



Fusion des indicateurs de sécurité routière : FUSAIN

Fusion der Indikatoren für Verkehrssicherheit: FUSAIN

Fusion of Safety Indicators: FUSAIN

**École Polytechnique Fédérale de Lausanne (EPFL)
Laboratoire des voies de circulation (LAVOC)
Minh Hai Pham, Dr.
Edward Chung, Prof.
André-Gilles Dumont, Prof EPF**

**French institute of science and technology for transport, development and networks (IFSTTAR)
Laboratoire Ingénierie Circulation Transport (LICIT)
Nour-Eddin El Faouzi, Dr.
Olivier de Mouzon, Dr.
Solenne Durand
Fouad Baouche**

Mandat de recherche ASTRA 2006/014 sur demande de l'Office fédéral de routes (OFROU)

Der Inhalt dieses Berichtes verpflichtet nur den (die) vom Bundesamt für Strassen beauftragten Autor(en). Dies gilt nicht für das Formular 3 "Projektabschluss", welches die Meinung der Begleitkommission darstellt und deshalb nur diese verpflichtet.

Bezug: Schweizerischer Verband der Strassen- und Verkehrsfachleute (VSS)

Le contenu de ce rapport n'engage que l' (les) auteur(s) mandaté(s) par l'Office fédéral des routes. Cela ne s'applique pas au formulaire 3 "Clôture du projet", qui représente l'avis de la commission de suivi et qui n'engage que cette dernière.

Diffusion : Association suisse des professionnels de la route et des transports (VSS)

Il contenuto di questo rapporto impegna solamente l' (gli) autore(i) designato(i) dall'Ufficio federale delle strade. Ciò non vale per il modulo 3 «conclusione del progetto» che esprime l'opinione della commissione d'accompagnamento e pertanto impegna soltanto questa.

Ordinazione: Associazione svizzera dei professionisti della strada e dei trasporti (VSS)

The content of this report engages only the author(s) commissioned by the Federal Roads Office. This does not apply to Form 3 'Project Conclusion' which presents the view of the monitoring committee.

Distribution: Swiss Association of Road and Transportation Experts (VSS)



Bundesamt für Strassen

Office fédéral des routes

Ufficio federale delle Strade

Fusion des indicateurs de sécurité routière : FUSAIN

Fusion der Indikatoren für Verkehrssicherheit: FUSAIN

FUSion of SAfety Indicators: FUSAIN

**École Polytechnique Fédérale de Lausanne (EPFL)
Laboratoire des voies de circulation (LAVOC)
Minh Hai Pham, Dr.
Edward Chung, Prof.
André-Gilles Dumont, Prof.**

**French institute of science and technology for transport, development
and networks (IFSTTAR)
Laboratoire Ingénierie Circulation Transport (LICIT)
Nour-Eddin El Faouzi, Dr.
Olivier de Mouzon, Dr.
Solenne Durand
Fouad Baouche**

**Mandat de recherche ASTRA 2006/014 sur demande de l'Office fédéral
de routes (OFROU)**

Impressum

Service de recherche et équipe de projet

Direction du projet

André-Gilles Dumont

Membres

Minh Hai Pham
Nour-Eddin El Faouzi
Olivier de Mouzon
Solenne Durand
Fouad Baouche
Edward Chung

Commission de suivi

Président

Patrick Maillard

Membres

Michel Bierlaire, Prof EPF
Gerhard Petersen
Pierre Schneider
Panos Tzieropoulos

Auteur de la demande

Office fédérale des routes (OFROU)

Source

Le présent document est téléchargeable gratuitement sur <http://partnershop.vss.ch>.

