typiquement lors du franchissement d'un ELS le véhicule se trouve en dehors des conditions normales d'usage sur l'infrastructure en situation de risque élevé d'accident mais sans atteindre une perte d'intégrité du véhicule sur l'infrastructure.

Nous considérerons d'abord plusieurs scénarii de défaillance du système qui sont indépendants les uns par rapport aux autres. Par exemple, parmi les 3 types de défaillances choisis, on peut sortir de la route sans avoir un inconfort important car le virage a été mal négocié. L'hypothèse d'indépendance entre les scénarii de défaillance implique plusieurs états limites indépendants. Par contre, les défaillances liées à l'interaction entre des états limites ne sont pas prises en compte. Dans le chapitre (II), les états limites nécessaires pour estimer les probabilités de défaillance des trajectoires seront proposés. Ensuite, après le calcul des probabilités de défaillance et en faisant une analyse de sensibilité par rap-

port aux critères choisis, nous vérifierons si des interactions existent entre ces états limites.

Après la définition des concepts fiabilistes, dans le paragraphe suivant, nous donnons les méthodes et les outils utilisés pour estimer les *probabilités de défaillance* du système.

I.5.3.6 Estimation de la probabilité de défaillance

La fiabilité d'un système, qu'il soit physique ou non, se mesure par sa probabilité de fonctionner sans défaillance pendant une durée donnée et dans un environnement spécifié Blanquart[7], Wasserman[104]. La notion temporelle de fiabilité se distingue par exemple par l'augmentation de la probabilité de défaillance en fonction du temps. Il existe en génie civil plusieurs travaux pour évaluer le risque qu'une structure soit défaillante.

Certains travaux considèrent une représentation probabiliste, dans laquelle les densités de probabilité sont statistiquement connues. L'emploi d'un outil de fiabilité adéquat pourrait conduire à un calcul précis de la probabilité de défaillance. D'autres représentations existent telles que les représentations imprécises, dans lesquelles les lois de probabilité ne correspondent pas forcément à la réalité (probabilité de défaillance imprécise), Utkin[100]. Il existe aussi des travaux considérant une représentation floue, pour laquelle, dans le cas d'une imprécision sur les densités de probabilité, des représentations par des ensembles flous sont utilisées. L'incertitude est alors munie d'une fonction d'appartenance, et les opérateurs flous sont employés pour donner une valeur approximative de la fiabilité, Reddy[80], Chen[20], Lei[58], Nachtmann[68]. Enfin, on peut également citer la représentation par intervalles, pour laquelle l'incertitude est déterminée par un simple intervalle (ou par un volume dans un espace de dimension élevée) définissant les plages de variation possibles des données (interval analysis). Cette représentation est utile dans le cas où l'on n'a pas d'information sur les incertitudes ou pour des raisons de gain de temps de calcul, Chang[18]. Dans certains travaux de la littérature, les bornes des intervalles ont été considérées comme des données incertaines, Hall[43] ainsi que la probabilité de défaillance, Tonon[99].

Le choix des outils d'analyse fiabiliste dépend fortement du problème posé. Il n'existe pas de méthode universelle permettant de faire face à la multiplicité des situations en pratique. Les décisions méthodologiques résultent donc de l'expérience acquise. Dans ce travail de thèse, nous utiliserons la technique de « représentation probabiliste » pour l'analyse des trajectoires défaillantes car les « densités de probabilité » sont statistiquement connues.

I.5. Aperçu de la solution envisagée

Plusieurs méthodes existent dans la littérature pour le calcul de cette probabilité : des méthodes de *simulation* et des méthodes d'*approximation*, Melchers[62], Lemaire[59], Fogli[36]. Les premières sont basées sur des simulations de Monte Carlo, mais le coût de calcul est souvent prohibitif. Les secondes sont basées sur des approximations de la fonction d'état-limite, ce qui permet une réduction importante du temps de calcul.

Plusieurs auteurs se sont attachés à donner une meilleure approximation de la probabilité de défaillance à partir d'une connaissance plus complète de la géométrie de la fonction d'état-limite au voisinage du point de conception. Les méthodes classiques sont : FORM(First Order Reliability Method)/SORM (Second Order Reliability Method), les techniques de développements asymptotiques et les méthodes de surfaces de réponse.

Nous n'utiliserons pas ces méthodes classiques de calcul de probabilité de défaillance car d'une part, la démarche n'est pas basée sur la connaissance du modèle *mécano-numérique* décrivant la dynamique du système. D'autre part, nous pourrions déterminer la *loi de probabilité des maxima* des processus stochastiques qui gouvernent le système une fois que les processus stochastiques sont identifiés. Et la connaissance de cette loi permet une estimation de la probabilité de défaillance de ce système plus simple et plus précise.

Comme application de la fiabilité dans le domaine de la sécurité routière, citons les travaux de Echaveguren[32]. Ces travaux consistent à prendre la différence entre la vitesse maximale réalisable et la vitesse de base comme une marge de sécurité. A partir de cette marge, il estime la probabilité de défaillance dans des conditions opérationnelles afin de déterminer les variables ayant un grand impact sur la fiabilité. Les résultats obtenus dans son étude ont démontré que le rayon de courbure, la résistance au glissement et la macrotexture sont des variables qui ont un grand impact sur la probabilité de défaillance. Citons également le travail de Sellami[91] dans le cadre de l'accidentologie des poids lourds. Ce travail consiste à estimer la probabilité de défaillance en fonction du critère de renversement noté LTR (Load Transfer Ratio). Ce critère est constitué des paramètres qui gouvernent la dynamique du véhicule. Nous citons également les travaux de Rey et al.[82].

Après avoir décrit les cadres probabiliste (construction et simulation des modèles) et fiabiliste (estimation des probabilités de défaillance) du problème posé, nous abordons les méthodes de reconnaissance nécessaires pour pouvoir associer toute trajectoire u à sa famille de trajectoires en entrée de virage.

I.5.4 Reconnaissance de la trajectoire par la méthode des SVMs

Nous proposons de construire un modèle de reconnaissance de trajectoires basé sur les *techniques d'apprentissage* de type Séparateurs à Vaste Marges (SVM). L'origine des SVMs remonte aux années 70 lorsque Vapnik et Chervonenkis proposèrent le principe du risque structurel et la VC-dimension. Ce n'est qu'en 1982 que Vapnik propose un premier classifieur basé sur la minimisation du risque structurel baptisé SVM, [103]. Ce modèle était linéaire et on ne savait pas encore traiter les problèmes avec des frontières de décision non linéaires.

En 1992, Boser et al proposent d'introduire des noyaux non-linéaires pour étendre le SVM au cas non-linéaire Boser[9]. En 1995, Cortes et al. proposent une version régularisée du SVM qui tolère les erreurs d'apprentissage tout en les pénalisant [23].

Depuis, différentes variantes de l'algorithme SVM ont été proposées par plusieurs chercheurs et appliquées à divers domaines. Par exemple, dans le domaine de la reconnaissance de chiffres manuscrits isolés on peut citer les travaux de Cortes [23], Scholkopf [90] et Burges [16]. Les travaux de Blanz [8] ont expérimenté les SVM sur des problèmes de reconnaissances d'objets. Les SVM sont appliqués dans d'autres domaines que la classification, notamment la regression Smola[92] et l'estimation de densité Vapnik[102], etc.

Le succès de cette méthode est justifié par les solides bases théoriques qui la soutiennent. Les SVM est une méthode de classification particulièrement bien adaptée pour traiter des données de grande dimension.

Avec un tel outil, nous pourrons utiliser le passé du système V-I-C par apprentissage pour prédire la famille (ou classe) d'appartenance de chaque trajectoire u en entrée de virage. Rappelons que dans la partie analyse fiabiliste nous proposons de construire des abaques de probabilité de défaillance par classe de trajectoires. Donc la connaissance de la classe de trajectoires permettra de connaître la probabilité de défaillance. A partir de cette probabilité de défaillance, nous proposons de construire un indicateur de risque.

Ceci est une indication pour le conducteur ou le gestionnaire d'infrastructure d'estimer la dangerosité de la trajectoire en cours de réalisation.

Nous allons exposer brièvement les grandes étapes de la *methodologie retenue*.

I.5. Aperçu de la solution envisagée

I.5.5 Résumé de l'approche proposée

La figure (I.8) représente les principales étapes de la première partie de la démarche. Cette première partie consiste de partir des trajectoires expérimentales jusqu'à l'estimation des probabilités de défaillance pour chaque comportement de conduite.

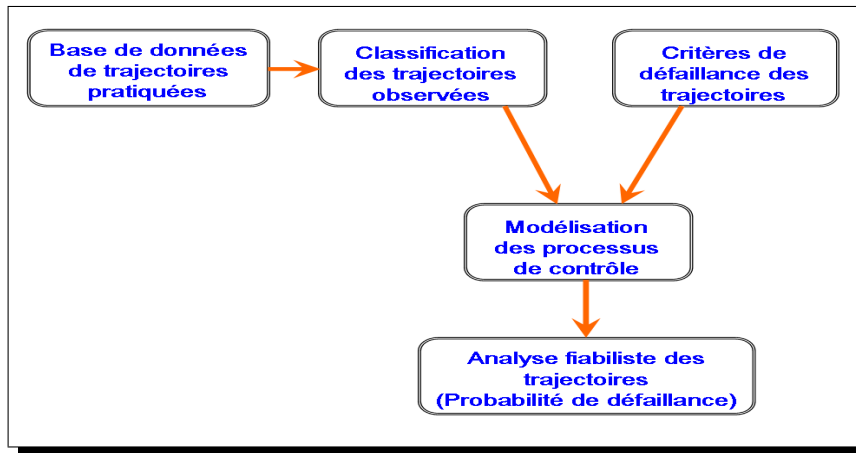


Figure I.8 – Architecture de la démarche proposée (1^{ère} partie)

La figure (I.9) représente l'affectation d'une trajectoire en cours de réalisation à sa famille d'appartenance dont on connaît sa probabilité de défaillance *via* la première architecture.

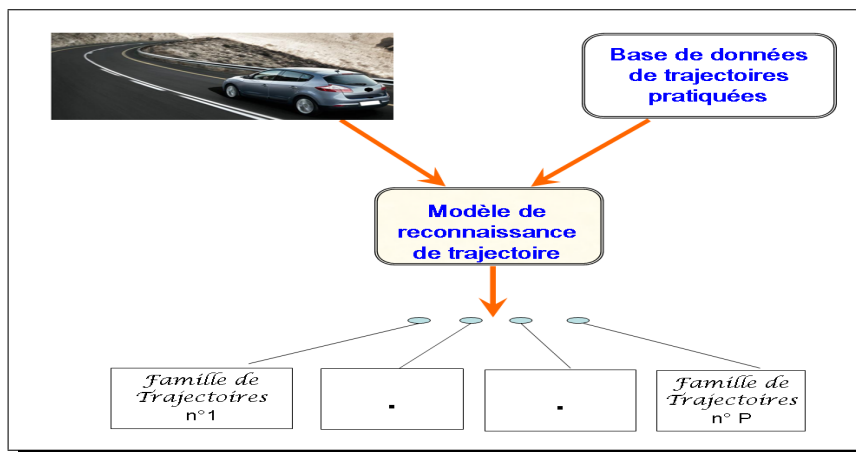


Figure I.9 – Architecture de la démarche proposée (2nd partie)

L'utilisation des deux architectures permettra de calculer l'indicateur de risque.

I.6 Conclusion

L'objet de ce chapitre (I) était de faire : (i) *une analyse accidentologique* des véhicules, (ii) *un état de l'art* sur le système V-I-C, (iii) une description de la *complexité* de la modélisation de ce système et (iv) un *aperçu de la méthodologie* proposée.

L'*analyse accidentologique* a montré d'une part que la France dispose d'une marge de manoeuvre pour réduire l'accidentologie des véhicules légers. D'autre part, qu'il existe des facteurs accidentogènes tels que la vitesse longitudinale ou l'accélération latérale. L'adaptation de ces facteurs se traduit par une perte de contrôle ou une sortie de route.

L'*état de l'art* a montré que des travaux scientifiques et industriels existent pour réduire ces accidents. Malgré ces résultats probants et significatifs, des phénomènes complexes tels que la visibilité, l'adhérence ou encore la perception sont difficilement modélisables par ces approches. Le système est souvent sous représenté dans ces travaux. La complexité du système et la non prise en compte des aléas nous oblige à proposer de nouvelles méthodes.

A partir de la description du système, nous avons proposé des formulations mathématiques permettant de le modéliser en intégrant toute la complexité. Le peu de connaissance sur le système ne nous permet pas d'avoir un modèle global utilisable en pratique.

Dans l'*aperçu de la méthodologie* proposée, des définitions et des outils ont été introduits pour la compréhension. La trajectoire est définie comme la réalisation d'un processus stochastique. Cette définition intègre les incertitudes (aléas) du système. Un modèle sera construit par identification de ce processus ou de sa transformation. Cette méthode de modélisation est proposée car d'une part, elle fait *abstraction* de toutes les interactions non linéaires qui sont difficilement modélisables. Et d'autre part, on dispose d'*instruments de mesures* de trajectoires discrètes avec une grande précision. La modélisation des données pour des processus stochastiques est une étape essentielle préalable à une analyse fiabiliste. Ainsi, les modèles probabilistes conçus serviront à alimenter l'analyse fiabiliste nécessaire pour calculer les probabilités de défaillance. Enfin, nous proposons d'élaborer un modèle de reconnaissance permettant d'associer à chaque trajectoire un niveau de risque préalablement estimé.

Le prochain chapitre décrit de façon détaillée la **méthodologie proposée** pour prédire les trajectoires à risque. Cette partie constitue le coeur de la thèse.

Méthodologie de prédiction de trajectoires à risque

II.1 Introduction

L'objet de ce chapitre II est de décrire la méthodologie proposée dans ce travail de thèse. Nous rappelons que cette méthodologie consiste, pour une configuration de virage donnée et d'observations représentatives de trajectoires pratiquées dans ce virage de prédire les trajectoires à risque en entrée de virage.

Dans un premier temps, nous proposons d'acquérir des trajectoires expérimentales et d'identifier différents comportements de conduite dans ces trajectoires pratiquées. Un comportement de conduite est caractérisé par un sous-ensemble de trajectoires appartenant à une même classe. Deux approches existent pour décrire une classe : générative et distinctive. La première décrit une classe par des propriétés caractéristiques d'objets qui la composent tandis que la deuxième décrit une classe par sa frontière avec ses voisins.

Ensuite, à partir des trajectoires expérimentales de chaque classe, nous construisons des modèles *probabilistes* robustes basés sur l'identification de processus stochastique. Le but de ces modèles est de simuler la dynamique du système V-I-C *vis-à-vis* de critères de défaillance. En effet, les modèles proposés sont spécifiques à des classes de trajectoires judicieusement identifiées.

En utilisant les simulations de ces processus stochastiques, nous évaluerons des probabilités de dépassement du seuil des critères de défaillance. Ceux-ci sont définis indépendamment des modèles. Le but de cette analyse fiabiliste est d'associer une probabilité de défaillance à chaque classe de trajectoires.

Enfin, pour compléter la méthodologie, nous élaborons un modèle de reconnaissance permettant d'affecter en entrée de virage chaque trajectoire à sa classe d'appartenance, et de lui attribuer un niveau de risque préalablement estimé à travers un indicateur de risque.

II.2 Classification des trajectoires u

L'identification des classes de trajectoires nécessite d'une part, de disposer d'un ensemble de trajectoires et d'autre part, d'utiliser des outils de classification de données.

II.2.1 Ensemble des trajectoires u observées

L'acquisition d'un tel ensemble de trajectoires se fait, soit par une expérimentation sur site réel, soit par l'utilisation d'un simulateur de conduite. Pour la pertinence de l'étude, nous proposons de faire une expérimentation sur site réel dans le but d'acquérir un important panel de trajectoires se rapprochant de celles qui seraient pratiquées en situation de conduite naturelle. Sur site réel, il faut disposer d'un *véhicule instrumenté* généralement plus précis en mesures ou d'un *observatoire de trajectoires* bord de voie plus riche en panel de conducteurs. Les deux systèmes de mesures (véhicule instrumenté et observatoire de trajectoires bord de voie) permettent de mesurer les coordonnées de la trajectoire u du véhicule à des instants discrets t_k .

Pour mener à bien l'acquisition de l'ensemble des trajectoires observées, un certain nombre de règles sont nécessaires. Entre autres, il faut garantir une certaine représentativité des trajectoires pratiquées, la précision des mesures avec une fréquence d'acquisition élevée, etc. Les observations discrètes collectées constituent l'ensemble des trajectoires observées. C'est une L -famille de la forme :

$$\mathcal{D} = \left(u_k^{(l)} \right)_{k \in I(l)}^{1 < l \leq L} \quad (\text{II.1})$$

où $u_k^{(l)} = u^{(l)}(t_k)$, L est le nombre total de trajectoires disponibles, $I(l) = \{1_l, 2_l, \dots, K_l\}$ et $K_l \in \mathbb{N}^*$ est le nombre d'observations de la $l^{\text{ème}}$ trajectoire.

II.2. Classification des trajectoires u

A tout $l \in \mathcal{L} = \{1, \dots, L\}$ est associé un unique passage d'automobiliste dans le virage :

$$u_k^{(l)} = \left(x_{1,k_l}^{(l)}, x_{2,k_l}^{(l)}, v_{1,k_l}^{(l)}, v_{2,k_l}^{(l)}, \gamma_{1,k_l}^{(l)}, \gamma_{2,k_l}^{(l)} \right)^T \in \mathbb{R}^6 ; k \in I(l) \quad (\text{II.2})$$

Pour tout $l \in \{1, \dots, L\}$ et pour tout $k \in I(l)$, $u_{k_l}^{(l)}$ est la valeur à l'instant t_k du l^{eme} passage du véhicule dans le virage. Pour tout $k_l \in I(l)$ le vecteur u_{k_l} est une observation de la trajectoire du véhicule (constituée des positions, vitesses et accélérations) dans le repère Galiléen de référence \mathcal{R}_0^A .

Il n'est pas évident, *a priori*, d'associer à un même processus stochastique les L trajectoires observées sur une configuration de virage donnée compte tenu, par exemple, des effets d'accoutumance ou de la perception des conducteurs ou encore de l'hétérogénéité des conducteurs. Remarquons toutefois qu'il est très difficile (voire impossible) de créer des conditions d'observations expérimentales d'un même processus. C'est la raison pour laquelle il est judicieux de rechercher des classes de trajectoires similaires sur la totalité du parcours.

Une classe \mathcal{C}_p sera constituée de L_p trajectoires expérimentales issues d'un même processus stochastique noté U_p défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R}^6 . Avec $p = \{1, \dots, P\}$ où P est le nombre de classes de trajectoires à identifier. Pour répartir l'ensemble des trajectoires de la L -famille \mathcal{D} en plusieurs classes ayant des propriétés statistiques homogènes, on utilisera des algorithmes de classification.

II.2.2 Méthodes de classification des trajectoires

L'objectif de la classification est de construire P classes $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_P$ contenant respectivement : L_1, L_2, \dots, L_P trajectoires (ou objets) de la L -familles \mathcal{D} . La classification est basée sur la mesure de proximité ou de dissemblance entre les trajectoires de \mathcal{D} .

En général, les méthodes de classification utilisent la technique d'apprentissage. Cette dernière est vue comme une fonction f_θ de \mathcal{D} dans l'espace de sortie \mathcal{Y} :

$$f_\theta : \mathcal{D} \rightarrow \mathcal{Y} \quad (\text{II.3})$$

où $\mathcal{Y} \subset \mathbb{N}$ représente les indices des classes \mathcal{C}_p de la partition de \mathcal{D} , et f_θ est une fonction spécifique à la méthode de classification choisie. La figure (II.1) est une illustration de la transformation de l'espace d'entrée en espace de sortie.

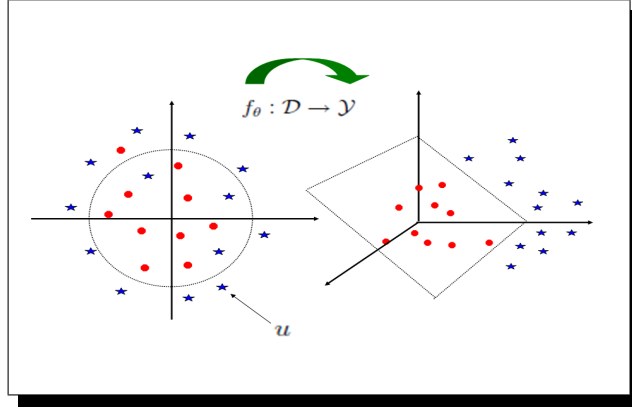


Figure II.1 – Transformation de l'espace d'entrée \mathcal{D} en espace de sortie \mathcal{Y}

La fonction de sortie peut être soit continue (problème de regression), soit discrète (problème de classification). Dans la littérature, il existe plusieurs méthodes de classification [30]. Parmi ces méthodes, nous avons choisi : les nuées dynamiques [70], le meanshift [21] et le Self Organising Maps [86] du fait de leur robustesse par rapport aux autres méthodes de classification de données. Ils sont également rapides en terme de temps de calcul.

II.2.2.1 Classification par nuées dynamiques

La méthode des **nuées dynamiques** due à Diday [27], utilise P noyaux (N_1, N_2, \dots, N_P) comportant chacun : L_1, L_2, \dots, L_P éléments de la population initiale de trajectoires. Chaque classe \mathcal{C}_p est caractérisée par son centre (ou noyau) qui correspond au barycentre ou à la moyenne des éléments qui le composent. Le but de cet algorithme est de trouver les centres de classes des objets (trajectoires) en minimisant la somme des variances intra-classes. L'algorithme est résumé en 4 étapes mais pour plus de détails voir [65].

Etape 1, On génère P noyaux, soit aléatoirement par manque d'avis d'expert du système, soit en les fixant *a priori* selon l'avis d'expert sur la connaissance physique du système.

Etape 2, On calcule la distance $D(u, N_p)$ pour tout objet u et pour tout noyau N_p . Cette distance représente la distance entre un objet et un groupe d'objets (les éléments du noyau).

$$D(u, N_p) = \frac{1}{L_p} \sum_{u^{(2)} \in N_p} d^2(u^{(1)}, u^{(2)}) \quad (\text{II.4})$$

où d une application de $\mathcal{D} \times \mathcal{D}$ dans \mathbb{R}_+ vérifiant les propriétés de symétrie, de séparation

II.2. Classification des trajectoires u

et d'inégalité triangulaire. La distance d entre un objet $u_p^{(1)}$ et un objet $u_p^{(2)}$ doit être choisie, par l'utilisateur, parmi l'ensemble des indices de distance utilisés pour une mesure de similarité. La notion de distance entre trajectoires se complique par rapport à la notion de distance usuelle du fait de la dimension de l'espace des phases, d'une part, et du fait de la dépendance des coordonnées de la trajectoire u , d'autre part. La notion naturelle de distance est associée à la notion d'écart en accélération et/ou en vitesse et/ou en position. Pour les raisons citées ci-dessus, la distance Euclidienne d_E , est mal adaptée pour bien caractériser l'écart entre les trajectoires. Dans une étude antérieure (cf. Koita et al.[52]), nous avons utilisé la **distance de Mahalanobis** pour tenir compte de la structure de corrélation entre les composantes de la trajectoire u . Cette distance de Mahalanobis est définie par :

$$d_M(u^{(1)}, u^{(2)}) = \sqrt{(u_k^{(1)} - u_k^{(2)})^T \tilde{\Sigma}^{-1} (u_k^{(1)} - u_k^{(2)})} \quad (\text{II.5})$$

où $\tilde{\Sigma}$ est l'estimée de la matrice de covariance du vecteur aléatoire :

$$(U_p(t_1), \dots, U_p(t_K))^T = (U_{p,1}, \dots, U_{p,K})^T \quad (\text{II.6})$$

Etape 3, On alloue tous les objets aux P noyaux. Un objet u sera alloué au noyau N_p , si sa distance $D(u, N_p)$ est la plus faible des distances à l'ensemble des noyaux. Un objet $u \in \mathcal{C}_p$ si :

$$D(u, N_p) = \min \{D(u, N_r) ; r = 1, \dots, p\} \quad (\text{II.7})$$

La classe \mathcal{C}_p est donc définie comme suit :

$$\mathcal{C}_p = \{u \in \mathcal{D} ; D(u, N_p) \leq D(u, N_r) , \forall p \neq r\} \quad (\text{II.8})$$

Etape 4, on recherche de nouveaux noyaux (N_1, N_2, \dots, N_p) en prenant les n_S objets les plus proches de la classe \mathcal{C}_p . Les éléments sont les objets qui minimisent une fonction \mathcal{R} distance d'un objet u à une classe \mathcal{C}_p . Pour une partition donnée en p classes et pour une classe \mathcal{C}_p de la partition, on a :

$$\mathcal{R}(u, p, \mathcal{P}) = D(u, \mathcal{C}_p). \quad (\text{II.9})$$

Remarque 1. La fonction qui fait correspondre à l'ensemble N des noyaux une partition \mathcal{P} est appelée par *Diday fonction d'agrégation* $f : \mathcal{P} = f(N)$.

Remarque 2. La fonction qui fait correspondre à la partition \mathcal{P} un ensemble N de noyaux est appelée *fonction d'écartement* $g : N = g(\mathcal{P})$.

Chapitre II. Méthodologie de prédiction de trajectoires à risque

Les étapes 2,3,4 sont réitérées jusqu'à l'obtention d'une convergence selon le critère :

$$W = \sum_{s=1}^p \sum_{u \in \mathcal{C}_p} \mathcal{R}(u, s, \mathcal{P}) \quad (\text{II.10})$$

Le test le plus strict de la convergence est la vérification que la composition du noyau est invariante d'une itération à une autre. En pratique on arrête le calcul, soit après un nombre maximum d'itérations, soit sur un seuil de décroissance :

$$\left| \frac{W^{(i)} - W^{(i+1)}}{W^{(i)}} \right| \leq \epsilon \quad (\text{II.11})$$

où $W^{(i)}$ la valeur du critère à l'itération i et $\epsilon > 0$ sert à définir un critère de proximité.

Remarque 3. *En minimisant W , l'algorithme des nuées dynamiques conduit à faire décroître la variance intra groupe, la variance des noyaux et à rapprocher chaque noyau du centre du groupe qui lui est associé.*

Pour illustrer nos propos, nous résumons la méthodologie des nuées dynamiques sous la forme d'une architecture.

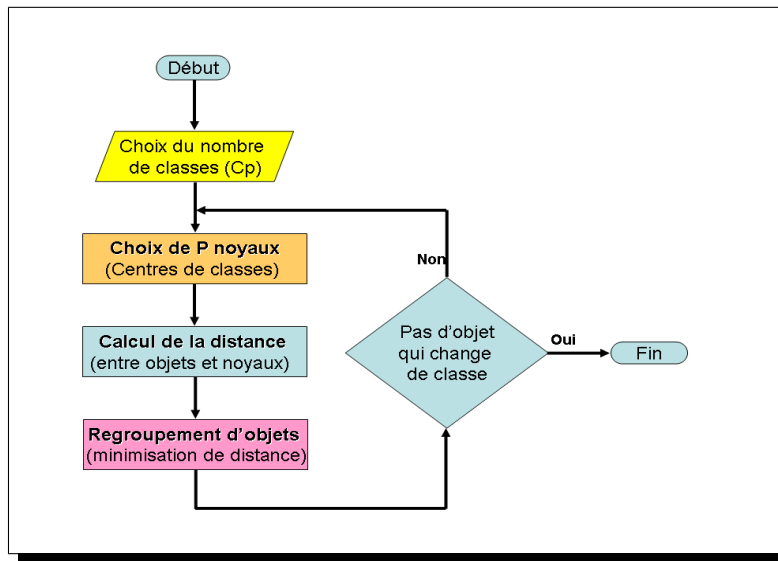


Figure II.2 – Illustration de la méthode des nuées dynamiques

Cet algorithme est souvent utilisé pour des raisons de simplicité de mise en oeuvre. Par contre, les résultats qu'il donne sont parfois très sensibles à la condition initiale donc pas d'unicité de la solution à cause de minimums locaux.

II.2. Classification des trajectoires u

II.2.2.2 Classification par Mean shift

La méthode **mean shift** a été introduite par Fukunaga and Hostetler en 1975 [38]. C'est une technique d'estimation itérative non paramétrique basée sur l'utilisation d'un noyau. Elle consiste à détecter les modes, i.e le maximum local, de la fonction de densité d'un nuage de points. L'algorithme mean shift est basé sur l'estimateur multidimensionnel de densité pour un ensemble de L objets $(u^{(l)})_{l=1,\dots,L}$ de l'espace \mathcal{D} . Le vecteur mean shift pour un noyau gaussien de covariance H est donné par :

$$m_H(u) = \frac{\sum_{l=1}^n u^{(l)} \exp\left(-\frac{1}{2}d_M^2(u, u^{(l)}, H)\right)}{\sum_{l=1}^n \exp\left(-\frac{1}{2}d_M^2(u, u^{(l)}, H)\right)} - u \quad (\text{II.12})$$

où d_M est la distance de Mahalanobis entre u et $u^{(l)}$.

Supposons maintenant que l'espace de dimension d puisse être décomposé en p espaces indépendants de dimensions d_p , avec $p \in \mathbb{N}$. Chacun associé à un type d'information différent (i.e position, vitesse, accélération). Les espaces étant indépendants, la matrice H peut s'écrire $H = \text{diag}[H_1, \dots, H_p]$ et le vecteur mean shift devient égal à :

$$m_H(u) = \frac{\sum_{l=1}^n u^{(l)} \prod_{p=1}^P \exp\left(-\frac{1}{2}D^2(u, u^{(l)}, H)\right)}{\sum_{l=1}^n \prod_{p=1}^P \exp\left(-\frac{1}{2}D^2(u, u^{(l)}, H)\right)} - u \quad (\text{II.13})$$

La classification mean shift est obtenue en appliquant à chaque itération l le déplacement donné par l'équation (II.13). Cet algorithme converge vers le mode local de la densité. La classification finale de l'ensemble des trajectoires de \mathcal{D} est obtenue en regroupant dans une même classe \mathcal{C}_p tous les objets ayant convergé vers le même mode.

Cette méthode de classification ne nécessite aucun *a priori* sur les densités de probabilité des objets de \mathcal{D} [22]. Sa principale limite réside dans le choix du paramètre d'échelle (largeur de bande) du noyau :

$$\hat{f}(u) = \frac{1}{n} \sum_{l=1}^n K_H(u - u^{(l)}) \quad (\text{II.14})$$

Le noyau K_H est une fonction symétrique, positive ou nulle, centrée sur zéro et dont l'intégrale vaut 1. Et H est la matrice de largeur de bande, symétrique, définie positive qui normalise le support du noyau.

II.2.2.3 Classification par Self Organizing Maps (SOM)

Les **cartes topologiques** ou auto-organisatrices ont été introduites pour la première fois par T.Kohonen en 1981 [51]. Elles font partie de la famille des modèles dits à apprentissage non supervisé. Cet algorithme partage avec les algorithmes classiques d'ordination (ACP, analyse des correspondances, ...), l'idée de représenter des données multidimensionnelles dans un espace de dimension inférieure (généralement un espace \mathbb{R}^2).

La méthode SOM consiste à placer non linéairement sur la carte de Kohonen les trajectoires $(u^{(l)})_{1 < l < L}$ de l'espace d'entrée \mathcal{D} en préservant les proximités, ainsi les trajectoires similaires devront se retrouver proches les unes des autres sur la grille (constituée de plusieurs prototypes). A cet effet, un apprentissage de la carte est nécessaire, il est réalisé de manière itérative et non supervisée à partir des observations de \mathcal{D} . La particularité des cartes auto-organisatrices est qu'elles rendent possible la comparaison des groupements qui ont été réalisés directement à partir des données. Une observation est affectée à un groupe qui est projeté sur un noeud de la carte. Les observations semblables ont la même projection sur la carte ; si les projections sont différentes, cela veut dire que les observations dans l'espace d'entrée sont différentes [28].

La figure (II.3) représente une illustration de la carte topologique. Celle-ci est consti-

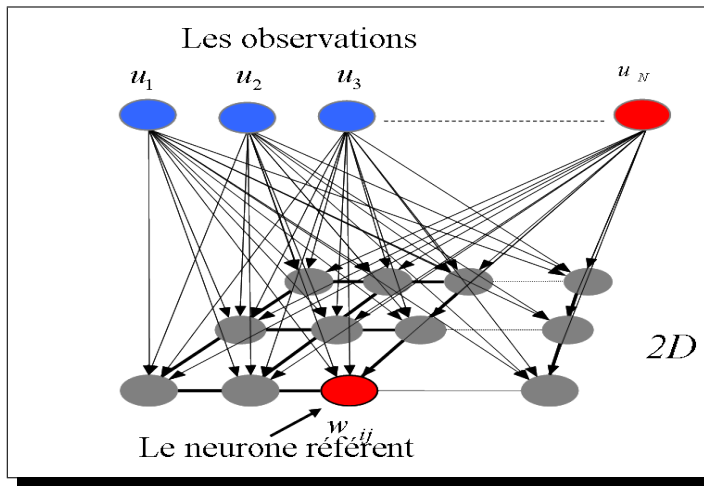


Figure II.3 – Carte topologique avec plusieurs neurones

tuée de plusieurs neurones. Chaque neurone a un vecteur référent qui le représente dans l'espace d'entrée \mathcal{D} . Après une initialisation aléatoire des valeurs de chaque neurone, on

II.2. Classification des trajectoires u

soumet une à une les observations à la carte. Selon les valeurs des neurones, il y en a un qui répondra le mieux au stimulus. Celui dont la valeur sera la plus proche de la donnée présentée sera considéré comme neurone gagnant. A chaque itération, l'algorithme cherche le neurone gagnant. C'est le plus proche dans l'espace d'entrée au sens de la similarité. Les poids du neurone gagnant ainsi que son voisinage sont modifiés. L'amplitude de la modification est un paramètre de l'algorithme. Moins un neurone est proche du neurone gagnant dans la grille, moins son déplacement est important. Les prototypes représentant des exemples proches, au sens de la similarité, seront proches sur la carte au sens des voisins. Une fois l'apprentissage terminé, on peut étiqueter les prototypes si on souhaite les associer à des classes $(C_p)_{1 < p < P}$. Les étapes suivantes résument l'algorithme SOM.

Etape 1, Initialisation des neurones de la carte, $\mathcal{W} = w_l$ avec $l = \{1, \dots, L\}$ et calcul de la distance de Mahalanobis d_M entre chaque objet $u^{(l)} \in \mathcal{D}$ et le prototype w_l .

Etape 2, Calcul de la distance minimale pour trouver le neurone gagnant w_g :

$$d_g(u^{(l)}, w_g) = \min (d_l(u^{(l)}, w_l))_{l=1, \dots, L}. \quad (\text{II.15})$$

Etape 3, Ajustement du poids du neurone gagnant

$$w_g(n+1) = w_g(n) + \eta(n) h_{ig}(n) (u^{(l)} - w_g(n)) \quad (\text{II.16})$$

où h_i est la fonction de voisinage gaussienne et $h_{ig}(t) = \exp(-\frac{d_{ij}}{2\sigma(t)})$ est la fonction de voisinage qui dépend de la distance d_{ij} entre un neurone i et le neurone gagnant g , la fonction h_{ij} est une fonction gaussienne. σ est le coefficient de voisinage. Son rôle est de déterminer un rayon de voisinage autour du neurone gagnant.

Etape 4, Répéter les étapes 2,3 et 4 jusqu'à atteindre le critère d'arrêt $\max_i \|\Delta(w_g)\| \leq \epsilon$. L'*avantage* de cette méthode est la possibilité de choisir la taille de la carte. Ceci permet donc de choisir le niveau de résolution souhaité, faisant ainsi apparaître plus ou moins de détails dans le jeu de données. L'*inconvénient* est d'avoir un voisinage fixe et une liaison entre neurones interchangeable même pour mieux représenter des données discontinues.

En dehors des 3 méthodes de classification décrites, on peut faire un *couplage de méthodes* pour le choix des conditions initiales. On fait d'abord une classification avec la méthode *SOM* puis introduire ce résultat comme condition initiale de la méthode des *Nuées dynamiques*. Cela permet à l'algorithme de converger rapidement.

Après la description des algorithmes (k-means, mean shift et SOM), il est nécessaire de choisir une méthode de représentation de la L -famille de \mathcal{D} afin de pouvoir utiliser les méthodes de classification citées ci-dessus.

II.2.3 Transformation 3D-2D des données de mesures

Rappelons que l'ensemble \mathcal{D} des trajectoires est une famille 3D de données discrètes. Celle-ci doit être transformée en famille 2D pour que les algorithmes de classification puissent être appliqués. Cette transformation consiste à caractériser autrement l'information contenue dans le système. La figure (II.4) est une illustration de cette transformation.

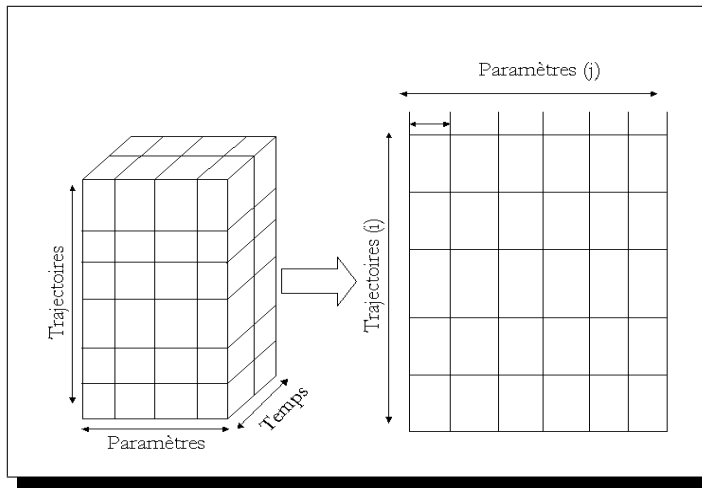


Figure II.4 – Transformation des données

Nous proposons 3 types de **caractérisation** qui sont : la moyenne, la variance et la médiane des observations discrètes des trajectoires $u^{(l)}$. Malgré une perte d'information due à la réduction de la dimension de l'espace de travail, les algorithmes choisis dans ce travail de thèse sont robustes pour distinguer les trajectoires entre elles.

Après la caractérisation de l'information, on utilise les algorithmes de classification décrits précédemment. Ces algorithmes permettront de regrouper les trajectoires ayant des propriétés statistiques voisines (variance inter-conducteurs faible pour une même classe \mathcal{C}_p). Chaque \mathcal{C}_p est considérée comme un comportement de conduite différent. Il existe dans la littérature des critères de régularité de la trajectoire permettant de définir chaque comportement de conduite de manière physique. Rappelons que ce n'est pas le but de cette thèse. On se contente pour la suite qu'il existe bien évidemment P classes de trajectoires

II.2. Classification des trajectoires u

correspondant à P comportements de conduite pour une configuration de virage donnée et un ensemble représentatif de trajectoires pratiquées sur ce virage. La figure (II.5) est une illustration de la répartition de \mathcal{D} en famille de trajectoires.

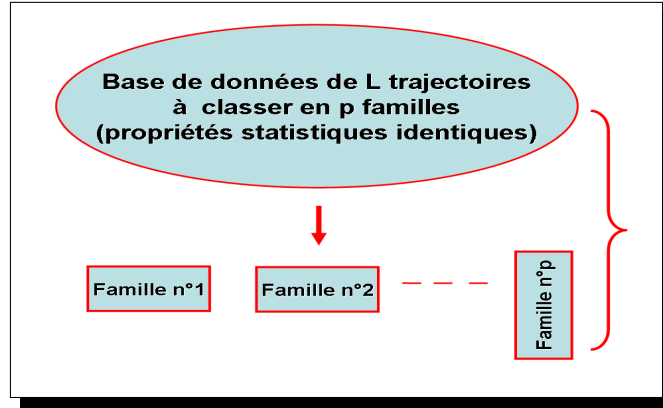


Figure II.5 – Répartition des trajectoires u observées en p familles

Nous allons décrire des critères permettant de choisir l’algorithme de classification et la méthode de représentation pour garantir une meilleure classification des trajectoires.

II.2.4 Validation des p classes de trajectoires identifiées

Les classes \mathcal{C}_p sont validées de deux manières différentes et complémentaires : la *convergence des algorithmes* de classification et l’utilisation de *critères de validité* des classes.

II.2.4.1 Convergence des algorithmes de classification

Chaque algorithme décrit précédemment admet son critère de convergence. On appelle convergence d’un algorithme, après plusieurs itérations, quand on retrouve à chaque itération la même partition de \mathcal{D} . Les classes de trajectoires ainsi obtenues sont stables. C’est-à-dire, les trajectoires affectées dans une classe ne changent pas de classe quelque soit l’itération. Il est **nécessaire** de déterminer les classes $(\mathcal{C}_p)_{p=1,\dots,P}$ avec une bonne précision car la performance des modèles probabilistes à développer en dépend fortement.

II.2.4.2 Critères de validité des classes \mathcal{C}_p

Une partition optimale présente des classes bien séparées par maximisation de la fonction « distances inter classes ». Avec les critères de validité, on évalue la performance des algorithmes (k-means, mean shift et SOM) ainsi que les méthodes de représentation (moyenne,

variance et médiane). Ces critères sont : la séparation, l'homogénéité et la dispersion : L'**homogénéité** permet d'avoir des classes les plus concises possibles, en maximisant la distance intra classes.

La **dispersion** minimise la fonction d'inertie (la somme des carrés des écarts à un centre d'objets).

La **séparabilité** linéaire des classes est mesurée à partir de la matrice de covariance des données de chaque \mathcal{C}_p et de la matrice des distances entre les moyennes de chaque \mathcal{C}_p , [66].

$$\sum_{\mathcal{C}_p} (i, j) = \frac{\left(u_i^{\mathcal{C}_p} - m_i\right) \left(u_i^{\mathcal{C}_p} - m_i\right)^T}{d - 1} \quad (\text{II.17})$$

où $m_i = \frac{1}{d} \sum_{j=1}^d u_{ij}^{\mathcal{C}_p}$.

Soit :

$$\begin{cases} S_w &= \sum_{c=1}^P \pi_c \sum_c \\ S_b &= \sum_{c=1}^P (\mu_c - M)^T (\mu_c - M) \end{cases} \quad (\text{II.18})$$

où $\mu_c = \sum_i u_i^c$ est la moyenne de chaque classe, $M = \sum_i u_i$ la moyenne de l'ensemble des observations et π_c est de chaque classe C , i.e. le nombre d'éléments dans la classe C comparée à la base de données. La séparabilité des classes est obtenue par l'équation :

$$S = \text{trace}(S_w^{-1} S_b) \quad (\text{II.19})$$

La représentation graphique de la fonction S permettra de choisir l'algorithme de classification et la méthode de représentation de \mathcal{D} . Le couple (algorithme/représentation) qui a une meilleure séparabilité entre les P classes assurera la classification la plus optimale.

Ce résultat permet d'avoir des classes \mathcal{C}_p stables *vis-à-vis* du nombre d'itérations. Chaque \mathcal{C}_p est considérée comme un comportement de conduite différent. Il est important d'identifier tous les comportements de conduite et que chacun soit représenté le mieux possible. Car des classes instables engendrent des problèmes d'*homogénéité* et par conséquent une mauvaise identification des processus stochastiques. Rappelons que les trajectoires de chaque \mathcal{C}_p sont considérées comme des réalisations discrètes d'un même processus stochastique U_p défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R}^6 . Cette hypothèse sera vérifiée dans la partie validation numérique des modèles au chapitre III.

La prochaine section consiste à faire une *modélisation probabiliste* par identification de processus stochastiques.

II.3 Modélisation stochastique du système V-I-C

L'objectif de cette section est de développer des *modèles probabilistes* robustes capables de reproduire l'état dynamique du système *vis-à-vis* de plusieurs critères de défaillance. Ces modèles sont spécifiques à des classes de trajectoires judicieusement identifiées dans la précédente section. Nous proposons cette méthode de modélisation car, d'une part, elle fait abstraction à des *interactions non linéaires* qui ne sont pas modélisables de manière précise, et d'autre part, nous disposons d'instruments de mesures performants capables d'acquérir des observations discrètes de trajectoires pratiquées.

Dans un premier temps, la méthode consiste à vérifier les hypothèses de stationnarité des processus stochastiques étudiés. Ensuite, nous caractérisons et identifions au mieux chaque processus. Enfin, nous décrivons des techniques de simulation de processus stochastiques stationnaires et non gaussiens nécessaires pour simuler les processus identifiés.

II.3.1 Processus d'état U du système

A chaque classe \mathcal{C}_p , $p \in \mathcal{P}$, précédemment définie, on associe un unique processus stochastique vectoriel $U_p = (U_p(t), t \in \mathbb{R})$ défini sur l'espace de probabilité $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R}^6 , tel que $\forall t \in \mathbb{R}$:

$$U_p(t) = (X_1(t), X_2(t), V_1(t), V_2(t), \Gamma_1(t), \Gamma_2(t))^T \quad (\text{II.20})$$

Ce processus U_p décrit, le comportement aléatoire de la dynamique du système V-I-C. Il est partiellement connu à travers les trajectoires de la classe \mathcal{C}_p à laquelle il est associé. Les composantes $X_1, X_2, V_1, V_2, \Gamma_1$ et Γ_2 sont des processus scalaires indexés sur \mathbb{R} à valeurs dans \mathbb{R} . Nous rappelons que ces composantes décrivent respectivement : les *positions*, *vitesse*s et *accélérations* du centre de masse (G) du véhicule relativement à la base fixe $\mathcal{B}_0^A = (\vec{e}_1, \vec{e}_2, \vec{e}_3)$. Pour des raisons de simplification de notations, on pose pour la suite du manuscrit :

$$U = U_p \quad (\text{II.21})$$

Remarque 4. *Compte tenu de la quantité d'observations expérimentales disponibles d'une part et des méthodes de simulation de processus stochastique d'autre part, il n'est pas envisageable de simuler le processus stochastique vectoriel U .*

Cependant, l'objet final de ce travail de thèse est d'évaluer le risque contenu dans le

système V-I-C par une analyse fiabiliste. Cette analyse conduit à choisir plusieurs critères de défaillance. Nous choisissons des critères qui portent sur des fonctionnelles scalaires du processus U , appelés **processus de contrôle**. Ceux-ci sont issus de la traduction du problème étudié en langage fiabiliste (projection fonctionnelle de U sur \mathbb{R}).

II.3.1.1 Processus de contrôle Z

Un processus de contrôle associé au processus stochastique U est un processus $Z = (Z(t), t \in \mathbb{R})$ défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R} tel que :

$$Z(t) = F(U(t)) , \forall t \in \mathbb{R} \quad (\text{II.22})$$

où F est une fonction de \mathbb{R}^6 dans \mathbb{R} . On suppose dans toute la suite que ce processus est du second ordre. Insistons bien sur le fait que F est relative au critère choisi pour définir la défaillance des trajectoires. Il a été dit au chapitre I que les critères de défaillance sont obtenus par traduction en langage fiabiliste des modes de défaillance suivants :

- La sortie du véhicule dans l'espace roulant.
- Une perte d'adhérence du véhicule qui rend celui-ci incontrôlable par le conducteur.

A partir de ces modes de défaillance, nous avons montré que la défaillance du système peut être caractérisée soit par une distance relative entre trajectoires, soit par une fonctionnelle des composantes de la trajectoire u dans l'espace des phases.

II.3.1.2 Critères de défaillance du système V-I-C

Dans cette étude, nous ne prétendons pas connaître les causes de la défaillance du système. Parce que des causes différentes peuvent engendrer le même type de défaillance. Nous considérons le système comme une boîte noire générant des trajectoires. Nous essayons d'associer le risque contenu dans ce système à des niveaux de défaillance portant sur la réponse du système (i.e. critère de défaillance portant sur les trajectoires). La traduction mathématique des 2 types de défaillance cités précédemment correspond respectivement aux critères de défaillance suivants : *distance entre trajectoires* et *accélération latérale* du véhicule.

II.3.1.2.1 Distance relative entre trajectoires Soit une trajectoire de référence u_{ref} connue et appartenant à l'espace roulant, cf. figure (II.6). Celle-ci est sûre par construction. La notion de trajectoire de référence est difficile à définir car étant sub-

II.3. Modélisation stochastique du système V-I-C

jective. Un choix possible de u_{ref} peut être la trajectoire moyenne u_m estimée à partir des trajectoires expérimentales observées telle que :

$$u_m(t_k) = \frac{1}{N} \sum_{l=1}^N u_k^{(l)} \quad (\text{II.23})$$

A partir de cette trajectoire moyenne, on peut calculer la distance Euclidienne d_E entre celle-ci et l'une quelconque des trajectoires observées u par :

$$d_E(t) = \sqrt{(u(t) - u_m(t))^T (u(t) - u_m(t))} ; \forall t \quad (\text{II.24})$$

Cependant, dans cette étude, nous proposons de calculer la distance géométrique D entre la ligne des centres (C_V) de la voie de circulation et le centre de masse (G) du véhicule :

$$D(t) = \underbrace{\min}_{M \in C_V} (d(G(t), M)) \quad (\text{II.25})$$

où d est la distance euclidienne usuelle dans l'espace affine \mathbb{R}^2 . L'ensemble des points $M \in C_V$ constituant la ligne des centres de voie (C_V) sont connus expérimentalement. La figure (II.7) représente une illustration de la projection de cette distance sur le plan $(x_1 O x_2)$. Nous rappelons qu'une meilleure estimation de la sortie de route serait la prise en compte de l'orientation du véhicule dans le calcul de la distance D .

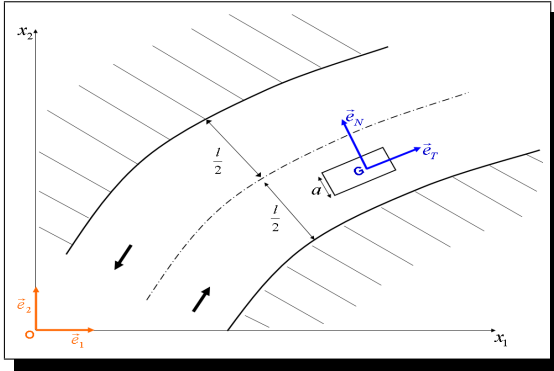


Figure II.6 – Espace roulant du véhicule.

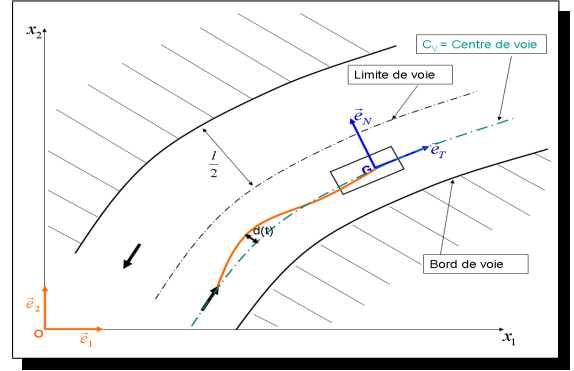


Figure II.7 – Distance relative $d(t)$.

La distance doit être bornée car on ne peut pas s'écarter indéfiniment de C_V tout en restant sur la route. La présence du véhicule sur l'espace roulant est évaluée par le critère :

$$\sup_{t \in T} |D(t)| > \delta^* \quad (\text{II.26})$$

Avec δ^* une valeur seuil qui est calculée selon ce critère comme suit.

$$\delta^* = \frac{l - 2a}{4} \quad (\text{II.27})$$

où $\frac{l}{2}$ est la largeur de la voie de circulation et a la largeur du véhicule. Ce critère de défaillance porte sur le processus de contrôle D . La distance $D = (D(t), t \in \mathbb{R})$ est un processus défini sur $(\Omega, \mathfrak{S}, \mathbb{P})$ à valeurs dans \mathbb{R}_+ .

Après la caractérisation de la défaillance par la *distance relative* entre trajectoires, nous allons décrire d'autres critères de défaillance qui portent sur des fonctionnelles de u .

II.3.1.2.2 Fonctionnelle des composantes de la trajectoire u Le critère de défaillance du système pourrait être défini aussi par une fonctionnelle des coordonnées de la trajectoire u dans l'espace des phases. Il en est ainsi par exemple pour le critère de défaillance portant sur l'*accélération transversale*.

Accélération transversale

L'accélération transversale et sa variation (le jerk) sont des critères souvent utilisés dans la littérature pour déterminer la **dangerosité des trajectoires**. Il en est proposé dans Kanayama et al[49] :

1. minimiser à vitesse constante l'intégrale du carré de la courbure le long du chemin, ce qui revient à minimiser les accélérations transversales ;
2. minimiser la dérivée de la courbure ρ le long du chemin, ce qui correspond à minimiser les variations de l'accélération transversale.

On trouve dans Hongo et al[97] un critère proche du second, consistant à minimiser l'intégrale du carré de la dérivée de l'accélération. Ce qui correspond encore une fois à minimiser les variations de l'accélération transversale. Elle est donc utilisée comme critère de défaillance car le dépassement d'un certain seuil (δ^*) peut engendrer une perte d'adhérence, c.f. Revue[81]. Ce critère de défaillance est défini par :

$$\sup_{t \in T} |\Gamma_N(t)| > \delta^* \quad (\text{II.28})$$

Il porte sur le processus de contrôle $\Gamma_N(t)$. Notons que $\Gamma_N = (\Gamma_N(t), t \in \mathbb{R})$ est défini sur $(\Omega, \mathfrak{S}, \mathbb{P})$ à valeurs dans \mathbb{R} . On peut le déterminer facilement car c'est une coordonnée de la trajectoire u dans le repère de Serret-Frénet. Nous avons montré le passage du repère Galiléen de référence \mathcal{R}_0^A au repère de Serret-Frénet \mathcal{R}_G^{SF} dans le chapitre I.

II.3. Modélisation stochastique du système V-I-C

Les 2 **critères de défaillance** introduits portent sur les processus de contrôle : $D(t)$ la *distance relative* et $\Gamma_N(t)$ l'*accélération transversale*. Les Sup de ces fonctions seront bornées par des seuils δ^* pour garantir la stabilité du véhicule sur la route. Ils nous semblent être des critères pertinents pour étudier la défaillance du système. Rappelons que $D(t)$ et $\Gamma_N(t)$ viennent d'être obtenus par la transformation du processus U via une fonction F selon le critère de défaillance choisi. Ces processus seront représentés par Z pour la suite.

$$\begin{cases} Z = D \\ Z = \Gamma_N \end{cases} \quad (\text{II.29})$$

Les observations expérimentales montrent que le processus Z n'est en général pas stationnaire. Dans le paragraphe suivant, nous allons construire une représentation permettant d'exprimer ce dernier comme une fonctionnelle affine d'un processus stationnaire.

II.3.1.3 Standardisation du processus de contrôle Z

Le processus Z est du second ordre mais n'est *a priori* pas stationnaire en m.o.d (moyenne d'ordre deux). Considérons alors le processus $X = (X(t), t \in \mathbb{R})$ défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R} tel que :

$$X(t) = \sigma_Z^{-1}(t) (Z(t) - \mu_Z(t)) \quad (\text{II.30})$$

où $\mu_Z(t) = \mathbb{E}[Z(t)]$ et $\sigma_Z(t) = \mathbb{E}[(Z(t) - \mathbb{E}[Z(t)])^2]$ sont respectivement la moyenne et l'écart-type temporels de Z . Par construction, le processus X est tel que $\mu_X(t) = 0$ et $\sigma_X(t) = 1$. Donc ces deux premiers moments sont constants. Ce processus est donc un bon candidat pour être supposé stationnaire en m.o.d. Pour qu'il en soit ainsi, il faut en toute rigueur que sa fonction d'autocorrélation R_X ne dépende que des accroissements temporels et non pas des couples d'instantanés (t_1, t_2) .

Une étude statistique de X , pour les exemples considérés, a montré (c.f. III.4.1.4) que cette propriété est approximativement vérifiée et donc que X peut être raisonnablement considéré comme stationnaire en m.o.d. Et, compte tenu de la relation (II.30), on obtient :

$$Z(t) = \mu_Z(t) + \sigma_Z(t)X(t) \quad (\text{II.31})$$

Si l'on dispose d'estimation satisfaisante de $\mu_Z(t)$ et $\sigma_Z(t)$, caractériser le processus Z reviendra à caractériser le processus X . Il s'agit donc maintenant de savoir caractériser

(à partir des données expérimentales disponibles) le processus X stationnaire en m.o.d.

II.3.2 Caractérisation du processus X

Un processus stochastique est un objet complexe à décrire. Par hypothèse, le processus X est centré, stationnaire en m.o.d et de variance unité. Compte tenu des données disponibles, nous choisirons pour le caractériser sa loi marginale d'ordre 1 dont nous supposons qu'elle admet une densité p_X , et sa fonction d'autocorrélation R_X , ou ce qui revient au même, sa densité spectrale de puissance S_X . Bien entendu, cette caractérisation n'est pas complète, car connaître complètement la loi d'un processus c'est connaître sa famille de lois finidimensionnelles, ce qui est loin d'être le cas pour le processus X . On s'attachera donc à rechercher des processus dont la description soit beaucoup plus simple afin de pouvoir les utiliser en pratique.

Pour le calcul des densités marginales d'ordre 1 et des densités spectrales de puissance, nous avons utilisé des échantillons expérimentaux. La densité marginale d'ordre 1 a été estimée à l'aide de l'estimateur standard et de l'estimateur à noyau gaussien (cf. annexes V.2). La densité spectrale de puissance a quand à elle été estimée à l'aide de l'estimateur de Welch avec fenêtre de Hamming (cf. annexes V.3). Enfin, pour la fonction d'autocorrélation, nous avons utilisé l'estimateur basé sur le périodogramme [93].

A partir des *estimées* p_X , F_X et S_X , nous avons cherché des lois théoriques par approximation nécessaires pour la simulation du processus stochastique X .

II.3.3 Approximation de la densité marginale d'ordre 1 de X

Pour cela deux grandes voies s'offrent à nous, soit on cherche une approximation dans la classe des lois usuelles Bouleau[12]. Soit on construit une approximation hilbertienne sur une base fonctionnelle à choisir. Nous avons retenu la base des polynômes d'Hermite.

II.3.3.1 Estimation par des lois théoriques usuelles

Rappelons que par construction la loi marginale d'ordre 1 à approximer est de moyenne $\mu_X = 0$ et de écart-type $\sigma_X = 1$, du fait de la normalisation de X . Par le biais de tests statistiques (Kolmogorov-Smirnov, Chi2, etc), nous avons testé l'adéquation de cette loi marginale d'ordre 1 à des lois théoriques connues. Parmi ces lois, nous avons la *loi normale* et la *loi logistique* qui peuvent convenir.

II.3. Modélisation stochastique du système V-I-C

La loi normale a pour espérance $\mathbb{E}(X) = \mu_X$ et variance $V(X) = \sigma_X^2$. L'utilisation de cette loi comme ajustement de la loi marginale d'ordre 1 de X impose obligatoirement de prendre $\mu_X = 0$ et $\sigma_X = 1$. Pour compléter les tests statistiques sur l'hypothèse *gaussienne* de la loi marginale d'ordre 1 de X , nous avons estimé les coefficients d'aplatissement K et d'asymétrie S (c.f. annexes V.4.3). Au cas où cette hypothèse est vérifiée, cela simplifierait l'étude de manière considérable.

La loi logistique a pour espérance $\mathbb{E}(X) = \mu$ et variance $V(X) = \frac{s^2\pi^2}{3}$. L'utilisation de cette loi comme approximation de la loi de probabilité empirique de X impose de prendre $\mu = 0$ et $s = (\frac{\sqrt{3}}{\pi})$.

Outre la méthode d'approximation par ces lois, nous proposons d'utiliser une approximation hilbertienne basée sur une écriture de la densité p_X sur la base orthonormée des polynômes d'Hermite Soize[93].

II.3.3.2 Approximation Hilbertienne sur la base des polynômes d'Hermite

Pour définir les polynômes d'Hermite, nous considérons l'espace de Hilbert $L^2(\mathbb{R}, \nu)$. C'est l'espace des fonctions réelles définies ν -presque partout sur \mathbb{R} , de carré intégrable par rapport à la mesure normale canonique $\nu(dx) = f(x)dx$, muni du produit scalaire :

$$((\varphi_1, \varphi_2)) = \int_{\mathbb{R}} \varphi_1(x)\varphi_2(x)f(x)dx \quad (\text{II.32})$$

où $f(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$, $x \in \mathbb{R}$ est la densité de probabilité d'une loi gaussienne standard.

Les polynômes d'Hermite $H_\alpha(x)$ sur \mathbb{R} , d'indice $\alpha \in \mathbb{N}$ sont définis par la formule de récurrence :

$$\begin{cases} H_0(x) & = 1 \\ H_{\alpha+1}(x) & = xH_\alpha - \frac{d}{dx}H_\alpha(x) \end{cases} \quad (\text{II.33})$$

La dérivée de $H_\alpha(x)$ par rapport à x s'écrit :

$$\frac{d}{dx}H_\alpha(x) = \alpha H_{\alpha-1}(x) \quad (\text{II.34})$$

Les polynômes d'Hermite sont également définis par :

$$H_m(x) = \int_{\mathbb{R}} (x + iy)^m f_1(x) dx \quad (\text{II.35})$$

Soit $p_X : \mathbb{R} \rightarrow \mathbb{R}_+$ la densité de probabilité cible qui a été estimée dans la partie (II.3.2) en utilisant les échantillons expérimentaux.

Nous cherchons p_X sous la forme :

$$p_X(x) = f(x) \sum_{\alpha=0}^{+\infty} q_{\alpha} h_{\alpha}(x) \quad (\text{II.36})$$

où $q_{\alpha} \in \mathbb{R}$ et $(h_{\alpha})_{\alpha}$ désigne la base orthonormée des polynômes d'Hermite normalisés, avec pour tout $\alpha \in \mathbb{N}$:

$$h_{\alpha}(x) = (\alpha!)^{-1/2} H_{\alpha}(x) \quad (\text{II.37})$$

Ils forment une base orthonormée de $L^2(\mathbb{R}, \nu)$ et $(\alpha!)^{-1/2}$ est le facteur de normalisation de manière à avoir $((h_j, h_k)) = \delta_{jk}$.

A partir de l'équation (II.36), nous pouvons effectuer une approximation d'ordre N de la forme :

$$p_X^N(x) = f(x) \sum_{\alpha=0}^N q_{\alpha} h_{\alpha}(x) \quad (\text{II.38})$$

où N est un entier ≥ 0 à déterminer.

Pour estimer les coefficients $(q_{\alpha})_{1 < \alpha < N}$, nous avons utilisé deux approches : l'une basée sur l'*estimation des moments* de la loi marginale d'ordre 1 et l'autre basée sur la *projection* de p_X sur la base des polynômes d'Hermite.

II.3.3.2.1 Estimation des coefficients q_{α} par estimation des moments

Une N -approximation du moment d'ordre $p \in \mathbb{N}$ de la loi marginale d'ordre 1 de X est donnée par :

Soit $p_X : x \rightarrow p_X(x)$ la densité de probabilité cible qui a été estimée dans la partie (II.3.2) en utilisant les échantillons expérimentaux.

$$\mu_X^p \approx \int_{\mathbb{R}} x^p p_X^N(x) dx = \int_{\mathbb{R}} x^p f(x) \sum_{\alpha=0}^N q_{\alpha} h_{\alpha}(x) dx = \sum_{\alpha=0}^N q_{\alpha} \int_{\mathbb{R}} x^p f(x) h_{\alpha}(x) dx \quad (\text{II.39})$$

II.3. Modélisation stochastique du système V-I-C

$\forall p \in \mathbb{N}$, μ_X^p peut-être estimé statistiquement.

$$\Rightarrow \mu_X^p \approx \sum_{\alpha=0}^N q_\alpha \lambda(p, \alpha) = \lambda(p, 0) + \sum_{\alpha=1}^N q_\alpha \lambda(p, \alpha) \quad (\text{II.40})$$

où $\lambda(p, \alpha) = \int x^p f(x) h_\alpha(x) dx$ est une constante à calculer pour chaque couple (p, α) donné.

Supposons que l'on connaisse une estimée statistique des N premiers moments de la loi marginale d'ordre 1 de X .

Soit $(\tilde{\mu}_X^p)_{1 < \alpha < N}$ ces N estimées d'après l'équation (II.40). Sachant que tous les couples $\lambda(p, \alpha)$ peuvent être calculés explicitement, les coefficients inconnus q_α peuvent alors être obtenus par résolution du système linéaire $N \times N$:

$$\begin{bmatrix} \tilde{\mu}_X^1 \\ \tilde{\mu}_X^2 \\ \tilde{\mu}_X^3 \\ \cdot \\ \cdot \\ \tilde{\mu}_X^N \end{bmatrix} = \begin{bmatrix} \lambda(1, 1) & \lambda(1, 2) & \cdot & \cdot & \lambda(1, N) \\ \lambda(2, 1) & \lambda(2, 2) & \cdot & \cdot & \lambda(2, N) \\ \lambda(3, 1) & \lambda(3, 2) & \cdot & \cdot & \lambda(3, N) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \lambda(N, 1) & \lambda(N, 2) & \cdot & \cdot & \lambda(N, N) \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ \cdot \\ \cdot \\ q_N \end{bmatrix} \quad (\text{II.41})$$

Soit $(\tilde{q}_\alpha)_{1 < \alpha < N}$ la solution de ce système, la N -approximation approchée de p_X est alors donnée par :

$$p_X^N \cong f(x) \left(1 + \sum_{\alpha=1}^N \tilde{q}_\alpha h_\alpha(x) \right) \quad (\text{II.42})$$

En raison d'un mauvais conditionnement de la matrice du système, les coefficients $(\tilde{q}_\alpha)_{\alpha=0, \dots, N}$ sont souvent mal estimés. Par conséquent l'approximation par l'estimation des moments peut-être de mauvaise qualité. C'est pourquoi, nous avons préféré estimer les coefficients $(q_\alpha)_{\alpha=1, \dots, N}$ par une autre méthode que nous détaillons dans le sous paragraphe suivant.

II.3.3.2.2 Estimation des coefficients par intégration de la loi marginale.

La loi marginale P_X peut s'écrire :

$$p_X(x) dx = q(x) \nu(dx) \quad (\text{II.43})$$

Avec $q(x) = p_X f^{-1}(x)$. Dans le cas particulier où $q \in L^2(\mathbb{R}, \nu)$, on peut écrire le développement de q sur la base hilbertienne $(h_\alpha)_\alpha$.

$$q(x) = \sum_{\alpha=0}^{+\infty} q_\alpha h_\alpha(x) \quad (\text{II.44})$$

La série du membre de droite de l'équation (II.36) est convergente dans $L^2(\mathbb{R}, \nu)$. Pour tout indice $\alpha \in \mathbb{N}$, le réel q_α est obtenu par projection sur la base $(h_\alpha)_\alpha$ de la forme :

$$\begin{aligned} q_\alpha &= ((q, h_\alpha)) \\ &= \int_{\mathbb{R}^N} q(x) h_\alpha(x) f(x) dx \end{aligned} \quad (\text{II.45})$$

Avec $h_\alpha(x) = 1$ pour $\alpha = 0$, la condition de normalisation $\int_{\mathbb{R}} p_X(x) dx = 1$ implique $q_\alpha = 1$ pour $\alpha = 0$.

Partant de la formule (II.45), les coefficients $(q_\alpha)_{1 < \alpha < N}$ peuvent être exprimés comme suit :

$$\bar{q}_\alpha = \int_{\mathbb{R}} \hat{p}_X(x) h_\alpha(x) dx \quad (\text{II.46})$$

où \hat{p}_X est une estimée de la densité p_X , obtenue à partir d'un échantillon statistique expérimental de la loi marginale de X . Pour obtenir la valeur numérique des $(q_\alpha)_{1 < \alpha < N}$, nous avons utilisé un schéma d'intégration numérique classique tel que Simpson [25].

Après avoir tronqué le développement d'Hermite à un ordre N raisonnable, nous avons obtenu une approximation p_X^N de la densité p_X de la loi marginale d'ordre 1 du processus X . Dans la grande majorité des cas traités l'ordre de troncature du développement d'Hermite n'a pas excédé $N = 7$. Nous allons maintenant décrire la procédure à suivre pour choisir la meilleure approximation de la densité p_X .

II.3.3.3 Comparaison des approximations de la loi marginale d'ordre 1 de X

On peut obtenir de bonnes approximations de p_X ce qui est dû à la régularité et/ou à la finesse de la densité p_X estimée. Si elle présente trop de fluctuations, l'approximation par Hermite peut ne pas donner de bons résultats malgré un ordre élevé de troncature.

Pour améliorer la qualité des approximations au cas où le développement sur la base des polynômes d'Hermite ne serait pas suffisant, on pourrait utiliser la méthode de développement sur la base d'Edgeworth. C'est une méthode qui sert à regrouper les termes

II.3. Modélisation stochastique du système V-I-C

$(q_\alpha)_{0 < \alpha < N}$ du même ordre de grandeur pour approximer au mieux les densités de probabilité cibles, Pugachev [77].

Toutes ces méthodes d'approximation utilisées présentent des avantages et des inconvénients. A titre d'exemple, pour les lois usuelles, on peut avoir la difficulté d'approximer les densités avec plusieurs fluctuations ; pour le developpement sur la base des polynômes d'Hermite, on a des problèmes d'instabilité pour un ordre de troncature élevé. Pour choisir la meilleure approximation des lois empiriques du processus X , nous proposons d'utiliser des critères de proximité.

Il existe plusieurs critères de mesures d'écart dans la littérature : la norme L^1 , la norme L^2 , ou encore les mesures étudiées par Bickel-Rosenblatt, Bickel[6].

Le critère de proximité L^1 noté I_1 est obtenu par :

$$\begin{aligned} I_1 &= \int_{\mathbb{R}} |\tilde{p}_X(x) - p_X^N(x)| dx \\ &= \sum_i (\tilde{p}_X(x_i) - p_X^N(x_i)) \Delta x_i \end{aligned} \quad (\text{II.47})$$

où Δx_i est le pas de discrétisation du schéma numérique.

Le critère de comparaison basé sur la norme L^2 noté I_2 est obtenu par :

$$\begin{aligned} I_2 &= \int_{\mathbb{R}} (\tilde{p}_X(x) - p_X^N(x))^2 dx \\ &= \sum_i (\tilde{p}_X(x_i) - p_X^N(x_i))^2 \Delta x_i \end{aligned} \quad (\text{II.48})$$

Pour une norme choisie, la meilleure méthode d'approximation est celle qui conduira à la plus petite erreur. Le choix se fera sur la méthode d'approximation qui donnera le plus petit écart entre la densité *cible* et la densité *approximée*.

Outre ces critères (I_1, I_2) de comparaison, on peut utiliser des *tests statistiques* tels que Kolmogorov-Smirnov, Chi2, etc. L'utilisation de ces tests permet d'accepter ou de rejeter l'hypothèse nulle H_0 pour chaque approximation de la densité. Chaque test a une p-value fixée par l'utilisateur pour comparer deux distributions [88]. Il est également possible d'utiliser le principe d'entropie maximale pour choisir la meilleure approximation.

On peut chercher également d'approximer les densités spectrales de puissance S_X pour la simulation. Sinon, on peut utiliser numériquement la densité spectrale estimée à partir des échantillons expérimentaux.

A partir de p_X et S_X , nous pouvons simuler le processus X . La prochaine section décrit des méthodes de simulation du processus X stationnaire du second ordre et non gaussien.

II.3.4 Simulation du processus X

La simulation du processus stochastique X consiste à générer numériquement des valeurs $X(t_1), \dots, X(t_n)$ selon la loi conjointe donnée par la loi marginale d'ordre n , ceci pour tout entier n et toute suite $(t_i)_{i=1, \dots, n}$. Ce qui est non réalisable en général car on n'a pas accès à la connaissance de ce système de lois marginales, sauf pour les processus gaussiens. Ceci n'est évidemment pas le cas pour les processus X dans les exemples traités.

Rappelons qu'à partir de l'équation (II.30) et des réalisations discrètes de Z , nous avons déduit les réalisations discrètes de X . Cette famille de réalisations a été utilisée dans les étapes précédentes pour :

- estimer les valeurs de la fonction d'autocorrélation R_X et vérifier que cette fonction conserve la même valeur pour tous les couples d'instants équidistants. En d'autres termes, on a montré que X vérifie l'hypothèse de stationnarité à l'ordre 2 ;
- estimer la densité p_X de la loi finidimensionnelle d'ordre 1 de X , la fonction de répartition F_X de cette loi, la Densité Spectrale de Puissance S_X (c.f. Fogli[36]) ;
- estimer les moments du processus stochastique X jusqu'à l'ordre p .

Dans tout ce qui suit nous notons $\tilde{p}_X, \tilde{F}_X, \tilde{S}_X, \tilde{R}_X$, les approximations respectives obtenues de ces grandeurs.

Comme nous l'avons expliqué précédemment, le but est de simuler une bonne approximation de X . Cette étape est cruciale dans notre démarche. Pour répondre à cette question, nous utilisons une méthode de simulation de processus stochastiques non gaussiens basée sur les polynômes d'Hermite (c.f. Poirion [75] et Puig [79]). Cette méthode ne nécessite que la connaissance de F_X et de R_X . Nous en rappelons les principales étapes.

Remarque 5. *Tout processus stationnaire admet des lois marginales d'ordre 1 qui sont isonomes à une loi donnée.*

Soit \mathbb{U} est une variable aléatoire uniforme sur $[0, 1]$. Alors $F_X^{-1}(\mathbb{U})$ est une variable aléatoire de même loi que X . Soit $F_{\mathcal{N}}$ la fonction de répartition d'une variable aléatoire réelle de loi gaussienne standard. On note $F_X^{-1} \circ F_{\mathcal{N}}$ la composée des fonctions F_X^{-1} et $F_{\mathcal{N}}$. Soit G une variable normale de loi uniforme sur $[0, 1]$. Donc la variable aléatoire $F_X^{-1}(F_{\mathcal{N}}(G))$ a même loi que X . Posons :

$$\mathbb{F} = F_X^{-1} \circ F_{\mathcal{N}}. \quad (\text{II.49})$$

II.3. Modélisation stochastique du système V-I-C

En pratique on ne connaît pas l'expression analytique de la fonction \mathbb{F} . On va utiliser une représentation de \mathbb{F} sur une base adaptée et construire ses coordonnées. La décomposition de la fonction \mathbb{F} sur la base des polynômes d'Hermite : lorsque X est une variable aléatoire du second ordre, la fonction \mathbb{F} appartient à cet espace. On suppose que $F_X^{-1} \circ F_{\mathcal{N}}$ appartient à l'espace de Hilbert :

$$L^2(\mathbb{R}) = \left\{ f : \mathbb{R} \rightarrow \mathbb{R} ; \int_{\mathbb{R}} |f(x)|^2 dx < +\infty \right\}. \quad (\text{II.50})$$

Cet espace est muni du produit scalaire :

$$\langle f, g \rangle = \int_{\mathbb{R}} f(x)g(x) \frac{\exp(-x^2/2)}{\sqrt{2\pi}} dx, \quad \forall f \in L^2(\mathbb{R}), \quad \forall g \in L^2(\mathbb{R}) \quad (\text{II.51})$$

Sous cette hypothèse \mathbb{F} peut être projetée sur la base orthonormée $(h_n)_{n \in \mathbb{N}}$ de $L^2(\mathbb{R})$. Cette base orthonormée est constituée des polynômes d'Hermite normalisés (cf. section (II.3.3.2)). Il existe donc une suite de réels $(f_n)_{n \in \mathbb{N}}$ tels que :

$$\mathbb{F} = F_X^{-1} \circ F_{\mathcal{N}} = \sum_{n \in \mathbb{N}} f_n h_n \quad (\text{II.52})$$

avec $\forall n \in \mathbb{N}$

$$f_n = \int_{\mathbb{R}} (F_X^{-1} \circ F_{\mathcal{N}})(x) h_n(x) \frac{\exp(-x^2/2)}{\sqrt{2\pi n!}} dx \quad (\text{II.53})$$

Soit G défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ un processus gaussien stationnaire centré à valeurs dans \mathbb{R} et de variance 1. Soit F_G la fonction de répartition de sa loi finidimensionnelle d'ordre 1 et R_G sa fonction d'autocorrélation (que l'on ne connaît pas *a priori*). Soit $V = (V(t), t \in \mathbb{R})$ le processus stationnaire du second ordre défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R} tel que $\forall t \in \mathbb{R}$:

$$V(t) = (F_X^{-1} \circ F_{\mathcal{N}})(G(t)) \quad (\text{II.54})$$

Ce processus possède les propriétés suivantes :

- La loi finidimensionnelle d'ordre 1 notée p_V de V a pour fonction de répartition F_X ,
- le processus V a les mêmes deux premiers moments que le processus X ,
- $V(t) = \sum_{n \in \mathbb{N}} f_n h_n(G(t))$, $\forall t \in \mathbb{R}$,
- la fonction d'autocorrélation du processus V doit coïncider avec la fonction d'autocorrélation imposée R_X

$$\begin{aligned} R_V(t) &= E(V(t+s)V(s)) \\ &= \sum_n \sum_m f_n f_m E[H_n(G(t+s))H_m(G(s))] \end{aligned} \quad (\text{II.55})$$

En utilisant la formule de Mehler : $E[H_n(G(s))] = n! ((E[G(u)G(s)])^n \delta_m^n$

$$R_V(t) = \sum_{n \in \mathbb{N}} f_n^2 (n!) (R_G(t))^n, \forall t \in \mathbb{R} \quad (\text{II.56})$$

Il est très difficile de trouver la fonction R_G connaissant la fonction R_X . C'est une équation polynômiale fonctionnelle non résolue en pratique. On la remplace par un problème d'optimisation : trouver une fonction d'autocorrélation R_G telle que R_V approche au mieux R_X :

$$\min_{R_G} \left\| R_X - \sum_n n! f_n^2 R_G(t)^n \right\|_{L^2} \quad (\text{II.57})$$

C'est un problème d'optimisation sous contraintes car la fonction d'autocorrélation R_G est de type positif. Cette contrainte ne peut pas être introduite telle quelle dans le problème. Le problème s'écrit en introduisant les mesures spectrales. Le théorème de Bochner est utilisé ; il lie la fonction d'autocorrélation et la densité spectrale et plus généralement fonction de type positif et mesure :

$$\begin{aligned} R_G(t) &= \int_{\mathbb{R}} e^{i\lambda t} S_G(\lambda) d\lambda ; \quad \forall t \in \mathbb{R} \\ R_X(t) &= \int_{\mathbb{R}} e^{i\lambda t} S_X(\lambda) d\lambda ; \quad \forall t \in \mathbb{R} \end{aligned} \quad (\text{II.58})$$

En substituant dans le problème d'optimisation on construit le nouveau problème :

$$\min_{R_G} \left\| R_X - \sum_n n! f_n^2 R_G(t)^n \right\|_{L^2} \quad (\text{II.59})$$

sous les contraintes : S_G fonction positive, paire, et telle que $\int_{\mathbb{R}} S_G(\lambda) d\lambda = 1$. A partir de la densité spectrale S_G , on peut chercher une représentation markovienne afin de pouvoir simuler le processus G construit précédemment. Par conséquent V est une approximation de X . En pratique, nous allons approximer X par un processus $V^M = (V^M(t), t \in \mathbb{R})$ de fonction d'autocorrélation \tilde{R}_X . Ce processus est défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ et $\forall t \in \mathbb{R}$:

$$V^M(t) = \sum_{n=0}^M f_n h_n(G(t)) \quad (\text{II.60})$$

II.3. Modélisation stochastique du système V-I-C

avec $M \in \mathbb{N}^*$ fixé, les coefficients $(f_n)_{n=0,\dots,M}$ donnés par (II.53), G le processus gaussien précédemment défini.

Pour approximer X nous simulons V^M en utilisant la relation (II.60). Le problème se ramène donc à celui de la simulation du processus gaussien G . Pour cela, nous utilisons une méthode classique de Poirion et al [75] basée sur l'utilisation de la densité spectrale de puissance S_G du processus G . Nous choisissons de prendre :

$$S_G \equiv \tilde{S}_X \quad (\text{II.61})$$

La génération de trajectoires du processus non gaussien X est alors immédiate dès qu'on a construit les trajectoires du processus gaussien G sous-jacent [78].

Nous rappelons les différentes étapes de la méthode de simulation choisie :

En pratique, S_G est uniquement connue aux points :

$$\lambda_j = -\lambda_L + (j + 1/2)\Delta\lambda, \quad j \in \{0, 1, \dots, N - 1\}, \quad (\text{II.62})$$

d'une N -partition régulière de l'intervalle $[-\lambda_L, \lambda_L]$, i.e.

$$S_G \equiv (S_G(\lambda_j), \quad j = 0, 1, \dots, N - 1) \quad (\text{II.63})$$

où λ_L est la fréquence de coupure, $\Delta\lambda$ est le pas de discrétisation. $\Delta\lambda$ et λ_L sont connus et imposés par le théorème d'échantillonnage de Shannon.

Ensuite, pour simuler le processus G , on peut utiliser soit la **représentation spectrale** soit la **représentation markovienne**. La méthode de la représentation spectrale est généralement plus performante que la méthode de représentation markovienne.

L'algorithme de simulation basé sur la représentation spectrale de G se présente comme suit. Soit $(\phi_j)_{j=0,1,\dots,N-1}$ une famille de variables aléatoires réelles définies sur $(\Omega, \mathfrak{F}, \mathbb{P})$ indépendantes dans leur ensemble, suivant une loi uniforme sur $[0, 2\pi]$. On peut montrer que le processus $G_N = (G_N(t), t \in \mathbb{R})$ défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ tel que $\forall t \in \mathbb{R}$:

$$G_N(t) = \sqrt{2\Delta\lambda} Re \left[\sum_{j=0}^{N-1} e^{i\lambda_j t} e^{i\phi_j} \sqrt{S_G(\lambda_j)} \right] \quad (\text{II.64})$$

Chapitre II. Méthodologie de prédiction de trajectoires à risque

converge en loi vers le processus G . L'algorithme de simulation s'obtient en échantillonnant l'approximation G_N de G aux noeuds $t_j = j\Delta t$, $j = 0, \dots, N - 1$ du domaine temporel T de simulation. Il fait intervenir la notion de FFT (Fast Fourier Transform).

Enfin, on reconstitue les trajectoires de X par la formule :

$$X = f_1 * G + f_2 * (G^2 - 1) + f_3 * \dots \quad (\text{II.65})$$

On peut également simuler le processus G en utilisant une représentation markovienne, (c.f. Fogli[36]). Cette méthode de simulation est basée sur le principe de réalisation markovienne des processus gaussiens centrés à spectre rationnel.

Définition 1. *On dit qu'un processus $X = (X(t), t \in \mathbb{R}_+)$ admet une réalisation markovienne de dimension finie en observant un processus à valeurs dans \mathbb{R}^n , s'il existe une matrice \mathbf{A} asymptotiquement stable, $\mathbf{Q} \in \mathbb{R}^n$ et $\mathbf{B} \in \mathbb{R}^n$ tels que :*

Le processus $X = (X(t), t \in \mathbb{R}_+)$ ainsi défini est donc le résultat d'un filtrage causal de bruit blanc (c.f. Krée[55]). En effet, soit $\psi(t) = \exp(t\mathbf{A})\mathbf{Q}$, filtre causal défini pour $t > 0$. Le processus $X = (X(t), t \in \mathbb{R}_+)$ s'obtient à l'aide du filtre causal $\phi = \mathbf{B}^T \psi$.