Table des matières

	Impressum	4
	Table des matières	5
	Résumé	7
	Zusammenfassung	16
	Summary	25
1	Introduction	32
1.1	Motivation	32
1.2	Objective	32
1.3	Project FUSAIN	32
1.4	Report Structure	33
2	Traffic Safety – State of the Art	34
2.1	Introduction.....	34
2.2	Human Factor.....	35
2.3	Risk Identification	35
2.3.1	Risk Indicators.....	35
2.3.2	Risk Identification Models	37
3	Methodology, Study Sites and Data	40
3.1	Methodology.....	40
3.2	Study Sites and Data	40
3.2.1	Swiss Motorways Network	40
3.2.2	Traffic Detectors	41
3.2.3	Meteorological Data	42
3.2.4	MeteoSwiss Stations	43
3.2.5	Summary	43
3.3	Preliminary analysis of accidents	46
3.3.1	Crash Distribution by Time of the Day and Day of the Week	46
3.3.2	Crash Distributions by Weather Conditions and Pavement Conditions	47
3.3.3	Crash Distributions by Crash Severity	48
3.3.4	Crash Statistics at Study Site CH023.....	49
3.4	Potential Risk Indicators.....	51
4	Sensitivity analyses of risk indicators using field data	53
4.1	Objective	53
4.2	Factors Influencing Traffic Safety.....	53
4.3	Sensitivity analyses	54
4.3.1	Non-crash Cases.....	54
4.3.2	Crash cases	63
4.4	Conclusions	72
5	Data Sampling & Traffic Regimes	73
5.1	Introduction.....	73
5.1.1	Motivation & Objective	73
5.1.2	Data Sampling Framework.....	73
5.1.3	Traffic Situation Characterization	73
5.1.4	Pre-crash and Non-crash Traffic Situations	75
5.1.5	Traffic Regimes	75
5.2	Traffic Regime Analyses	77
5.2.1	Overview	77
5.2.2	Traffic States on Two Lanes	79
5.2.3	Traffic State Variations and Non-traffic Characteristics	81
5.2.4	Examples of Traffic Regimes	82
5.2.5	Summary	83
5.3	TS Transitions	84
5.3.1	Introduction.....	84
5.3.2	NTS Patterns.....	84
5.3.3	PTS Patterns	85

5.4	Summary.....	87
6	Fusion of Risk Indicators	88
6.1	Introduction of data fusion.....	88
6.2	Fusion Schemes	89
6.2.1	Feature-Based Inference	90
6.2.2	Cognitive-based models	94
6.3	Risk Indicators Fusion.....	95
6.3.1	Dempster -Shafer technique of data fusion	95
6.3.2	Implementation of the Dempster Shaffer technique	97
6.3.3	Fusion of risk indicators with Dempster-Shaffer Framework	98
6.4	Results	102
6.4.1	Regime-based Risk Identification Models (RIM).....	102
6.4.2	Crash Risk Prediction Model.....	103
7	Conclusions.....	104
7.1	Summary.....	104
7.2	Applications.....	104
7.3	Perspectives.....	105
	Acknowledgement.....	107
	Annexes	108
	Abréviations.....	157
	Bibliographie.....	159
	Clôture du projet.....	161
	Index des rapports de recherche en matière de route	164

Résumé

La présente étude vise à contribuer à l'amélioration de la sécurité autoroutière par l'élaboration de modèles capables d'identifier les risques de trafic en temps réel. L'approche consiste à classer des indicateurs de risque potentiels et à les tester avec des données historiques sur, notamment, la circulation, les conditions météorologiques et les collisions. En premier lieu, les indicateurs de risque sont sélectionnés sur la base de leur sensibilité éprouvée par rapport aux données réelles. On distingue ensuite deux types de conditions de trafic : celles qui mènent à des collisions, appelées « situations pré-accident », et celles où aucune collision ne se produit, appelées « situations sans accident ». Les situations pré-accident et sans accident sont ensuite mises en correspondance pour que, dans chaque groupe, une situation pré-accident et une sans accident soient comparables. Chacun de ces groupes est considéré comme un régime de trafic. On élabore ensuite un modèle à une seule variable pour chaque régime de trafic, et la fusion des résultats de chacun de ces modèles permet de réaliser l'étape suivante, soit le modèle définitif.

Une dernière étape visant à prédire les risques de collision à court terme est effectuée en combinant le risque de plusieurs situations récentes de trafic en temps réel. La confrontation avec les données historiques montre que le modèle prédictif ainsi élaboré présente un haut degré d'exactitude. L'écart de vitesse apparaît comme le plus important facteur de risque de collision. Selon la densité du trafic, la différence de vitesse entre deux voies peut être réduite pour diminuer le risque de collision. Il ressort en outre qu'un fort pourcentage de collisions avérées aurait pu être repéré au moins 5 minutes avant l'accident.

Bien que des essais plus étendus soient nécessaires, le système de prédiction à court terme des risques d'accident de la route démontre son potentiel pour aider les opérateurs de la circulation à adopter un système de gestion du trafic orienté sur les risques.

Mots-clés: *indicateur de risque, modèle d'identification des risques, mesures préventives, prévention des accidents, indicateur de sécurité*

Introduction

Généralités

La sécurité routière devient une préoccupation majeure dans de nombreux pays. Bien que la Suisse compte parmi les pays les plus sûrs dans ce domaine, on déplore toujours de nombreux morts et blessés sur ses routes. Afin de contribuer à une amélioration de la sécurité routière en Suisse, la présente étude vise à établir des indicateurs de risque capables de fournir aux opérateurs de la circulation et aux usagers de la route des informations sur les risques d'accident qu'ils encourent. Les données de plusieurs indicateurs de risque sont fusionnées en un unique modèle de détection des risques apte à identifier des informations sur les risques en temps réel et de prédire les risques de la circulation à court terme.