Le processus admet alors une densité spectrale.

$$S_X(\lambda) = \frac{1}{2\pi} \left| \int_{\mathbb{R}^+} e^{-i\lambda x} \phi(x) dx \right|^2 \quad (\text{II.66})$$

D'autre part, on a :

$$\int_{\mathbb{R}^+} e^{-i\lambda x} \phi(x) dx = R(i\lambda)Q^{-1}(i\lambda) \quad (\text{II.67})$$

où Q est le polynôme caractéristique de \mathbf{A} et R est un polynôme dont les coefficients sont donnés par les vecteurs \mathbf{B} et \mathbf{Q} . On a en particulier le théorème suivant :

Théorème 1. *$G = (G(t), t \in \mathbb{R})$ est à réalisation markovienne de dimension finie ssi G admet une densité spectrale de type rationnel :*

$$S_G(\lambda) \approx \left| \frac{R(i\lambda)}{Q(i\lambda)} \right| \quad (\text{II.68})$$

où R et Q des polynômes avec les racines de $Q(X)$ dans $\mathcal{R}(z) < 0$ tels que :

$$\begin{cases} R(X) &= a_q X^q + \dots + a_1 X + a_0; \\ Q(X) &= b_r X^r + \dots + b_1 X + b_0; r < q \end{cases} \quad (\text{II.69})$$

II.3. Modélisation stochastique du système V-I-C

II.3.4.1 Principe de réalisation markovienne

La classe de Hardy $H_+^2(\mathbb{C})$ est l'ensemble des transformées de Laplace des éléments de $L^2(\mathbb{R}_+, \mathbb{C})$. Une fraction rationnelle de $L^2(\mathbb{R}, \mathbb{C})$ appartient à $H_+^2(\mathbb{C})$ ssi ses pôles sont dans le demi-plan $Re(z) < 0$, ce qui est bien le cas dans le théorème précédent. D'autre part une fonction f quelconque appartient à $H_+^2(\mathbb{C})$ ssi elle vérifie la condition de Paley-Wiener :

$$\int_{\mathbb{R}} \frac{\ln(|f(\lambda)|)}{(1 + \lambda^2)} d\lambda > -\infty \quad (\text{II.70})$$

Le principe de markovianisation approchée se fonde sur le résultat de densité suivant :

Théorème 2. *L'ensemble des fonctions rationnelles $\omega \rightarrow R(i\lambda)/Q(i\lambda)$ qui sont dans $H_+^2(\mathbb{C})$ est dense dans $H_+^2(\mathbb{C})$.*

On peut se reporter à Krée[54] pour la démonstration.

Soit $G = (G(t), t \in \mathbb{R})$ processus gaussien stationnaire centré du second ordre admettant pour densité spectrale :

$$S_G(\lambda) = |G(i\lambda)|^2 ; \quad \forall \lambda \in \mathbb{R} \quad (\text{II.71})$$

avec $G \in H_+^2(\mathbb{C})$. Le principe de markovianisation approchée est donc, grâce au résultat de densité du théorème, d'approcher G par une fraction rationnelle de $H_+^2(\mathbb{C})$. On cherche ainsi à approcher $X = (X(t), t \in \mathbb{R}_+)$ au sens de l'énergie, c'est-à-dire : on cherche $V = (V(t), t \in \mathbb{R}_+)$ processus gaussien de densité spectrale S_V telle que :

$$\|S_X - S_V\|_1^2 < \epsilon \quad (\text{II.72})$$

En ce qui concerne la simulation numérique, on ne va donc pas simuler le processus $X = (X(t), t \in \mathbb{R}_+)$ qui modélise le phénomène physique mais un processus $V = (V(t), t \in \mathbb{R}_+)$ de densité spectrale rationnelle aussi proche que l'on veut de $X = (X(t), t \in \mathbb{R}_+)$ au sens de l'énergie. Un algorithme pour déterminer ce spectre rationnel est proposé dans l'article de Bernard et Bonnemoy[4].

Pour cela, on utilise la technique de recuit simulé. C'est est une métaheuristique inspirée d'un processus utilisé en métallurgie. Ce processus alterne des cycles de refroidissement lent et de réchauffage (recuit) qui tendent à minimiser l'énergie du matériau. Elle est aujourd'hui utilisée en optimisation pour trouver les extrema d'une fonction.

Chapitre II. Méthodologie de prédiction de trajectoires à risque

Soit, la fonction F non convexe donc plusieurs minimas locaux. Il faut éviter de rester piégé autour de ces minimas locaux. Il faut faire une exploration aléatoire de l'espace d'état en favorisant les descentes et en permettant les remontées dans certaines situations.

Plus précisément, soit une chaîne de Markov sur l'espace d'état, et on accepte ou on refuse une transition avec une probabilité 1 si F décroît, et si F croît avec une probabilité $e - \frac{-\Delta F}{T}$. Où T est un paramètre, appelé température par analogie avec la mécanique statistique. Plus la température est grande, plus seront facilitées les transitions ascendantes. En revanche, à la limite $T = 0$, on obtient un algorithme de descente. Tout au long de l'algorithme, on fait décroître T , ni trop vite pour ne pas rester bloqué autour d'un minimum local, ni trop lentement si on veut avoir un résultat en un temps raisonnable.

Le lecteur trouvera plus de détail sur la représentation markovienne d'un processus dans Bernard[5] et Krée[55], Fogli[37].

La simulation de S_G et l'utilisation de la formule (II.65) ont permis de reconstruire les trajectoires de X . Ce résultat nous permet d'obtenir des simulations du processus X . En utilisant la formule (II.31), on obtient des réalisations du processus Z par simulation. Les grandes étapes de cette modélisation probabiliste du système V-I-C sont résumées dans le schéma (II.8).

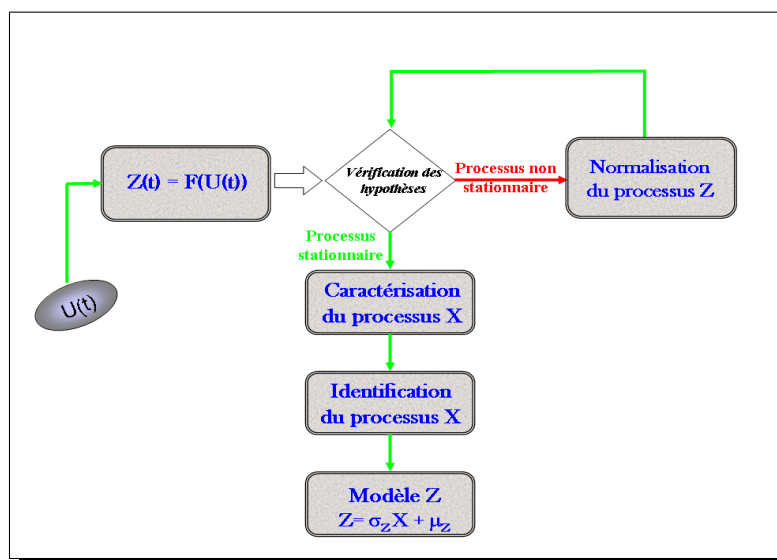


Figure II.8 – Résumé des grandes étapes de la modélisation probabiliste proposée

II.3. Modélisation stochastique du système V-I-C

Les simulations sont utilisées dans un premier temps pour valider les modèles Z . Dans un second temps, elles serviront à estimer les probabilités de dépasser le seuil des critères de sécurité choisis par classe de trajectoires.

II.3.5 Validation des modèles Z

Une fois les modèles Z conçus, pour les valider numériquement, il est nécessaire de comparer les propriétés statistiques des observations mesurées à celles des simulations de Z . A partir de ces simulations, on estime quelques caractéristiques statistiques des données simulées telles que les moments μ^p , la densité de la loi marginale d'ordre 1 p_X et la densité spectrale de puissance S_X de chaque processus X . Ensuite, on compare ces caractéristiques statistiques à celles obtenues à travers les mesures. Avoir des caractéristiques statistiques proches entre les données issues de mesures et celles issues de simulation permet de valider numériquement les modèles probabilistes Z .

Les modèles ainsi conçus sont capables de simuler la dynamique du système V-I-C *vis-à-vis* des deux critères de défaillance.

Ce choix de modélisation permet de prendre en compte la grande variabilité du triplet V-I-C. Il nous a également permis de nous affranchir d'éventuelles difficultés dans l'alimentation des paramètres gouvernant la défaillance du système. Nous avons noté au chapitre I que cette approche de modélisation stochastique va considérablement simplifier l'étape d'analyse fiabiliste. Elle permet d'éviter les problèmes sous-jacents liés à l'utilisation des méthodes fiabilistes classiques telles que FORM/SORM, etc.

La **prochaine section** consiste à utiliser les simulations issues des modèles Z pour faire l'analyse fiabiliste du système V-I-C.

II.4 Analyse fiabiliste du système V-I-C

Les grandes étapes d'une démarche fiabiliste classique ont été introduites au chapitre I. L'objet de cette section est de décrire la méthode d'estimation de la probabilité de défaillance du système, à partir d'observations représentatives de trajectoires pratiquées dans une configuration de virage donnée. Rappelons que 2 critères de défaillance ont été définis dans la section précédente. Ils sont notés K_1 et K_2 pour la suite de l'étude. Ces critères portent respectivement sur les processus de contrôle D (*distance*) et Γ_N (*accélération transversale*). Ces fonctions représentées par Z sont des processus de contrôle scalaires définis sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R} . Dans la partie (II.3), le processus Z a été partiellement caractérisé et identifié.

A partir de Z , nous allons définir les variables de contrôle, les marges de sécurité et les états limites associés. Ensuite, nous identifierons la loi de probabilité des extrêmes de chaque processus Z et par classe \mathcal{C}_p de trajectoires. Enfin, nous estimerons les probabilités de défaillance P_f relatives aux 2 critères de défaillance et à chaque classe \mathcal{C}_p .

II.4.1 Variable de contrôle et critères de sécurité

C'est une variable telle que les valeurs qu'elle prend au delà d'un certain seuil admissible noté δ^* caractérise l'état de défaillance du système. Soit Y une variable de contrôle définie sur l'espace de probabilité $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R} tel que :

$$Y = \sup_{t \in T} |Z(t)| \quad (\text{II.73})$$

Rappelons que $Z(t) = F(U(t))$, $\forall t \in \mathbb{R}_+$, où $F : \mathbb{R}^6 \rightarrow \mathbb{R}_+$ est une fonctionnelle qui opère sur le processus d'état U . Elle représente un opérateur de projection qui permet de se ramener au cas scalaire et qui intervient dans la fonction d'état limite. L'identification partielle du processus Z dans la section (II.22) a permis de simuler Z . Ces simulations permettent d'avoir des réalisations de Y à travers la relation (II.73).

La variable aléatoire Y représente le point de défaillance le plus probable. Elle modélise les paramètres incertains de la dynamique du système. Cette formulation de Y a pour but d'éliminer la notion temporelle le long du virage étudié et de raisonner en terme de franchissement d'un seuil (δ^*) donné. L'hypothèse de stationnarité vérifiée précédemment prend tout son sens. Car si tel n'était pas le cas une telle formulation ne serait pas possible. A partir de la variable de contrôle Y , on associe une *marge de sécurité*.

II.4. Analyse fiabiliste du système V-I-C

II.4.2 Marge de sécurité M

A chaque critère de sécurité est associé une marge de sécurité qu'on note M . Cette grandeur scalaire est une variable aléatoire qui mesure en un point du système mécanique à étudier, l'écart entre la variable de contrôle Y et une valeur admissible δ^* de cette dernière. Une fois que le seuil δ^* est fixé, nous obtenons les réalisations de M par la relation :

$$M = \delta^* - Y \quad (\text{II.74})$$

Ces réalisations sont connues car les observations de Y sont disponibles. La marge de sécurité est aussi à la base de la définition des événements de sécurité noté E_s et de défaillance noté E_f que nous allons définir dans le paragraphe suivant.

II.4.3 Fonction d'état limite G

La fonction G sert à définir la frontière linéaire ou non-linéaire entre le domaine de sécurité et celui de la défaillance. Autrement dit, elle définit les événements de sécurité E_s et de défaillance E_f associés aux états **fiable** et **défaillant** du modèle. Elle est définie par :

$$G(y) = \delta^* - y \quad (\text{II.75})$$

Pour tout $y \in \mathbb{R}$, $G(y) > 0$ caractérise l'état fiable et $G(y) < 0$ l'état défaillant. Par conséquent, E_s et E_f sont définis par :

$$E_s = \{\omega \in \Omega : G(Y(\omega)) > 0\} \quad ; \quad E_f = \{\omega \in \Omega : G(Y(\omega)) < 0\} \quad (\text{II.76})$$

et vérifient :

$$E_s \cup E_f = \Omega \quad ; \quad E_s \cap E_f = \emptyset \quad (\text{II.77})$$

A ces deux événements sont associés les domaines de sécurité D_s et de défaillance D_f du système. Ils sont définis par :

$$D_s = \{y \in \mathbb{R} : G(y) > 0\} \quad ; \quad D_f = \{y \in \mathbb{R} : G(y) \leq 0\} \quad (\text{II.78})$$

et qui vérifient :

$$D_s \cup D_f = \mathbb{R} \quad ; \quad D_s \cap D_f = \emptyset \quad (\text{II.79})$$

Remarque 6. D'après (II.74) et (II.75), $M = G(Y)$.

Les états limites considérés peuvent être considérés de deux manières différentes : indépendants et dépendants. Dans un *premier temps*, ils seront considérés indépendants. En revanche, nous n'aborderons pas la généralisation de la théorie des valeurs extrêmes qui traite des phénomènes où plusieurs paramètres sont nécessaires pour caractériser un phénomène. Car, il faut chercher mathématiquement à estimer la probabilité d'occurrence simultanée de plusieurs facteurs. Dans un *second temps*, nous les considérerons dépendants afin d'estimer la probabilité de défaillance P_f liée à leur interaction. La marge de sécurité associée à cette fonction d'état limite est $M = \delta^* - Y$.

L'étape suivante consiste à identifier la loi de cette variable aléatoire M , nécessaire pour estimer la probabilité de défaillance des trajectoires.

II.4.4 Identification de la loi de la v.a M

La variable aléatoire M est définie par la loi P_M sur (\mathbb{R}, β) , laquelle sera supposée admettre une densité p_M par rapport à la mesure de Lebesgue dy sur (\mathbb{R}, β) :

$$P_M(dy) = p_M(y)dy \quad (\text{II.80})$$

C'est-à-dire

$$\forall B \in \mathcal{B}_p, P_M(B) = \int_B p_M(y)dy \quad (\text{II.81})$$

Pour des raisons évidentes liées à (II.81), cette densité doit vérifier :

$$\text{Supp}(p_M) \subset D_G \quad (\text{II.82})$$

où $\text{Supp}(p_M)$ désigne le support de p_M , D_G le domaine de définition de G . Nous utilisons une estimation de la densité de probabilité p_M de M à travers les réalisations de M . Cette loi de probabilité correspond à la loi des maxima du processus scalaire Z . Il faut maintenant trouver une approximation correcte de la loi de probabilité p_M .

La loi de p_M peut être approximée dans un premier temps par des lois de valeurs extrêmes telles que : Gumbel, Weibull ou Fréchet. Ces lois asymptotiques de valeurs extrêmes sont stables. Par contre, ces 3 types de loi asymptotique d'extrême ne sont pas exhaustifs. D'autres lois ne vérifiant pas le principe de stabilité existent dans certains cas (rare en pratique). Leur intérêt serait de surcroît limité du fait de l'instabilité de leur forme.

II.4. Analyse fiabiliste du système V-I-C

Au cas où ces lois de valeurs extrêmes ne seraient pas adaptées, on peut faire une approximation par un développement sur la base orthonormée des polynômes d'Hermite comme dans la partie (II.3.3.2). Malgré la possibilité de trouver une très bonne approximation de la loi par la méthode Hermite, l'inconvénient se trouve dans la robustesse de cette approximation. Il est donc préférable d'utiliser une des trois lois de valeurs extrêmes citées ci-dessus et de trouver les bons paramètres de cette loi.

A partir d'une approximation correcte de la loi de M , nous estimons la *probabilité de défaillance* $P(E_f)$.

II.4.5 Estimation de la probabilité de défaillance P_f

Le calcul de la **probabilité de défaillance** notée P_f nécessite la simulation de M et donc par conséquent du modèle Z . Notons que la simulation de Z a été faite dans la section (II.3.4). En fixant $\delta^* \in \mathbb{R}_+^*$, on peut estimer la probabilité de défaillance P_f *via* :

$$P_f = P(M < 0) = \int_{-\infty}^0 p_M(x) dx \quad (\text{II.83})$$

L'estimation de cette intégrale peut se faire en utilisant une *méthode de Monte Carlo*. Ces méthodes reposent sur la loi des grands nombres : en répétant un grand nombre de fois une expérience de façon (théorique indépendante), on obtient une approximation de plus en plus précise de (II.83). On peut calculer directement (II.83) dès que la *fonction de répartition* de M est connue. Par exemple si M suit la loi de Gumbel de type (I).

alors F_M a pour expression :

$$F_M(y) = \exp\left(-e^{-\frac{y-\alpha}{\beta}}\right), \quad y \in \mathbb{R} \quad (\text{II.84})$$

et par conséquent :

$$P_f = F_M(0) = \exp\left(-e^{\frac{\alpha}{\beta}}\right) \quad (\text{II.85})$$

où α et β sont deux réels strictement positifs :

$$\begin{aligned} \alpha &= \mu_M - \frac{\gamma\sqrt{6}}{\pi}\sigma_M \\ \beta &= \frac{\sqrt{6}}{\pi}\sigma_M \end{aligned} \quad (\text{II.86})$$

On peut également estimer P_f à partir d'une *approximation numérique* de l'intégrale (II.83) en utilisant un schéma numérique standard si l'on dispose directement d'une bonne

approximation de la loi de p_M (cf. figure(II.9)).

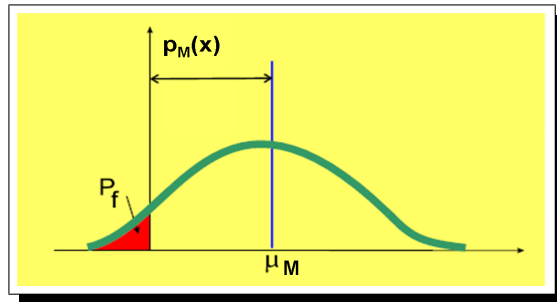


Figure II.9 – Estimation de la probabilité P_f via l'aire sous la courbe.

Notons qu'une des **difficultés** sous-jacente pour la partie applicative est le calage d'un seuil de sécurité δ^* . Nous proposons de faire varier δ^* dans un intervalle de valeurs admissibles et d'estimer la valeur de la probabilité P_f .

II.4.6 Variation de P_f en fonction du seuil δ^*

L'une des caractéristiques probabilistes les plus importantes est la caractéristique dite opérationnelle. C'est la fonction $\pi(\delta^*)$ correspondant à la probabilité d'accepter une classe de trajectoires contenant un nombre de trajectoires u à risque dépassant le seuil de sécurité δ^* .

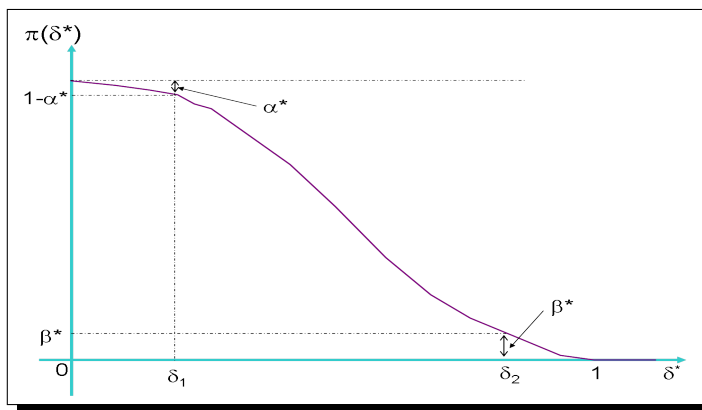


Figure II.10 – Variation de la probabilité P_f en fonction du seuil δ^* .

où α^* est la probabilité d'accepter une classe \mathcal{C}_p contenant un nombre de trajectoires défectueuses, β^* est la probabilité de rejeter une classe contenant un nombre de trajectoires sécurisées. Un choix judicieux de α^* et β^* permet d'identifier la zone linéaire avec précision.

II.4. Analyse fiabiliste du système V-I-C

La fonction $\pi(\delta^*)$ permet de déterminer la *rupture géométrique* (zone de variation linéaire et non linéaire) pour chaque classe de trajectoires identifiée. Une zone de fonctionnement linéaire correspond généralement à une conduite sécurisée pour un groupe de conducteurs. Tandis qu'une zone non linéaire correspond à une situation de conduite instable pour le même groupe de conducteurs.

A cette probabilité de défaillance, nous associons un *intervalle de confiance* à chaque décision car il n'est pas toujours possible de diviser de façon catégorique à l'aide d'un seul nombre δ^* , les trajectoires sécurisées et celles défaillantes. Il apparaît souvent une zone d'indétermination, où nous ne pouvons pas dire exactement si la classe de trajectoires est à risque ou pas. Dans de tels cas la classe est estimée plus sécurisée si $\delta < \delta_1$ et plus à risque si $\delta > \delta_2$, avec $\delta_2 > \delta_1$. Comme la fonction $\pi(\delta^*)$ est monotone décroissante et du fait que les décisions erronées sur l'acceptation ou le rejet sont inévitables, il est souhaitable que les probabilités des décisions erronées soient faibles.

Ainsi, les exigences envers l'étude peuvent consister en ce que les probabilités des erreurs, soient respectivement non supérieures à α^* et β^* , (c.f. figure (II.10)).

En choisissant le critère de défaillance portant sur l'*accélération latérale* et en se référant au paragraphe (I.2.3), obtenir un intervalle d'appartenance de la zone linéaire d'une classe inférieur à $4m/s^2$ signifie une classe de conduite apaisée. Par contre, obtenir une zone linéaire qui dépasse le seuil de $6m/s^2$ signifie une classe de conduite risquée. Ces hypothèses devraient être vérifiées dans le chapitre III concernant l'application de la méthodologie.

La zone linéaire de la fonction $\pi(\delta^*)$ est limitée par la probabilité seuil. En se référant au génie civil, l'évaluation de la **probabilité seuil** (P_{seuil}) pour une structure repose sur les conséquences engendrées par les facteurs suivants : (i) la ruine , (ii) la durée de vie envisagée, (iii) la valeur de remplacement, (iv) le rôle économique de l'ouvrage. Cet ensemble de facteurs peut être estimé avec une certaine probabilité et une confiance sur l'estimation. C'est-à-dire, limiter la probabilité d'atteindre une défaillance à une valeur acceptable. Nous précisons que le choix de la valeur finale de la *probabilité seuil* reste du ressort du législateur. C'est avant tout un choix politique, qui implique de faire accepter la non-sûreté absolue des constructions (c.f. Eurocodes[33]).

Nous venons de décrire l'influence de δ^* sur l'estimation de P_f . Ce résultat permet de fixer la probabilité seuil à partir de la zone de fonctionnement linéaire de chaque classe \mathcal{C}_p .

En utilisant la fonction $\pi(\delta^*)$, nous pouvons comparer les probabilités de défaillance P_f par critère de sécurité K_i et par classe \mathcal{C}_p de trajectoires.

II.4.6.1 Comparaison des fonctions de probabilités $\pi(\delta^*)$

Nous avons 2 critères de défaillance pour P classes de trajectoires. Ce qui fait $2P$ fonctions de la forme $\pi(\delta^*)$. Nous avons fait deux analyses intéressantes pour la comparaison des probabilités qui sont : l'effet des critères de défaillance sur une classe \mathcal{C}_p donnée et l'effet des classes sur un critère donné.

L'expérience nous a montré que les classes de trajectoires ont des niveaux de défaillance différents quelque soit le critère de défaillance considéré. La fonction $\pi(\delta^*)$ décroît rapidement avec l'accroissement de la valeur seuil δ^* pour une conduite apaisée. Tandis qu'elle décroît moins vite avec l'accroissement de la valeur du seuil pour une conduite rapide.

L'expérience a également montré qu'il existe une relation d'ordre entre les critères en terme d'influence sur l'estimation de la probabilité de défaillance. A partir de ce résultat, nous avons mis en évidence une relation d'ordre entre les classes de trajectoires également.

Pour chaque classe \mathcal{C}_p , il a été associé 2 probabilités de défaillance (P_{f1} et P_{f2}) selon les deux critères de défaillance (K_1 et K_2). Certes, cette méthode d'estimation de la défaillance associant plusieurs P_f à une classe \mathcal{C}_p est une démarche intéressante pour des états limites indépendants. Mais, il est également intéressant d'associer une seule probabilité P_f à une classe \mathcal{C}_p en utilisant la fiabilité en série avec le critère du Sup.

Pour cela, soit un ensemble d'événements de défaillance E_i , de probabilités $P(E_i)$. Le système formé d'événements E_i est un système en série si l'occurrence d'un seul événement entraîne la défaillance du système. La probabilité de défaillance du système P_{f-sys} est alors la probabilité de l'union des événements de défaillance :

$$P_{f-sys} = Prob(\cup E_i), \quad i = 1, 2 \quad (\text{II.87})$$

Cette probabilité de défaillance est obtenue en utilisant le critère du Sup.

Par contre, pour des états limites dépendants, ce qui correspond à la fiabilité des systèmes où l'analyse du système est constitué de combinaisons logiques de composants fiabilistes. On a plusieurs états limites en interactions les uns par rapport aux autres. Cette approche nécessite la mise en place d'une méthode de simulation d'un processus stochastique vec-

II.4. Analyse fiabiliste du système V-I-C

toriel défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R}^3 . On peut utiliser en effet la méthode des *copules*.

L'utilisation des copules peut-être envisagée lorsqu'on connaît les lois marginales de variables aléatoires dépendantes et qu'on souhaite caractériser ou modéliser un résultat dépendant de ces variables, [71].

A la probabilité de défaillance P_{f-sys} , on associe un *indice de fiabilité conventionnel*.

II.4.7 Indice de fiabilité conventionnel β_c

Outre la **probabilité de défaillance**, on définit l'**indice de fiabilité conventionnel** β_c par la relation :

$$\beta_c = -\Phi^{-1}(P_{f-sys}) \quad (\text{II.88})$$

où Φ est la fonction de répartition d'une variable gaussienne centrée réduite.

Un indice élevé signifie un risque faible de défaillance de la trajectoire. Lorsque la fonction d'état limite est affine et la variable aléatoire est gaussienne, on a une relation de la forme :

$$\beta_c = \beta_{HL} \quad (\text{II.89})$$

où β_{HL} désigne l'indice de Hasofer-Lind. La figure (II.11) représente une interprétation géométrique de cet indice.

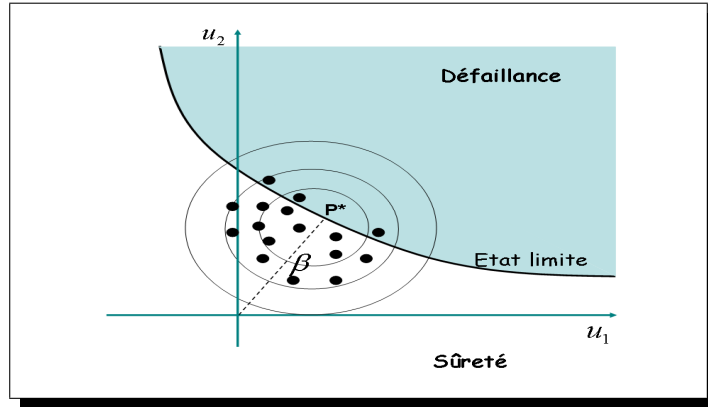


Figure II.11 – Illustration de l'indice de fiabilité β_{HL} .

Dans le cas où la surface de rupture n'est pas linéaire, elle est approchée par un développement de Taylor au premier ordre au point P^* . Pour plus de détails, le lecteur peut consulter la référence dans l'ouvrage de Lemaire [59], Fogli[36].

Chapitre II. Méthodologie de prédiction de trajectoires à risque

Cet indice de fiabilité β_c représente la valeur cumulée de la probabilité de défaillance du système. On montre que la probabilité de défaillance est liée à l'indice de fiabilité β_c selon le tableau suivant pour les Eurocodes0 (EC0) à titre d'exemple.

P_{f-sys}	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
β_c	1.28	2.32	3.09	3.72	4.27	4.75	5.20

Tableau II.1 – Probabilité de défaillance en fonction de l'indice de fiabilité β_c pour EC0.

Ainsi, les *états limites ultimes* correspondent à une probabilité d'occurrence (de défaillance) de l'ordre de 10^{-3} à 10^{-6} et les *états limites de service*, quant à eux, correspondent à une probabilité de l'ordre de 0.5 à 10^{-2} pour les EC0.

Nous venons de mettre en oeuvre une fonction reliant les classes de trajectoires à la probabilité de défaillance. Cette fonction correspond à une table de probabilité P_f en fonction du seuil δ^* par classe de trajectoires \mathcal{C}_p .

Une des difficultés sous-jacentes des méthodes fiabilistes classiques est l'estimation de la probabilité de défaillance en temps réel. Cette difficulté est due à la recherche du point de conception dont l'unicité n'est jamais assurée, Sudret[96]. Une fois le point de défaillance le plus probable P^* obtenu, le calcul de la dérivée seconde pour la méthode FORM/SORM peut s'avérer très fastidieux en grande dimension.

Alors que l'approche d'évaluation de la probabilité de défaillance des trajectoires que nous proposons dans cette thèse est basée sur la lecture d'une table de probabilité. A partir de cette table, il suffit de connaître la classe d'appartenance d'une trajectoire pour déduire sa probabilité de défaillance.

La prochaine section consiste à décrire la méthode de *prédiction de la classe d'appartenance* de chaque trajectoire u .

II.5 Méthode de reconnaissance des trajectoires u

L'objet de cette section est de décrire la mise en oeuvre d'un *modèle de reconnaissance* robuste. Ce modèle noté \mathcal{M} permet de prédire la classe d'appartenance \mathcal{C}_p de chaque trajectoire u en entrée de virage. Nous rappelons qu'à la section précédente, nous avons établi une relation entre les classes de trajectoires $(\mathcal{C}_p)_{p=1,\dots,P}$ et la probabilité de défaillance P_f en fonction du seuil de sécurité δ^* . Les classes de trajectoires sont homogènes en probabilité de défaillance. La prédiction de la classe d'appartenance des trajectoires permettra de connaître la probabilité de défaillance de chaque trajectoire u .

Nous avons utilisé les techniques d'apprentissage de type Séparateurs à Vaste Marges (SVM). C'est une technique qui consiste à extraire de la connaissance à partir d'une base d'apprentissage pour classer de nouvelles données. Le lecteur peut se référer à l'ouvrage de Vapnik, « The nature of statistical learning theory »[101].

Le choix des SVM se justifie à plusieurs points. Un algorithme performant du à l'introduction de la notion de Marge. Il fournit de bonnes performances de classification à partir d'un nombre réduit d'exemples d'apprentissage. Une autre caractéristique des SVMs est leur capacité à traiter des problèmes de grandes dimensions.

L'*originalité* de cette méthode de reconnaissance réside non seulement dans l'utilisation d'une base d'apprentissage en entrée de virage mais également en utilisant les coordonnées les plus appropriées pour décrire la dynamique du système V-I-C, i.e. les coordonnées de la trajectoire u tel qu'elle est définie dans l'équation I.11.

Dans un premier temps, nous identifions la zone d'apprentissage nécessaire pour construire le modèle. Dans un second temps, nous mettrons en oeuvre un modèle multi-classes en utilisant les algorithmes SVM sur la zone d'apprentissage identifiée.

II.5.1 Identification de la zone d'apprentissage

La méthode SVM est basée sur la technique d'apprentissage. Il est donc nécessaire de disposer de données d'observations en entrée de virage. A partir de celles-ci, pour identifier la *zone d'apprentissage*, nous proposons de suivre l'algorithme suivant :

Etape 1 : Soit N le nombre de points d'échantillonnage de la trajectoire u sur tout le virage. On choisit k , la taille de la fenêtre des observations discrètes avec $k = [1, \dots, N_1]$

et $N_1 < N$. Soit le vecteur d'observations suivant :

$$x^v \triangleq \left(x_{1,k}^{(l)}, x_{2,k}^{(l)}, v_{1,k}^{(l)}, v_{2,k}^{(l)}, \gamma_{1,k}^{(l)}, \gamma_{2,k}^{(l)} \right)^T \quad (\text{II.90})$$

Ce vecteur d'observations représente les coordonnées de la trajectoire à des instants $(t_k)_{k=1,\dots,N_1}$ différents.

Etape 2 : On construit la base d'apprentissage en utilisant ces observations discrètes enregistrées entre t_0 et t_k à l'entrée du virage. L'ensemble des données de mesures collectées dans cette zone d'apprentissage est défini par :

$$\mathcal{D}' \triangleq \{x_1, x_2, \dots, x_L\} \quad (\text{II.91})$$

où L est le nombre de trajectoires observées. \mathcal{D}' est un ensemble de trajectoires partielles.

Etape 3 : On implémente l'algorithme de classification des nuées dynamiques en utilisant la base d'apprentissage \mathcal{D}' . Nous obtenons les classes $\mathcal{C}_{p'}$ contenant chacune un certain nombre de trajectoires. Avec $p' \in \mathcal{P}' = \{1, \dots, P'\}$ où P' est le nombre de classe de trajectoires pour la classification de l'ensemble \mathcal{D}' . Rappelons qu'en considérant la totalité de la trajectoire, nous avons obtenu une partition avec P classes de trajectoires (cf. section(II.2)).

Etape 4 : On calcule le nombre de trajectoires communes entre les classes $(\mathcal{C}_p)_{p=1,\dots,P}$ obtenues avec la partition de \mathcal{D} et les classes $(\mathcal{C}_{p'})_{p'=1,\dots,P'}$ obtenues avec la partition de \mathcal{D}' .

Soient $(\mathcal{I}_i)_{i=1,\dots,P}$ et $(\mathcal{I}_j)_{j=1,\dots,P'}$ représentant respectivement le vecteur des indices des trajectoires classées dans $(\mathcal{C}_p)_{p=1,\dots,P}$ et $(\mathcal{C}_{p'})_{p'=1,\dots,P'}$. On obtient le nombre de trajectoires communes entre les classes égal à :

$$N_{co}(i, j) = \text{Card}(\mathcal{I}_i \cap \mathcal{I}_j) \quad (\text{II.92})$$

où N_{co} est une matrice de dimension $P \times P'$.

A partir de ce résultat, on cherche le nombre maximum de trajectoires communes entre une classe \mathcal{C}_p donnée et les classes $(\mathcal{C}_{p'})_{p'=1,\dots,P'}$. On obtient :

II.5. Méthode de reconnaissance des trajectoires u

$$N_{co}^{max}(j) = \max(N_{co}(1, j), N_{co}(2, j), \dots, N_{co}(P, j)) \quad (\text{II.93})$$

où $N_{co}^{max} \in \mathbb{N}^P$ est un vecteur.

Etape 5 : Pour chaque itération k de cet algorithme, nous calculons le pourcentage Pr de trajectoires communes entre les deux partitions :

$$Pr(k) = \frac{\sum_{j=1}^{P'} N_{co}^{max}(j)}{L} \quad (\text{II.94})$$

où L est le nombre de trajectoires classées dans la partition de \mathcal{D} . On réitère les étapes 1, 2, 3, 4, 5 jusqu'à l'obtention de la convergence. On appelle convergence lorsque Pr est invariant d'une itération à une autre. En pratique, on arrête le calcul, soit après un nombre maximum d'itérations, soit sur un seuil de décroissance :

$$W_x = \left| \frac{Pr^{(k+1)} - Pr^{(k)}}{Pr^{(k)}} \right| \leq \epsilon \quad (\text{II.95})$$

Cet algorithme conduit à faire décroître la différence entre les classes des deux partitions. Le résultat idéal serait d'obtenir une fonction Pr qui tend vers la valeur 1. Cela signifie que les deux partitions sont similaires. Ce qui revient à réduire toute l'information contenue dans les observations du virage entier à la zone d'apprentissage en entrée de virage.

En appliquant l'algorithme ci-dessus sur des trajectoires expérimentales, le résultat est représenté par la fonction $k \rightarrow Pr(k)$. Plus la taille de cette zone est réduite (proche de l'entrée du virage), plus c'est mieux car le conducteur aura plus de temps pour corriger la trajectoire de son véhicule. Il est nécessaire de faire un compromis entre le minimum d'observations discrètes k et le maximum de performance Pr . L'amélioration de la valeur de Pr revient à augmenter la précision du modèle d'apprentissage à construire.

La prochaine étape consiste à construire un *modèle multi-classes* en utilisant les algorithmes SVM et les observations issues de la zone d'apprentissage que nous venons d'identifier.

II.5.2 Construction du modèle de reconnaissance

Soit un problème de classification multi-classes, nous considérons le formalisme suivant. Soit l'ensemble \mathcal{S} tel que :

$$\mathcal{S} \triangleq \{(\mathbf{x}_i, y_i) \in \mathbb{R}^v \times \{1, \dots, p\}\}_{i=1}^L.$$

A partir de la base d'apprentissage, nous allons décrire l'utilisation d'un classifieur linéaire. Un classifieur est dit linéaire lorsqu'il est possible d'exprimer sa fonction de décision par une fonction linéaire en fonction de \mathbf{x} . On peut exprimer une telle fonction comme :

$$\mathbf{x} \mapsto f_{\boldsymbol{\theta}}(\mathbf{x}) = \boldsymbol{\theta}^T \mathbf{x} + b = \langle \boldsymbol{\theta}, \mathbf{x} \rangle + b, \quad (\text{II.96})$$

où $\boldsymbol{\theta}, \mathbf{x} \in \mathbb{R}^v$ et $b \in \mathbb{R}$. Pour décider à quelle catégorie y un exemple \mathbf{x} appartient, il suffit de voir le signe de la fonction de décision : $y = \text{sign}(f_{\boldsymbol{\theta}}(\mathbf{x}))$. Géométriquement, cela revient à considérer un hyperplan qui est le lieu des points \mathbf{x} satisfaisant la condition : $\langle \boldsymbol{\theta}, \mathbf{x} \rangle + b = 0$ et de considérer un côté pour lequel les exemples sont classés positivement et l'autre côté pour lequel les exemples sont classés négativement. La figure (II.12) donne une représentation imagée de la situation dans \mathbb{R}^2 . On voit que le vecteur $\boldsymbol{\theta}$ définit la pente de l'hyperplan : $\boldsymbol{\theta}$ est perpendiculaire à l'hyperplan.

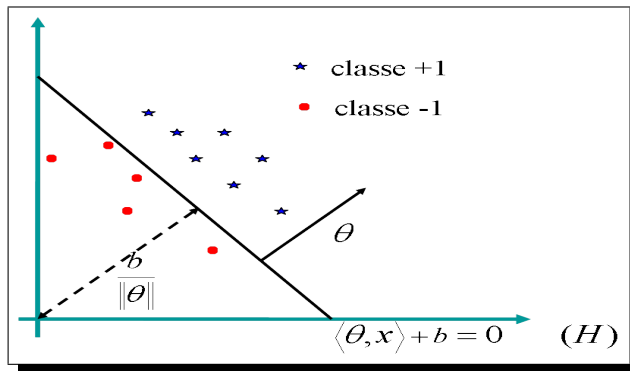


Figure II.12 – Hyperplan séparateur d'un classifieur linéaire. La quantité $\frac{b}{\|\boldsymbol{\theta}\|}$ exprime la distance entre l'hyperplan séparateur (H) et l'origine.

Une description détaillée de la méthode SVM se trouve en annexes(V.5). Une étude comparative sur les différentes techniques d'apprentissage et les méthodes de classification se trouve dans, Boubezoul [10].

Nous allons présenter la procédure de *validation* de ce modèle de reconnaissance conçu par apprentissage sur les observations de \mathcal{D}' .

II.5.3 Validation du modèle \mathcal{M} de reconnaissance

La validation du *modèle de reconnaissance* a été faite par une vérification *a posteriori* de ses performances. Le principe est simple mais efficace (après avoir construit le modèle, on estime ses performances de généralisation). La question de la sélection de modèle est abordée dans le sens de la sélection de l'architecture optimale, c'est-à-dire de la complexité de la famille de fonctions utilisée. On construit un nombre de modèles de structures différentes, on réalise l'apprentissage des paramètres optimaux θ pour chacun d'entre eux, et on sélectionne au final la structure (à travers ses paramètres optimaux) qui donne la plus faible erreur de généralisation.

Il est crucial de mesurer l'erreur de généralisation sur des exemples qui n'ont pas servi lors de la construction du modèle. Pour cela, la méthode de validation croisée a été utilisée. Elle consiste à diviser l'ensemble des trajectoires partielles de \mathcal{D}' en 2 parties :

- un sous ensemble d'apprentissage (\mathcal{S}) dont les données serviront à l'apprentissage (ou construction) du modèle ;
- un sous ensemble de validation (\mathcal{V}) dont les données seront utilisées uniquement pour évaluer la performance du modèle entraîné.

Ensuite, nous répétons cette procédure de rééchantillonnage (apprentissage/test) pour 100 itérations de manière indépendante. Nous avons à chaque itération un taux de reconnaissance et un taux d'erreur. Le résultat du test en terme de prédiction des classes est présenté par une matrice de confusion (cf. tableau(II.2)).

Tableau II.2 – Matrice de confusion SVM Multi-classes.

%	$\hat{\mathcal{C}}_1$	$\hat{\mathcal{C}}_p$
\mathcal{C}_1				
...				
...				
\mathcal{C}_p				

où $(\hat{\mathcal{C}}_1)_{p=1,\dots,P}$ sont les classes prédites par le modèle de reconnaissance. Cette matrice de confusion montre à la fois le *taux de reconnaissance* et le *taux d'erreur* de chaque classe prédite. La meilleure prédiction des classes par le modèle \mathcal{M} correspond à 100% de **taux de reconnaissance** et 0% de **taux d'erreurs** pour chaque classe dans le tableau (II.2).

Chapitre II. Méthodologie de prédiction de trajectoires à risque

Comme nous l'avons fait sous-entendre au chapitre (I), l'erreur dépend d'un certain nombre de facteurs : le nombre de points d'apprentissage N , la corrélation entre les paramètres, le nombre et la pertinence de ces paramètres, etc.. Dans cette étude, avec les observations numériques, nous avons fait varier ces paramètres afin d'illustrer l'efficacité des méthodes de rééchantillonnage.

Le principal *avantage* de la méthode de reconnaissance proposée est l'utilisation de la trajectoire dans la base d'apprentissage pour tenir compte de l'ensemble de la dynamique du système V-I-C. Un autre avantage consiste à utiliser peu d'information en entrée du virage pour donner des résultats avec une bonne performance. Par contre, l'*inconvenient* de cette méthode est qu'il soit sensible au nombre de trajectoires observées et à la représentativité de ces trajectoires dans les classes \mathcal{C}_p .

Le modèle \mathcal{M} permet d'affecter chaque trajectoire u à une classe d'appartenance $\hat{\mathcal{C}}_p$ de trajectoires. Cette affectation implique la connaissance de la probabilité de défaillance P_f de la classe $\hat{\mathcal{C}}_p$ une fois que le seuil de sécurité δ^* est choisi pour une configuration de virage donnée. A partir de cette probabilité P_f , nous avons proposé un *indicateur de risque* qui permet d'évaluer la dangerosité de la trajectoire en cours de réalisation.

II.6 Indicateur de risque du système V-I-C

L'objet de cette section est de construire un indicateur de risque de défaillance de la trajectoire. Nous allons d'abord définir la notion du risque.

II.6.1 Définition du risque

La perception du risque est cruciale. Pour lutter contre un risque, encore faut-il en être conscient ! Cela appelle donc des procédures qui permettent de l'identifier, de le mesurer, et de définir des mesures de prévention si on estime que le risque est trop grand. Dans bien des domaines, le risque est défini par :

Risque = fréquence/probabilité potentielle d'occurrence \times intensité/sévérité du risque.

Dans des systèmes complexes, le risque résulte d'une multitude de facteurs. Il faut donc généraliser la notion de risque en considérant les causes liées à des actions individuelles.

Risque = \sum probabilité de l'événement \times son intensité \times facteur de pondération.

Ces multiples définitions montrent que le risque est subjectif ou plus exactement, nous avons une perception subjective du risque. Chacun de nous a une conscience plus ou moins aiguë du risque qu'il prend ou bien auquel il est soumis. Cette prise de conscience est plus ou moins une bonne représentation du risque réel. La figure (II.13) représente le diagramme de Farmer présenté en 1967 qui dit que plus la gravité de l'événement est grande plus la probabilité d'occurrence est faible.

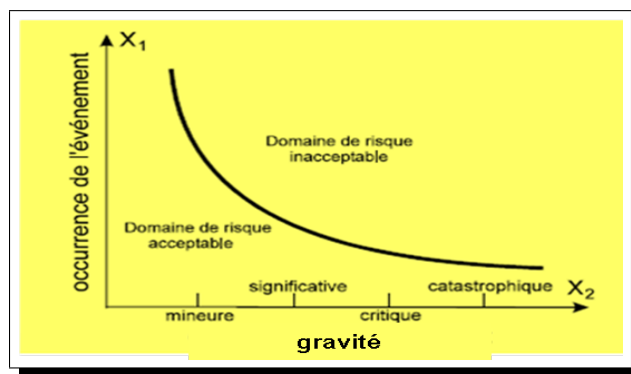


Figure II.13 – Diagramme de Farmer : probabilité-gravité.

La perception du risque est fonction du contexte socio-économique dans lequel nous évoluons. Toute évaluation du risque est basée sur une estimation. Le choix de la mesure du

Activités	Taux moyen de décès par heure d'exposition ($\times 10^{-9}$)	Exposition (h/année)	Taux de décès par année ($\times 10^{-6}$)
Escalade	30000 – 40000	50	1500 – 2000
Tabac	2500	400	1000
Transport aérien	1200	20	24
Transport automobile	700	300	200
train	2.32	80	15
Travail en chantier	70 – 200	2200	150 – 440
Incendie	1 – 3	3.09	8 – 24
Défaillances structurales	0.02	6000	0.01

Tableau II.3 – Risque associé à quelques activités.

risque n'est pas anodin car il conditionne l'interprétation des résultats qu'on en tire. Le tableau (II.3) présente le risque associé à quelques activités.

Le paragraphe suivant décrit l'indicateur de risque proposé dans ce travail de thèse.

II.6.2 Evaluation du risque du système

Pour évaluer un risque, il est nécessaire d'introduire une fonction coût. Cette dernière est définie sur la base des conséquences en perte de vies humaines, d'ordre économique, social et environnemental.

$$\mathbf{Risque} = P_f \times (\text{gravité}) \quad (\text{II.97})$$

où la **gravité** est un scalaire calculé selon le critère de sécurité utilisé. A titre d'exemple la vitesse est un facteur aggravant. La violence des chocs et les conséquences corporelles des accidents en sont considérablement augmentés.

Un tel indicateur de risque noté I_{risque} permet en dehors de la probabilité de défaillance calculée pour une classe de trajectoire donnée, d'associer un risque à chaque trajectoire u de \mathcal{C}_p . Par exemple, soient deux trajectoires $u^{(1)}$ et $u^{(2)}$ appartenant à une même classe \mathcal{C}_p . Elles ont la même probabilité P_f mais leurs dynamiques étant différentes, elles n'auront pas la même gravité en cas d'accident donc des niveaux de risque différents.

Cependant, pour que cet indicateur fonctionne en pratique, il faut mesurer les coordonnées de u et disposer de la probabilité de défaillance P_f de la trajectoire en cours de réalisation.

II.6. Indicateur de risque du système V-I-C

Pour la mesure des coordonnées de la trajectoire u , plusieurs techniques plus ou moins précises sont décrites à la section II.2.1. Quant à la probabilité P_f , nous avons établi au paragraphe (II.4.6) une fonction (abaque) reliant chaque classe \mathcal{C}_p à une probabilité P_f selon un seuil de sécurité δ^* donné. A partir de ces abaques, il suffit de reconnaître une trajectoire en cours de réalisation et de l'affecter à sa classe d'appartenance. La connaissance de cette classe donne directement la probabilité P_f . Nous avons développé à la section précédente un modèle de reconnaissance de trajectoire permettant de prédire la classe d'appartenance de chaque trajectoire u . Donc l'indicateur I_{risque} pourrait être connu.

Au delà de cet indicateur de risque qui est lié à un comportement de conduite donné, on peut établir aussi des classes de fiabilité afin d'effectuer du diagnostic d'infrastructure.

II.6.3 Classes de fiabilité du système

Nous considérons les classes \mathcal{C}_p comme des classes de conséquences. Ceci revient à faire un classement des comportements de conduite sur la base des conséquences en perte de vies humaines, d'ordre économique, social et environnemental. Cette classification par exemple en 3 **classes de conséquences (CC)** est résumée dans le tableau suivant :

Classes de conséquences	Description	Exemples
CC3 Elevée	Conséquences élevées en termes de perte de vies humaines ; conséquences économiques, sociales ou environnementales très importantes	\mathcal{C}_3
CC2 Moyenne	Conséquences moyennes en termes de perte de vies humaines ; conséquences économiques, sociales ou environnementales considérables	\mathcal{C}_2
CC1 Basse	Conséquences faibles en termes de perte de vies humaines ; conséquences économiques, sociales ou environnementales faibles	\mathcal{C}_1

Tableau II.4 – Classes de conséquences du système V-I-C.

Le tableau fait le lien entre les classes de conséquences et les p classes de trajectoires identifiées. On peut faire le même lien entre les classes de conséquences et les configurations de virage. Un tel résultat permet de faire du diagnostic d'infrastructure en classant les infrastructures entre elles en terme de dangérosité.

II.7 Conclusion

L'objectif de ce chapitre (II) était de décrire la *méthodologie proposée*. Cette méthodologie consiste, pour une configuration de virage donnée, et d'observations représentatives de trajectoires pratiquées dans ce virage, à prédire les trajectoires jugées dangereuses.

Après la description du recueil de trajectoires expérimentales, nous avons présenté les techniques de classification les plus adaptées pour regrouper les trajectoires observées en classe de trajectoires avec des propriétés statistiques homogènes. Cette classification est nécessaire car les conducteurs n'ont pas le même comportement de conduite. Des outils sont développés pour garantir des classes avec une séparabilité optimale. Nous avons fait l'hypothèse que les trajectoires d'une classe sont issues d'un même processus stochastique. Cette hypothèse est autant plus vraie, au fur et mesure que les classes sont bien séparées.

Ensuite, à chaque classe de trajectoires \mathcal{C}_p est associée un modèle probabiliste judicieusement construit. Les modèles représentent la dynamique du système *vis-à-vis* de critères de défaillance. Ils sont obtenus par identification de processus stochastiques scalaires normalisés, faiblement stationnaires, ergodiques et non gaussiens. Ce choix de modélisation permet de prendre en compte la grande variabilité du système V-I-C. Chaque processus est partiellement caractérisé par sa loi marginale d'ordre 1 et sa densité spectrale de puissance. La loi marginale est approximée par une technique basée sur le développement de la densité sur la base des polynômes d'Hermite. Les modèles permettent de faire des simulations ayant des propriétés statistiques proches de celles issues de l'expérimentation.

Puis, ces simulations ont servi à estimer la probabilité de défaillance par classe de trajectoires selon les critères de défaillance K_1 et K_2 .

Enfin, nous avons élaboré un modèle de reconnaissance à partir d'observations partielles permettant d'affecter chaque trajectoire u à une classe \mathcal{C}_p dont on connaît préalablement sa probabilité de défaillance. Au-delà de la probabilité de défaillance par classe \mathcal{C}_p , nous avons proposé un indicateur de risque I_{risque} . La connaissance de la probabilité de défaillance P_f , la vitesse V du véhicule permet d'estimer le risque I_{risque} de défaillance de la trajectoire u en cours de réalisation.

Le **prochain chapitre** consiste à appliquer cette méthodologie de prédiction des trajectoires à risque sur le virage Infrasure (LCPC/Nantes).

Application de la méthodologie sur le virage (LCPC/Nantes)

III.1 Introduction

L'objet de ce chapitre est d'appliquer la **méthodologie** décrite dans le chapitre (II). Cette méthodologie consiste à évaluer le risque contenu dans le système V-I-C *vis-à-vis* de critères de défaillance afin de prédire les trajectoires potentiellement dangereuses à partir d'une configuration de virage donnée et d'observations représentatives de trajectoires pratiquées dans ce virage. Nous rappelons que le système V-I-C est défini par les interactions du triptyque : *Véhicule-Infrastructure-Conducteur*.

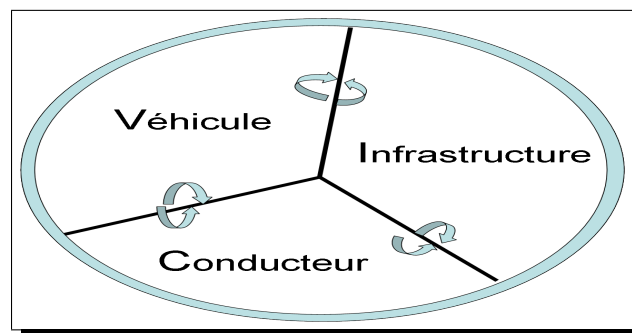


Figure III.1 – Représentation simplifiée du système V-I-C.

Nous avons montré au chapitre I que les interactions entre les trois sous systèmes sont mal représentées et présentent des incertitudes (aléas). Ce qui implique une modélisation

partielle de ce système par les approches *déterministes*. A titre d'exemple, des phénomènes complexes comme l'adhérence, la perception ou encore la visibilité sont difficilement modélisables par ces approches. C'est une des raisons pour lesquelles, nous avons proposé une approche d'analyse et de modélisation probabiliste basée sur l'observation de trajectoires représentatives pratiquées sur une configuration de virage donnée. Elle permet de considérer le véhicule et son environnement.



Figure III.2 – Observation des trajectoires u pratiquées sur ce virage

Une autre raison et non la moindre est le progrès enregistré ces dernières années sur les dispositifs de mesures des trajectoires u surtout pour les véhicules instrumentés en précision et en fréquence d'acquisition.

Dans ce chapitre, après cette introduction, nous commencerons par décrire le *plan d'expérience* qui a été mis en oeuvre pour recueillir les trajectoires pratiquées. A partir de ces trajectoires, comme exposé à la section (II.2.2), nous effectuerons une analyse statistique pour déterminer P classes de trajectoires avec une séparabilité optimale entre ces classes.

Les trajectoires u de chacune de ces classes \mathcal{C}_p sont considérées comme des réalisations d'un même processus stochastique U défini sur $(\Omega, \mathfrak{S}, \mathbb{P})$ à valeurs dans \mathbb{R}^6 . A partir des observations expérimentales de U , nous déterminerons les réalisations du processus de contrôle Z défini sur $(\Omega, \mathfrak{S}, \mathbb{P})$ à valeurs dans \mathbb{R} . Rappelons que dans le chapitre II, deux critères de défaillance ont été choisis pour étudier la défaillance du système. Ces critères portent sur les processus de contrôle suivant : $D(t)$ la *distance relative* à une trajectoire de référence et $\Gamma_N(t)$ l'*accélération transversale* du centre de masse (G) du véhicule. Ces processus de contrôle sont représentés par Z dans cette étude.

III.1. Introduction

Comme décrit dans le chapitre II, Z sera normalisé pour obtenir un nouveau processus noté X défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R} . Le processus X est par définition de moyenne nulle et de variance unité. Pour compléter l'analyse statistique, les hypothèses de *stationnarité* au second ordre de X seront vérifiées. Nous vérifierons notamment les propriétés de sa fonction d'autocorrélation R_X .

Ensuite, dans la **partie approximation** du processus X , nous construirons un processus qui converge en lois vers X . Nous estimerons certaines caractéristiques statistiques de X telles que la loi marginale d'ordre 1 de p_X , la densité spectrale de puissance S_X pour chaque critère de défaillance K_i et par classe de trajectoires \mathcal{C}_p . Un choix de méthodes d'approximation de lois sera effectué en utilisant d'une part des critères de proximité (I_1, I_2) et d'autre part, des tests statistiques d'ajustement de loi. Les modèles seront validés par comparaison des propriétés statistiques (μ^p, p_X, S_X) entre les données mesurées issues de ce virage et celles simulées issues des modèles. Les simulations serviront également à alimenter la partie analyse fiabiliste (estimation du niveau de risque).

Concernant la **partie analyse fiabiliste**, nous associerons la probabilité de défaillance P_f à chaque classe de trajectoires \mathcal{C}_p selon le critère de défaillance K_i . Nous interpréterons le niveau de risque de chaque classe \mathcal{C}_p . Les abaques construits serviront à calculer l'indicateur de risque I_{risque} .

Pour compléter cette méthodologie de prédiction des trajectoires à risque, les résultats du **modèle \mathcal{M} de reconnaissance des trajectoires** seront présentés. Ce modèle permettra de prédire la classe d'appartenance de chaque trajectoire u à l'abord du virage. Il sera validé à l'aide d'une matrice de confusion consistant à présenter le taux de reconnaissance et le taux d'erreur des trajectoires affectées à une classe \mathcal{C}_p .

Enfin, nous calculerons le niveau de risque I_{risque} de défaillance de chaque trajectoire u à partir de sa classe d'appartenance \mathcal{C}_p et de sa probabilité de défaillance P_f correspondante.

III.2 Recueil de données expérimentales

L'objet de cette section est d'acquérir un ensemble de trajectoires u se rapprochant de celles qui seraient pratiquées en situation de conduite naturelle. Pour cela, il faut disposer d'un système de mesures, d'une configuration de virage et d'un protocole expérimental.

III.2.1 Choix d'un dispositif métrologique

Le LCPC dispose d'un véhicule instrumenté appelé VERT (Renault Clio du LCPC/MACS) permettant de mesurer des paramètres cinématiques du véhicule à temps discret à une fréquence de $100Hz$. Le véhicule VERT est équipé d'un dispositif de mesures, appelé MRT (Mesure de Référence des Trajectoires). Ce dispositif de mesures est un système d'acquisition couplé à des capteurs précis (GPS centimétrique et centrale inertielle). Il enregistre les positions, les vitesses, les accélérations et les trois angles d'Euler. La figure (III.3) représente à la fois le véhicule VERT et le système d'acquisition de mesures.

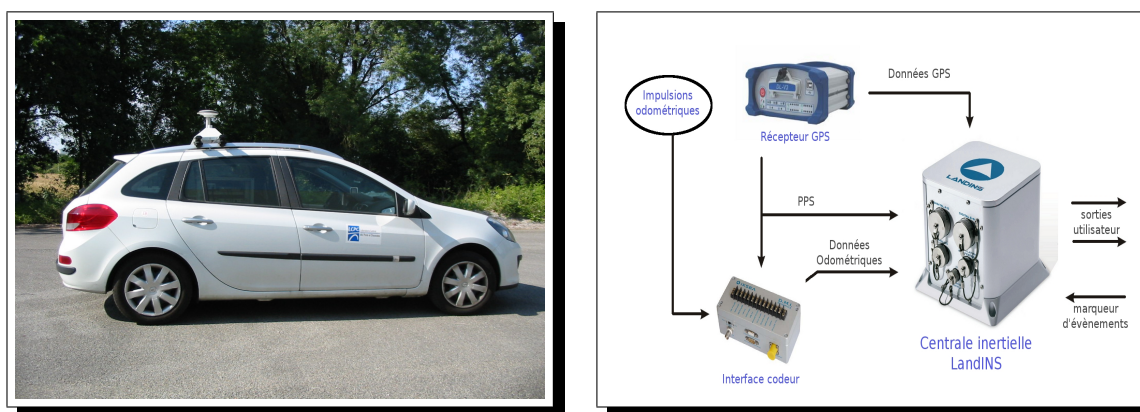


Figure III.3 – Véhicule instrumenté et centrale inertielle.

Grâce aux capteurs de grande qualité utilisés, la centrale inertielle couplée avec un GPS différentiel de phase permet de mesurer des trajectoires avec une précision RMS (Root Mean Square) de 0.01° sur les angles, $3.5cm$ en position longitudinale, $5cm$ en position latérale et $0.07km/h$ en vitesse. Pour plus de détails sur le véhicule VERT, il faut se référer à Ortiz [72]. Nous utilisons ce véhicule car il est non seulement plus précis mais dispose plus de champ d'observation que les *observatoires de trajectoires bord de voie* [42].

III.2.2 Descriptif de l'environnement du véhicule

En raison de la particularité de ce véhicule lourdement instrumenté et par soucis de ne pas gêner la circulation du trafic routier, il a été utilisé sur un circuit fermé. Le site qui a servi

III.2. Recueil de données expérimentales

à cette expérimentation est la piste d'essais du LCPC à Nantes. Cette **piste** est constituée d'un virage similaire à ceux que l'on peut rencontrer sur certaines routes nationales en France. La figure (III.4) représente une vue aérienne de ce virage.



Figure III.4 – Vue aérienne du virage (LCPC/Nantes).

La configuration géométrique de ce virage correspond à une double clothoïde. C'est donc un virage intéressant pour étudier les comportements de conduite. Nous allons maintenant décrire le protocole mis en place pour recueillir des trajectoires observées.

III.2.3 Protocole expérimental

Les **conducteurs** ont été sélectionnés *via* l'emploi d'un *questionnaire* comportant plusieurs critères : tranche d'âge, sexe, nombre d'années du permis de conduire, expérience de la conduite, etc. Nous ne présentons pas la répartition des conducteurs en fonction des critères de sélection car le but de cette étude ne consiste pas à faire de la représentation sociale des conducteurs. On se contente d'obtenir un échantillon représentatif de trajectoires pratiquées qui sont effectuées par une population de conducteurs.

L'accident relatif au nombre de kilomètres parcourus est un phénomène rare donc une observation classique dans le temps n'aurait pas suffi pour observer l'ensemble des situations accidentogènes ou du quasi-accident. Afin de se rapprocher de ces situations accidentogènes et dans le but d'assurer une certaine variabilité des trajectoires observées, 2 consignes cibles à respecter ont été données : **Conduite apaisée** et **Conduite rapide**.

1. La 1^{ère} consigne est destinée à maximiser le confort des passagers donc des trajectoires se rapprochant de la conduite sécurisée.
2. La 2^{ème} consigne avait pour but de minimiser le temps de parcours. Contrairement à la première, cette conduite se situe en état de stress (donc situation à risque).

Ces consignes sont alternées dans un ordre déterminé pour éviter les effets d'accoutumance à une consigne donnée. Après un repérage du trajet en tant que passager, chacun des 32

conducteurs a réalisé 10 passages. L'ensemble des trajectoires collectées est noté \mathcal{D} :

$$\mathcal{D} = \left(u_k^{(l)} \right)_{k \in I(l)}^{1 < l \leq L} \quad (\text{III.1})$$

où $L = 232$ est le nombre total de trajectoires disponibles ; $I(l) = \{1, 2, \dots, k(l)\}$.

III.2.4 Illustration d'une trajectoire expérimentale

Rappelons que la trajectoire u définie par la formule (II.2) décrit respectivement la trace et la loi horaire du centre de masse (G) du véhicule dans le repère Galiléen de référence \mathcal{R}_0^A . Les figures (III.5) à (III.7) représentent une illustration de l'évolution des coordonnées d'une quelconque des trajectoires u dans la base fixe \mathcal{B}_0^A .

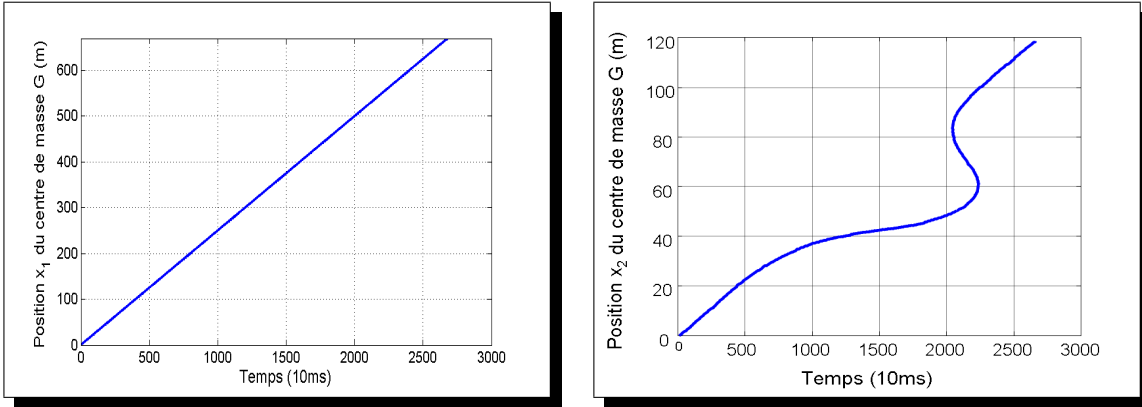


Figure III.5 – Evolution temporelle des coordonnées (x_1 et x_2) d'une trajectoire u .

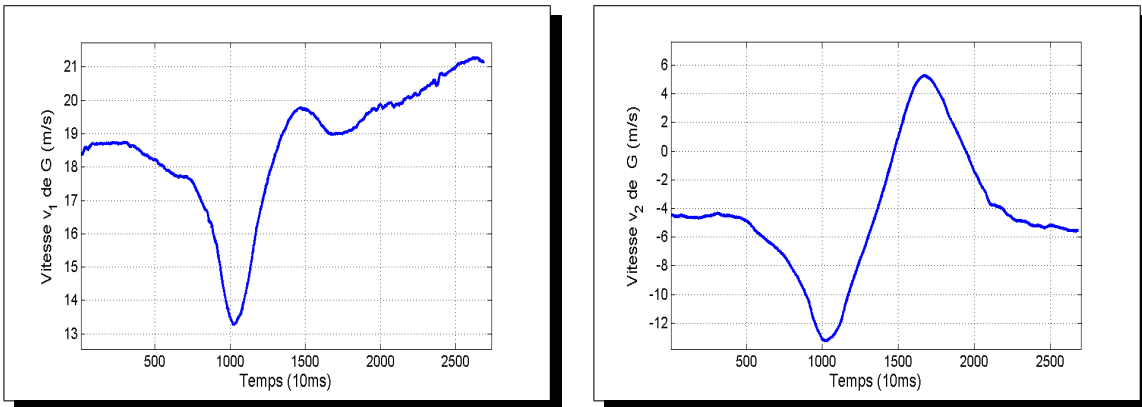


Figure III.6 – Evolution temporelle des coordonnées (v_1 et v_2) d'une trajectoire u .

III.2. Recueil de données expérimentales

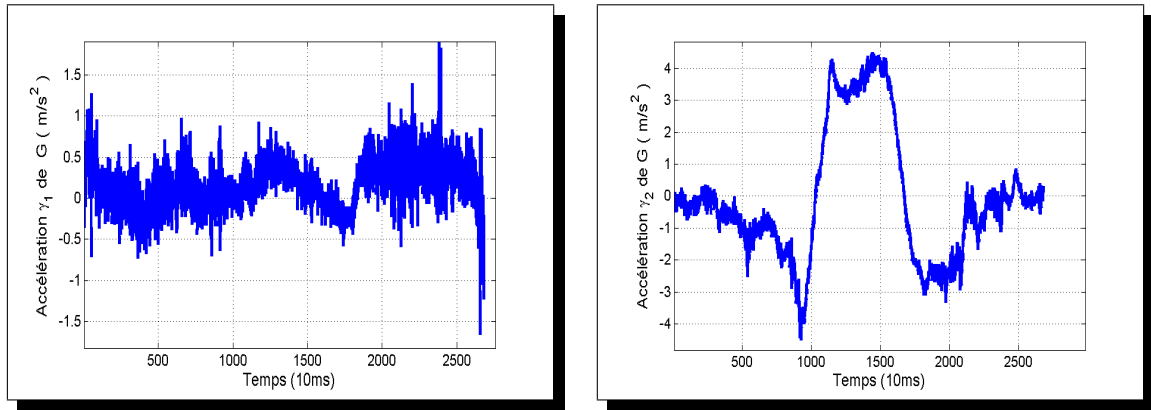


Figure III.7 – Evolution temporelle des coordonnées (γ_1 et γ_2) d'une trajectoire u .

Nous constatons à travers ces figures que la mesure des coordonnées de la trajectoire u est d'une bonne précision. Cette dernière est nécessaire pour l'ensemble des étapes de la méthodologie. Notons qu'une des forces et faiblesses de la méthodologie proposée réside dans la précision des données.

Nous avons expliqué dans le chapitre II que les conducteurs n'ont pas le même comportement de conduite sur la totalité du parcours, il est donc nécessaire de regrouper les trajectoires u de l'ensemble \mathcal{D} dans des classes \mathcal{C}_p .

La prochaine section consiste à présenter les résultats de la classification des trajectoires observées sur cette configuration de virage.

III.3 Classification des trajectoires u observées

L'objet de cette section est de présenter les résultats de la classification des trajectoires. Nous rappelons que l'ensemble des trajectoires \mathcal{D} obtenu dans la précédente section est composé de 232 trajectoires expérimentales. Les méthodes (nuées dynamiques, meanshift et SOM) décrites au paragraphe (II.2.2) ont été utilisées pour répartir ces trajectoires dans des classes \mathcal{C}_p différentes avec des propriétés statistiques homogènes.

III.3.1 Résultat de la classification

Dans un premier temps, la L -famille \mathcal{D} est répartie en 2 classes représentant les **trajectoires rapides** et les **trajectoires lentes**, selon les 2 consignes données. Les 2 classes obtenues admettent une grande variance intra-classe, de ce fait les algorithmes ne convergent pas pour trouver un résultat identique en répétant l'expérience. En essayant une partition avec $P = 4$ classes, les algorithmes de classification convergent en donnant des classes stables *vis-à-vis* du nombre d'itération. Pour $P > 4$, les algorithmes peuvent toujours converger mais il faut un compromis entre la convergence et le nombre minimum de classes de trajectoires représentatives. Car un nombre élevé de classes \mathcal{C}_p identifiées conduit à des classes moins représentatives en trajectoires pratiquées. Dans cette étude, le seuil ϵ de décroissance pour la convergence est fixé à 10^{-2} , (cf. II.2.2).

Rappelons que cette classification est multi-dimensionnelle. Pour des raisons de représentation graphique, le résultat a été projeté sur les composantes *vitesse* v_1 et *accélération* γ_2 du centre de masse (G) du véhicule, c.f. figure (III.8).

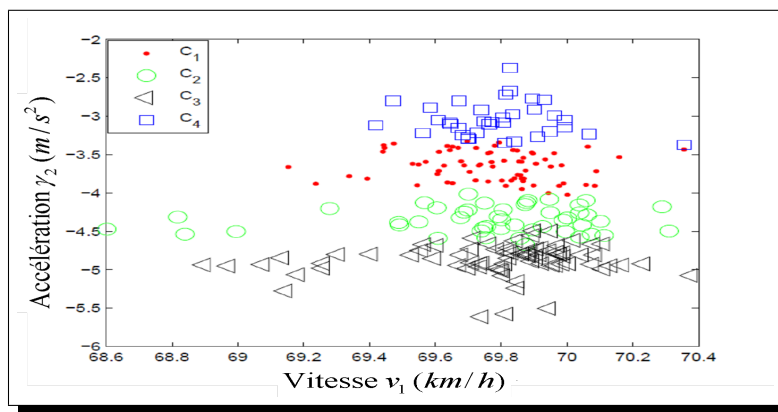


Figure III.8 – Projection dans \mathbb{R}^2 du résultat de la classification des trajectoires u .

Nous distinguons sur la figure (III.8), une séparation *non linéaire* entre les classes \mathcal{C}_1 et

III.3. Classification des trajectoires u observées

\mathcal{C}_2 . Ce résultat s'explique par le fait que la consigne n'a pas été comprise de la même façon par tous les conducteurs. Ce qui se traduit parfois par une conduite hésitante. Par contre, entre les classes \mathcal{C}_2 et \mathcal{C}_3 , nous avons une séparation *linéaire*. Ce qui est du à la différence des consignes données pour réaliser ces deux classes de trajectoires.

Malgré la convergence des algorithmes de classification, nous allons vérifier le critère de séparabilité S entre les 4 classes \mathcal{C}_p dans le but d'obtenir une partition optimale.

III.3.2 Validation des 4 classes de trajectoires identifiées

Il est important de déterminer les classes avec une grande précision car la performance des modèles probabilistes à développer dépend également de la précision de ce résultat. Afin d'assurer une classification *optimale*, nous avons utilisé des critères de validité des algorithmes de classification tels que la séparabilité entre les classes, c.f. figure (III.9).

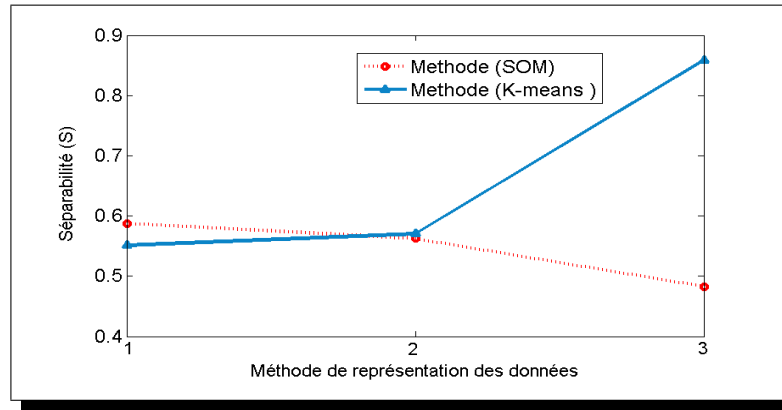


Figure III.9 – Séparabilité entre les 4 classes de trajectoires.

Sur cette figure, les méthodes de représentation sont indiquées par les indices suivants : 1 =moyenne, 2 =variance et 3 =médiane sur l'axe des abscisses. En ordonnée, on a la valeur de la séparabilité entre les quatre classes de trajectoires. Cette valeur est comprise entre 0 et 1. Elle est obtenue par la formule (II.19).

On a expliqué au chapitre II que les données de l'ensemble \mathcal{D} doivent être représentées par un ensemble en $2D$ pour que les algorithmes de classification puissent être utilisés.

Le résultat de la figure (III.9) a permis de choisir l'algorithme k -means comme méthode de classification et la médiane comme méthode de représentation des données. Car ce couple donne une meilleure séparabilité entre les 4 classes de trajectoires identifiées.

Chapitre III. Application de la méthodologie sur le virage (LCPC/Nantes)

Une classification optimale correspond à une meilleure séparabilité des classes, cf. section(II.2.4).

Cette classification a permis de regrouper les trajectoires ayant des propriétés statistiques voisines (variance inter-conducteurs faible pour une même classe \mathcal{C}_p).

Le tableau (III.1) donne la répartition du nombre de trajectoires observées de l'ensemble \mathcal{D} par classe \mathcal{C}_p .

Classes	\mathcal{C}_1	\mathcal{C}_2	\mathcal{C}_3	\mathcal{C}_4
Nombre de trajectoires	35	73	47	77

Tableau III.1 – Répartition en nombre de trajectoires

III.3.3 Interprétation physique des classes

Pour donner une interprétation physique aux 4 classes \mathcal{C}_p de trajectoires identifiées, nous avons associé un comportement de conduite à chacune de ces classes.

1. La 1^{ère} classe \mathcal{C}_1 est la classe des trajectoires lentes avec la conduite apaisée, elle correspond à une conduite lente et hésitante,
2. La 2^{ème} classe \mathcal{C}_2 est la classe des trajectoires rapides avec la consigne apaisée, elle correspond à une conduite prudente,
3. La 3^{ème} classe \mathcal{C}_3 est la classe des trajectoires lentes de la consigne rapide, elle correspond à une conduite performante,
4. La 4^{ème} classe \mathcal{C}_4 est la classe des trajectoires rapides de la conduite rapide, elle correspond à une conduite risquée.

Toujours dans l'interprétation de ces 4 classes de trajectoires, la vitesse v_1 et l'accélération γ_2 ont une amplitude qui augmente de la classe \mathcal{C}_1 à la classe \mathcal{C}_4 . Ce qui signifie que les 4 classes n'ont pas la même dynamique pour un observateur terrestre.

Rappelons que l'identification des classes est une étape décisive car la méthodologie proposée est basée sur celles-ci. Il est donc important de vérifier l'hypothèse que les classes $(\mathcal{C}_p)_{p=1,\dots,4}$ sont physiquement liées au système V-I-C étudié. Pour cela, nous avons fait une seconde expérimentation dans les mêmes conditions que la première expérimentation.

III.3. Classification des trajectoires u observées

Les propriétés statistiques (moyenne μ , écart-type σ , etc.) des nouvelles classes de trajectoires sont proches de celles des anciennes classes. Le schéma (III.10) résume la démarche suivie pour la vérification de l'existence physique des classes $(\mathcal{C}_p)_{p=1,\dots,4}$.

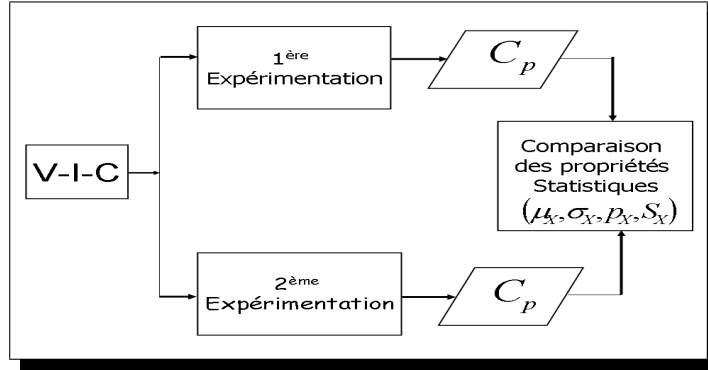


Figure III.10 – Schéma de validation des classes \mathcal{C}_p obtenues.

La figure (III.11) montre la comparaison entre la première et la seconde expérimentation. Cette comparaison correspond à la moyenne temporelle de l'accélération γ_2 du centre de masse (G) du véhicule dans le repère fixe \mathcal{R}_0^A .

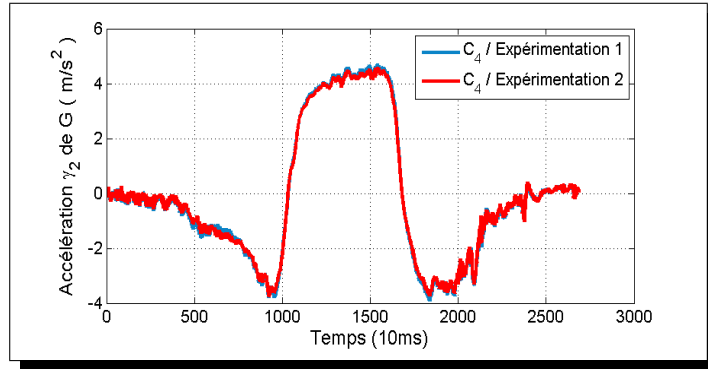


Figure III.11 – Accélération γ_2 pour la 1^{ère} et 2^{ème} expérimentation.

Les deux courbes étant proches, cela signifie que les classes sont liées au système V-I-C étudié. Après cette vérification, nous rappelons que les trajectoires de chaque classe \mathcal{C}_p sont assimilées à des réalisations d'un unique processus stochastique vectoriel U_p pour tout $p \in \{1, \dots, 4\}$.

La prochaine section consiste à vérifier les hypothèses de *régularité des processus* stochastiques étudiés.

III.4 Analyse statistique des processus stochastiques

L'objet de cette section est de vérifier l'hypothèse de stationnarité du processus d'état U , du processus de contrôle Z et du processus normalisé X . Il est important de vérifier cette hypothèse car les modèles probabilistes que nous mettrons en oeuvre dans la prochaine section sont basés sur l'identification de processus stochastiques stationnaires.

III.4.1 Vérification des hypothèses de régularité de U , Z et X

Rappelons que $U_p = (U_p(t), t \in \mathbb{R})$ est le processus d'état du système V-I-C, défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R}^6 , tel que $\forall t \in \mathbb{R}$:

$$U_p(t) = (X_1(t), X_2(t), V_1(t), V_2(t), \Gamma_1(t), \Gamma_2(t))^T \quad (\text{III.2})$$

Le processus U_p est noté U afin d'alléger les notations. On a $u(t) = U(t, \omega)$, $\forall t \in T$ où $\omega \in \Omega$.

Les composantes de U sont respectivement : les *positions*, les *vitesses* et les *accélérations* du centre de masse (G) du véhicule selon la base $\mathcal{B}_0^A = (\vec{e}_1, \vec{e}_2, \vec{e}_3)$ du repère fixe \mathcal{R}_0^A .

III.4.1.1 Processus d'état U

Le processus U est *non stationnaire* sur la totalité du virage car la moyenne temporelle de ses composantes est fortement dépendante du temps, c.f. figure (III.12). Par contre,

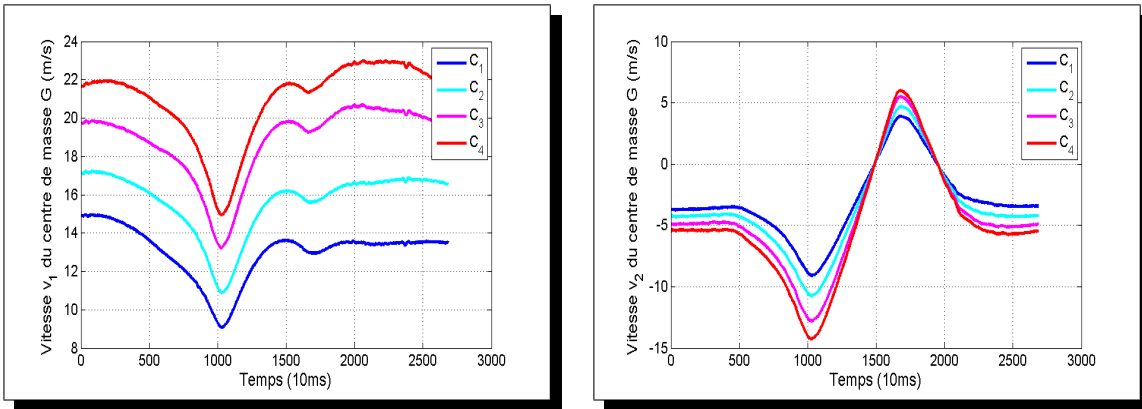


Figure III.12 – Moyenne temporelle des coordonnées (v_1 et v_2) de u .

la figure (III.13) montre une **stationnarité par morceaux** pour les composantes : *accélération* $\gamma_1(t)$ et *accélération* $\gamma_2(t)$ du centre de masse (G) du véhicule. Malgré cette stationnarité par morceaux, il est difficile de caractériser U . Car cela nécessite d'une part,

III.4. Analyse statistique des processus stochastiques

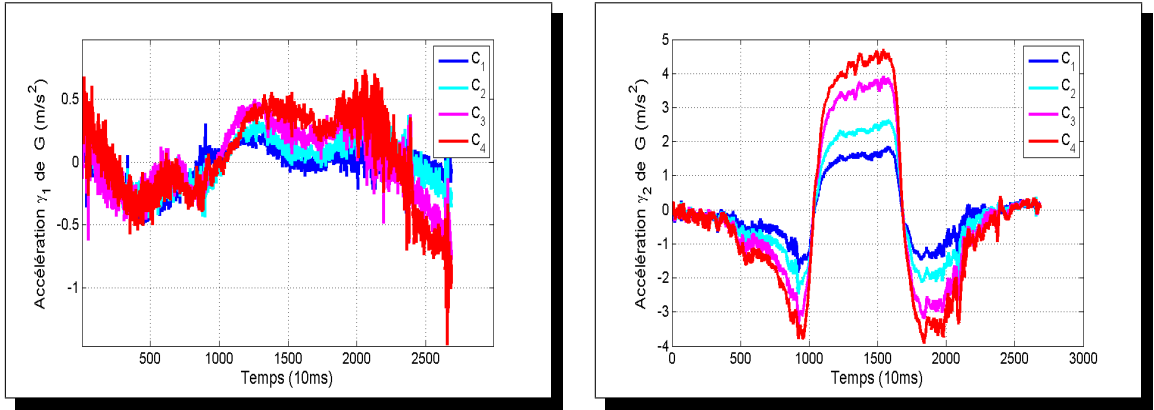


Figure III.13 – Moyenne temporelle des coordonnées (γ_1 et γ_2) de u .

de faire un découpage du virage en zones stationnaires. Et d'autre part, il faut relier les différentes zones de découpage après la caractérisation de chacune de ces zones. Ce qui n'est pas envisageable vu la dimension du processus U et la forme des signaux. A partir des observations de U , nous obtenons les réalisations du processus Z défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R} .

III.4.1.2 Processus de contrôle Z

Le processus de contrôle $Z = F(U)$ est obtenu par la transformation du processus d'état U , où F désigne une fonctionnelle affine. Les deux cas de Z choisis dans le chapitre II pour étudier la défaillance du système V-I-C sont :

- $D = (D(t), t \in \mathbb{R}_+)$, processus de contrôle représentant la *distance* entre le centre de la voie (C_V) et la position latérale du centre de masse (G) dans le repère route. Rappelons que D est obtenue par :

$$D(t) = \underbrace{\min}_{M \in C_V} (d(G(t), M)) \quad (\text{III.3})$$

- $\Gamma_N = (\Gamma_N(t), t \in \mathbb{R}_+)$, processus de contrôle représentant l'*accélération transversale* du véhicule dans le repère de Serret-Frénet. Nous avons montré au chapitre I, le passage du repère galiléen fixe \mathcal{R}_0^A au repère mobile de Serret-Frénet \mathcal{R}_G^{SF} .