La présente étude commence par recenser les indicateurs de risque documentés dans la littérature, en vue d'élaborer de nouveaux indicateurs. Tous les indicateurs de risque sont ensuite testés sur des données de terrain issues du réseau routier suisse. Une fois mesurée la performance de chaque indicateur, ceux-ci sont ensuite fusionnés en des modèles uniques afin d'en optimiser l'efficacité globale. L'étude suggère enfin des mesures préventives adéquates pour limiter les risques d'accident.

Contexte

Les études sur la sécurité routière peuvent être rangées dans trois grandes catégories selon que l'objet de la recherche est l'infrastructure, l'être humain ou les véhicules. Les facteurs humains jouent habituellement un rôle important dans les cas d'accident. Certains de ces facteurs ne se limitent pas à l'état du conducteur juste avant l'accident (sous l'influence de l'alcool ou de drogues, par exemple), mais sont dus aux limites générales de

l'être humain, comme l'incapacité à prédire les futures conditions de circulation. Les nouvelles technologies contribuent donc à pallier ces carences de l'être humain.

Dans cette mesure, l'une des tendances en sécurité routière est de quantifier les situations à risque sur la route au moyen d'indicateurs de risque. Cette mesure peut se baser sur des données effectives obtenues par des capteurs sur la route tels les caméras de surveillance, détecteurs à boucle, etc. Un grand nombre d'indicateurs de risque a été développé dans ce but. Parmi les indicateurs largement utilisés, citons le temps de collision TTC (*Time-To-Collision*), l'indicateur de collision potentielle en cas de freinage d'urgence PICUD (*Potential Index for Collision with Urgent Deceleration*), le TIT (*Time-Integrated-TTC*), le TET (*Time-Exposed-TTC*), etc. Au départ, il s'agissait d'indicateurs théoriques créés en laboratoire. Par la suite, avec le développement de la technologie de détection du trafic, ces indicateurs ont pu être appliqués à des données de circulation réelles. Cette nouvelle technologie a en outre ouvert la voie à une nouvelle tendance d'études désagrégées permettant d'étudier les accidents individuels au moyen des données de trafic récoltées.

Plusieurs études ont été menées pour tenter de détecter en temps réel les situations à risque sur l'autoroute. Celles-ci visent en principe à isoler les conditions de circulation qui précèdent immédiatement une collision, dites situations pré-accident ou PTS (*pre-crash traffic situations*). Ces PTS sont considérés comme situations à risque car elles ont donné lieu à une collision. Les autres conditions de circulation où aucune collision n'est observée sont appelées situations sans accident ou NTS (*non-crash traffic situations*). Les modèles de détection des situations à risque sur la route sont élaborés par comparaison entre les situations NTS et PTS.

Méthodologie et zone d'étude

Méthodologie

L'examen d'études désagrégées met au jour plusieurs inconvénients. La méthodologie proposée pour la présente étude vise à atteindre les buts de celle-ci en répondant aux problèmes suivants :

- Le déséquilibre entre populations de NTS et de PTS. Les accidents étant rares, le nombre de PTS est nettement inférieur à celui des PTS.
- L'arbitraire du processus d'échantillonnage des NTS pour comparaison avec les PTS. Pour élaborer un modèle capable de comparer NTS et PTS, les cas NTS sont habituellement réduits à une population plus faible pour assurer que le nombre de NTS corresponde à celui de PTS. Toutefois, dans les études existantes, ce processus d'échantillonnage est arbitraire.

Conformément aux objectifs de ce projet, l'approche comporte les étapes suivantes :

- 1-Identification des indicateurs de risque applicables. Les indicateurs choisis sont testés avec des données réelles afin de vérifier leur aptitude à différencier les conditions de circulation menant à des collisions (situations à risque) des autres conditions.
- 2-Application d'un processus d'échantillonnage des données pour les cas sans accident. En effet, les collisions sont rares, et celles par l'arrière ou latérales ne se produisent que dans des circonstances particulières.
- 3-Elaboration de modèles d'identification des risques à une seule variable.
- 4-Fusion des résultats des modèles à une seule variable pour obtenir des modèles agrégés plus efficaces.
- 5-Prédire les risques de circulation à court terme en utilisant les résultats des modèles fusionnés dans un cadre en temps réel.