Nous rappelons que ces 2 processus de contrôle portent respectivement les critères de défaillance K_1 et K_2 . Les figures (III.14) et (III.15) montrent une variation importante de la moyenne temporelle des processus Z pour chaque classe \mathcal{C}_p et pour chacun des critères $(K_i)_{(i=1,2)}$ considérés. Cela signifie que Z n'est également pas stationnaire comme le processus stochastique U .

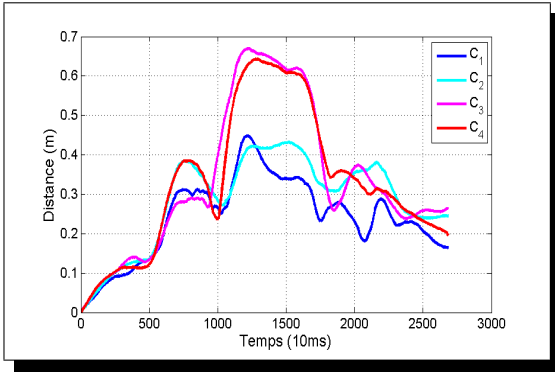


Figure III.14 – Moyenne temporelle de Z selon le critère K_1 pour les 4 classes.

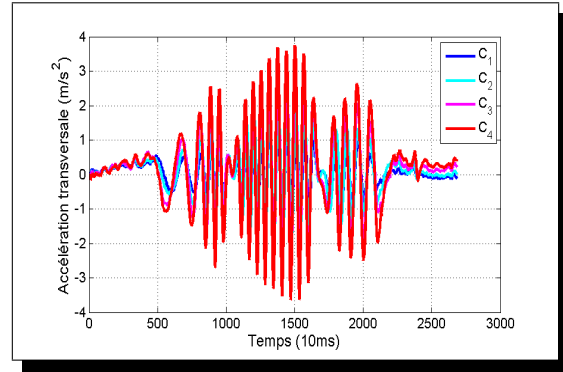


Figure III.15 – Moyenne temporelle de Z selon le critère K_2 pour les 4 classes.

Rappelons que les méthodes de simulation utilisées dans ce travail de thèse sont adaptées pour des processus stationnaires en moyenne d'ordre deux.

L'étape suivante de notre démarche (c.f. chapitre II) consiste à chercher une *représentation standardisée* de Z que l'on a noté X .

III.4.1.3 Représentation standardisée du processus Z

Le processus scalaire X défini sur $(\Omega, \mathfrak{F}, \mathbb{R})$ à valeurs dans \mathbb{R} est obtenu par la relation (II.30). Ce processus admet par construction une moyenne $\mu_X = 0$ et un écart-type $\sigma_X = 1$.

Les figures (III.16) et (III.17) représentent une ω -réalisation du processus X relative aux 4 classes \mathcal{C}_p de trajectoires et à chacun des 2 processus de contrôle $D(t)$ et $\Gamma_N(t)$.

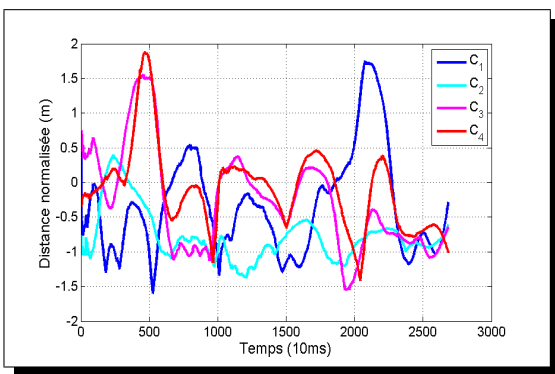


Figure III.16 – Une ω -réalisation de X du processus de contrôle $D(t)$.

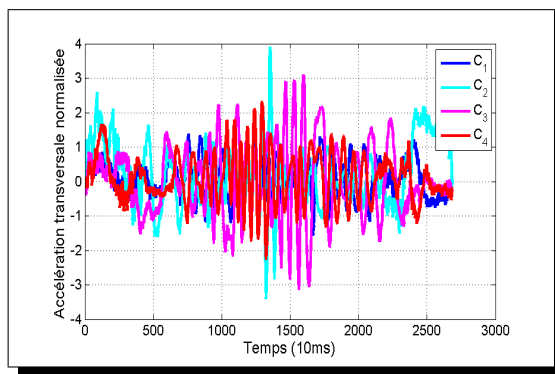


Figure III.17 – Une ω -réalisation de X du processus de contrôle $\Gamma_N(t)$.

Nous allons maintenant vérifier que le processus X est stationnaire au second ordre.

III.4. Analyse statistique des processus stochastiques

III.4.1.4 Fonction d'autocorrélation X

Par construction le processus X est centré et de variance unité. Mais l'hypothèse de stationnarité au second ordre suppose aussi que la fonction d'autocorrélation R_X de X ne dépend que de l'écart τ entre deux instants t_1 et t_2 , (cf. annexesV.1). Les figures (III.18) et (III.19) représentent les fonctions d'autocorrélation $\tau \rightarrow R_X(\tau)$ relatives aux 4 classes de trajectoires et à chacun des 2 critères de défaillance K_1 et K_2 .

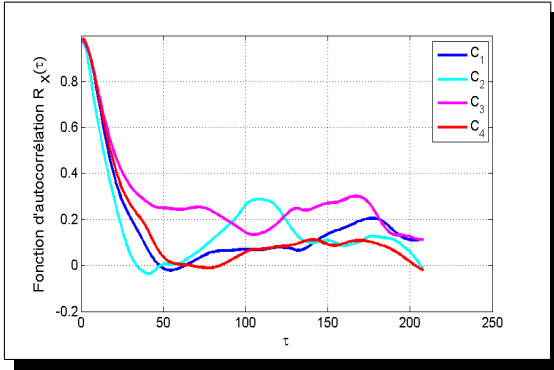


Figure III.18 – Graphes de R_X selon le processus de contrôle D .

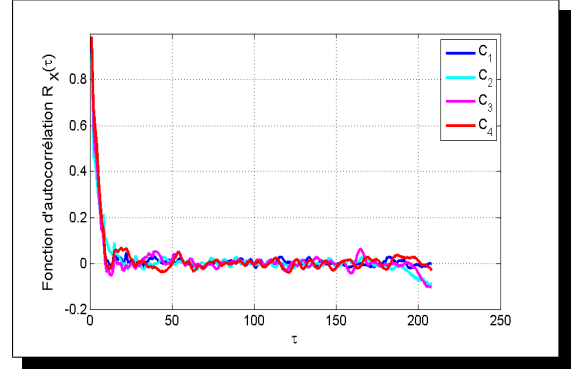


Figure III.19 – Graphes de R_X selon le processus de contrôle Γ_N .

Selon la classe \mathcal{C}_p de trajectoires et le critère K_i , on a une décroissance rapide de la fonction d'autocorrélation R_X . On constate sur ces figures que la variation de la fonction R_X dépend de l'écart entre deux instants $\tau = (t, t + k\Delta)$.

Outre la vérification des conditions de stationnarité à l'ordre 2 du processus X par la fonction d'autocorrélation $R_X(\tau)$, on a utilisé aussi des tests statistiques de stationnarité.

Nous avons utilisé le test de substitués *temps-fréquence*, Xiao[105]. Le principe de ce test est d'utiliser des caractéristiques *temps-fréquence* pour donner un sens statistique précis à un signal. Ce test est relatif à une durée d'observation et peut s'appliquer à des bandes de fréquence, donnant un cadre à ce qui est souvent fait en pratique. Un degré de non-stationnarité peut être attaché au test, ainsi que, le cas échéant, une échelle caractéristique de la non-stationnarité détectée. En appliquant ce test sur le processus X , nous remarquons que l'hypothèse de stationnarité n'est pas rejetée. Ceci confirme, le résultat obtenu avec la fonction d'autocorrélation $R_X(\tau)$.

Après cette vérification des hypothèses de stationnarité du processus X , nous allons à présent construire des *modèles probabilistes* spécifiques à chaque classe \mathcal{C}_p de trajectoires.

III.5 Modélisation stochastique du processus X

L'objet de cette section est de construire des modèles probabilistes par identification de processus stochastiques. Dans un premier temps, nous allons estimer la loi marginale d'ordre 1 p_X et la densité spectrale de puissance S_X en utilisant les observations expérimentales disponibles dans chacune des classes \mathcal{C}_p . Ces estimées seront nécessaires pour simuler le processus X et par conséquent le modèle probabiliste Z . Ensuite, nous chercherons des approximations adéquates aux estimées numériques p_X et S_X . Enfin, nous présenterons les résultats de la validation numérique des modèles Z .

III.5.1 Loi marginale d'ordre 1 de X

III.5.1.1 Estimées de la densité p_X et de la fonction de répartition F_X

Rappelons que les densités marginales d'ordre 1 ont été estimées à l'aide de l'estimateur standard et de l'estimateur à noyau gaussien, cf. annexes(V.2). Les figures (III.20) et (III.21) montrent les estimées de la densité p_X de la loi marginale d'ordre 1 relatives aux 4 classes de trajectoires identifiées $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4$ et aux 2 critères de défaillance considérés K_1, K_2 .

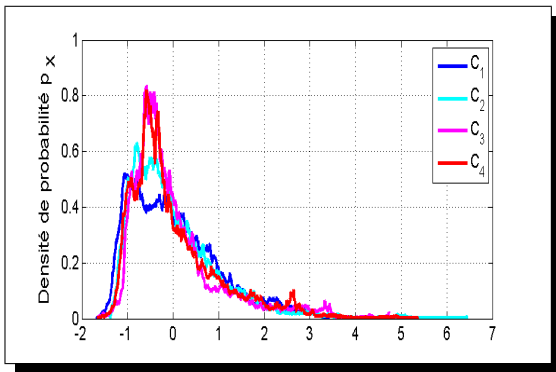


Figure III.20 – Graphes de p_X pour le critère K_1 .

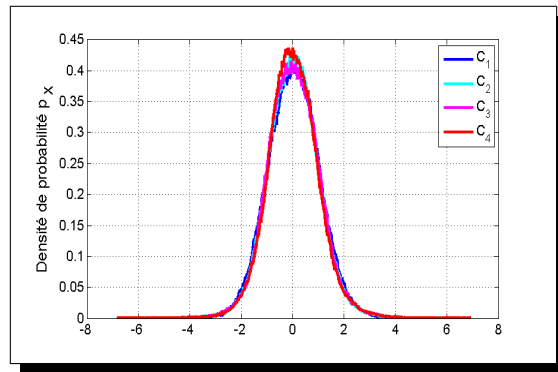


Figure III.21 – Graphes de p_X pour le critère K_2 .

La densité p_X relative au critère K_1 montre plus de variation (fluctuation) que celle du critère K_2 . Ceci s'explique par la variabilité de chaque critère de défaillance.

En observant la figure (III.21), on pourrait penser que les lois obtenues pour le critère d'*accélération transversale* sont proches à des lois gaussiennes. Cependant, les valeurs des coefficients d'aplatissement K et d'asymétrie S relatifs à ces lois ne confirment pas cette

III.5. Modélisation stochastique du processus X

hypothèse, cf. tableau (III.2).

Coefficients	\mathcal{C}_1	\mathcal{C}_2	\mathcal{C}_3	\mathcal{C}_4
Aplatissement (K)	3.25	3.38	3.32	3.54
Asymétrie (S)	0.01	-0.02	-0.01	0.11

Tableau III.2 – Coefficients d’aplatissement et d’asymétrie de X .

Pour une loi gaussienne, on a $K = 3$ et $S = 0$.

Nous allons présenter les approximations trouvées pour les 8 densités de probabilité p_X estimées ci-dessus. Ces densités correspondent aux 4 classes de trajectoires pour chacun des 2 critères de défaillance K_1 et K_2 .

Au chapitre II, nous avons proposé d’approximer les lois cibles dans un premier temps par les lois usuelles (loi normale et loi logistique) et dans un second temps par le développement de la densité p_X sur la base des polynômes d’Hermite. Nous avons proposé cette seconde méthode car les lois classiques ne sont pas souvent adaptées. Au cas où ces deux méthodes ne seraient pas adaptées, il faut utiliser le principe d’entropie maximale.

III.5.1.2 Approximation de la loi marginale d'ordre 1 de X

Les figures (III.22) et (III.23) représentent les approximations \tilde{p}_X de la densité cible p_X pour le critère de *distance relative* selon les classes \mathcal{C}_1 et \mathcal{C}_2 , puis \mathcal{C}_3 et \mathcal{C}_4 .

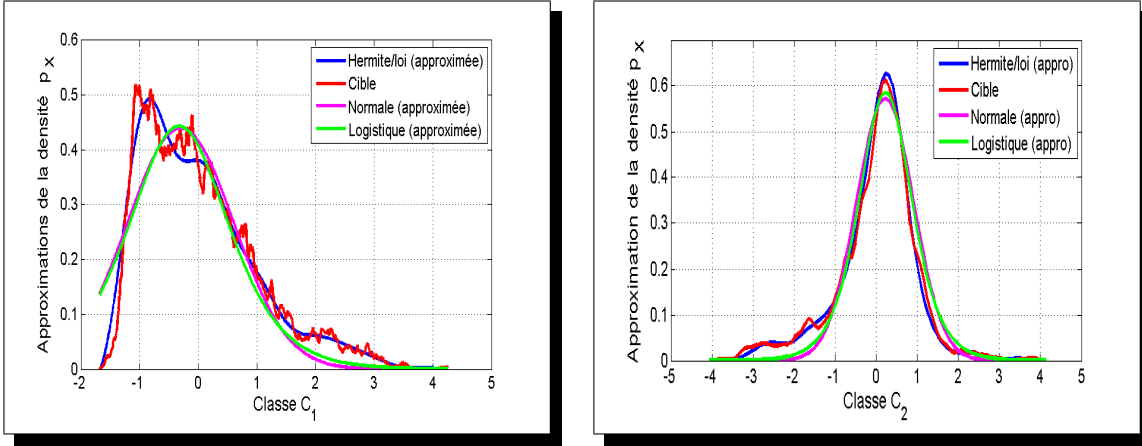


Figure III.22 – Approximations de p_X pour \mathcal{C}_1 et \mathcal{C}_2 selon le critère K_1

La figure (III.22) montre une bonne approximation de p_X avec le développement sur la base des polynômes d'Hermite. Par contre, celles obtenues par les lois usuelles (loi normale et loi logistique) sont très mauvaises quelque soit la classe de trajectoires.

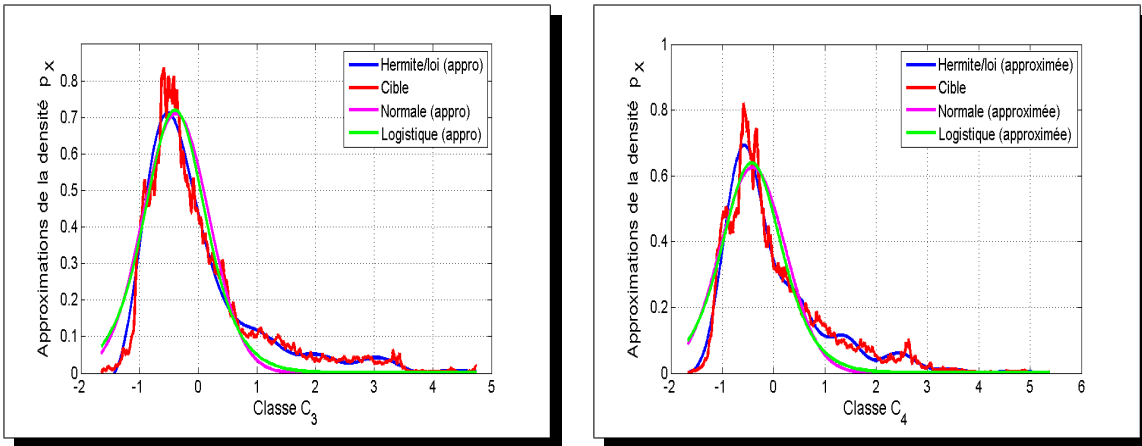


Figure III.23 – Approximations de p_X pour \mathcal{C}_3 et \mathcal{C}_4 selon le critère K_1

La figure (III.23) montre également une bonne approximation par la méthode Hermite et une très mauvaise approximation par les lois usuelles. Ce résultat est confirmé par les critères de proximité (I_1 et I_2) et les tests statistiques d'adéquation de lois. D'où l'intérêt d'utiliser ces méthodes complexes pour garantir la qualité des approximations.

III.5. Modélisation stochastique du processus X

Critère K_2 :

Quant aux figures (III.24) et (III.25), elles montrent une approximation \tilde{p}_X de la loi cible p_X pour le critère d'accélération latérale selon les classes \mathcal{C}_1 et \mathcal{C}_2 , puis \mathcal{C}_3 et \mathcal{C}_4 .

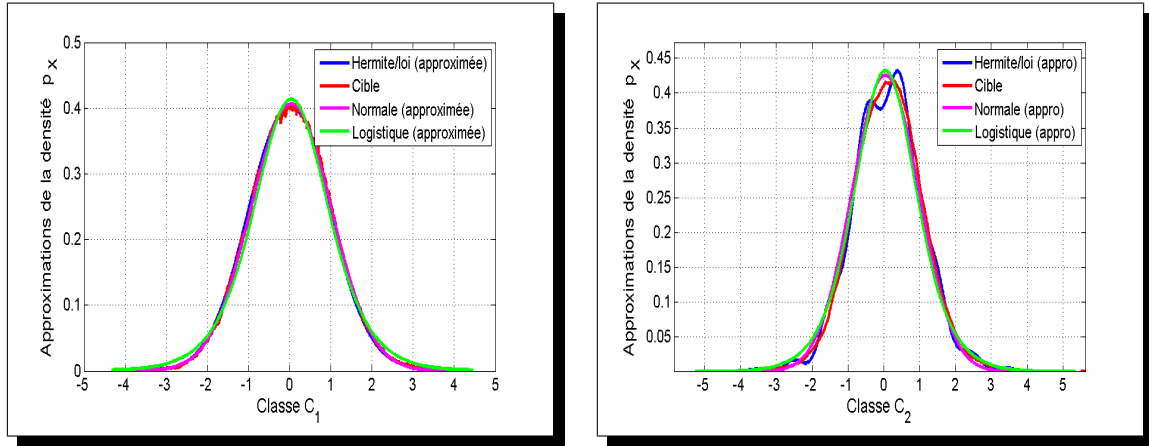


Figure III.24 – Approximations de p_X pour \mathcal{C}_1 et \mathcal{C}_2 selon le critère K_2

La partie gauche de la figure (III.24) montre une densité p_X cible bien approximée avec la méthode Hermite. Alors que les lois usuelles donnent des approximations moins précises surtout en queue de distribution. La précision des modèles dépend aussi bien de la tendance centrale que des queues de distribution. Par contre la partie droite montre une mauvaise approximation par la méthode Hermite. L'utilisation de cette méthode ne garantit pas une approximation de p_X positive surtout à un ordre de troncature α élevé du fait de la multiplicité des polynômes d'Hermite entrant en jeu.

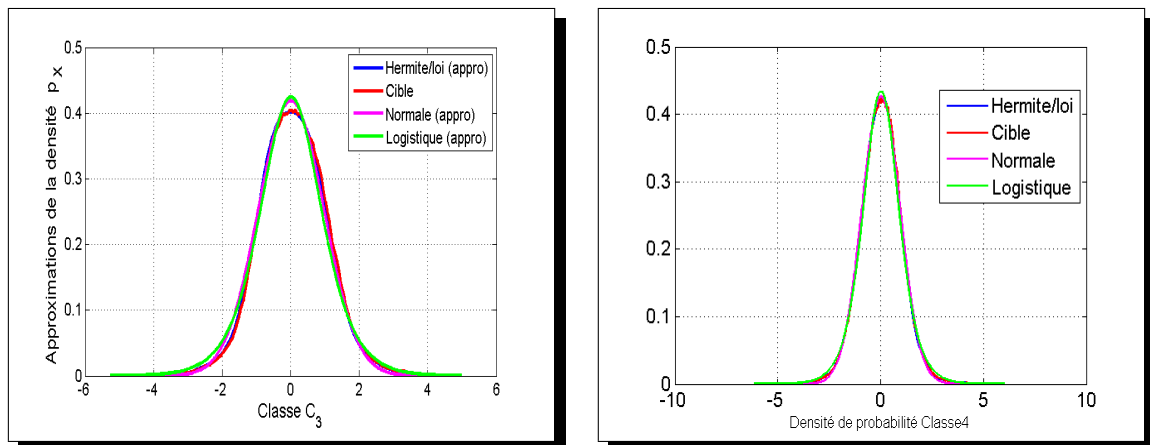


Figure III.25 – Approximations de p_X pour \mathcal{C}_3 et \mathcal{C}_4 selon le critère K_2

A travers les figures ci-dessus, on remarque que l'approximation par des lois usuelles est

plus précise pour une densité p_X fine que si elle comporte des fluctuations. L'estimée de la loi relative au critère K_1 est un exemple de densité avec fluctuations, c.f. figure (III.22).

III.5.1.3 Choix de la méthode d'approximation

Outre l'aspect graphique, l'*acceptation* ou le *rejet* d'une approximation se fait en utilisant les critères de proximité (I_1 et I_2) de la section (II.3.3.3) et des tests statistiques. Les tests d'ajustement utilisés sont des tests classiques inspirés des résultats de Glivenko et Kolmogorov, (cf. Saporta[88]). L'erreur de première espèce (ou p-value) est fixée à 5%.

Le tableau (III.3) montre une comparaison entre les méthodes d'approximation utilisées : *Hermite*, loi *normale* et loi *logistique*. La comparaison se fait par : les critères de proximité I_1 et I_2 , l'hypothèse nulle H_0 , la p-value du test, la moyenne et l'écart-type de chaque approximation. Les résultats concernent la classe \mathcal{C}_3 pour le critère de défaillance K_1 .

critères de validité	Hermite	Loi normale	Loi logistique
I_1	0.1656	0.46	0.51
I_2	0.01	0.03	0.04
H_0	0	1	1
$p - value$	0.75	12.10^{-10}	14.10^{-18}
μ	0.10	0.10	0.13
σ	0.14	0.12	0.17

Tableau III.3 – Comparaison des méthodes d'approximation de la loi marginale de p_X .

Les autres tableaux de comparaison relatifs au couple (classe \mathcal{C}_p /critère K_i) ne sont pas présentés car la démarche étant identique. Dans une très large majorité des cas traités, on a choisi le développement de la densité sur la base des polynômes d'Hermite. Car, le test ne rejette pas l'hypothèse nulle H_0 à 5% des cas environ. Ce qui confirme les valeurs faibles de I_1 et I_2 pour cette méthode par rapport à l'approximation par des lois usuelles.

Outre les critères de proximité (I_1 et I_2), le couple (μ, σ) le plus proche des caractéristiques (moyenne et écart-type) de la loi cible est encore la loi d'ajustement par la méthode Hermite. Pour la loi cible, on a $\mu = 0.10$ et $\sigma = 0.14$. Malgré les bons résultats de la méthode, il faut une certaine prudence. Il existe des cas particuliers où une loi classique peut être adaptée(cf. figure III.24). Nous cherchons la meilleure approximation car la précision des modèles dépend de l'erreur qu'on peut commettre sur ces approximations. Dans la prochaine étape, nous cherchons à estimer la *densité spectrale de puissance* S_X .

III.5. Modélisation stochastique du processus X

III.5.2 Densité spectrale de puissance S_X

Rappelons que le processus X défini sur $(\Omega, \mathfrak{S}, \mathbb{P})$ à valeurs dans \mathbb{R} est centré, stationnaire en moyenne d'ordre deux. Nous allons chercher d'une part une estimation de S_X et d'autre part, une approximation théorique de S_X si possible.

III.5.2.1 Estimées de la densité spectrale de puissance

Les densités spectrales de puissance S_X ont été estimées à l'aide de l'estimateur de Welch avec fenêtre de Hamming (cf. annexesV.3). Les figures (III.26) et (III.27) montrent les estimées de S_X relatives aux 4 classes \mathcal{C}_p de trajectoires et à chacun des 2 critères de défaillance (K_1 et K_2).

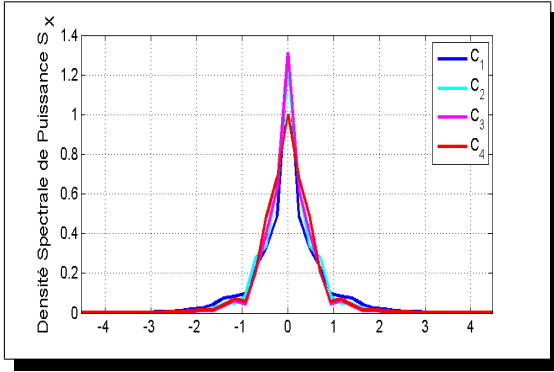


Figure III.26 – Graphes de S_X pour les 4 classes selon le processus de contrôle $D(t)$.

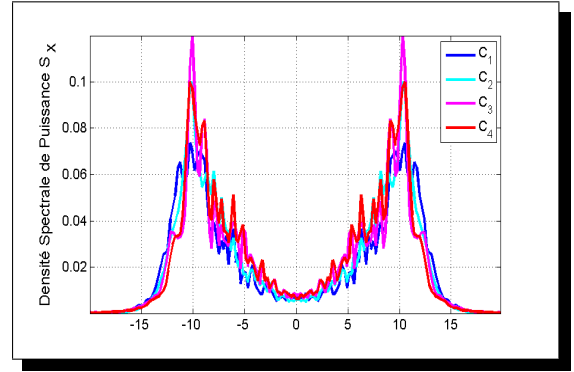


Figure III.27 – Graphes de S_X pour les 4 classes selon le processus de contrôle $\Gamma_N(t)$.

Nous distinguons des différences importantes au sens de l'énergie entre ces densités spectrales de puissance S_X . Ceci explique également la différence entre les classes $(\mathcal{C}_p)_{p=1, \dots, P}$ en terme de dynamique.

Avec les estimées p_X et S_X , il est possible de manière numérique de simuler le processus X en utilisant la méthode décrite dans la section (II.3.4). Il est également possible de chercher une *approximation de la densité spectrale* S_X sous la forme d'une fraction rationnelle.

III.5.2.2 Approximation des densités spectrales S_X

Essayons d'approximer les estimées de S_X par des DSP théoriques en utilisant une **approximation markovienne** de X . Afin de se placer sous les hypothèses relatives à la

Chapitre III. Application de la méthodologie sur le virage (LCPC/Nantes)

réalisation markovienne de dimension finie, soient R et Q de la forme suivante :

$$\begin{cases} R(x) &= (\alpha_1 + \beta_1 x)(\alpha_2 + \beta_2 x) \\ Q(x) &= (x^2 + 2\gamma_1 \delta_1 x + \delta_1^2)(x^2 + 2\gamma_2 \delta_2 x + \delta_2^2) \end{cases} \quad (\text{III.4})$$

où $(\alpha_1, \beta_1), (\alpha_2, \beta_2) \in \mathbb{R} - \{(0, 0)\}$ et $\gamma_1, \gamma_2, \delta_1, \delta_2$ sont des réels strictement positifs. On cherche alors à déterminer les 8 coefficients $(\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2, \delta_1, \delta_2)$ qui minimisent la quantité :

$$\sum_l \left(R_X - \sum_n n! f_2^n \left(\sum_k \left| \frac{R(i\omega_k)}{Q(i\omega_k)} \right|^2 e^{i\omega_k t_l} \right) \right)^2 \quad (\text{III.5})$$

Bien que ce problème de minimisation soit de petite taille (8 paramètres à déterminer), la minimisation n'en est pas pour autant plus rapide. Ceci est sans doute dû à deux points : premièrement, il n'est pas possible d'utiliser un algorithme de Transformation de Fourier Rapide, et en second lieu nous n'avons aucune idée de la valeur des coefficients des polynômes R et Q tandis que dans la méthode spectrale en partant de $\sigma_k = S(\omega_k)$, l'expérience montre que l'on est déjà près de la solution.

Remarquons que la classe fonctionnelle est très contrainte car les pôles des racines doivent être dans le demi plan gauche $\{z \in \mathbb{C} : \text{Re}(z) < 0\}$, (c.f. Poirion [75]).

Pour cette étude, la méthode d'approximation n'a pas donné de meilleur résultat surtout pour les S_X relatives au critère K_2 . Ceci doit être dû aux multiples pics que contiennent ces fonctions de densité spectrale.

Nous avons donc utilisé numériquement la densité spectrale de puissance S_X obtenue par les données de mesures expérimentales pour simuler le processus X .

A partir de ces simulations, il faut valider les modèles Z pour que les données issues de ces modèles puissent être utilisées dans la partie *analyse fiabiliste*. Nous rappelons que les modèles Z sont définis par la relation suivante :

$$Z(t) = \sigma_Z(t) \times X(t) + \mu_Z(t) \quad (\text{III.6})$$

où μ_Z et σ_Z sont des fonctions temporelles déterministes connues par estimation statistique. Le processus X vient d'être partiellement caractérisé et identifié. Avec Z , nous simulons l'état dynamique du système V-I-C *vis-à-vis* des 2 critères de défaillance K_1 et

III.5. Modélisation stochastique du processus X

K_2 selon chaque classe \mathcal{C}_p de trajectoires.

La validation des modèles Z est l'occasion également de vérifier l'hypothèse que toutes les trajectoires expérimentales d'une classe \mathcal{C}_p sont des réalisations d'un même processus stochastique U . La figure (III.28) représente le principe de validation des modèles Z .

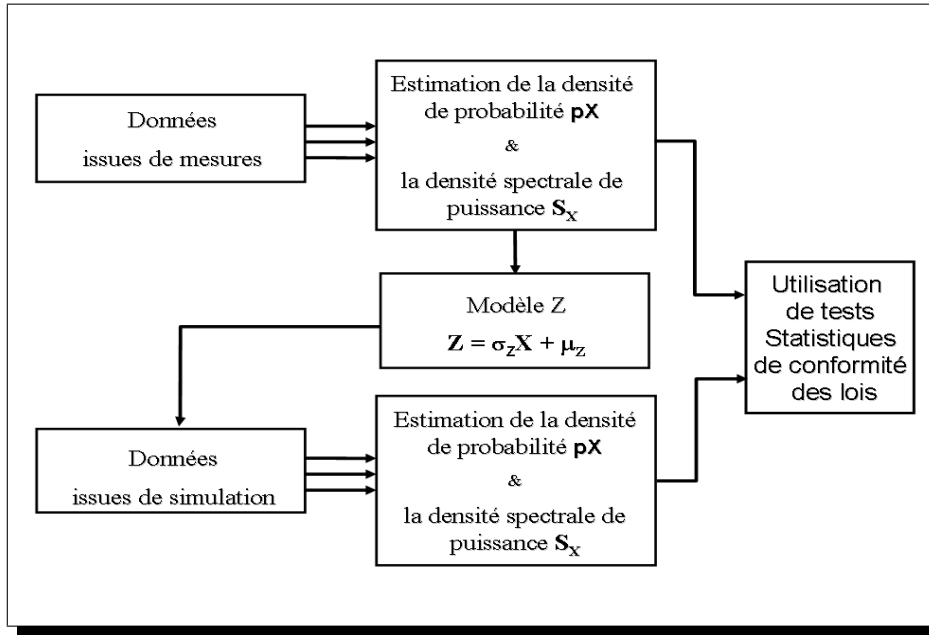


Figure III.28 – Principe de validation du modèle probabiliste Z .

Il est important et nécessaire d'avoir des modèles Z validés. Car les probabilités de défaillance que nous allons calculer dans la partie *analyse fiabiliste* sont très sensibles aux erreurs de modélisation.

Le prochain paragraphe présente cette étape de *validation des modèles* en comparant les lois issues des *données mesurées* à celles issues des *données simulées*.

III.5.3 Validation numérique des modèles probabilistes Z

Dans un premier temps, nous simulons 100000 réalisations de taille $N=2685$ du processus X en utilisant les estimées p_X et S_X selon la méthode développée par Poirion et Puig [75], cf. section (II.3.4). A partir de ces réalisations, nous estimons la loi marginale d'ordre 1 \tilde{p}_X et la densité spectrale de puissance \tilde{S}_X des données simulées.

Ensuite, nous proposons de comparer l'estimée de la loi marginale d'ordre 1 de p_X obtenue à partir de la famille d'observations à celle obtenue par les simulations du processus X .

Enfin, nous comparons les estimées des densités spectrales de puissance S_X issues des données mesurées à celles des données simulées.

III.5.3.1 Comparaison de lois marginales d'ordre 1 (*cibles et simulations*)

La validation des modèles Z de manière quantitative nécessite non seulement le calcul des critères de proximité I_1 et I_2 mais aussi l'utilisation de test statistique. Le test utilisé est celui de Kolmogorov-Smirnov pour une erreur de première espèce à 5%. Les résultats de ce test et les valeurs des critères I_1 et I_2 sont présentés dans le tableau (III.4) pour chaque classe \mathcal{C}_p selon le critère d'accélération transversale K_2 .

critères de validité	\mathcal{C}_1	\mathcal{C}_2	\mathcal{C}_3	\mathcal{C}_4
I_1	0.20	0.24	0.27	0.13
I_2	0.01	0.03	0.07	0.006
H_0	0	0	0	0
$p - value$	0.12	0.08	0.06	0.32

Tableau III.4 – Critères de validité des simulations à partir des modèles Z .

L'hypothèse nulle H_0 de conformité n'est pas rejetée pour la densité de probabilité p_X associée à chaque classe de trajectoires \mathcal{C}_p . Ceci explique les valeurs faibles des critères de proximité I_1 et I_2 . Ces résultats confirment la conformité des deux lois marginales d'ordre 1 de p_X obtenue par mesures et \tilde{p}_X obtenue par simulations.

Nous ne présentons pas les autres tableaux de comparaison car la démarche étant identique d'une part et les performances de la méthode sont comparables, d'autre part.

Outre la comparaison (classe \mathcal{C}_p /critère K_i) des résultats du test, on présente aussi une comparaison graphique des densités selon les critères de défaillance K_1 et K_2 .

III.5. Modélisation stochastique du processus X

Critère de défaillance K_1 :

Les figures (III.29) et (III.30) montrent la comparaison entre la loi marginale d'ordre 1 de p_X obtenue par estimation statistique des observations et celle obtenue par la simulation. Cela concerne les classes \mathcal{C}_1 et \mathcal{C}_2 , puis \mathcal{C}_3 et \mathcal{C}_4 selon le critère *distance relative* K_1 .

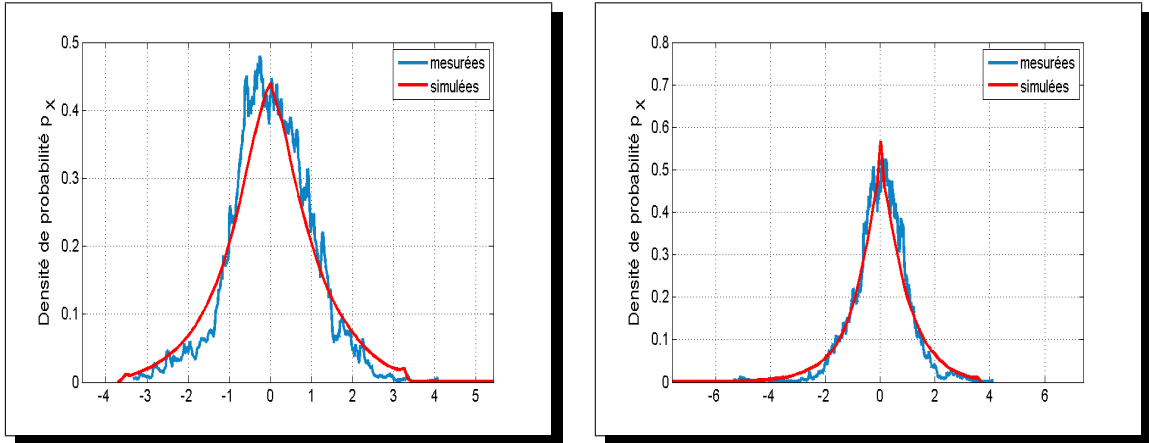


Figure III.29 – Comparaison des graphes de p_X pour \mathcal{C}_1 et \mathcal{C}_2 selon le critère K_1

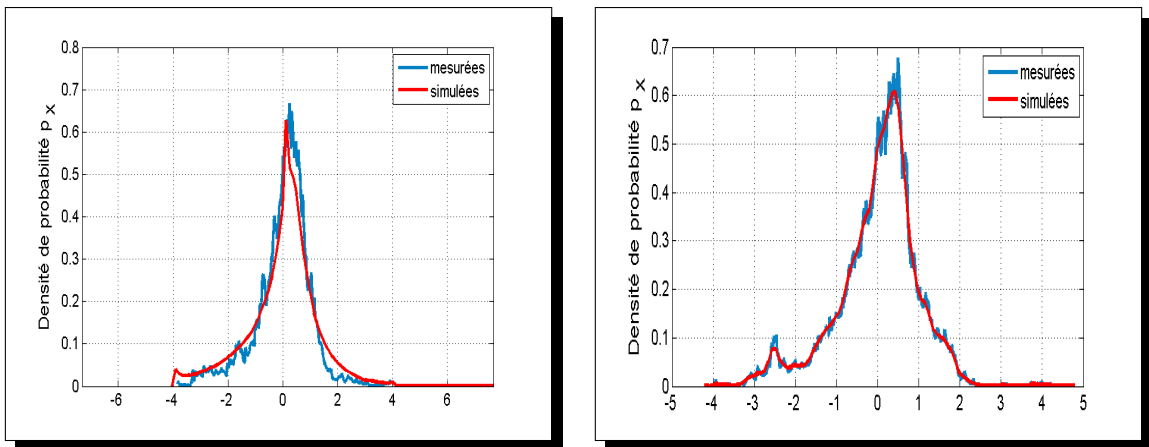


Figure III.30 – Comparaison des graphes de p_X pour \mathcal{C}_3 et \mathcal{C}_4 selon le critère K_1

L'hypothèse nulle H_0 de conformité n'est pas rejetée pour les densités relatives aux 4 classes de trajectoires. Ce résultat est confirmé par la représentation graphique. Nous constatons toute fois une meilleure superposition pour certaines classes que pour d'autre. Ceci doit être du à la représentativité de ces classes en trajectoires pratiquées. Justement une des faiblesses de cette méthode de modélisation reste bien entendu la représentativité des classes *vis-à-vis* de trajectoires pratiquées.

Critère de défaillance K_2 :

Les figures (III.31) et (III.32) montrent une comparaison des densités marginales d'ordre 1 pour le critère d'accélération latérale selon les classes \mathcal{C}_1 et \mathcal{C}_2 , puis \mathcal{C}_3 et \mathcal{C}_4 .

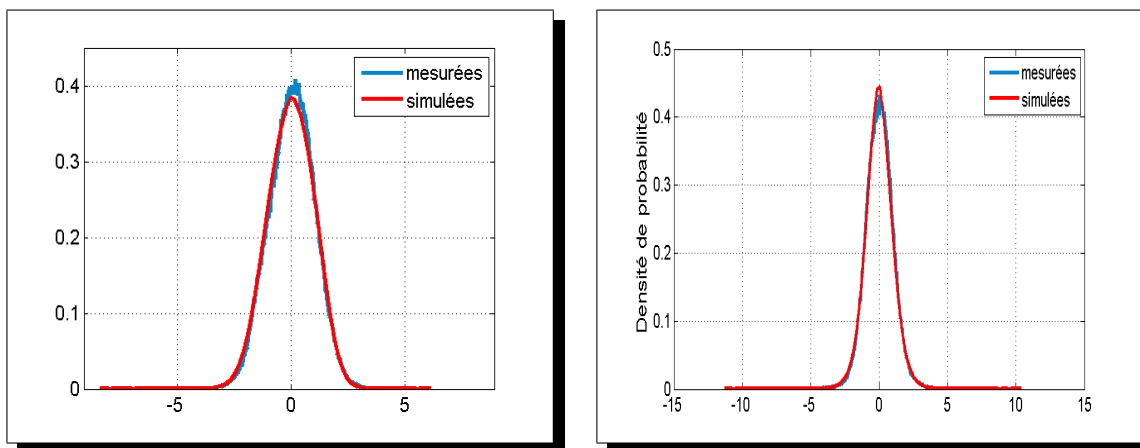


Figure III.31 – Comparaison des graphes de p_X pour \mathcal{C}_1 et \mathcal{C}_2 selon le critère K_2

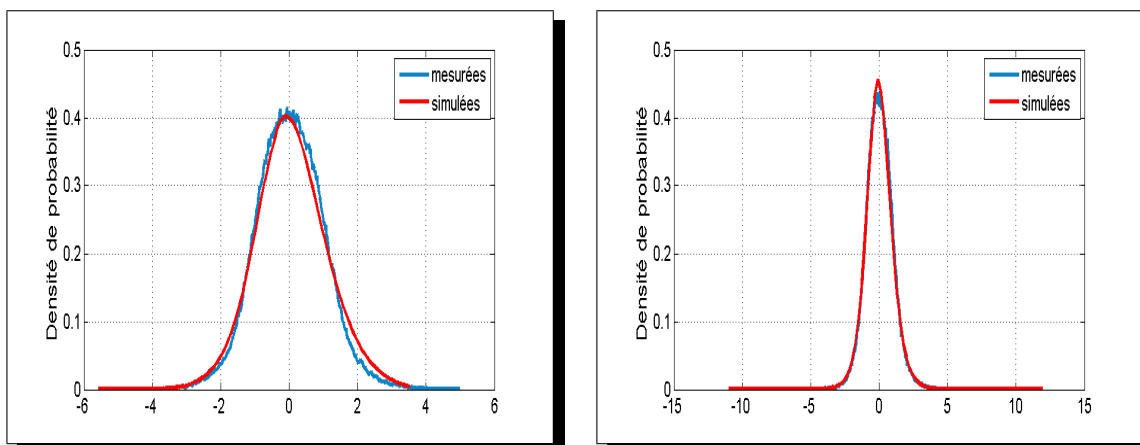


Figure III.32 – Comparaison des graphes de p_X pour \mathcal{C}_3 et \mathcal{C}_4 selon le critère K_2

La superposition des graphes confirme les résultats obtenus du tableau (III.4). En se référant à ce tableau, l'hypothèse nulle H_0 n'est pas rejetée pour les lois obtenues à partir des observations des 4 classes. Nous pouvons considérer les données simulées à partir des modèles Z comme des réalisations issues des 4 classes de trajectoires. Ce résultat vérifie aussi une hypothèse posée au début de l'étude qui consiste à supposer que toutes les trajectoires d'une classe \mathcal{C}_p sont issues d'un même processus stochastique U .

Nous allons présenter la comparaison entre les densités spectrales S_X des mesures à celles des simulations.

III.5. Modélisation stochastique du processus X

III.5.3.2 Comparaison des densités spectrales S_X (*cibles et simulations*)

Critère de défaillance K_1 :

Les figures (III.35) et (III.36) montrent la comparaison des graphes de S_X entre les mesures et les simulations pour chaque classe de trajectoires selon le critère de distance.

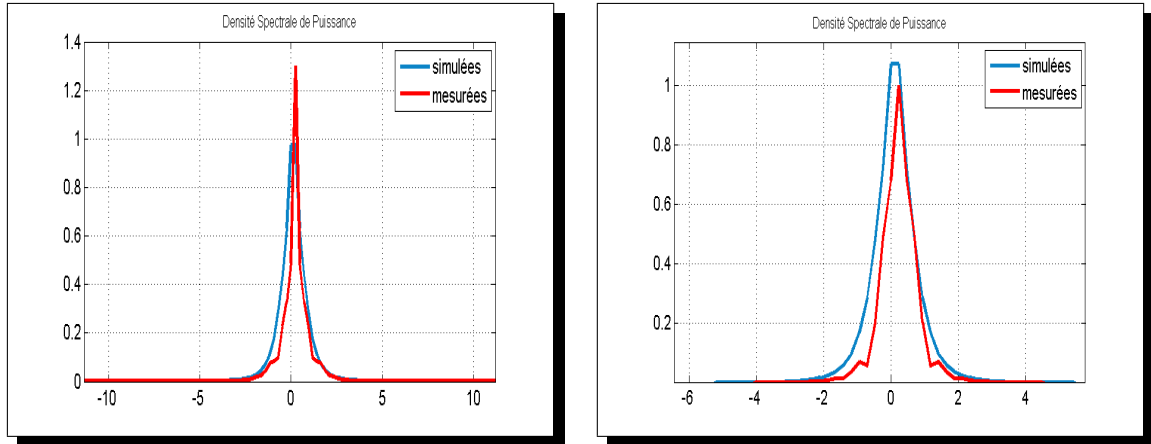


Figure III.33 – Comparaison des graphes de S_X pour C_1 et C_2 selon le critère K_1

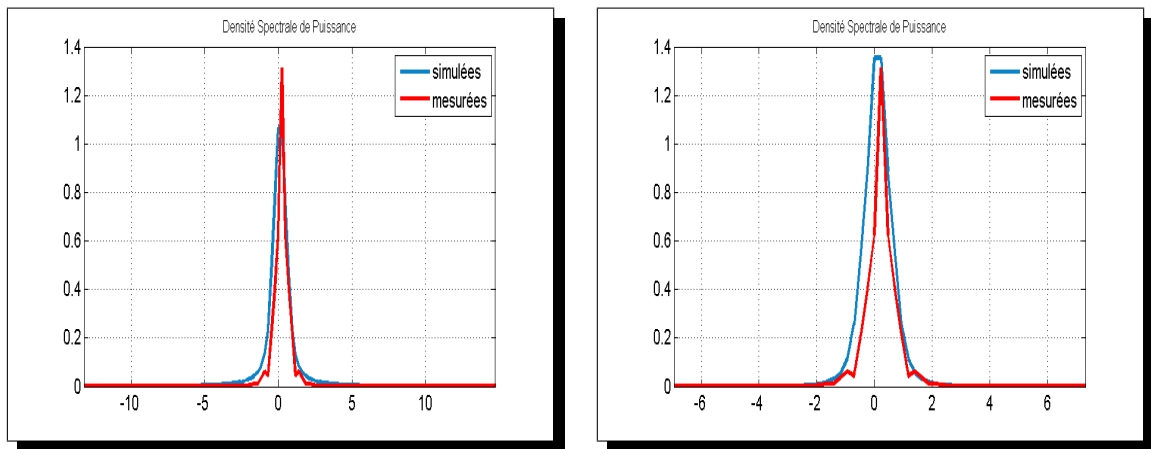


Figure III.34 – Comparaison des graphes de S_X pour C_3 et C_4 selon le critère K_1

Nous constatons une superposition des densités spectrales S_X entre les mesures et les simulations comme pour les densités de probabilité p_X . Cette superposition est encore plus précise pour les classes C_2 et C_4 que pour les classes C_1 et C_3 . Ceci est encore du à la représentativité de ces classes en trajectoires pratiquées. Plus les classes sont représentatives de trajectoires u pratiquées, plus les caractéristiques statistiques des mesures sont proches de celles des simulations.

Critère de défaillance K_2 :

Les figures (III.35) et (III.36) montrent la comparaison des graphes de S_X entre les données mesurées et les données simulées pour le critère d'*accélération transversale* et selon chaque classe \mathcal{C}_p .

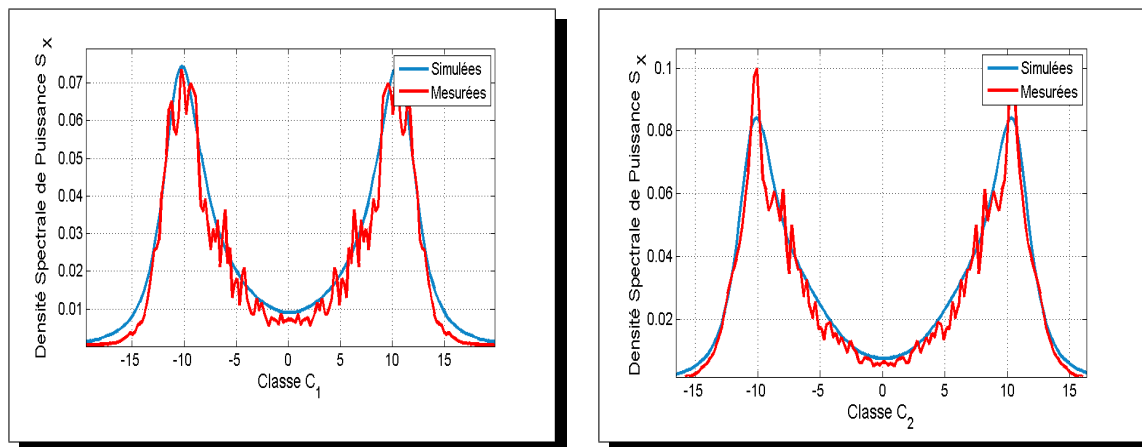


Figure III.35 – Comparaison des graphes de S_X pour \mathcal{C}_1 et \mathcal{C}_2 selon le critère K_2

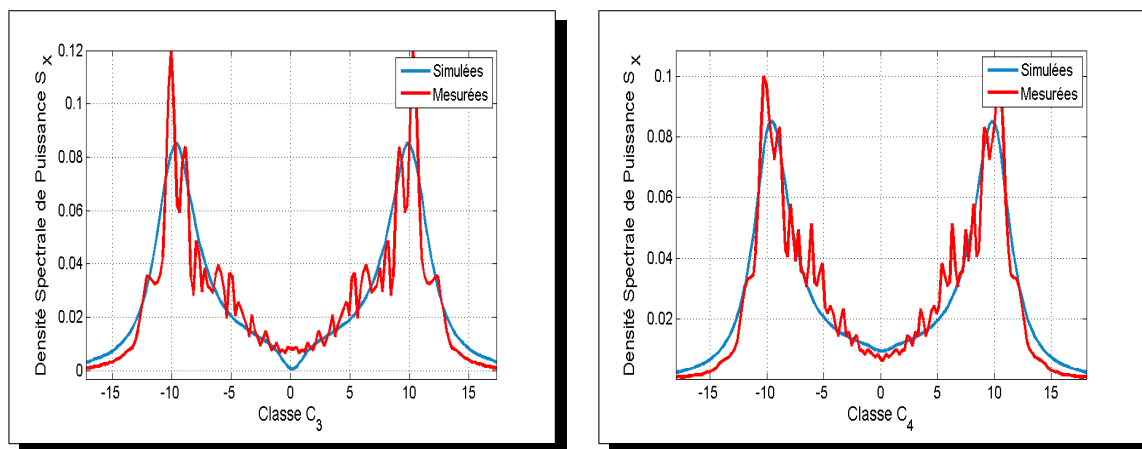


Figure III.36 – Comparaison des graphes de S_X pour \mathcal{C}_3 et \mathcal{C}_4 selon le critère K_2

On distingue une différence visible entre les mesures et les simulations pour la classe \mathcal{C}_3 . L'écart entre les graphes de S_X pour les classes \mathcal{C}_1 et \mathcal{C}_2 est plus faible que celui de la classe \mathcal{C}_3 . Par contre, la classe \mathcal{C}_4 donne une bonne superposition des densités spectrales S_X entre les mesures et les simulations. Ces résultats montrent l'importance de la représentativité des classes \mathcal{C}_p en trajectoires pratiquées. Notons que des pics existent pour les S_X obtenues par mesure alors que celles obtenues par simulation ne les considèrent pas. Cela ouvre une discussion entre le fait de considérer ces pics comme un *bruit de mesures* ou de

III.5. Modélisation stochastique du processus X

modélisation. Les résultats précédents et les tableaux (III.5, III.6) ont permis de clore cette discussion. Les pics des S_X issues des données mesurées sont des bruits de mesures.

III.5.3.3 Comparaison des moments d'ordre p (mesures et simulations)

Les tableaux (III.5) et (III.6) présentent la comparaison des moments d'ordre p entre les estimations statistiques issues de mesures et celles issues de simulations.

Classe	μ_X^1	μ_X^1	μ_X^2	μ_X^2	μ_X^3	μ_X^3	μ_X^4	μ_X^4
	mesure	simulation	mesure	simulation	mesure	simulation	mesure	simulation
\mathcal{C}_1	0	0.06	1	1.12	1.51	1.67	6.22	5.95
\mathcal{C}_2	0	0.05	1	1.14	2.15	1.98	10.79	9.13
\mathcal{C}_3	0	0.07	1	1.14	1.58	1.44	5.88	5.07
\mathcal{C}_4	0	0.11	1	1.20	1.94	1.66	8.49	6.54

Tableau III.5 – Comparaison des moments μ^p entre mesures et simulations pour K_1 .

Classe	μ_X^1	μ_X^1	μ_X^2	μ_X^2	μ_X^3	μ_X^3	μ_X^4	μ_X^4
	mesure	simulation	mesure	simulation	mesure	simulation	mesure	simulation
\mathcal{C}_1	0	0.0001	1	0.99	-0.03	-0.02	2.66	2.68
\mathcal{C}_2	0	0.0001	1	1.001	0.068	0.072	3.33	3.32
\mathcal{C}_3	0	-0.0002	1	1.001	-0.09	-0.08	2.97	2.77
\mathcal{C}_4	0	-0.0001	1	0.997	0.01	0.05	4.01	3.01

Tableau III.6 – Comparaison des moments μ^p entre mesures et simulations pour K_2 .

A travers ces tableaux, on note que les moments μ_X^p des données de mesures sont assez proches de ceux des données simulées. En conclusion, les performances du procédé de simulation du processus X sont acceptables.

La **prochaine section** consiste à utiliser les données simulées à partir des modèles Z pour calculer les probabilités de défaillance P_f .

III.6 Analyse fiabiliste du système V-I-C

L'objet de cette section est de construire des abaques de probabilité de défaillance par classe de trajectoires. Rappelons que ces abaques $\pi(\delta^*)$ sont spécifiques au virage étudié. Nous allons d'abord identifier la loi des maxima du processus Z . A partir de ces lois, nous estimerons les probabilités de défaillance P_f relatives à chaque classe \mathcal{C}_p et selon les critères de défaillance K_1 et K_2 . Enfin, nous ferons une étude de sensibilité de ces probabilités par rapport aux 4 classes \mathcal{C}_p et aux 2 critères de défaillance.

III.6.1 Identification de la loi de Y

Rappelons que la variable aléatoire Y obtenue par la relation (III.7) représente la loi de probabilité des maxima du processus de contrôle Z .

$$Y = \text{Sup}_{t \in T} |Z(t)| \quad (\text{III.7})$$

Il est équivalent de connaître la loi de probabilité de p_Y pour les maxima et celle de p_M pour la marge. Nous rappelons que la marge de sécurité est obtenue par :

$$M = \delta^* - Y \quad (\text{III.8})$$

En utilisant les simulations de Z et en fixant le seuil δ^* , on obtient les réalisations de la marge M . L'évènement $M \leq 0$ caractérise l'état défaillant tandis que l'évènement $M > 0$ caractérise l'état fiable. A partir des réalisations de M , nous avons estimé et approximé la loi p_M de M . Si on se réfère à la section (II.4.4), en général les lois de probabilité des valeurs extrêmes suivent les 3 lois suivantes : Gumbel, Frechet et Weibull.

Le test d'ajustement de Kolmogorov-Smirnov a été utilisé pour accepter ou rejeter l'adéquation de la densité p_M par rapport aux 3 lois de valeurs extrêmes. Pour la plupart des cas étudiés, la loi de frechet et la loi Weibull sont rejetées.

Nous allons d'abord présenter les résultats de ces tests statistiques. Pour chaque cas considéré, le test statistique a été utilisé pour une erreur de première espèce à 5%. Ensuite, nous présenterons également les résultats graphiques de ces ajustements pour les 4 classes \mathcal{C}_p et selon chaque critère de défaillance K_i . Les ajustements seront précis, car les probabilités de défaillance que nous allons estimer à partir des lois p_M sont faibles donc un mauvais ajustement implique des erreurs d'estimation considérables.

III.6. Analyse fiabiliste du système V-I-C

III.6.1.1 Résultats des tests statistiques d'ajustement de loi

Nous présentons les résultats des tests statistiques d'ajustement de la loi de p_M par les lois de valeurs extrêmes telle que la loi de Gumbel pour les critères de *distance* K_1 et d'*accélération transversale* K_2 .

Critère de défaillance K_1 :

critères de validité	\mathcal{C}_1	\mathcal{C}_2	\mathcal{C}_3	\mathcal{C}_4
Hypothèse nulle H_0	0	0	0	0
$p - value$	0.40	0.3	0.13	0.11

Tableau III.7 – Validation des ajustements de la loi p_M des maxima de Z .

L'hypothèse nulle H_0 n'est pas rejetée pour les 4 classes \mathcal{C}_p . La loi de p_M est approximable par la loi de Gumbel. Le calcul direct est donc possible pour estimer la probabilité de défaillance P_f . La qualité de l'estimation de la densité p_M permet également d'estimer P_f par la méthode Monte carlo ou par un schéma d'intégration numérique classique.

Critère de défaillance K_2 :

critères de validité	\mathcal{C}_1	\mathcal{C}_2	\mathcal{C}_3	\mathcal{C}_4
Hypothèse nulle H_0	0	0	0	0
$p - value$	0.10	0.52	0.08	0.166

Tableau III.8 – Critères de validité des ajustements de la loi des max.

Le test statistique ne rejette pas l'hypothèse nulle H_0 pour les 4 classes. La loi de p_M est donc approximable par la loi de Gumbel. On peut calculer directement la probabilité de défaillance P_f pour ce critère car on connaît la fonction de répartition de la loi de Gumbel, cf. (II.4.5). Ce résultat est essentiel car il nous permet d'extrapoler nos observations pour déduire les quantiles extrêmes.

Nous tenons à préciser que la qualité de l'ajustement de la loi p_M dépend également du niveau de validation des modèles Z . Les résultats du test montre le bon fonctionnement des étapes précédentes de la méthodologie proposée.

Critère de défaillance K_1 :

Outre les résultats du test, les figures (III.37) et (III.38) montrent l'estimée de la loi des maxima de p_M et une approximation de cette loi par la loi de Gumbel pour les 4 classes selon le critère K_1 .

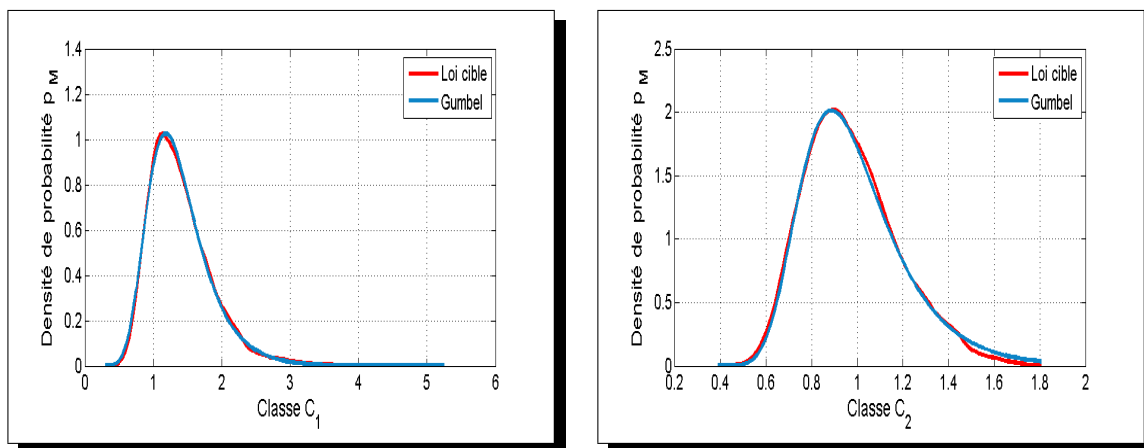


Figure III.37 – Approximation de la loi de p_M pour C_1 et C_2 selon le critère K_1

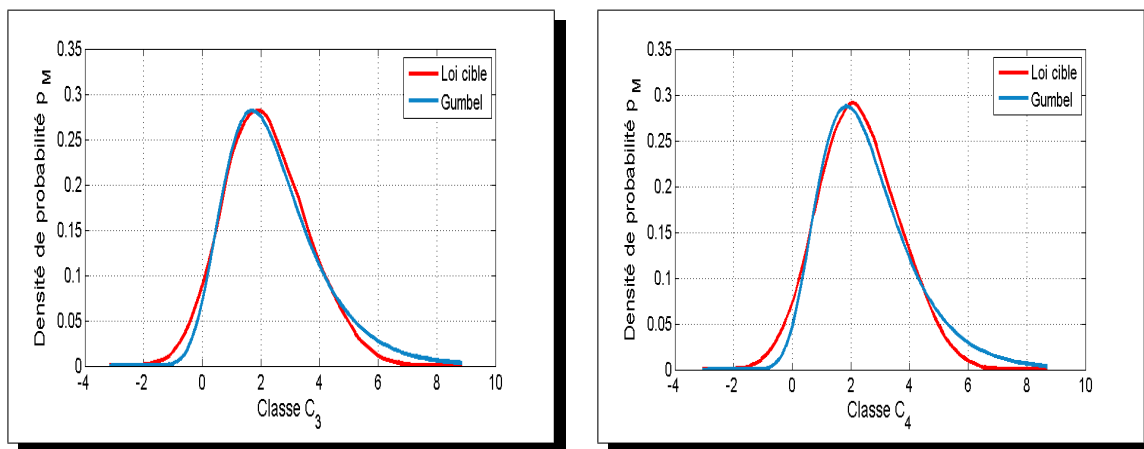


Figure III.38 – Approximation de la loi de p_M pour C_3 et C_4 selon le critère K_1

Après l'acceptation de l'adéquation par le test statistique, la comparaison graphique montre également que la *loi cible* peut être approximée par la loi de Gumbel.

Nous remarquons que pour le critère K_1 correspondant à la « distance relative entre trajectoires », en fixant le seuil δ^* à $1,5m$ du centre de la voie (C_V), seules les classes C_3 et C_4 contiennent des trajectoires défaillants.

III.6. Analyse fiabiliste du système V-I-C

Critère de défaillance K_2 :

Les figures (III.39) et (III.40) montrent l'approximation de la loi de p_M pour les classes C_1 et C_2 , puis C_3 et C_4 selon le critère d'accélération transversale.

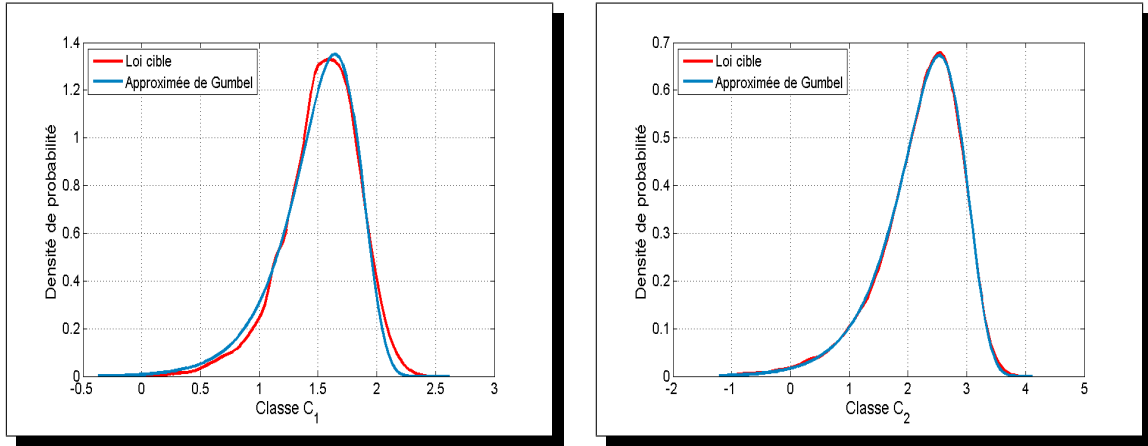


Figure III.39 – Approximation de la loi de p_M pour C_1 et C_2 pour le critère K_2

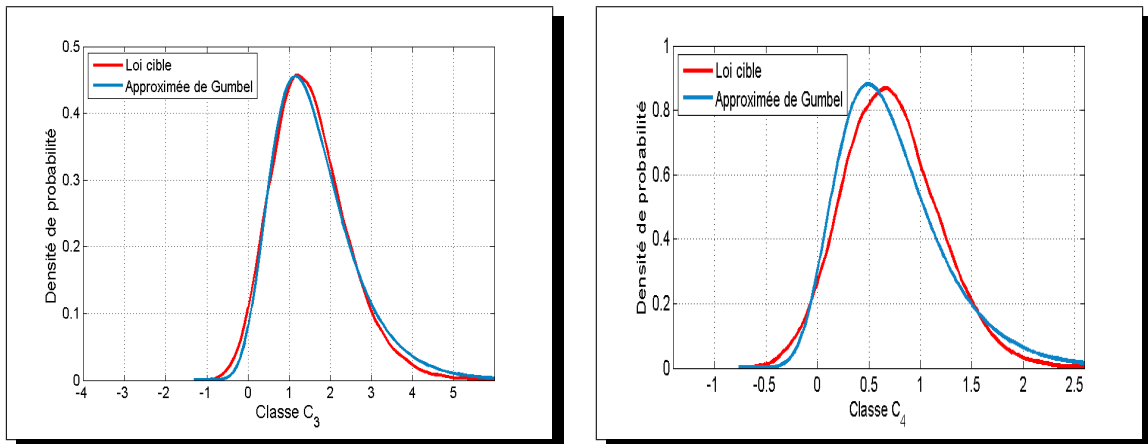


Figure III.40 – Approximation de la loi de p_M pour C_3 et C_4 selon le critère K_2

En général, la loi des extrêmes d'un processus gaussien est une loi de Gumbel. Les approximations \tilde{p}_M des figures (III.39) et (III.40) correspondent à des lois de Gumbel. Ce qui confirme l'aspect d'une distribution normale que nous avons constaté pour le processus X relatif au critère d'accélération latérale. Les allures proches entre la loi cible et la loi de Gumbel confirment les résultats du test statistique. Malgré la possibilité d'obtenir une bonne approximation avec la méthode Hermite, l'inconvénient se trouve dans la robustesse (cf. II.3.3.3). Les densités de probabilité p_M identifiées servent à estimer la probabilité de défaillance P_f et l'indice de fiabilité conventionnel β_c associé.

III.6.2 Estimation de la probabilité de défaillance P_f

Rappelons qu'en fixant $\delta^* \in \mathbb{R}_+^*$, on peut estimer la probabilité de défaillance P_f par :

$$P_f = P(\delta^* \leq Y) = P(M \leq 0) = \int_{-\infty}^0 p_M(x) dx \quad (\text{III.9})$$

Analysons l'influence de la variation du seuil (δ^*) de sécurité sur le calcul de la **probabilité de défaillance** P_f et de l'**indice de fiabilité** β_c relatifs aux 4 classes de trajectoires selon chacun des critères de défaillance K_1 et K_2 . On remarque bien entendu que plus le seuil de sécurité δ^* augmente, plus la probabilité de défaillance P_f diminue.

En fonction de ce résultat, plusieurs questions sont posées. Entre autres, est-ce que les niveaux de risque sont identiques pour les 4 classes identifiées ? Existe-t'il une zone *linéaire* et *non linéaire* pour une classe \mathcal{C}_p donnée en faisant varier P_f en fonction du seuil δ^* ? Les critères K_1 et K_2 ont ils la même sensibilité par rapport à une classe \mathcal{C}_p donnée ? Nous apporterons des réponses précises à toutes ces interrogations.

III.6.2.1 Probabilités de défaillance en fonction du seuil de sécurité δ^*

Les figures (III.41) et (III.42) représentent la variation de P_f en fonction de δ^* selon les critères K_1 et K_2 . On constate que la probabilité P_f décroît rapidement avec l'accroissement de la valeur du seuil δ^* pour les classes \mathcal{C}_1 et \mathcal{C}_2 (**conduite apaisée**). Tandis qu'elle décroît moins vite pour \mathcal{C}_3 et \mathcal{C}_4 (**conduite rapide**).

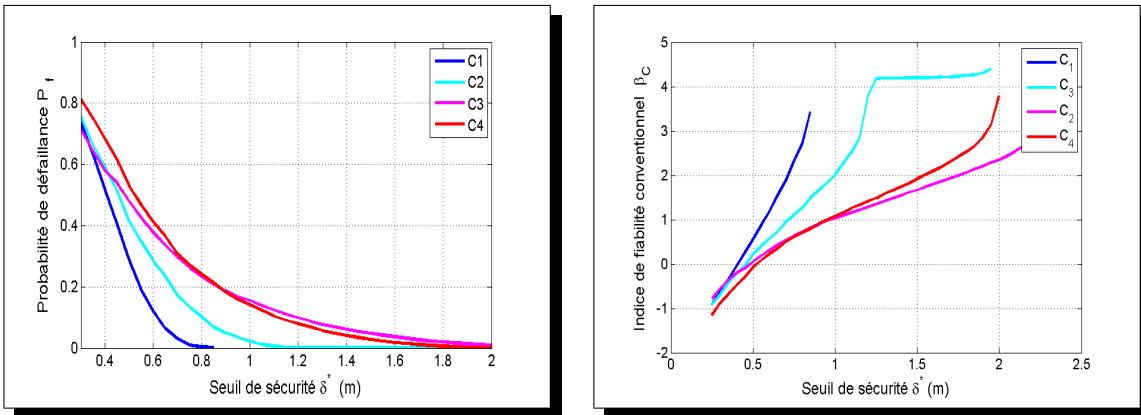


Figure III.41 – Graphes de $\delta^* \rightarrow P_f(\delta^*)$ et $\delta^* \rightarrow \beta_c(\delta^*)$ pour le critère K_1 .

Les conducteurs de \mathcal{C}_1 sont ceux qui se rapprochent le plus du centre de la voie (C_V). Tandis que les conducteurs de \mathcal{C}_3 et \mathcal{C}_4 sont ceux qui s'éloignent le plus de C_V . Pendant

III.6. Analyse fiabiliste du système V-I-C

l'expérimentation, des conducteurs de \mathcal{C}_4 ont emprunté la voie de gauche afin de minimiser encore plus le temps de parcours. Notons que plus la trace du véhicule se rapproche du bord de voie, plus il y a de chance de sortir de la voie.

Nous allons vérifier une hypothèse utilisée dans la littérature qui considère l'accélération latérale limite $\bar{\Gamma}_N$ à $3m/s^2$ pour une trajectoire sécurisée en virage. Des études récentes [24] considèrent la limite de la zone de fonctionnement linéaire du pneumatique à $4m/s^2$. La figure (III.42) montre une rupture géométrique pour la classe \mathcal{C}_2 à $3.5m/s^2$.

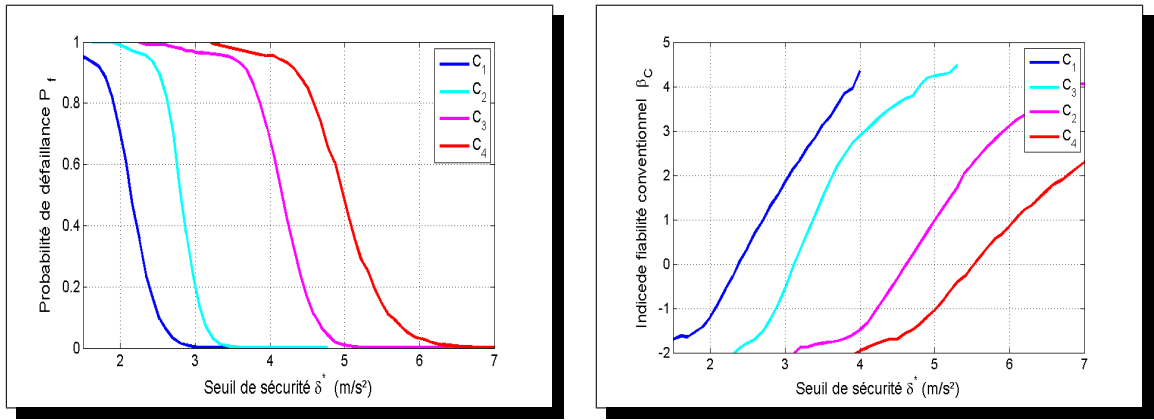


Figure III.42 – Graphes de $\delta^* \rightarrow P_f(\delta^*)$ et $\delta^* \rightarrow \beta_c(\delta^*)$ pour le critère K_2 .

Les 2 classes de la **conduite apaisée** ont une rupture géométrique inférieure à cette valeur. Par contre, à la section(I.2.3), nous avons montré qu'une accélération latérale supérieure à $6m/s^2$ est considérée comme dynamiquement instable. On constate que 10% des trajectoires de la classe \mathcal{C}_4 dépasse cette valeur donc elles sont instables.

Nous concluons sur cette partie que le niveau de défaillance est différent d'une classe de trajectoires à une autre. De la classe \mathcal{C}_1 à \mathcal{C}_4 , la probabilité augmente de la forme :

$$\mathcal{C}_1 \prec \mathcal{C}_2 \prec \mathcal{C}_3 \prec \mathcal{C}_4 ; \forall K_i \text{ avec } i = 1, 2 \quad (\text{III.10})$$

Nous allons montrer l'effet des critères de défaillance sur l'estimation de P_f par classe \mathcal{C}_p .

III.6.2.2 Etude de sensibilité en fonction des critères de défaillance

Une analyse de sensibilité classique cherche à calculer les paramètres statistiques des variables de sortie en fonction des paramètres statistiques des données d'entrée. Dans cette étude, on cherche à caractériser la sensibilité de la probabilité de défaillance $\frac{\partial P_f}{\partial K_i}$ d'une

Chapitre III. Application de la méthodologie sur le virage (LCPC/Nantes)

classe \mathcal{C}_p par rapport aux critères $(K_i)_{(i=1,2)}$. Pour comparer les critères sur une même échelle, nous avons utilisé la normalisation « min-max » sur chaque critère de défaillance. Le seuil δ^* *normalisé* varie donc dans $[0, 1]$. On a fait l'hypothèse au chapitre II que les 2 critères sont *pertinents* pour analyser la défaillance du système V-I-C. Pour vérifier cette hypothèse, nous rajoutons un 3^{ème} critère lié à l'orientation (ψ) du véhicule sur la route. Ce critère supplémentaire noté K_3 est défini par :

$$Y = \sup_{t \in T} |\psi(t)| > \bar{\psi} \quad (\text{III.11})$$

où $\bar{\psi} = \delta^*$ est une valeur seuil de l'orientation du véhicule à choisir dans un intervalle $[\psi_{min}, \psi_{max}]$. Les figures (III.43) à (III.46) montrent la comparaison des fonctions $\pi(\delta^*)$. Ces fonctions représentent la variation de P_f en fonction du seuil δ^* normalisé pour les 3 critères K_1 , K_2 et K_3 selon la classe \mathcal{C}_1 , puis \mathcal{C}_2 , \mathcal{C}_3 , et \mathcal{C}_4 .

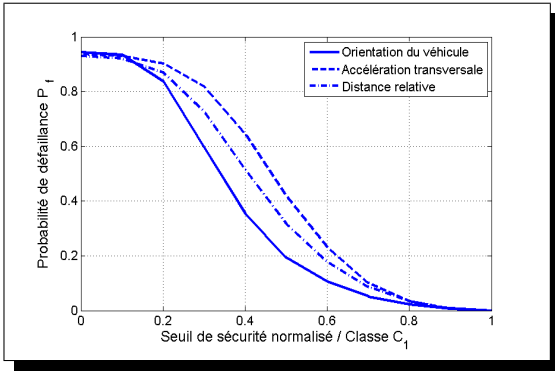


Figure III.43 – Evolution de P_f en fonction de δ^* selon la classe \mathcal{C}_1 .

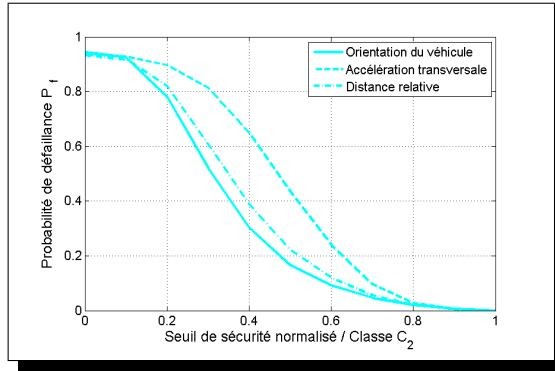


Figure III.44 – Evolution de P_f en fonction de δ^* selon la classe \mathcal{C}_2 .

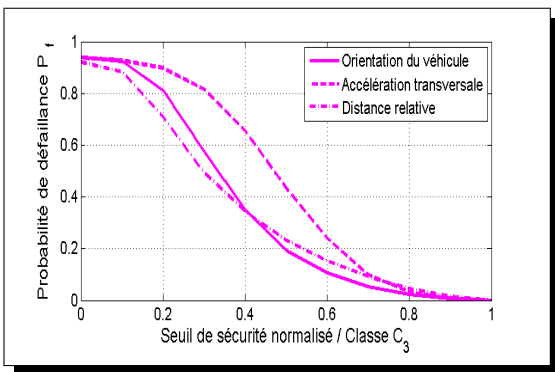


Figure III.45 – Evolution de P_f en fonction de δ^* selon la classe \mathcal{C}_3 .

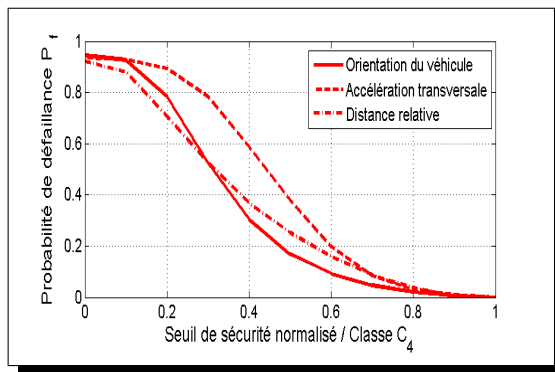


Figure III.46 – Evolution de P_f en fonction de δ^* selon la classe \mathcal{C}_4 .

III.6. Analyse fiabiliste du système V-I-C

On distingue que les critères n'ont pas la même variabilité. L'écart le plus important entre les critères se situe dans la classe \mathcal{C}_2 entre le critère d'*accélération transversale* et le critère d'*orientation du véhicule*. Ce résultat correspond à une situation de conduite où les conducteurs de cette classe génèrent une accélération transversale forte en minimisant l'orientation du véhicule par rapport à l'axe de la voie de circulation.

On constate également que les critères $(K_i)_{i=1,\dots,3}$ sont non seulement hiérarchisés *vis-à-vis* de P_f mais ils présentent aussi des corrélations surtout pour K_1 et K_3 selon les classes \mathcal{C}_3 et \mathcal{C}_2 . On en déduit la relation d'ordre suivante :

$$K_3 \prec K_1 \prec K_2 \quad (\text{III.12})$$

Nous concluons que le critère d'*orientation* K_3 a moins d'influence sur la probabilité de défaillance P_f que les deux autres. Ce qui implique la pertinence des critères de défaillance K_1 et K_2 . La relation (III.12) permet de mieux distinguer les 4 classes *vis-à-vis* de P_f .

Ce résultat est considéré comme une étude de sensibilité de P_f d'une classe \mathcal{C}_p par rapport à l'effet des 3 critères de défaillance. Nous retenons que le critère d'*accélération transversale* est le plus pertinent pour évaluer la défaillance du système V-I-C. Par contre, le critère d'*orientation du véhicule* est le moins pertinent que les critères de *distance* et d'*accélération transversale* pour cette application.

Le reproche fait à ce résultat est de ne pas calculer la probabilité de l'interaction entre les critères K_1 et K_2 . Pour calculer une telle probabilité, il faut simuler ensemble les processus de contrôle D et Γ_N . Nous avons proposé comme perspective d'utiliser la méthode des copules pour simuler un processus stochastique vectoriel à valeurs dans \mathbb{R}^3 .

Sinon, pour répondre au reproche, nous avons proposé dans la méthodologie au chapitre II, de calculer la probabilité de défaillance en introduisant le *critère du Sup*.

III.6.2.3 Probabilités de défaillance en utilisant le critère du Sup

Nous faisons l'hypothèse d'étudier la défaillance du système V-I-C en utilisant l'approche de la fiabilité des composants en série. Cela signifie dès qu'un critère K_i n'est pas respecté le système devient défaillant. Pour cela, on a calculé la probabilité de l'union P_{f-sys} .

$$P_{f-sys} = Prob(\cup E_i), \quad i = 1, 2 \quad (\text{III.13})$$

Les figures (III.47) et (III.48) représentent la comparaison des probabilités de défaillance selon les critères K_1 , K_2 et le critère du Sup.

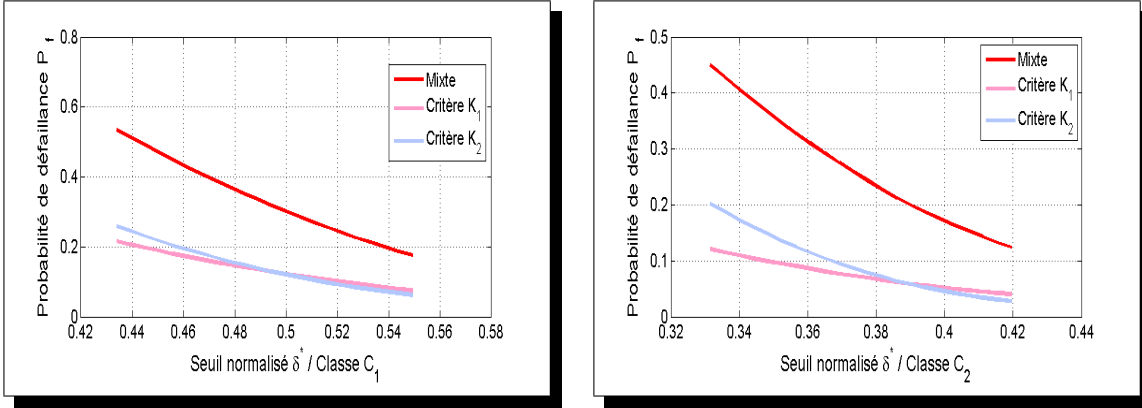


Figure III.47 – Probabilité de défaillance avec critère du Sup pour C_1 et C_2

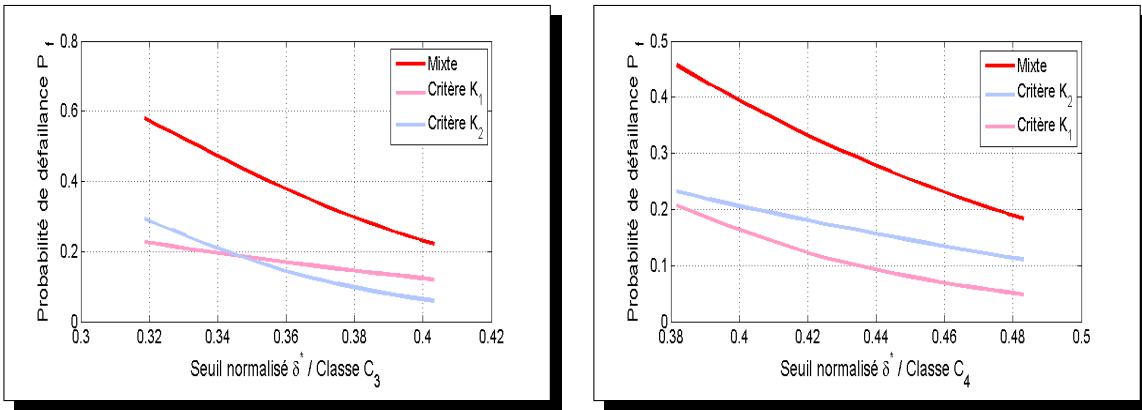


Figure III.48 – Probabilité de défaillance avec critère du Sup pour C_3 et C_4

Le critère du Sup est donc plus contraignant que chacun des critères K_1 et K_2 pris séparément. Ceci permet de mieux gérer les fausses alertes et la non détection.

Nous venons d'associer une probabilité P_f à chaque classe C_p de trajectoires. Pour cette configuration de virage, en fixant δ^* , on connaît la probabilité P_f par classe C_p . Ceci correspond à la construction d'abaques de probabilité de défaillance. En fonction de ces abaques $\pi(\delta^*)$, on peut faire aussi du diagnostic d'infrastructure. Par exemple, pour un δ^* donné, une valeur élevée de P_f pour toutes les classes signifierait une infrastructure inadaptée à la circulation. En utilisant $\pi(\delta^*)$, pour avoir la probabilité P_f d'une trajectoire u en cours de réalisation, il suffit de reconnaître sa classe d'appartenance C_p .