La première étape est importante en raison du grand nombre d'indicateurs de risque, dont certains ne sont pas directement applicables en fonction des données disponibles. A la deuxième étape, l'objectif est de mettre en relation les conditions de circulation pré-accident avec les conditions sans accident correspondantes. En effet, les collisions sont des événements rares et il faut un contexte particulier pour les collisions par l'arrière ou latérales. En échantillonnant les données sans accident, seules les conditions sans accidents

comparables à des conditions pré-accident sont retenues pour traitement. Le processus d'échantillonnage des données crée des groupes de conditions de circulation appelés régimes de trafic, afin que pour chaque régime on dresse une liste de conditions sans accident et de conditions pré-accident. La troisième étape est appliquée à chaque régime de trafic dont les situations pré-accident et sans accident sont pertinentes. Les variables peuvent être les indicateurs de risque choisis ou les paramètres de circulation. Chaque variable a une performance donnée pour l'identification des risques de collision. L'étape 5 est également menée pour chaque régime de trafic dans le but de combiner les performances de toutes les variables en fusionnant les modèles à une seule variable. En dernier lieu, afin de prédire le risque de collision de la prochaine situation de trafic, on évalue les risques de plusieurs situations de trafic antérieures par le biais des modèles de détection agrégés afin de déterminer si la prochaine situation de trafic est à risque ou non.

Zone d'étude

La Suisse dispose de plus de 300 doubles détecteurs de trafic capables de fournir des données sur des véhicules individuels. La distance entre ces détecteurs est telle que la zone d'étude doit se restreindre à un seul emplacement. La zone d'étude est choisie selon la procédure suivante :

- Sur la liste des plus de 300 capteurs fournissant des données sur la circulation, compter pour chaque emplacement le nombre de collisions satisfaisant aux conditions dans un rayon tampon de 1 km, soit les accidents considérés. Mettre la liste par ordre descendant du nombre d'accidents considérés.
- Les accidents de la route étant du ressort des polices cantonales (26 cantons en Suisse), les rapports d'accident dans plusieurs cantons, comme Vaud ou le Tessin par exemple, peuvent omettre de détailler l'emplacement des collisions. Les emplacements de ces cantons sont donc éliminés de la liste.
- Eliminer de la liste ainsi ordonnée tous les capteurs qui ne disposent pas d'une station météorologique dans un rayon de 15 km. Ceux qui restent figurent sur la liste restreinte des emplacements.
- Sélectionner le premier emplacement de la liste restreinte.

Le site finalement sélectionné figure en 27^{ème} position sur la liste considérée. La zone d'étude porte le nom de la station de détection de trafic : le site CH023 est situé sur l'autoroute A1 reliant Berne à Zürich. S'il ne s'agit pas de l'emplacement répertoriant le plus grand nombre de collisions, le site CH023 a été choisi en raison de la disponibilité combinée de données sur la circulation, météorologiques et sur les accidents, ce qui est essentiel pour l'élaboration de modèles d'identification des risques avec la méthodologie proposée pour cette étude.

Résultats et analyses

Analyse préliminaire des accidents

Sur la base des données d'accidents, les observations suivantes s'imposent quant aux collisions sur les autoroutes suisses :

- Le nombre le plus élevé d'accidents a lieu le vendredi, puis le samedi. Les accidents sont plus fréquents en fin d'après-midi le vendredi (entre 16h et 19h) qu'à tout autre moment. Bien que la fréquence des accidents soit élevée le samedi, ceux-ci ne se produisent pas à une heure particulière. Les accidents tôt le dimanche matin sont plus fréquents qu'à la même heure de tout autre jour de la semaine.
- Plus de 3/4 des accidents se produisent pendant des précipitations (pluie ou neige). Près de 2/3 des accidents ont lieu sur route sèche.
- En termes de gravité de l'accident, le nombre d'objets impliqués, le nombre de personnes impliquées et le nombre de blessures sont trois représentatifs de la gravité. On observe qu'il y a des accidents individuels impliquant des nombres élevés de véhicules et de personnes.
- Dans la zone d'étude, il y a un nombre élevé de collisions par l'arrière, suivi par le nombre de collisions impliquant un seul véhicule. La plupart des accidents dans la zone d'étude se sont produits l'après-midi à l'heure de pointe, par forte densité du trafic.

Sensibilité des indicateurs de risque

Indicateurs de risque potentiels

En fonction des caractéristiques de la zone d'étude qui dictent la disponibilité de données réelles, plusieurs indicateurs de risque sont considérés adéquats pour la présente étude. En définitive, ce sont trois indicateurs qui sont retenus : TTC, PICUD, le risque de temps de freinage individuel IBTR (*Individual Braking Time Risk*), le risque de temps de freinage du peloton PBTR (*Platoon Braking Time Risk*), et la vitesse au-delà de la limite autorisée SOSL (*Speed Over Speed Limit*).