III.7 Reconnaissance des trajectoires u observées

L'objet de cette section est de présenter les résultats de la méthode de reconnaissance des trajectoires. Nous avons utilisé une technique d'apprentissage de type Support Vector Machine. Cette technique s'applique sur les trajectoires expérimentales issues du virage étudié. Dans un *premier temps*, nous montrons les performances du modèle \mathcal{M} de reconnaissance en fonction de la zone d'apprentissage considérée. Dans un *second temps*, nous présenterons les résultats de la prédiction de la classe d'appartenance \mathcal{C}_p de chaque trajectoire u en entrée de virage.

III.7.1 Identification de la zone d'apprentissage \mathcal{S}

Comme nous l'avons souligné au chapitre I, l'erreur de généralisation (c'est-à-dire la qualité du modèle) dépend d'un certain nombre de facteurs. Parmi ces facteurs, on trouve notamment le nombre N d'échantillons de la base d'apprentissage \mathcal{S} . Il est intéressant de faire varier ce paramètre ainsi que la proportion des différentes populations afin d'illustrer leurs impacts sur l'erreur de généralisation.

A partir des trajectoires expérimentales, nous avons identifié la base d'apprentissage (cf. figure III.49) nécessaire pour construire le modèle \mathcal{M} . Pour cela, l'algorithme décrit au paragraphe (II.5.1) a été utilisé. Le choix de cette base d'apprentissage se justifie dans la figure (III.50), où nous comparons la performance du modèle de reconnaissance pour 3 zones d'apprentissage de tailles différentes avec $k = \{400, 800, 1000\}$ observations.

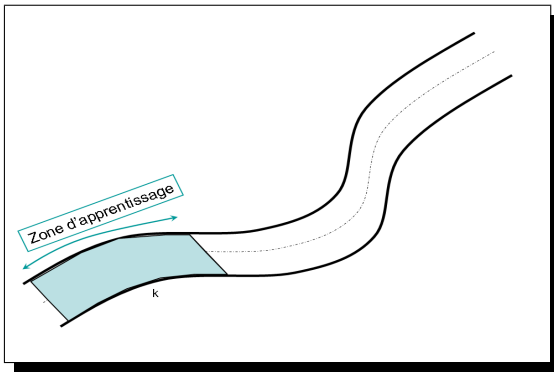


Figure III.49 – Zone d'apprentissage du virage.

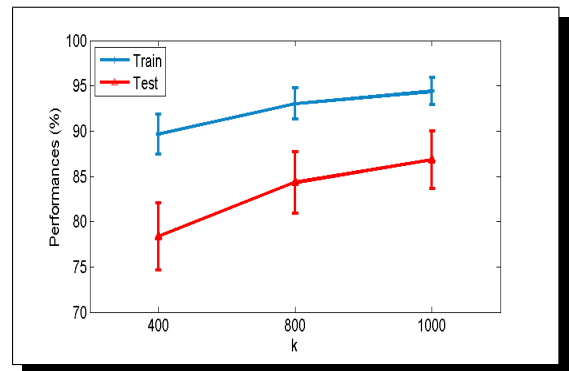


Figure III.50 – Choix de la base d'apprentissage optimale du virage.

Sur cette figure, nous donnons la moyenne μ et l'écart-type σ du taux d'apprentissage (train) et du taux de reconnaissance (test) pour chaque zone. Pour cette configuration

de virage et en fonction des trajectoires u observées, la zone d'apprentissage optimale correspond à 25% du virage. En utilisant cette zone comme base d'apprentissage \mathcal{S} , le conducteur dispose 75% du temps pour corriger sa trajectoire au cas où elle serait dangereuse.

III.7.2 Résultats de l'affectation des trajectoires

A partir de la base d'apprentissage précédemment identifiée, nous utilisons la méthode Bootstrap consistant à diviser l'échantillon des 232 trajectoires en deux sous ensembles (70% pour l'apprentissage et 30% pour le test). Ensuite, nous avons implémenté l'algorithme multi-classes des SVMs. Le résultat de la prédiction des classes d'appartenance \mathcal{C}_p pour les 30% de trajectoires de la base de test se trouve dans le tableau (III.9).

%	$\hat{\mathcal{C}}_1$	$\hat{\mathcal{C}}_2$	$\hat{\mathcal{C}}_3$	$\hat{\mathcal{C}}_4$
\mathcal{C}_1	79.2730	0	20.6497	0
\mathcal{C}_2	0	89.6662	17.5617	12.7721
\mathcal{C}_3	9.8303	10.2295	78.8922	1.0479
\mathcal{C}_4	0.0773	7.4839	0.3871	92.1290

Tableau III.9 – Validation du modèle Multi-class (SVM) par une matrice de confusion.

On remarque que les trajectoires u de la classe \mathcal{C}_4 sont les mieux affectées avec un taux de reconnaissance de 92.12%. Ce résultat a été confirmé tout au long de l'étude à travers l'estimation des lois de probabilité ou la validation des modèles Z . Notons que c'est une classe qui est bien représentée en trajectoires pratiquées. D'où la nécessité de la représentativité des classes que nous avons souligné tout le long de ce travail de thèse.

Nous avons 7.4% des trajectoires de \mathcal{C}_4 qui sont affectées dans la classe \mathcal{C}_3 . Ces 2 classes sont obtenues avec la même consigne « conduite rapide ». Il n'y a que 0.4% de trajectoires qui sont affectées aux classes \mathcal{C}_1 et \mathcal{C}_2 « conduite apaisée ». Ce résultat montre également la bonne performance du modèle \mathcal{M} .

Malgré que la classe \mathcal{C}_3 soit supérieure en nombre de trajectoires par rapport à la classe \mathcal{C}_1 , la représentativité de cette dernière fait que son taux de reconnaissance est supérieur à celui de la classe \mathcal{C}_3 . Certes, le nombre de trajectoires u dans une classe \mathcal{C}_p est important mais ce n'est pas le seul paramètre à considérer.

III.7. Reconnaissance des trajectoires u observées

Nous avons remarqué aussi que les trajectoires de la classe \mathcal{C}_3 sont difficiles à identifier par le modèle de reconnaissance par rapport aux autres classes. C'est la classe qui a le plus petit taux de reconnaissance (78.89%) et le plus grand taux d'erreur (37%).

Globalement, cette approche de reconnaissance de trajectoire conduit à un taux d'erreur (i.e. mauvaise prédiction de la classe d'appartenance) inférieur à 8% pour une classe \mathcal{C}_p représentative.

L'originalité de cette méthode de reconnaissance réside non seulement dans l'utilisation d'une base d'apprentissage d'observations partielles mais aussi dans l'utilisation des coordonnées pertinentes de la trajectoire u telles que l'équation I.11. Cette démarche permet de tenir compte de toute la dynamique qui existe pour décrire le système V-I-C.

Construction de l'indicateur de risque

Les résultats de cette application ont permis de créer des abaques $\pi(\delta^*)$ de risque de défaillance des trajectoires u pour cette configuration de virage. Les classes de conséquences associées aux classes \mathcal{C}_p de trajectoires sont :

- La classe \mathcal{C}_1 : une conduite lente, hésitante, débutante
- La classe \mathcal{C}_2 : une conduite sécurisée, prudente, fiable
- La classe \mathcal{C}_3 : une conduite performante, faible risque,
- La classe \mathcal{C}_4 : une conduite rapide, risque élevé,

A partir de la prédiction de la classe d'appartenance de chaque trajectoire u , on connaît la probabilité P_f à travers les abaques $\pi(\delta^*)$. La connaissance de P_f et la mesure des coordonnées de u a permis de calculer l'indicateur de risque I_{risque} .

$$\mathbf{Risque} = P_f \times (\text{gravité}) \quad (\text{III.14})$$

Rappelons que la **gravité** est calculée en fonction du critère de défaillance K_i considéré. En terme d'illustration, pour le critère d'*accélération transversale*, nous avons normalisé l'accélération latérale de la trajectoire en cours de réalisation par l'accélération latérale obtenue avec le v_{85} des trajectoires observées. Le v_{85} représente la vitesse longitudinale au dessous de laquelle roulent 85% des conducteurs (ce qui permet d'exclure les vitesses considérées comme atypiques et extrêmes), c.f. Louah et Violette[60].

Chapitre III. Application de la méthodologie sur le virage (LCPC/Nantes)

A partir de cette *gravité* et en fixant le seuil δ^* à $5m/s^2$, la probabilité P_f par classe \mathcal{C}_p est représentée dans le tableau III.10 :

%	\mathcal{C}_1	\mathcal{C}_2	\mathcal{C}_3	\mathcal{C}_4
P_f	0	0	0.01	0.17

Tableau III.10 – Probabilité de défaillance par classe de trajectoires.

Par conséquent, le risque de dépasser ce seuil est nul pour les classes de la « consigne apaisée ». Il est égal à 0.6 pour la classe \mathcal{C}_3 et 0.79 pour la classe \mathcal{C}_4 .

Au-delà de cette validation de la méthodologie avec des observations de trajectoires d'une précision de 100Hz issues du véhicule VERT, dans le cadre du projet DIVAS, nous avons utilisé des trajectoires issues d'un observatoire de trajectoires bord de voie (Télémètre laser, avec moins de précision).

Dans un *premier temps*, nous avons récupéré les coordonnées d'une trajectoire en temps réel sous forme de données numériques. En fonction de l'état dynamique du véhicule, le modèle \mathcal{M} de reconnaissance a affecté cette trajectoire à la classe \mathcal{C}_4 . A partir de ce résultat, en se référant aux abaques $\delta^* \rightarrow \pi(\delta^*)$ de la section (III.6.2) et en fixant le seuil de sécurité δ^* à $5.5m/s^2$ pour le critère d'**accélération transversale**, on obtient une probabilité $P_f = 0.12$. Enfin, la valeur du risque associé à cette trajectoire u est calculée. Cette trajectoire qui appartient à la classe \mathcal{C}_4 avec un risque supérieur à 0.8 est considérée comme une *conduite rapide à risque élevé* selon la classe des conséquences.

Il faut remarquer que la méthode de reconnaissance est robuste car malgré l'utilisation de données de mesures du télémètre de moins bonne qualité par rapport au véhicule instrumenté VERT qui a servi à faire la base d'apprentissage \mathcal{S} , on obtient un taux de reconnaissance de 84%. Les résultats sont encourageants pour une première validation en pratique. Pour améliorer ce résultat, il est nécessaire de mesurer les coordonnées de la trajectoire u avec plus de précision et de fréquence d'acquisition des mesures.

Cette validation sur site réel dans le cadre de ce projet a été l'occasion de montrer la faisabilité technique de la méthodologie proposée pour prédire des trajectoires potentiellement dangereuses. La figure III.51 présente la méthodologie proposée sous forme d'une architecture afin d'avoir une vision globale.

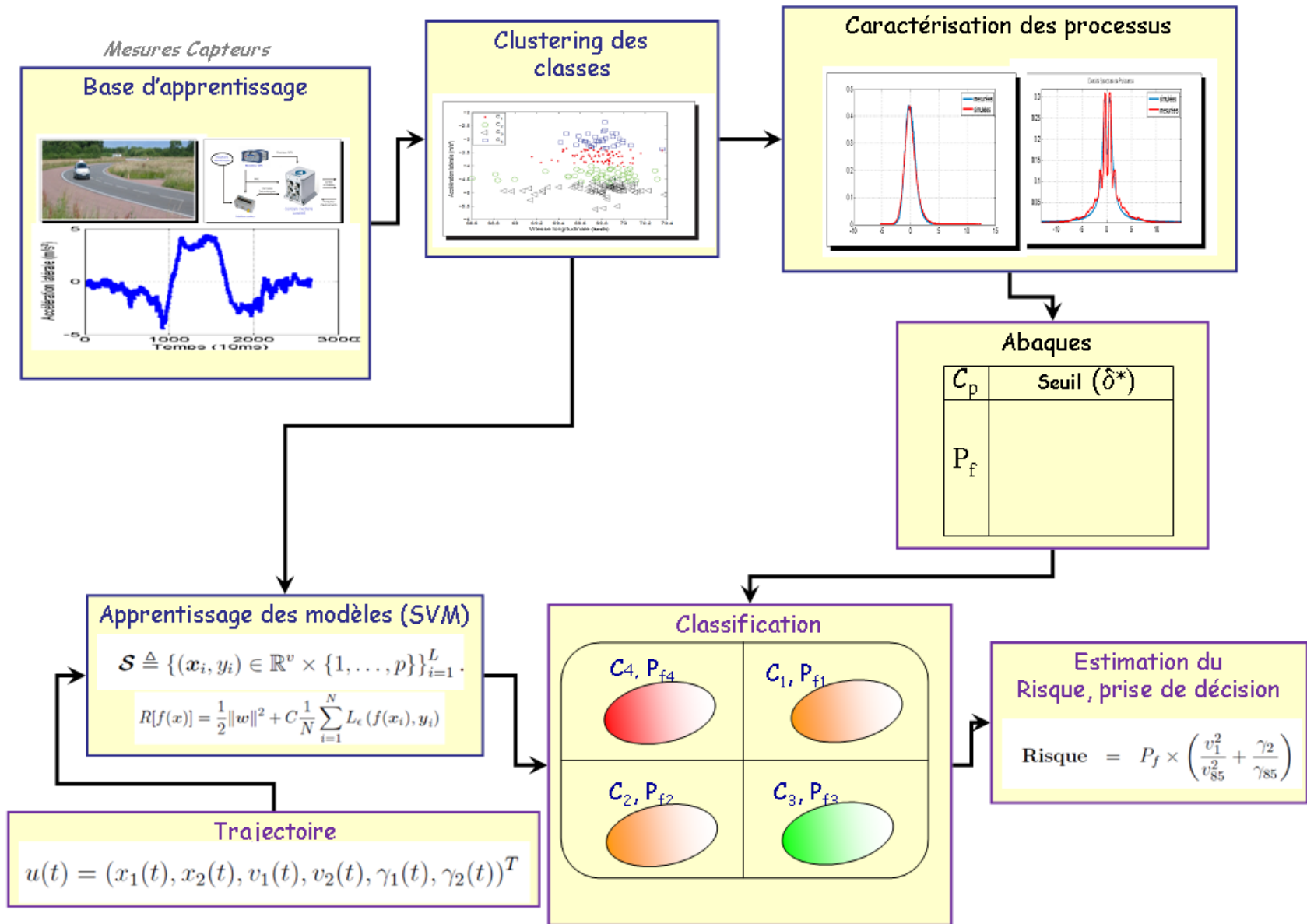
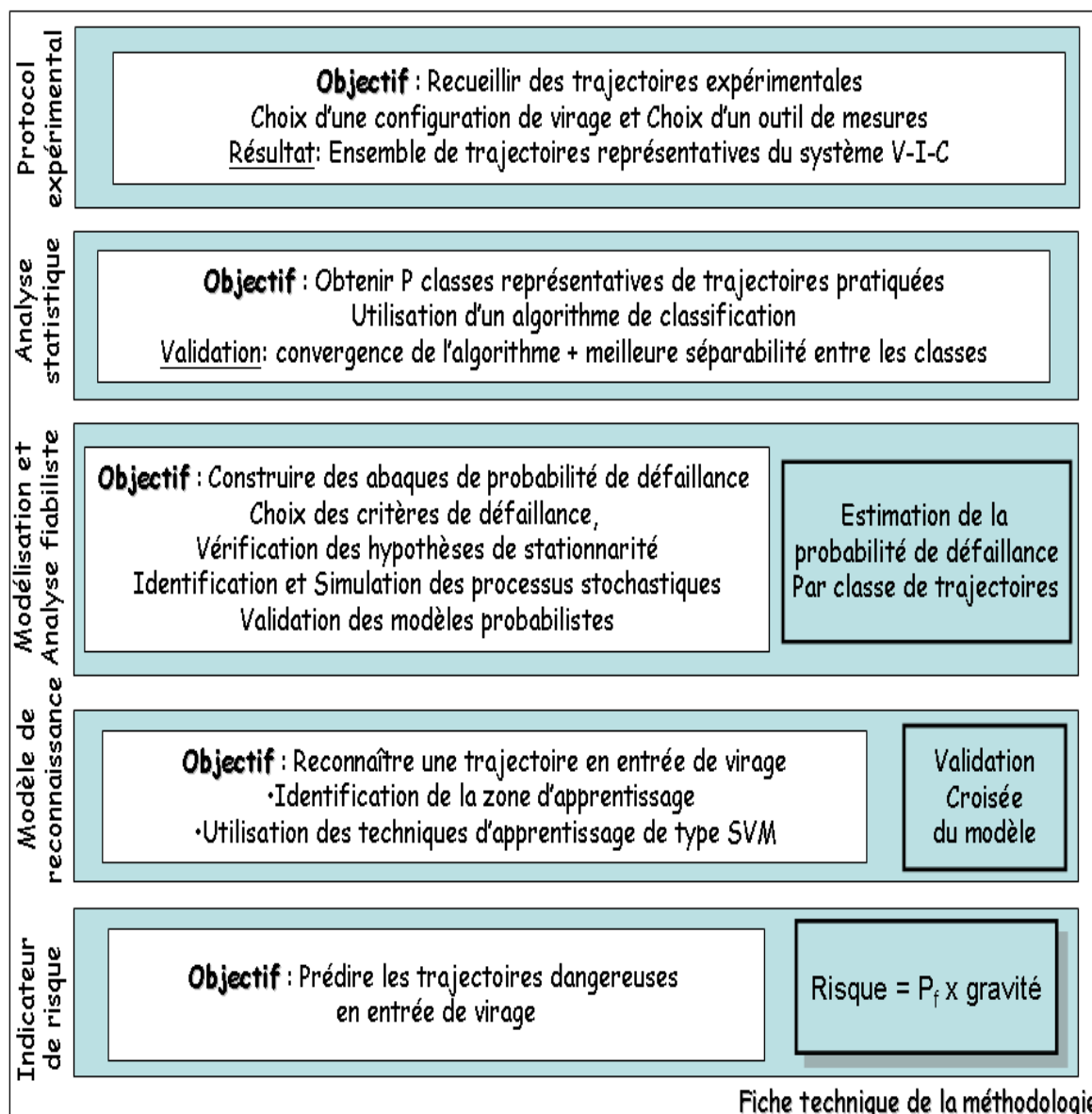


Figure III.51 – Résumé de la méthodologie pour cette configuration de virage.

III.8 Fiche technique d'utilisation de la méthodologie

Pour mieux comprendre la méthodologie décrite dans ce travail de thèse, nous avons résumé les grandes étapes dans une fiche technique.



Pour la généralisation, il suffit de choisir une configuration de virage et d'acquérir des observations représentatives de trajectoires pratiquées dans ce virage.

III.9 Conclusion

L'objet de ce chapitre III était d'appliquer la méthodologie proposée sur des données expérimentales issues du virage (LCPC/Nantes) afin de vérifier la faisabilité en pratique.

D'abord, nous avons identifié 4 classes de trajectoires à partir des observations expérimentales. Les trajectoires observées de chaque classe sont considérées comme des réalisations d'un même processus U . A travers ces réalisations, nous avons vérifié que le processus d'état U et le processus de contrôle Z ne sont pas stationnaires. Par contre, les hypothèses de stationnarité au second ordre sont vérifiées pour le processus X .

Ensuite, les approximations adéquates ont été faites pour le processus X . Pour la plupart des cas traités, nous avons choisi le développement sur la base des polynômes d'Hermite. Avec les estimées de la loi marginale d'ordre 1 et la densité spectrale de puissance, des simulations du processus X ont été obtenues. Nous avons comparé les caractéristiques statistiques des mesures à celles des simulations. Les critères de proximité et les tests statistiques ont permis de valider les modèles probabilistes Z proposés.

Les simulations ont servi à estimer la probabilité de défaillance des 4 classes de trajectoires. La comparaison des P_f par critère K_i et par classe \mathcal{C}_p a permis d'associer un niveau de risque par comportement de conduite. Un comportement de conduite est assimilé à une classe de trajectoires avec des propriétés statistiques homogènes. Pour chaque classe, on constate que les critères ont une relation d'ordre. Ceci permet de mieux distinguer les 4 classes *vis-à-vis* du risque encouru. Nous avons montré qu'il existait une relation d'ordre entre les classes \mathcal{C}_p .

Ce résultat permet de conclure sur cette application que quelque soit le critère de défaillance K_i , la classe \mathcal{C}_1 est la plus sécurisée de toutes les classes identifiées, la classe \mathcal{C}_4 est la plus risquée et la classe \mathcal{C}_3 a une probabilité de défaillance plus élevée que la classe \mathcal{C}_2 . Ces résultats sont conformes aux consignes données pendant l'expérimentation.

Dans cette configuration de virage (III.3), nous sommes capables de reconnaître une trajectoire dangereuse en entrée de virage. Cette information est une indication pour le conducteur pour agir sur les commandes du véhicule ou au moins d'être vigilant sur sa conduite. L'information peut également intéresser un gestionnaire d'infrastructure qui souhaiterait identifier des usagers dangereux dans le trafic.

CHAPITRE IV

Conclusion générale

Dans un contexte de développement de stratégies pour améliorer la sécurité routière, ce travail de thèse répond à la problématique suivante : proposer une méthodologie d'aide à la détection et à la prédiction des trajectoires à risque à partir d'une configuration de virage donnée et d'observations de trajectoires pratiquées sur ce virage.

Pour mener à bien cette réflexion, une étude accidentologique des véhicules légers a d'abord été effectuée. Les résultats de cette étude ont montré que la France dispose d'une marge de manoeuvre pour réduire l'accidentologie des véhicules en virages, notamment les *sorties de route* et les *pertes de contrôle* auxquelles nous nous sommes intéressés dans cette thèse. On s'est intéressé exclusivement à l'accidentologie des véhicules légers en virage car des études du SETRA et de l'INRETS ont montré que le risque d'accident en virage est 5 à 10 fois plus élevé qu'en alignement droit. L'étude accidentologique a montré qu'il existe plusieurs facteurs de risque tels qu'une *vitesse excessive* ou une grande variation de l'*accélération latérale*, etc.

Ensuite, nous avons constaté à travers un état de l'art qu'un nombre important de travaux scientifiques et industriels existent dans ce domaine. Malgré des résultats probants et significatifs, la complexité du système de conduite nous oblige à améliorer les méthodes existantes ou à proposer de nouvelles méthodes. Cet état de l'art a montré que les anciens travaux ne considéraient pas le système V-I-C dans sa globalité. Généralement, ils

cherchent à modéliser une sous partie. Le système contient également plusieurs *incertitudes* qui ne peuvent pas être prises en compte par les méthodes déterministes.

Premièrement, nous avons proposé dans cette thèse de considérer le système de la conduite comme un système V-I-C. Ce dernier est composé de l'ensemble des interactions du triptyque Véhicule-Infrastructure-Conducteur. Ces interactions sont *non linéaires* et difficilement modélisables de manière précise. Différentes formulations mathématiques de ce système ont été présentées en tenant compte des incertitudes (aléas) du système. Cependant, il existe plusieurs difficultés techniques pour accéder aux équations qui gouvernent ce système dans sa globalité. En revanche, nous disposons des trajectoires qui sont la réponse de ce système à des excitations extérieures. Ces excitations sont aussi mal représentées d'où la complexité de ce système.

Par ailleurs, nous avons proposé une solution originale à la problématique posée. Elle consiste à effectuer une analyse fiabiliste du système à partir d'observations de trajectoires pratiquées. D'une part, cette méthode d'analyse et de modélisation fait abstraction à toutes les interactions non linéaires qui ne sont pas modélisables de manière précise. Et d'autre part, on dispose actuellement d'instruments de mesure des trajectoires très précis.

La méthodologie proposée consiste d'abord à recueillir des trajectoires représentatives pratiquées sur une configuration de virage donnée. Ces trajectoires observées sont réparties dans des classes de trajectoires en utilisant des algorithmes de classification. Les classes de trajectoires obtenues sont judicieusement constituées. Car une mauvaise classification entraîne des problèmes d'hétérogénéité des trajectoires d'une classe.

Deuxièmement, nous avons proposé de considérer les trajectoires u d'une classe \mathcal{C}_p comme des réalisations d'un même processus stochastique U défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R}^6 . A partir de ces trajectoires, une stratégie fiabiliste du système a été proposée. Elle consiste d'abord à choisir des critères de défaillance. Nous avons montré que la *distance* et l'*accélération latérale* sont des critères pertinents pour étudier la défaillance du système VIC. Ces critères portent sur des transformations de la forme $Z = F(U)$. Le processus de contrôle Z défini sur $(\Omega, \mathfrak{F}, \mathbb{P})$ à valeurs dans \mathbb{R} est non stationnaire comme le processus U . Cependant, la modélisation probabiliste proposée suppose que les processus stochastiques utilisés soient stationnaires en m.o.d. Nous avons proposé une représentation standardisée de Z qui vérifie l'hypothèse de stationnarité au second ordre.

Ensuite, à partir des réalisations de X , celui-ci a été partiellement caractérisé et identi-

fié. La loi marginale d'ordre 1 de X et la densité spectrale de puissance ont permis de simuler le processus X . A cet effet, une technique de simulation de processus stationnaire non gaussien a été utilisée. Les simulations obtenues ont permis de valider les modèles Z par comparaison entre les *données mesurées* et les *données simulées*. Pour cette comparaison, des *critères de proximité* heuristiques et des *tests statistiques* de conformité de lois ont été utilisés. L'hypothèse nulle H_0 n'est pas rejetée pour la conformité des lois relatives aux mesures et aux simulations pour une erreur de première espèce à 5%. Ce résultat vérifie également l'hypothèse que toutes les trajectoires d'une classe \mathcal{C}_p sont issues d'un même processus U_p . Les simulations ont également servi à estimer la probabilité de défaillance P_f relative aux 4 classes de trajectoires et à chacun des 2 critères de défaillance.

Troisièmement, nous avons identifié la loi p_M des maxima du processus Z afin d'estimer la probabilité de défaillance P_f du système. La probabilité P_f peut être calculée directement si p_M est approximée par une des 3 lois de valeurs extrêmes. Sinon, une bonne estimation de cette loi permet également de faire une estimation correcte de P_f . Par contre, la difficulté de la mesure du risque dans le système VIC est le choix du seuil de sécurité δ^* . Pour cela, on a fait varier δ^* dans un intervalle de valeurs admissibles. Ensuite, on a comparé les fonctions de probabilité $\pi(\delta^*)$ relatives aux 4 classes de trajectoires et à chacun des 2 critères de défaillance. Cette comparaison a montré que les 4 classes \mathcal{C}_p n'ont pas le même niveau de risque pour un critère K_i donné. Elle a également montré que le niveau de risque n'est pas le même selon le critère pour une classe \mathcal{C}_p donnée. L'utilisation de ces résultats a permis de faire un classement par ordre croissant des classes de trajectoires en terme de probabilité de défaillance.

Au-delà de l'estimation de la probabilité P_f pour chaque classe \mathcal{C}_p , nous avons introduit la notion du risque. Le calcul du risque I_{risque} proposé a pour but de différencier les niveaux de risque de deux trajectoires $u^{(1)}$ et $u^{(2)}$ appartenant à une même classe \mathcal{C}_p .

Quatrièmement, pour compléter la méthodologie, un modèle \mathcal{M} de reconnaissance des trajectoires a été développé. Ce modèle est basé sur les observations en entrée de virage. Il a permis d'affecter une trajectoire u à une classe \mathcal{C}_p dont on a préalablement calculé ses probabilités de défaillance P_f . Ensuite, en fonction des coordonnées de la trajectoire u et de sa probabilité P_f , on lui associe un niveau de risque I_{risque} .

Avec une telle méthodologie et dans cette configuration de virage (III.3), on peut reconnaître une trajectoire dangereuse à l'entrée du virage. Cette information est une indication

pour le conducteur pour agir sur les commandes du véhicule ou au moins d'être vigilant sur sa conduite. Cette information peut concerner également un gestionnaire d'infrastructure qui souhaiterait identifier des usagers dangereux dans le trafic. Au-delà des résultats significatifs et probants de cette *méthodologie*, celle-ci présente plusieurs avantages considérables qui sont entre autres : son utilisation ne nécessite que la connaissance de trajectoires représentatives pratiquées, la facilité de mise en oeuvre numérique, pas de difficultés d'identifier les paramètres pertinents du système contrairement aux modèles de dynamique véhicule, et surtout prise en compte des incertitudes du système, etc.

Malgré les bons résultats des modèles élaborés et les avantages de cette méthodologie, néanmoins elle reste sensible aux données de mesures. Pour appliquer correctement la méthodologie proposée, il est nécessaire de disposer d'une *bonne représentativité* des trajectoires pratiquées et une *bonne précision* des trajectoires mesurées car chacune des étapes de la méthodologie repose sur ces conditions. Nous sommes confiants dans ce domaine, car les outils de mesures de trajectoires seront performants et accessibles dans un futur proche.

En perspectives, afin d'améliorer la précision de cette méthodologie, nous proposons d'enrichir la base de données expérimentales. Par ailleurs, plusieurs pistes peuvent être envisagées pour améliorer les performances du classifieur. Ceci permet d'éviter les fausses alertes ou la non détection des trajectoires potentiellement dangereuses. On peut penser à généraliser cette méthodologie dans d'autres *configurations de virage* ou avec d'autres *systèmes de mesures* moins performants que le véhicule VERT. Cette généralisation aura pour but d'identifier probablement d'autres comportements de conduite, niveaux de risque, etc.

Il est également possible de faire une classification des infrastructures routières en terme de dangerosité en utilisant des indices de fiabilité conventionnels comme ce qui existe pour les structures de génie civil. Une des possibilités serait d'évaluer la probabilité de collision d'un véhicule avec des véhicules circulant dans le sens inverse. Ce résultat pourra contribuer à l'analyse de la dangerosité de certains virages et de certains comportements de conduite en virage. On peut envisager d'adapter cette méthodologie à d'autres domaines de la sécurité routière tels que l'identification des comportements dangereux des conducteurs de deux-roues.

V.1 Stationnarité et ergodicité d'un processus X

La stationnarité joue un rôle important dans la prédiction de séries temporelles. La translation de l'origine des temps d'une durée ne doit pas affecter la loi conjointe. Plusieurs types de stationnarité existent : au sens strict, au sens faible, etc. La notion de stationnarité au sens strict est rarement utilisable en pratique car l'on dispose rarement de toutes les densités de probabilité d'un signal aléatoire.

V.1.1 Stationnarité en moyenne quadratique de X

On dit qu'un processus stochastique X est stationnaire à l'ordre 2 lorsque la moyenne et la variance ne dépendent pas du temps, et que la fonction d'autocovariance ne dépend que de la différence entre les deux arguments t_1 et t_2 Bouleau[11], soit :

$$\begin{cases} \mathbb{E}[X(t)] = \mu_X \text{ (constant)} ; \forall t \in T \\ \mathbb{E}[(X(t) - \mathbb{E}[X(t)])^2] = \sigma_X \text{ (constant)} ; \forall t \in T \\ \text{Cov}(X(t_1), X(t_2)) = C_{XX}(t_2 - t_1) = C_{XX}(\tau); \forall t_1 \forall t_2; \tau = t_2 - t_1 \end{cases} \quad (\text{V.1})$$

Remarque 7. *La stationnarité au second ordre est suffisante pour assurer la stationnarité forte lorsque le processus peut être supposé gaussien, hypothèse souvent utilisée, parfois fautive de mieux.*

Par construction, $\mu_X = 0$ et $\sigma_X = 1$ car le processus X a été normalisé en se référant à l'équation (II.3.1.3). Quant à la fonction d'autocorrélation R_X de X , celle-ci ne doit dépendre que de l'écart

τ entre deux instants t_1 et t_2 . Ces conditions sont complétées par des tests statistiques pertinents [105].

V.1.2 Hypothèse d'ergodicité de X

Lorsque l'on utilise l'opérateur d'espérance mathématique \mathbb{E} , son application stricte suppose de recourir à ce que l'on appelle une moyenne d'ensemble. Par exemple, la moyenne d'une variable aléatoire est l'intégrale sur toutes les valeurs possibles, notées x , pondérées par la densité de probabilité de la v.a. Imaginons maintenant que l'on souhaite estimer la moyenne d'un processus stochastique $X(t)$ si l'on dispose de L_p réalisations de ce processus :

$$\begin{cases} \mu_X &= \mathbb{E}(X(t)) \\ G_T &= \frac{1}{T} \int_0^T X(t) dt \end{cases} \quad (\text{V.2})$$

Le paramètre μ_X est la moyenne de $X(t)$ et représente une moyenne d'ensemble (moyenne probabiliste) ; G_T représente une moyenne sur la trajectoire du processus X qui est stationnaire. La notion d'ergodicité correspond au fait que pour $T \rightarrow \infty$, la suite de v.a. G_T tend en un certain sens vers μ_X , c'est-à-dire que cette propriété permet de remplacer la moyenne d'ensemble par la moyenne sur une trajectoire. Le processus X est dit ergodique par rapport à μ_X dans $L^2(\sigma, \mathbb{R})$ si G_T tend vers μ_X dans $L^2(\sigma, \mathbb{R})$ quand $T \rightarrow \infty$: $\lim_{T \rightarrow \infty} \mathbb{E}(|G_T - \mu_X|^2) = 0$, Soize[93].

En général on se contente de supposer l'ergodicité pour les moments d'ordre 1 et 2 soit : μ_X , σ_X^2 et $C_{XX}(t)$. La vérification de l'hypothèse de stationnarité est nécessaire avant de procéder à la caractérisation des processus X .

V.2 Estimations de la densité de probabilité p_X

On dispose d'un N-échantillon ordonné $\varepsilon_N = \{x_1, \dots, x_N\}$ de réalisations du processus X absolument continu. Avec ε_N étant un empilement de L_p réalisations de X . On veut construire à partir de ε_N une estimée de la densité de probabilité p_X et de la fonction de répartition F_X .

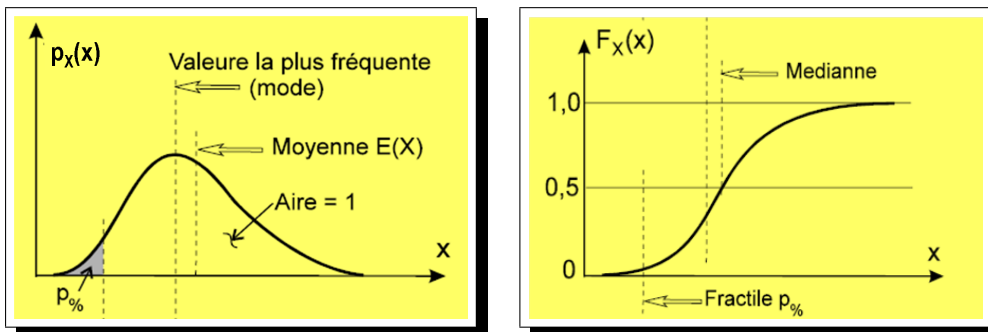


Figure V.1 – Approximations de p_X et F_X

V.2. Estimations de la densité de probabilité p_X

Bien entendu, plus le nombre d'épreuves est grand, plus les intervalles de partitionnement du domaine des valeurs possibles du processus peuvent être fins et, par conséquent, plus l'histogramme peut donner une meilleure approximation de la densité de probabilité (c.f. Bouleau [12], Fogli[36]).

Une estimation convenable de p_X peut être la densité relative des points expérimentaux. Le domaine d'estimation est l'intervalle fermé \mathbb{D} de \mathbb{R} tel que $\mathbb{D} = [x_1, x_N]$. On se donne un entier strictement positif M petit devant N et on subdivise \mathbb{D} en M intervalles D_1, \dots, D_M de même longueur δ , tel que, $\forall j \in J_M = 1, \dots, M$:

$$D_j^\delta = \begin{cases} [x_1 + (j-1)\delta, x_1 + j\delta] & \text{si } j < M \\ [x_N - \delta, x_N] & \text{si } j = M \end{cases} \quad (\text{V.3})$$

On a donc :

$$\mathbb{D} = \cup_{j \in J_M} D_j^\delta, \quad \mathbb{D}_i^\delta = \cap_{j \in J_M, i \neq j} D_j^\delta = \emptyset \quad (\text{V.4})$$

et, $\forall j \in J_M$:

$$\delta = |D_j^\delta| = \frac{x_N - x_1}{M} \quad (\text{V.5})$$

Notons que le choix d'une partition régulière de \mathbb{D} n'est pas impératif. Il n'a en fait d'autre raison que de simplifier la mise en oeuvre informatique des estimées. Si l'échantillon ε_N présente une forte irrégularité dans la répartition de ses éléments, il est d'ailleurs préférable d'avoir recours à un fractionnement non régulier de \mathbb{D} . A chaque intervalle D_j^δ de la partition choisie de \mathbb{D} , on associe le nombre N_j^δ d'éléments de ε_N lui appartenant.

Les estimées cherchées de la densité de probabilité p_X et de la fonction de répartition F_X sont alors définies respectivement par :

$$p_X^N(x) = \sum_{j \in J_M} \frac{N_j^\delta}{N\delta} \mathbf{1}_{D_j^\delta}(x) \quad ; \quad x \in \mathbb{D} \quad (\text{V.6})$$

$$F_X^N = \sum_{j \in J_M} \frac{\sum_{k=1}^j N_k^\delta}{N} \mathbf{1}_{D_j^\delta}(x) \quad ; \quad x \in \mathbb{D} \quad (\text{V.7})$$

où $\mathbf{1}_{D_j^\delta}$ est la fonction indicatrice de D_j^δ . L'estimation (V.6) de p_X n'est bien entendu pas unique. On peut en définir bien d'autres qui sont d'ailleurs en général de meilleure qualité.

Parmi ces estimateurs, on a l'**estimation de Rosenblatt** basée sur le principe de la fenêtre mobile. L'estimée de rosenblatt de p_X relative à l'échantillon ε_N est alors définie par :

$$p_X^N(x) = \frac{1}{N\delta} \sum_{i=1}^N \mathbf{1}_{\Delta_{1/2}} \left(\frac{x - x_i}{\delta} \right), \quad x \in \mathbb{R} \quad (\text{V.8})$$

avec $\Delta_{1/2} = [-\frac{1}{2}, \frac{1}{2}]$.

On peut également faire une **estimation à noyau de Parzen**. Elle est également basée sur le principe de la fenêtre mobile, cette estimation est définie par :

$$p_X^N(x) = \frac{1}{N\delta} \sum_{i=1}^N K\left(\frac{x-x_i}{\delta}\right), \quad x \in \mathbb{R} \quad (\text{V.9})$$

où K , appelé noyau de l'estimation, est une densité de probabilité sur \mathbb{R} de support : $\Delta_1 = [-1, 1]$ et δ est un réel > 0 à choisir, appelé paramètre de lissage.

Les **noyaux** les plus couramment utilisés dans la littérature sont les suivants :

- le noyau triangulaire

$$K(u) = (1 - |u|) \mathbf{1}_{\Delta_1}(u), \quad u \in \mathbb{R} \quad (\text{V.10})$$

- le noyau d'Epanechnikov

$$K(u) = \frac{3}{4} (1 - u^2) \mathbf{1}_{\Delta_1}(u), \quad u \in \mathbb{R} \quad (\text{V.11})$$

- le noyau de Turkey

$$K(u) = \frac{15}{16} (1 - u^2)^2 \mathbf{1}_{\Delta_1}(u), \quad u \in \mathbb{R} \quad (\text{V.12})$$

- le noyau sinusoidal

$$K(u) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right) \mathbf{1}_{\Delta_1}(u), \quad u \in \mathbb{R} \quad (\text{V.13})$$

On peut également utiliser des noyaux portés par \mathbb{R} , comme par exemple le noyau gaussien :

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right), \quad u \in \mathbb{R} \quad (\text{V.14})$$

La constante de lissage δ détermine la régularité de l'estimée p_X^N : un δ trop grand lisse trop, alors qu'un δ trop petit conduit à une estimation très chaotique.

V.3 Estimation densité spectrale de puissance S_X

Soit le processus stochastique X , centré, continu en moyenne quadratique, stationnaire et ergodique. $L_p \in \mathbb{N}$, on note $x^{(l)}$ avec $l \in \{1, \dots, L_p\}$, une famille de L_p réalisations indépendantes de $X(t)$. Soit $S_X : \mathbb{R} \rightarrow \mathbb{R} : \omega \rightarrow S_X(\omega)$ la DSP de X . L'estimateur statistique $\hat{S}_{L,T}$ de Welch a été utilisé pour estimer S_X (c.f.[54]). C'est en fonction de la qualité de l'estimateur que dépend

V.4. Approximations de lois de probabilité

celle des estimées. Il importe donc d'étudier soigneusement les propriétés (biais, efficacité, consistance) de l'estimateur. Pour plus de détail concernant cette question, nous renvoyons le lecteur à l'ouvrage de Soize[93]. Nous nous contenterons juste de donner les expressions de l'estimateur de Welch avec la fenêtre de Hamming.

$$\hat{S}_{L,T}(\omega) = \frac{1}{2\pi} \frac{1}{L_p} \sum_{l=1}^{L_p} |X^{(l)}|^2 \quad (\text{V.15})$$

$$X_p^{(l)}(\omega) = \int_0^T W_T(t) X_p^{(l)}(t) \exp(-i\omega t) dt \quad (\text{V.16})$$

où W_T est la fenêtre temporelle relative à l'intervalle d'acquisition $t = [0, T]$. Pour cette fenêtre, nous choisissons le modèle de Hamming qui s'écrit, $\forall t \in \mathbb{R}$:

$$W_T(t) = \frac{1}{\sqrt{T}} 1.5865 \left(0.54 - 0.46 \cos \left(\frac{2\pi t}{T} \right) \right) 1_T(t) \quad (\text{V.17})$$

V.4 Approximations de lois de probabilité

V.4.1 Lois usuelles

Rappelons que la loi normale de paramètres $\mu \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+^*$ est une loi de probabilité de support $x \in]-\infty, +\infty[$ et dont la densité de probabilité ou fonction de masse est :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (\text{V.18})$$

Sa fonction de répartition est de la forme :

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(u-\mu)^2}{2\sigma^2}} du \quad (\text{V.19})$$

Rappelons également que la loi logistique de paramètre $\mu \in \mathbb{R}$ et pour $s > 0$ est une loi de probabilité de support $x \in]-\infty, +\infty[$ et dont la densité ou la fonction de masse est :

$$f(x) = \frac{\exp\left(-\frac{(x-\mu)}{s}\right)}{s(1 + \exp\left(-\frac{(x-\mu)}{s}\right))^2} \quad (\text{V.20})$$

Sa fonction de répartition est :

$$F(x) = \frac{1}{(1 + \exp\left(-\frac{(x-\mu)}{s}\right))} \quad (\text{V.21})$$

V.4.2 Développement de la densité sur la base des polynômes d'Hermite

Ces polynômes peuvent être construits par récurrence :

$$\begin{cases} H_0(x) &= 1 \\ H_{m+1}(x) &= xH_m - \frac{d}{dx}H_m(x) \end{cases} \quad (\text{V.22})$$

Par exemple, voici les 5 premiers :

$$\begin{cases} H_0(x) = 1 \\ H_1(x) = x \\ H_2(x) = x^2 - 1 \\ H_3(x) = x^3 - 3x \\ H_4(x) = x^4 - 6x^2 + 3 \\ H_5(x) = x^5 - 10x^3 + 15x \end{cases} \quad (\text{V.23})$$

A partir des polynômes H_m , on obtient les fonctions normalisées h_m sur \mathbb{R} à valeurs dans \mathbb{R} définies par :

$$h_m(x) = (m!)^{-1/2}H_m(x) \quad ; \quad \forall m \in \mathbb{N} \quad (\text{V.24})$$

Ils forment une base orthonormée de $L^2(\mathbb{R}, \nu_1)$ et $(m!)^{-1/2}$ est le facteur de normalisation de manière à avoir $((h_j, h_k)) = \delta_{jk}$. La dérivée de $H_m(x)$ par rapport à x s'écrit :

$$\frac{d}{dx}H_m(x) = mH_{m-1}(x) \quad (\text{V.25})$$

La transformée de Fourier de H_m par rapport à la mesure $\nu_1 dx$ s'écrit pour tout u dans \mathbb{R} .

$$\begin{aligned} (tF_\nu H_m)(u) &= \int_{\mathbb{R}} \exp(iux)H_m(x)f_1(x)dx \\ &= (iu)^m \exp(-\frac{1}{2}u^2) \end{aligned} \quad (\text{V.26})$$

Pour $m > 0$, ces polynômes ont une moyenne nulle :

$$\mathbb{E}[H_m(Y(x))] = \int H_m(x)g(x)dx = 0 \quad (\text{V.27})$$

une variance unité :

$$\text{Var}[H_m(Y(x))] = E[H_m(Y(x))^2] = 1 \quad (\text{V.28})$$

et vérifient, pour $p \neq m$, positifs ou nuls :

$$\text{Cov}[H_p(Y(x)), H_m(Y(x))] = \mathbb{E}[H_p(Y(x))H_m(Y(x))] = 0 \quad (\text{V.29})$$

V.4. Approximations de lois de probabilité

A partir de la définition et des propriétés citées ci-dessus, nous allons montrer comment développer une fonction en polynôme d'Hermite.

Développement d'une fonction en polynômes d'Hermite

Toute fonction $f[Y(x)]$ peut se développer en polynômes d'Hermite :

$$\begin{aligned} f[Y(x)] &= f_0 + f_1 H_1[Y(x)] + f_2 H_2[Y(x)] + f_3 H_3[Y(x)] \\ &= \sum_{n=0}^{\infty} f_n H_n[Y(x)] \end{aligned} \quad (\text{V.30})$$

Du fait de l'orthogonalité des polynômes : $\mathbb{E}[f[Y(x)]H_m[Y(x)]] = f_n$.

En effet :

$$\begin{aligned} \mathbb{E}[f[Y(x)]H_m[Y(x)]] &= \mathbb{E}\left(\left[\sum_{p=0}^{\infty} f_p H_p[Y(x)]\right] H_m[Y(x)]\right) \\ &= \sum_{p=0}^{\infty} f_p \mathbb{E}[H_p(Y(x))H_m(Y(x))] \\ &= f_m \end{aligned} \quad (\text{V.31})$$

Pour toute fonction f donnée, ceci permet de calculer ces coefficients :

$$\begin{aligned} f_m &= \mathbb{E}[f[Y(x)]H_m[Y(x)]] \\ &= \int f'(x)H_m(x)g(x)dx \end{aligned} \quad (\text{V.32})$$

Avec $f_0 = \mathbb{E}[f[Y(x)]]$.

V.4.3 Coefficients d'aplatissement et d'asymetrie

Le coefficient d'**aplatissement** est une statistique qui mesure le degré de vraisemblance des événements extrêmes. Plus, il est grand, plus les queues de distribution sont épaisses par rapport à celles de la loi normale, c'est à dire plus des événements extrêmes peuvent potentiellement survenir. Ce coefficient, noté K se calcule à partir de l'estimateur suivant :

$$\hat{K} = \frac{\hat{\mu}_X^{(4)}}{\left(\hat{\mu}_X^{(2)}\right)^4} \quad (\text{V.33})$$

où $\hat{\mu}_X^{(r)}$ est l'estimateur du moment centré d'ordre r , tel que :

$$\hat{\mu}_X^{(r)} = \frac{\sum_{l=1}^L [X^{(l)}(t)]^r}{L}, \forall t \in T \quad (\text{V.34})$$

Si $K > 0$, la distribution est plus allongée que la loi normale, et les queues de distribution sont plus épaisses. Si $K = 0$, la distribution est quasi normale, Si $K < 0$, la distribution est plus aplatie que la loi normale, les queues de distribution sont moins épaisses.

Le coefficient d'**asymétrie** mesure le degré d'asymétrie de la distribution. Ce coefficient, noté S se calcule à partir de l'estimateur suivant :

$$\hat{S} = \frac{\hat{\mu}_X^{(3)}}{\left(\hat{\mu}_X^{(2)}\right)^3} \quad (\text{V.35})$$

Pour une distribution parfaitement symétrique (par exemple, une loi normale), le coefficient d'asymétrie $S = 0$. Si le coefficient $S < 0$, la distribution est asymétrique à gauche : il y a une forte probabilité qu'un événement soit au dessus de la moyenne qu'en dessous (la valeur modale est au dessus de la moyenne). Si, $S > 0$ la distribution est asymétrique à droite : il y a une plus faible probabilité qu'un événement soit au dessus de la moyenne qu'en dessous (la valeur modale est au dessous de la moyenne).

V.5 Généralités et définitions sur les méthodes SVM

Soit un problème de classification binaire, nous considérons le formalisme suivant. Soit l'ensemble \mathcal{S} tel que :

$$\mathcal{S} \triangleq \{(\mathbf{x}_i, y_i) \in \mathbb{R}^v \times \{-1, 1\}\}_{i=1}^N.$$

V.5.1 Classifieur linéaire

Un classifieur est dit linéaire lorsqu'il est possible d'exprimer sa fonction de décision par une fonction linéaire en fonction de \mathbf{x} . On peut exprimer une telle fonction comme :

$$\mathbf{x} \longmapsto f_{\boldsymbol{\theta}}(\mathbf{x}) = \boldsymbol{\theta}^T \mathbf{x} + b = \langle \boldsymbol{\theta}, \mathbf{x} \rangle + b, \quad (\text{V.36})$$

où $\boldsymbol{\theta}, \mathbf{x} \in \mathbb{R}^v$ et $b \in \mathbb{R}$. Pour décider à quelle catégorie y un exemple \mathbf{x} appartient, il suffit de voir le signe de la fonction de décision : $y = \text{sign}(f_{\boldsymbol{\theta}}(\mathbf{x}))$. Géométriquement, cela revient à considérer un hyperplan qui est le lieu des points \mathbf{x} satisfaisant la condition : $\langle \boldsymbol{\theta}, \mathbf{x} \rangle + b = 0$ et de considérer un côté pour lequel les exemples sont classés positivement et l'autre côté pour lequel les exemples sont classés négativement. La figure (V.2) donne une représentation imagée de la situation dans \mathbb{R}^2 . On voit que le vecteur $\boldsymbol{\theta}$ définit la pente de l'hyperplan : $\boldsymbol{\theta}$ est perpendiculaire à l'hyperplan.

Données linéairement séparable et marge

Il est utile d'introduire la notion de *données linéairement séparable*.

V.5. Généralités et définitions sur les méthodes SVM

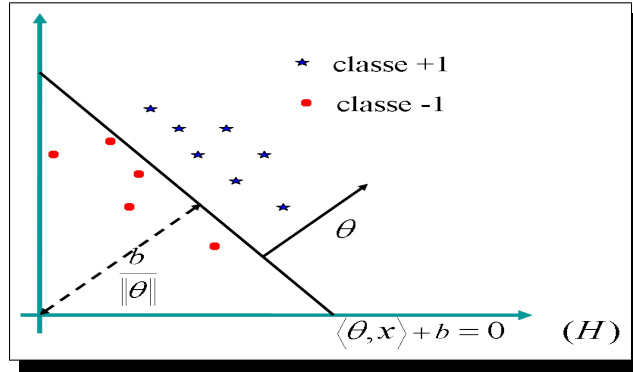


Figure V.2 – Hyperplan séparateur d'un classifieur linéaire. La quantité $\frac{b}{\|\theta\|}$ exprime la distance entre l'hyperplan séparateur (H) et l'origine.

Définition 2. (*linéairement séparable*) Un ensemble $\{(\mathbf{x}_i, y_i) \in \mathbb{R}^v \times \{-1, 1\}\}_{i=1}^N$ est linéairement séparable ssi :

$$\exists \theta \in \mathbb{R}^v, b \in \mathbb{R} : y_i(\langle \theta, \mathbf{x}_i \rangle + b) \geq 0, \quad \forall i = 1, \dots, N. \quad (\text{V.37})$$

La définition consiste à dire qu'il doit exister un hyperplan permettant de séparer les exemples positifs des exemples négatifs. Dans le cas de données linéairement séparables, il existe plusieurs méthodes pour trouver un tel hyperplan. La plus connue et la plus ancienne parmi ces méthodes est l'algorithme du perceptron de Rosenblatt [84]. Il existe généralement une infinité d'hyperplans qui permettent de séparer les données d'apprentissage mais tous ne permettent pas de bien généraliser, c'est-à-dire de bien classer les données à venir et pas encore vues dans la base d'apprentissage.

L'objectif est alors de déterminer parmi ces hyperplans celui qui sépare d'une manière « optimale » les données d'apprentissage, autrement dit, il faut choisir l'hyperplan qui permettra une meilleure généralisation. Dans le domaine des SVM un hyperplan peut être considéré comme optimal dans le sens où il maximise la marge de séparation. Nous n'allons plus seulement nous contenter de trouver un hyperplan séparant les deux classes, mais nous allons en plus chercher parmi eux celui qui sépare au mieux les classes (au sens de la maximisation de la marge).

Intuitivement, cela revient à chercher l'hyperplan le plus sûr. Le fait d'avoir une marge plus large procure plus de « sécurité » lorsque l'on classe un nouvel exemple. Comme le montre la figure (V.3), la partie gauche nous montre le cas où on a un hyperplan optimal avec une grande marge, ce qui nous permet de bien classer un nouvel exemple malgré le fait qu'il soit tombé dans la marge.

Cependant, sur la partie droite malgré le fait que l'on ait trouvé un hyperplan optimal, l'exemple

se voit mal classé, du fait que, la marge soit petite. Nous présentons maintenant la notion de la

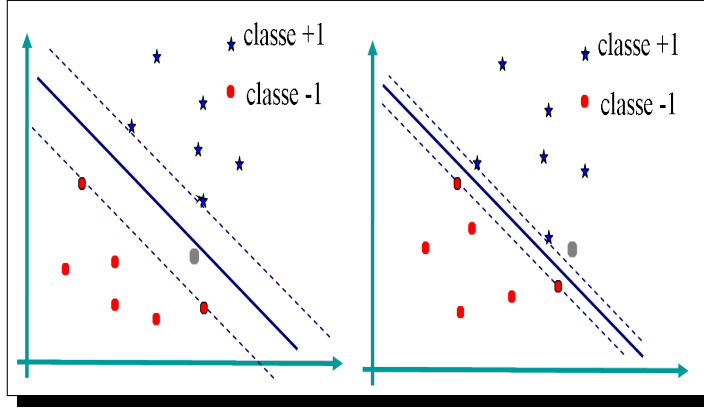


Figure V.3 – Impact de la maximisation de la marge sur la capacité de généralisation.

marge qui est une notion très importante pour les SVMs. La notion de marge peut être relative à un exemple particulier ou à un ensemble d'exemples. La marge géométrique représente la distance Euclidienne prise perpendiculairement entre l'hyperplan et l'exemple \mathbf{x}_i .

Définition 3. (*marge géométrique d'un exemple*) La marge géométrique d'un exemple \mathbf{x}_i , par rapport à l'hyperplan définie par $\{\boldsymbol{\theta}, b\}$ est la quantité :

$$\Psi_{\boldsymbol{\theta}, b}(\mathbf{x}_i, y_i) \triangleq y_i \left(\frac{\langle \boldsymbol{\theta}, \mathbf{x}_i \rangle + b}{\|\boldsymbol{\theta}\|} \right). \quad (\text{V.38})$$

Définition 4. (*marge géométrique de l'ensemble d'apprentissage*) La marge de l'ensemble d'apprentissage par rapport à l'hyperplan définie par $\{\boldsymbol{\theta}, b\}$ est la quantité :

$$\Phi_{\boldsymbol{\theta}, b} = \min_{i=1, \dots, N} \{\Psi_{\boldsymbol{\theta}, b}(\mathbf{x}_i, y_i)\} \quad (\text{V.39})$$

Ainsi, la marge a l'expression mathématique suivante :

$$\begin{aligned} M &= \min_{\mathbf{x}_i | y_i = +1} \left\{ \frac{\langle \boldsymbol{\theta}, \mathbf{x}_i \rangle + b}{\|\boldsymbol{\theta}\|} \right\} - \max_{\mathbf{x}_i | y_i = -1} \left\{ \frac{\langle \boldsymbol{\theta}, \mathbf{x}_i \rangle + b}{\|\boldsymbol{\theta}\|} \right\}, \\ &= \frac{1}{\|\boldsymbol{\theta}\|} - \frac{-1}{\|\boldsymbol{\theta}\|}, \\ &= \frac{2}{\|\boldsymbol{\theta}\|}. \end{aligned} \quad (\text{V.40})$$

V.5.1.1 Hyperplans canoniques

Dans le cadre de la classification par maximisation de la marge d'un ensemble d'apprentissage linéairement séparable, l'hyperplan séparateur correspond à l'hyperplan permettant la séparation des exemples de la bases d'apprentissage en deux catégories. On peut définir deux plans

V.5. Généralités et définitions sur les méthodes SVM

se trouvant de part et d'autre de l'hyperplan et parallèles à celui-ci, sur lesquels se trouvent les exemples les plus proches. Dans notre définition de l'hyperplan en (*cf.* section V.5.1), il est possible que différents hyperplans correspondent à cette définition (correspondant mathématiquement à l'équation d'un hyperplan (voir la figure (V.4))). Il est donc possible de modifier les

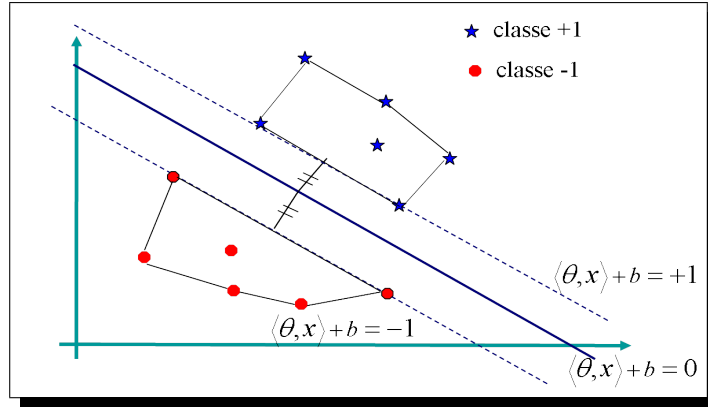


Figure V.4 – Hyperplans de séparation.

paramètres θ et b de telle façon que les deux plans parallèles aient respectivement pour équation :

$$\begin{cases} \langle \theta, x \rangle + b = 1 \\ \langle \theta, x \rangle + b = -1. \end{cases} \quad (\text{V.41})$$

Ces deux hyperplans sont appelés hyperplans canoniques (voir la figure (V.5) pour une représentation visuelle de ces hyperplans canoniques). Notons que la marge des hyperplans canoniques est $\frac{1}{\|\theta\|}$. Le lien entre la maximisation de la marge et le principe de minimisation du risque

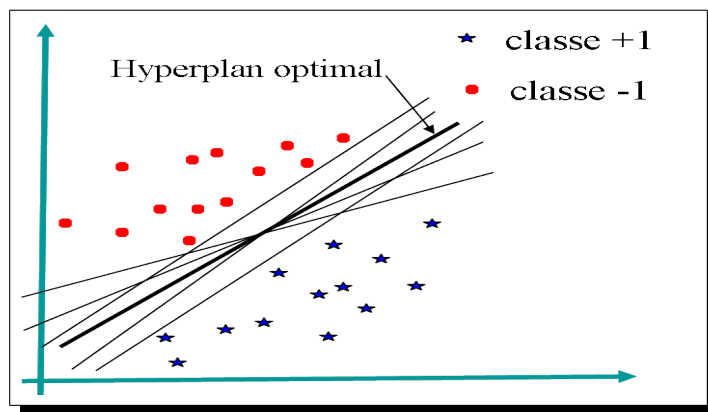


Figure V.5 – Hyperplans canoniques.

structurel dans la théorie des SVM se fait à travers le théorème suivant (voir [90]) :

Théorème 3.

(Dimension-VC d'un classifieur linéaire) Soit l'ensemble des hyperplans $\langle \boldsymbol{\theta}, \mathbf{x} \rangle + b = 0$ où $\boldsymbol{\theta}$ est normalisé de tel sorte que les exemples les plus proches se trouvent sur les hyperplans canoniques : $\min_{i=1, \dots, N} \|\langle \boldsymbol{\theta}, \mathbf{x}_i \rangle + b\| = 1$. La famille des fonctions de décision $f_{\boldsymbol{\theta}}(\mathbf{x}) = \langle \boldsymbol{\theta}, \mathbf{x} \rangle + b$; telle que $\|\boldsymbol{\theta}\| \leq \Lambda$ possède une dimension-VC satisfaisant l'inégalité : $h \leq R^2 \Lambda$. Où R est le rayon de la plus petite sphère contenant les exemples de l'ensemble d'apprentissage centrée à l'origine.

- Remarque 8.** – Ce théorème nous montre qu'en diminuant la borne Λ sur $\|\boldsymbol{\theta}\|$, c'est-à-dire en augmentant (indirectement) la marge, la dimension VC du classifieur diminue. On peut donc contrôler la dimension VC en agissant sur la marge.
- Un point très important est le fait que ce résultat est indépendant de la dimension des données de l'ensemble d'apprentissage. On peut donc obtenir une dimension VC contrôlable dans un espace à très grande dimension.

V.5.2 Classifieur à marge maximale

Maintenant que nous avons défini les notions de marges et d'hyperplans canoniques, la recherche d'un hyperplan optimal consiste à maximiser cette marge, autrement dit, à minimiser $\|\boldsymbol{\theta}\|$. Le problème peut se formaliser de la façon suivante :

$$\begin{cases} \min_{\boldsymbol{\theta}} & \frac{1}{2} \|\boldsymbol{\theta}\|^2 \\ \text{sous les contraintes} & y_i(\langle \boldsymbol{\theta}, \mathbf{x}_i \rangle + b) \geq 1, \forall i = 1, \dots, N. \end{cases} \quad (\text{V.42})$$

Il s'agit d'un problème quadratique sous contraintes linéaires dont la fonction objectif est à minimiser. Cette fonction objectif est le carré de l'inverse de la marge. L'unique contrainte se traduit par le fait que les exemples doivent être bien classés et qu'ils ne doivent pas dépasser les hyperplans canoniques.

Dans cette formulation, que l'on appelle *formulation primale*, les variables à fixer sont les composantes $\boldsymbol{\theta}$ et b . Cette écriture du problème implique le réglage de $v + 1$ paramètres, v étant la dimension du vecteur \mathbf{x}_i . Cette formulation, pose problème dans le cas où nous aurons à traiter des données de grandes dimensions (fléau de la dimensionalité¹). Pour éviter cela, Il est alors judicieux de reformuler le problème dans une forme dite *formulation duale*.

¹Curse of dimensionality en anglais.

V.6 Liste des publications

Ouvrages

- A. Koita, D. Daucher : "Distance Evaluation Between Vehicle Trajectories and risk indicator". Chapter in book : Modelling, Simulation and Optimization, ISBN 978-953-307-048-3, Publisher :INTECH, pages :148-166 ,Février 2010.

Revue internationale

- A. Koita, A. Boubezoul, D. Daucher : "Predictive model of vehicle failure trajectories using Support Vectors Machines ". Control Engineering Practice, Elsevier, (soumis 2010) .
- D. Daucher, A. Koita, M. Fogli : " Stochastic modelling of vehicle trajectories in a bend by identification techniques ". IEEE Transportation Research - Part B, (soumis 2010) .

Conférences internationales avec comité de sélection

- A. Koita, D. Daucher, " Stochastic Analysis of Vehicle Trajectories in Bend : toward a risk indicator construction", in : Proceedings of the 10th International Conference on Structural Safety and Reliability (ICOSSAR 09), , Osaka (Japan) , September 2009.
- D. Daucher, A. Koita, "Modélisation de trajectoires de véhicules de véhicules en virage par analyse de données expérimentales - application à la sécurité routière", in : Proceedings of the 19th Congrès Français de Mécanique (CFM 09), Marseille (France), Août 2009.

- A. Koita, D. Daucher, A. M. Caucci, "Distance Evaluation Between Vehicle Trajectories in the Functional Space", in : Proceedings of International Association of Science and Technology for Developments, IASTED Applied Simulation and Modelling, Corfu (Greece) , June 2008.
- D. Daucher, B. Jacob, A. Koita, "Trajectoires de véhicules et construction d'indicateur de risque", in : Proceedings of Congrès International ATEC-ITS France, Paris (France), Février 2009.
- D. Daucher, A. Koita, M. Fogli, "Stochastic analysis of trajectories of vehicles in bend - Estimation of probability of failure", the 8th International Conference on Structural Dynamics (Eurodyn 11), Leuven (Belgium), July 2011.
- A. Koita, D. Daucher, M. Fogli, " Vehicle Risk Level Estimation By Using Experimental Trajectories", International Mechanical Engineering Congress et Exposition, ASME IMECE 2011, USA.
- A. Boubezoul, A. Koita , D. Daucher, "Vehicle Trajectories Classification using Support Vectors Machines for Failure Trajectory Prediction", in : Proceedings of the 8th IEEE International Conference on Advances in Computational Tools for Engineering Applications (ACTEA 09), Beyrouth (Liban), July 2009.

Conférences nationales avec comité de sélection

- A. Koita, D. Daucher, M. Fogli, "Prédiction des trajectoires défailtantes en utilisant des techniques d'identification stochastique", dans les Proceedings de la Conférence Sécurité Routière : Prévention des Risques et Aides à la Conduite, PRAC 2010, Paris (France), Mai 2010.
- A. Koita, D. Daucher, M. Fogli, "Analyse et modélisation stochastiques des trajectoires de véhicule en virage en vue de la construction des indicateurs de risque : Application au diagnostic d'infrastructure routière" , GTAA-GdR MACS CNRS, Mulhouse (France), Mai 2010.

Bibliographie

- [1] M. Basset, J. Baujon, and G.-L. Gissinger, *Intelligent driver aid system based on trajectory reconstruction for active safety.*, IFAC World Congress **1** (1999), 467–472.
- [2] M. Basset, B. Zami, G.L. Gissinger, and P. Baggio, *Problématique de l'identification de la liaison véhicule-sol*, Journées Nationales de la Recherche en Robotique (2005).
- [3] P. Bernard, *Simulation numérique de processus et champs aléatoires*, Cours du Collège de Polytechnique (2009).
- [4] P. Bernard and C. Bonnemoy, *An algorithm for spectral factorization using random search techniques*, Probabilistic Engineering Mechanics **4** (1989), 67–72.
- [5] P. Bernard, M. Fogli, P. Bressolette, and M. Lemaire, *A stochastic algorithm of simulation by approximate markovianisation. application to random mechanics*, Journal de Mécanique théorique et Application **3,6** (1984), (905–950).
- [6] P.J. Bickel and M. Rosenblatt, *On some global measures of the deviations of density function estimates*, Ann. Statist **1** (1973), 1071–1095.
- [7] P. Blanquart and J.C. Roncin, *Fiabilité*, Technique de l'ingénieur (1981).
- [8] V. Blanz, B. Scholkopf, H. Bulthoff, C. Burges, V. Vapnik, and T. Vetter, *Comparison of view based object recognition algorithms using realistic 3d models*, Proceedings of the International Conference on Artificial Neural Networks ICANN96 **1** (1996), 251–256.
- [9] B.E. Boser, I. Guyon, and V. Vapnik, *A training algorithm for optimal margin classifiers*, Computational Learning Theory **1** (1992), 144–152.
- [10] A. Boubezoul, *Système d'aide au diagnostic par apprentissage : application aux systèmes microélectroniques*, Ph.D. Thesis, 2008.
- [11] N. Bouleau, *Processus stochastiques et application* (Hermann, ed.), 1988.
- [12] ———, *Probabilités de l'ingénieur : Variables aléatoires et simulation* (Hermann, ed.), 2002.
- [13] M. Bouteldja, *Modélisation des interactions dynamiques poids lourd/infrastructures pour la sécurité et les alertes*, Ph.D. Thesis, 2005.

Bibliographie

- [14] T. Brenac, *Note technique sur les relations vitesse courbe devers. liaisons interurbaines*, SETRA, France (1984).
- [15] J.-P. Brossard, *Vehicle dynamic and complex system modelling*, 2006.
- [16] C. Burges and B. Scholkopf, *Improving the accuracy and speed of support vector machines 9*, Computational Learning Theory **1** (1997), 375–381.
- [17] C. Canudas-De-Wit, P. Tsiotras, E. Velenis, M. Basset, and G. Gissinger, *Dynamic friction models for road/tire longitudinal interaction*, Vehicle System Dynamic **3** (1996), 189–226.
- [18] D.-S. Chang, *Reliability bounds for the stress-strength model*, Computers ind. Engineering (1995).
- [19] K. Chao, *Measurements of path and others parameters in motor vehicle dynamics tests and their errors*, Vehicle System Dynamic **26** (1996), 321–342.
- [20] S. Chen, E. Nikolaidis, and H.-H. Cudney, *Comparison of probabilistic and fuzzy set methods for designing under uncertainty*, American Institute of Aeronautics and Astronautics, 99-1579 (1999).
- [21] D. Comaniciu and P. Meer, *Mean shift analysis and applications*, Proceedings of the International Conference on Computer Vision **2**, p.1197 (1999).
- [22] D. Comaniciu, V. Ramesh, and P. Meer, *The variable bandwidth mean shift and data-driven scale selection*, 2001.
- [23] C. Cortes and V. Vapnik, *Support-vector networks* **20,3** (1995), 273–297.
- [24] J. Dakhlallah, *Risque intégré de sortie de route et proposition d'une assistance à la conduite*, Ph.D. Thesis, LCPC-INRETS, 2009.
- [25] P.J. Davis and P. Rabinowitz, *Methods of numerical integration*, Academic Press, New York (1975).
- [26] Y. Delanne, M.-T. Do, M. Gothié, and G. Delalande, *Potentiel d'adhérence du couple pneumatique/chaussée*, Collection Etudes et recherches des LPC-série Routes- **43** (2006).
- [27] E. Diday, *Nouvelle méthode en classification automatique et reconnaissance des formes. la méthode des nuées dynamiques*, Revue de Statistiques Appliquées **XX,2** (1971), 19–33.
- [28] G. Dreyfus, J.M Martinez, M. Samuelides, M. B. Gordon, F. Badran, S. Thiria, and L. Héroult, *Réseaux de neurones : méthodologie et applications*, Eyrolles, 2nd edition (2004).
- [29] DSCR, *Direction de la sécurité et de la circulation routières, bilan de l'insécurité routière* (2008).
- [30] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, John Wiley et Sons (2001).
- [31] G. Dupré, O. Floris, and L. Patte, *Amélioration de la sécurité des virages des routes principales en rase campagne*, CETE Normandie - Centre et SETRA (2002).
- [32] T. Echaveguren, M. Bustos, and H. de Solminihac, *Assessment of horizontal curves of an existing road using reliability concepts*, Ph.D. Thesis, 2004.
- [33] Eurocodes, *Bases de calcul des structures*, [http](http://eurocodes.org) : (2008).
- [34] M. Faure, *Cours de routes à l'entpe*, Aléas (1971).
- [35] F. Fernandez and J. Yerpez, *Caractéristiques routières et sécurité : reconnaissance de la contribution des facteurs route dans la genèse des accidents*, Rapport technique, INRETS.
- [36] M. Fogli, *Cours de mécanique aléatoire*, DEA, Université Blaise Pascal de Clermont-Ferrand (2000).
- [37] ———, *Simulation numérique de processus et champs aléatoires gaussiens vectoriels stationnaires et non stationnaires*, Cours du Collège de Polytechnique (2009).

Bibliographie

- [38] K. Fukunaga and L. Hostetler, *The estimation of the gradient of a density function, with applications in pattern recognition.*, IEEE Transactions on Information Theory **21(1)** (1975), 32–40.
- [39] T.-D. Gillespie, *Fundamentals of vehicle dynamic*, SAE (1992), 487–496.
- [40] Y. Girard, *Bilan de l'étude détaillée des accidents : fiche de synthèse*, INRETS, Institut National de Recherche sur les Transports et leur Sécurité, France (2002).
- [41] S. Glaser, *Modélisation et contrôle d'un véhicule en trajectoire limitée : application au développement d'un système d'aide à la conduite*, Ph.D. Thesis, 2004.
- [42] Y. Goyat, *Estimation des trajectoires de véhicule par un système optique*, Ph.D. Thesis, 2008.
- [43] J.-W. Hall and J. Lawry, *Generation, combination and extension of random set approximations to coherent lower and upper probabilities.*
- [44] H. Imine, *Observation d'état d'un véhicule pour l'estimation du profil dans les traces de roulement*, Thèse Université de Versailles St-Quentin (2003).
- [45] IRTAD, *International traffic safety data and analysis : Roads accident statistics*, [http](http://www.irtad.net) : (2001).
- [46] B. Jacob, *Lois de valeurs extrêmes des variables aléatoires et des processus*, Cours ENPC (1989).
- [47] C. Joly, D. Bétaille, and F. Peyret, *Étude comparative des techniques de filtrage non-linéaire appliquées à la localisation 2d d'un véhicule en temps réel*, Revue Traitement du signal (2008).
- [48] Y. Kanayama and N. Miyake, *Trajectory generation for mobile robots*, Robotic Research : The Third International Symposium (1986), 333–340.
- [49] Y.-J. Kanayama and B.-I. Hartman, *Smooth local-path planning for autonomous vehicles*, International Journal of Robotic Research **16** (1997), no. 3, 263–284.
- [50] P.E. Kloeden and E. Platen, *Numerical solution of stochastic differential equation*, Springer, Application of Mathematics (1999).
- [51] T. Kohonen, *Learning vector quantization for pattern recognition*, Technical Report (1986).
- [52] A. Koita and D. Daucher, *Stochastic analysis of vehicle trajectories in bend : toward a risk indicator construction*, Icosar 2009, international conference on structural safety and reliability, 2009.
- [53] K. Komoriya and K. Tanie, *Trajectory control and design of wheel-type mobile robot using b-spline curve*, IEEE/RSJ International Workshop on Intelligent Robots and Systems (1989), 398–405.
- [54] P. Krée and C. Soize, *Mécanique aléatoire* (Dunod, ed.), 1983.
- [55] P. Kree and C. Soize, *Markovianization of non linear oscillators with colored input.* (1982).
- [56] J. Ph. Lauffenburger, M. Basset, F. Coffin, and G. Gissinger, *A vehicle path-planning algorithm using polar polynomials optimised through fuzzy-logic*, Ifac conference on mechatronic systems, darmstadt, allemagne, 2000, pp. 139–144.
- [57] J.Ph. Lauffenburger, M. Basset, F. Coffin, and G. Gissinger, *Driver-aid system using path-planning for lateral vehicle control*, Journal of Control Engineering Practice **11-2** (2003), 215–229.
- [58] X.-R. Lei, Z. Ren, W.-Y. Huang, and Bi-Y. Chen, *Fuzzy reliability analysis of distribution systems accounting for parameters uncertainty.*
- [59] M. Lemaire, *Fiabilité des structures, couplage mécano-fiabiliste statique*, Hermes, 2005.
- [60] G. Louah, O. Menacer, and E. Violette, *La vitesse pratiquée ou v85, formule de calcul*, Note d'information du Sétra - Série Conception Sécurité Equipements Exploitation - 127 (2008).

Bibliographie

- [61] S. Mammarr, *Contrôle latéral assisté et automatisé des véhicules : approches par commande robuste*, HDR (2001), 7–21.
- [62] R.-E. Melchers, *Structural reliability analysis and prediction* (Wiley Second Edition, ed.)
- [63] R.-A.-R. Mendoza, *Sur la methode et la commande des véhicules automobiles*, Ph.D. Thesis, 2004.
- [64] Z. Messaoudene, *Contribution au développement d'un générateur de trajectoires pour véhicules routiers*, Ph.D. Thesis, 1999.
- [65] T. Mitchell, *Machine learning*, McGraw Hill, New York (1997).
- [66] P. Mitra, C. Murthy, and S. Pal, *Unsupervised feature selection using feature similarity*, IEEE Transactions on Pattern Analysis and Machine Intelligence (1979).
- [67] Glossaire MTT, *Métronologie des trajectoires et du trafic*, Opération de recherche du LCPC (2010).
- [68] H. Nachtmann and J.-R. Chimka, *Fuzzy reliability in conceptual design*, Annual Reliability and Maintainability Symposium (2003).
- [69] M. Nadji, *Adéquation de la dynamique du véhicule à la géométrie des virages routier : Apport pour la sécurité*, Ph.D. Thesis, 2007.
- [70] J.P. Nakache and J. Confais, *Approche pragmatique de la classification*, Technip **1** (2005).
- [71] R.B. Nelsen, *An introduction to copulas* (Lewis Springer Series in Statistics and Clark, eds.), 2006.
- [72] M. Ortiz, F. Peyret, and D. Betaille, *Mesures de référence des trajectoires : description et performances*, LCPC/MI, Nantes (2008).
- [73] H. Pacejka and E. Bakker, *The magic formula tyre model*, Colloque on tyre models for vehicle dynamics analysis **1** (1991), 1–18.
- [74] H. Peng and M. Tomizuka, *Preview control for vehicle lateral guidance in highway automation*, Journal of dynamic systems, measurement and control **1** (1993), 679–686.
- [75] F. Poirion, B. Puig, and C. Soize, *Non-gaussian simulation using hermite polynomial expansion*, Computational Stochastic mechanics **1** (2003).
- [76] F. Poirion and C. Soize, *Simulation numérique de champs vectoriels stochastiques gaussiens homogènes et non homogènes*. **1** (1989), 41–61.
- [77] V. Pugachev, *Théorie des probabilités et statistique mathématique*, 1982.
- [78] B. Puig, *Modélisation et simulation de processus non gaussien*, Ph.D. Thesis, 2003.
- [79] B. Puig, F. Poirion, and C. Soize, *Non-gaussian simulation using hermite polynomial expansion : convergences and algorithms* **17** (2002), 253–264.
- [80] R.-K. Reddy and A. Haldar, *A random-fuzzy reliability analysis*, 1st Int. Symp. on Uncertainty Modeling and Analysis **7** (1990), 161–166.
- [81] A. Revue, *Contribution des systèmes d'informations géographiques à la sécurité routière : Evaluation de configurations à risques*, Ph.D. Thesis, 2003.
- [82] G. Rey, D. Clair, and M. Fogli, *Reliability analysis of roadway departure risk using stochastic processes*, Mechanical Systems and Signal Processing, Elsevier (2010).
- [83] A. Riedel, *A model of the real driver for use in vehicle dynamics simulation models*, SAE Journal **1** (1993), 94–109.

Bibliographie

- [84] F. Rosenblatt, *The perceptron : A probabilistic model for information storage and organization in the brain*, Psychological Review **65** (1958), 386–408.
- [85] J.-A. Rothengatte, H. Alm, M.-J. Kuiken, J.-A. Michon, and W.-B. Verwey, *The driver, generic intelligent driver support*, Taylor and Francis Editions, 1993.
- [86] P. Rousset, *Application des algorithmes d'auto-organisation à la classification et à la prévision*, Université Paris-Panthéon Sorbonne (1999).
- [87] D. Sammier, *Sur la modélisation et la commande de suspension de véhicules automobile*, Ph.D. Thesis, 2001.
- [88] G. Saporta, *Théorie et méthodes de la statistique* (Paris Technip, ed.), 1978.
- [89] A.-R. Savkoor and S. Ausejo, *Analysis of driver's steering and speed control strategies in curve negotiation*, Vehicle System Dynamics **33** (1999), 94–109.
- [90] B. Scholkopf, C. Burges, and A. Smola, *Introduction to support vector learning, advances in kernel methods*, Support Vector Learning **1** (1998), 1–22.
- [91] Y. Sellami, *Approche fiabiliste et mécanique pour la prédiction de risques d'accident de poids lourds*, Ph.D. Thesis, 2008.
- [92] A. Smola and B. Scholkopf, *A tutorial on support vector regression*, Technical Report NeuroCOLT NC-TR-98-030 (1998).
- [93] C. Soize, *Méthodes mathématiques en analyse du signal*, 1993.
- [94] ———, *The fokker-planck equation for stochastic dynamical systems and its explicit steady state solutions*, World Scientific, Singapore (1994).
- [95] J. Stephant, A. Charara, and D. Meizel, *Contact roue-sol comparaison de modèles d'efforts* (2001).
- [96] B. Sudret, *Eléments finis stochastiques spectraux et fiabilité*, Congrès Français de Mécanique (2001).
- [97] A. Takahashi, T. Hongo, and Y. Ninomiya, *Local path planning and motion, control for agv in positioning*, IROS **1 1** (1996), 392–395.
- [98] M. Tomizuka and C. Chen, *Modeling and control of articulated vehicles*, California PATH Research Report UCB-ITS-PRR-97-42 (1997).
- [99] F. Tonon, *A search algorithm for calculating validated reliability bounds*, REC (2004).
- [100] L.-V. Utkin and F.-P.-A. Coolen, *Imprecise reliability : An introductory overview*, Departement of Computer Science, Russia and Dep. Mathematical Sciences, Durham, England (2005).
- [101] V. Vapnik, *An overview of statistical learning theory*, IEEE Transactions on Neural Networks.
- [102] V. Vapnik and S. Mukherjee, *Support vector method for multivariate density estimation*, Advances in Neural Information Processing Systems **12 2** (2000), 659–665.
- [103] V. N. Vapnik, *Estimation of dependences based on empirical data*, Springer-Verlag (1982).
- [104] G.-S. Wasserman, *Reliability verification, testing, and analysis in engineering design*, Marcel Dekker, Inc. New York (2003).
- [105] J. Xiao, P. Borgnat, and P. Flandrin, *Testing stationarity with time-frequency surrogates*, 15th European Signal Processing Conference EUSIPCO-2007, Pologne (2007).
- [106] M.A. Zamora-Izquierdo, D. Bétaille, F. Peyret, and C. Joly, *Comparative study of extended kalman filter, linearised kalman filter and particle filter applied to low-cost gps-based hybrid positioning system for land vehicles*, Int. J. Intelligent Information and Database Systems (2008).

RESUME :

Situé dans le contexte général de la sécurité routière, et plus particulièrement axé sur la sécurité des véhicules légers (VL) en virages, ce travail de thèse a pour objet de proposer une méthodologie fiabiliste de prédiction de trajectoires à risque, basée sur le traitement statistique et la modélisation probabiliste de trajectoires réelles de VL en virages.

La *première partie* du travail concerne la construction de modèles probabilistes simples et robustes représentatifs des trajectoires réelles observées. Ces modèles sont des transformées de processus aléatoires scalaires normalisés du second ordre, faiblement stationnaires, ergodiques et non gaussiens, et permettent de décrire de façon réaliste la variabilité aléatoire observée du triptyque Véhicule-Infrastructure-Conducteur. Ils permettent aussi, par construction, de s'affranchir d'éventuelles difficultés dans l'alimentation des paramètres dominants qui les gouvernent.

La *seconde partie* est consacrée au développement et à la mise en oeuvre d'une stratégie fiabiliste destinée à associer un niveau de risque à chaque trajectoire en entrée de virage. Basée sur l'emploi conjoint de méthodes probabilistes pour la modélisation des incertitudes, fiabilistes pour l'évaluation des niveaux de risque et statistiques pour la classification et le traitement des trajectoires, cette approche est une réponse réaliste au problème posé.

De par sa conception et ses possibilités, la méthodologie fiabiliste proposée est une contribution significative au développement de procédures d'alerte destinées à réduire notablement le nombre d'accidents en virages.

ABSTRACT :

This PhD thesis tackles the general context of road safety, focussing on the safety of light vehicles (LV) in bends. A reliability engineering methodology is proposed to predict dangerous trajectories, based on the statistical processing and probabilistic modelling of actual trajectories in a bend.

In the *first part* of this work, simple and robust probabilistic models are built to describe trajectories measured in an instrumented bend. The models are transforms of scalar normalized second order stochastic processes which are slightly stationary, ergodic and non-Gaussian. They offer a realistic description for the observed random variability of the Vehicle-Infrastructure-Driver system. They also inherently circumvent possible difficulties in the identification of the dominant parameters which control the system.

The *second part* of this work is devoted to the development and implementation of a reliability engineering strategy intended to associate a risk level to each trajectory at a bend entry. Based on the joint use of probabilistic methods for modelling uncertainties, reliability engineering for assessing risk levels and statistics for classifying and processing the trajectories, this approach provides a realistic answer to the tackled problem.

From its design and its possibilities, the proposed reliability engineering methodology constitutes a significant contribution to the development of warning procedures the deployment of which are expected to notably reduce the number of accidents in bends.