Analyses de sensibilité

Le principal objectif de l'analyse de sensibilité est de déterminer la performance des indicateurs de risque avec des données historiques, notamment sur la circulation, les conditions météorologiques et les accidents. Le test de sensibilité doit permettre de dégager une évaluation de la performance des indicateurs qui serve de ligne directrice pour l'usage ultérieur d'indicateurs de risque aux étapes suivantes de l'étude. Deux facteurs sont susceptibles d'influencer la performance des indicateurs de risque choisis : l'écoulement du trafic et les conditions météorologiques.

L'approche consiste à tester les indicateurs en conditions sans accident et pré-accident, avec et sans précipitations, afin de déterminer la différence de distribution des indicateurs de risque dans chaque cas.

Les indicateurs PBTR, IBTR, et PICUD qui comportent des éléments météorologiques dans leur formule (coefficient de décélération) sont sensibles aux conditions météorologiques. A l'exception du TTC, la plupart des indicateurs sont sensibles à la gamme de débit du trafic. Les indicateurs sont également sensibles aux voies de circulation. Le teste de sensibilité montre que le PICUD est moins fiable, car il indique que la circulation par beau temps est plus dangereuse que par temps de pluie. Le PICUD ne doit donc plus être utilisé dans cette étude.

Le meilleur moyen d'éprouver la performance des indicateurs de risque est de contrôler leur évolution dans les conditions de circulation qui précèdent immédiatement une collision – là où on est sûr qu'il existe un risque avéré de collision. Pour établir si une situation de trafic est risquée ou non avec les cas d'accident testés, il est recommandé de tenir compte non seulement des valeurs liées des indicateurs de risque, mais aussi du pourcentage de véhicules.

Régimes de trafic

Echantillonnage des données

Les collisions étant des événements rares, le nombre de PTS est nettement inférieur à celui des NTS. Les accidents considérés dans la présente étude sont exclusivement des collisions par l'arrière ou latérales. Ce sont des collisions qui ne se produisent que sous certaines conditions, comme par exemple par trafic dense qu'on constate habituellement à l'heure de pointe. Ainsi, pour chaque PTS, de nombreuses situations NTS ne sont pas vraiment adéquates pour la comparaison, comme les NTS qui surviennent de nuit avec peu de véhicules sur la route.

L'échantillonnage des NTS ne sert pas uniquement à réduire le nombre de cas, mais également à mettre en relation les NTS avec les PTS les plus comparables. Le processus d'échantillonnage des données comporte les étapes suivantes :

- 1-Définition de la situation de trafic en général, caractérisée par des variables.
- 2-Définition des situations sans accident et pré-accident.
- 3-Agrégation des situations sans accident en groupes.
- 4-Classement des situations pré-accident dans les groupes ainsi obtenus.

Au terme du processus on obtient des groupes de situations de trafic. Chaque groupe comporte une liste de situations pré-accident et sans accident. Chaque groupe constitue

un régime de trafic car il représente des conditions de circulation similaires.

Une situation de trafic se définit comme étant caractérisée par 22 variables qui représentent l'instantanéité (2 variables), l'état de la circulation sur chaque voie (14 variables), l'écart de vitesse entre les voies (1 variable), l'évolution de la circulation entre les situations de trafic consécutives (4 variables) et les conditions de circulation (1 variable). Une situation de trafic est classée PTS ou NTS selon qu'elle précède ou non un accident. Le temps de collision permet d'identifier les 30 minutes précédant la collision comme la période pré-accident. La période tampon se situe juste avant la période pré-accident et sert à séparer les périodes de trafic précédentes de la période pré-accident. Une période post-accident indique aussi le temps juste après la collision où la circulation revient à la normale. La période tampon pré-accident et la période post-accident ne sont pas considérées dans la présente étude. Toutes les situations de trafic qui ne se situent pas dans les périodes pré-accident, tampon pré-accident ou post-accident sont qualifiées de périodes sans accident.

Le processus de regroupement se fonde sur une méthode de k-moyennes où le nombre de groupes (*clusters*) K est choisi sur la base du taux d'erreur : en augmentant le nombre de groupes, lorsque l'erreur de classification ne baisse pas de façon significative, on prend le nombre précédent de groupes. En définitive, on obtient 8 groupes – huit régimes de trafic nommés de A à H.

En présence des huit régimes de trafic obtenus à partir du processus de regroupement des NTS, les PTS sont classés dans les régimes de trafic en fonction de la similitude entre les PTS et les centres des huit groupes.

Analyse des régimes de trafic

En examinant chacune des 22 variables sous chaque régime de trafic, on aboutit aux constatations suivantes :

- Les régimes A, E et F représentent des conditions de trafic fluide. Ces régimes sont caractérisés par une très faible densité de circulation sur les deux voies.
- Le régime H représente les conditions de haute densité et de congestion. Sous ce régime, la densité de circulation est élevée sur les deux voies. On observe parfois de la congestion, avec une réduction significative de la vitesse moyenne.
- Les régimes B et G représentent des densités de circulation élevées à moyennes. Sous ces régimes, le trafic est à peine moins dense que celui du régime H. Sur la base des variables d'évolution du trafic, on observe que sous le régime B, la circulation augmente, alors qu'elle diminue sous le régime G. Ainsi, le régime B peut être considéré comme un régime de trafic pré-congestion, alors que le régime G peut être qualifié de post-congestion.
- Les régimes C et D représentent des densités de circulation faibles à moyennes. La différence la plus significative entre le régime C et le régime D est la présence de précipitations. Alors que la plupart des situations de trafic classées sous le régime C se produisent dans des conditions météorologiques favorables, des précipitations sont habituellement présentes sous le régime D.

Comme pour les résultats de classification des PTS en régimes de trafic, le nombre de PTS classées sous le régime H est plus important que pour tout autre régime. Après cela les régimes G et B présentent le deuxième, respectivement le troisième nombre le plus important de PTS. Il n'y a en particulier aucune PTS classée sous les régimes de circulation fluide A et F. Seul un petit nombre de PTS tombe sous le régime de faible densité E. Entre les régimes C et D, il y a davantage de PTS sous le régime D que sous le C. Il est possible que les conducteurs observent une plus grande distance par temps de pluie afin d'éviter les collisions par l'arrière ou latérales.

Le taux de risque d'un régime de trafic est le nombre de PTS divisé par le nombre de NTS sous ce régime. Le taux de risque des régimes H et G est plus élevé que pour les autres régimes. Ceci pourrait indiquer que les conditions de circulation sont plus risquées dans des périodes de haute densité ou post-congestion.

Transitions de situations de trafic

La transition de situation de trafic est un changement de situation d'un régime à un autre. Le régime suivant peut être le même ou différent du régime actuel. Comme on utilise des PTS de 5 minutes pour quantifier la circulation précédant une collision, une période pré-accident de 30 minutes se compose donc de six PTS. Un ensemble de six PTS consécutives est donc considéré comme un scénario. Cette section vise à considérer tous les scénarios possibles pour établir lesquels mènent souvent à un accident.

Il y a au total $6^6 = 1'679'616$ scénarios possible. Certaines transitions ne peuvent se produire, comme par exemple de A à H à A (évolution d'une situation sans aucun véhicule vers une circulation très dense puis à nouveau vers une route sans aucun véhicule, dans un laps de temps de 15 minutes, par exemple). C'est pourquoi le nombre total de scénarios de transition NTS sur l'ensemble de la période étudiée est de 11'782 sur les 1'050'372 scénarios disponibles (il faut au moins 30' de données consécutives pour produire un scénario). Certains scénarios se répètent plus souvent que d'autres. Par exemple, le scénario DDDDDD, où le trafic reste sous régime D pendant 30 minutes, est le plus fréquent avec 86'624 répétitions, soit 8.25% des scénarios disponibles. Après cela, les scénarios EEEEE, FFFFFFF, BBBBB et CCCCC représentent respectivement 4.30%, 2.40%, 2.28% et 1.54% des scénarios disponibles. Ce résultat est conforme à un pourcentage élevé de transitions de et vers le même régime pour les régimes B, C, D, E et F. Il s'agit donc de scénarios stables.

Si l'on considère les scénarios de PTS, sur 120 collisions il n'y a que 72 scénarios de PTS. Le scénario HHHHHH est le plus fréquent avec huit occurrences, ce qui signifie que même en cas de circulation stable sous le régime H (où la congestion dure 30 minutes), le risque de collision devient élevé. On compte 64 accidents (soit 53.3%) où le trafic se situe uniquement dans les régimes G et H, à l'exception de tout autre régime. Il s'agit de la fluctuation de trafic entre les régimes H et G.

En comparant les scénarios de PTS et de NTS, deux scénarios de PTS, BBBBB et CCCCC, sont les plus fréquents parmi les scénarios de NTS, avec un taux respectif de 1/23'913 et 1/16'171. Cela signifie que le risque est faible lorsque la circulation obéit à ce scénario. Toutefois, certains scénarios peuvent induire un taux plus élevé de situations pré-accident, comme HGBCHH (un scénario PTS pour quatre scénarios NTS), HHHBHB (deux scénarios PTS pour dix scénarios NTS), CGH-CHC (un scénario PTS pour sept scénarios NTS), et GDDGHB (un scénario PTS pour huit NTS). Si ces scénarios sont rares, ils sont toutefois à haut risque.

Fusion des indicateurs de risque

Fusion de données

Le Data Fusion Technical Committee de la Geoscience and Remote Sensing Society de l'IEEE définit la fusion de données comme un processus d'alliance de données spatiales et temporelles indexées provenant de divers instruments et sources afin d'améliorer le traitement et l'interprétation de ces données.

Les principaux objectifs du projet FUSAIN (Fu sion of Sa fety I ndicators) sont d'améliorer la détection des risques et l'aptitude des indicateurs à caractériser les conditions de sécurité routière. La spécificité du processus de fusion de données pour les indicateurs de sécurité est que la plupart des indicateurs de sécurité/risque sont tirés des mêmes données de circulation, mais sous des formes différentes. Le processus de combinaison est donc effectué sur les mêmes données fournies par une seule source, ce qui correspond à la nouvelle pratique dite de *bagging* des estimateurs. Les avantages attendus en sont non seulement une amélioration de la précision, mais aussi une plus grande fiabilité de l'ensemble du processus de détection des risques.

Parmi les diverses techniques de fusion de données, comme les méthodes d'inférence basées sur des descripteurs (p.ex. approche bayésienne, théorie de l'évidence, modèles paramétriques, réseaux neuronaux artificiels, algorithmes de regroupement, méthodes de vote, etc.) et les modèles cognitifs (modèles logiques, systèmes experts basés sur la connaissance, etc.) à disposition, c'est la méthode de Dempster-Shafer qui a été retenue pour

la présente étude. Suite à l'échantillonnage des situations NTS, il y a des régimes qui ne comportent aucune PTS. La technique de fusion de données n'est donc appliquée qu'aux régimes B, C, G et H, ceux qui comportent un nombre élevé de PTS. Pour chacun de ces régimes, un modèle fusionné à même de détecter les risques du trafic est élaboré sur la base de modèles à une seule variable. Les variables utilisées pour l'élaboration de ces modèles à une seule variable comptent 22 variables caractérisant les situations de trafic ainsi que trois indicateurs de risque : TTC, PBTR et IBTR.

Résultats

Il existe un modèle fusionné d'identification des risques pour chacun des quatre régimes B, C, G, et H. En termes d'identification des risques, ces modèles se révèlent très performants pour différencier les NTS des PTS. Les données (y compris NTS et PTS) sous chaque régime sont partagées en jeux de données d'apprentissage et de validation. Les données d'apprentissage sont utilisées pour former les modèles alors que celles de validation servent à valider les modèles. Pour tous les régimes, les modèles fusionnés d'identification des risques fonctionnent bien avec les données d'apprentissage (identifiant toutes les PTS et au moins 92% de l'ensemble des NTS). Avec les données de validation, bien que les modèles fonctionnent moins bien qu'avec les données d'apprentissage, les pourcentages de NTS et PTS identifiées restent très élevés (au moins 85% des NTS et au moins 77% des PTS). Sur l'ensemble des quatre régimes, 85% des PTS sont correctement identifiées, et le pourcentage des NTS est de 90%. Si l'on considère les situations de trafic des quatre autres régimes (A, D, E, et F), le pourcentage global de NTS identifiées est plus élevé.

Avec ces modèles fusionnés d'identification des risques, une étape supplémentaire est entreprise en vue de prédire l'état des risques de la situation de trafic suivante. Elle est basée sur un concept de longueur de mémoire du risque (*length of risk memory*), qui consiste à se servir, en temps réel, de l'état du risque des précédentes situations à risque pour estimer l'état du risque de la prochaine situation de trafic. La longueur de mémoire du risque est le nombre des situations de trafic précédentes dont l'état du risque identifié est mémorisé. Cette mesure induit un compromis entre le taux de fausse alarme et le taux de dépistage manqué. Avec l'augmentation de la longueur de mémoire du risque, le taux de fausse alarme diminue mais le taux de dépistage manqué augmente, et vice-versa : lorsque la longueur de mémoire du risque se réduit, le taux de fausse alarme augmente alors que celui de dépistage manqué baisse. Il n'y a donc pas de longueur optimale de mémoire du risque qui réduise tant les fausses alarmes que les dépistages manqués. Il est important de noter que la longueur de mémoire du risque est positive et inférieure à 5.

Conclusions

Les travaux présentés ici introduisent l'élaboration de modèles d'identification des risques au moyen de techniques de fusion de données. Les résultats obtenus permettent les conclusions suivantes :

- Les collisions par l'arrière ou latérales se sont produites plus souvent aux heures de pointe, surtout après congestion ou lorsque la circulation fluctue beaucoup entre différents régimes de trafic.
- Séparément, les indicateurs de risque sont moyennement aptes à identifier les risques de la circulation. Or, lorsqu'ils sont fusionnés et qu'on y ajoute d'autres variables sur l'écoulement du trafic, les modèles ainsi obtenus (modèles fusionnés d'identification des risques basés sur le régime de trafic) sont très performants.
- Les modèles fusionnés d'identification des risques peuvent être utilisés en temps réel pour prédire les risques de circulation à court terme. Cependant, un compromis entre les taux de fausse alarme et de dépistage manqué ne permet pas de réduire simultanément ces deux taux. Pour mettre ces modèles en pratique, il est important de réduire le taux de fausse alarme afin d'éviter que les usagers de la route ne perdent confiance dans le système de gestion du trafic.
- La variation de vitesse ressort comme le facteur le plus important qui influence le risque de collision. Selon la densité du trafic, l'écart de vitesse entre deux voies peut être réduit afin de limiter le risque d'accident.

- On observe en outre qu'un pourcentage élevé de collisions avérées pourrait avoir été détecté au moins 5 minutes avant l'accident.
- Les modèles élaborés peuvent fournir aux gestionnaires un outil supplémentaire d'aide à la décision pour la gestion de la circulation. Les informations fournies par ces modèles peuvent également être intéressantes pour orienter les usagers de la route sur leur choix d'itinéraire.

Zusammenfassung

Ziel der folgenden Studie ist, den Strassenverkehr durch die Entwicklung von Modellen sicherer zu machen, die in der Lage sind, Gefahren im aktuellen Verkehrsgeschehen vorherzusagen. Dies erfolgt über eine Klassifizierung möglicher Gefahrenindikatoren und deren Abgleich mit zuvor erhobenen Verkehrs-, Wetter- und Unfalldaten. Zunächst werden die Risikoindikatoren ausgewählt, die in einer Wechselbeziehung zu erhobenen Daten stehen. Danach werden zwei Verkehrssituationen unterschieden: Situationen, in denen Unfälle auftreten, so genannte "pre-crash cases" und Situationen, die nicht zu Unfällen führen, so genannte „non-crash cases“. Pre-Crash-Cases und Non-Crash-Cases werden in Gruppen geordnet, die es erlauben, die Pre-Crash-Situationen mit denen der Non-Crash-Situationen zu vergleichen. Jede dieser Gruppen wird als ein spezifisches Verkehrsgeschehen (Traffic-Regime) betrachtet. Anschliessend werden für jedes Verkehrsgeschehen Modelle mit einer einzigen Variablen entwickelt. Diese dienen als Grundlage für das integrierte Modell, das im nächsten Schritt für jedes Traffic-Regime entwickelt wird und in das die Ergebnisse der Modelle mit einer einzigen Variablen einfließen.

Ein weiterer Schritt, um Verkehrsunfallgefahren kurzfristig vorherzusagen, besteht in der Kombination des Risikoniveaus verschiedener aktueller Verkehrssituationen in Echtzeit. Der Abgleich mit zuvor erhobenen Daten zeigt, dass das daraus entwickelte Modell sehr genaue Vorhersagen ermöglicht. Geschwindigkeitsunterschiede, so hat sich herausgestellt, sind der Faktor, der den stärksten Einfluss auf das Unfallrisiko hat. Je nach Verkehrsdichte kann durch die Verringerung des Geschwindigkeitsunterschieds zwischen zwei Fahrstreifen das Unfallrisiko vermindert werden. Die Studie hat ebenfalls ergeben, dass viele der dokumentierten Verkehrsunfälle mindestens fünf Minuten vor ihrem Eintritt hätten vorhergesagt werden können.

Auch wenn weitere Arbeiten notwendig sind, stellt das Modell durch die Vorhersage von Verkehrsunfällen in Echtzeit seine Nützlichkeit für Verkehrsleitstellen im Rahmen einer risikoabhängigen Verkehrssteuerung unter Beweis.

Schlüsselbegriffe: Risikoindikator, Gefahrenanalysemodell, vorbeugende Massnahme, Unfallverhütung, Sicherheitsindikator

Einführung

Überblick

In vielen Ländern wird der Verkehrssicherheit immer mehr Bedeutung zugemessen. In punkto Verkehrssicherheit gilt die Schweiz zwar als eines der sichersten Länder, dennoch werden auf Schweizer Strassen noch immer viele Menschen getötet oder verletzt. Um die Verkehrssicherheit in der Schweiz zu verbessern, sucht die vorliegende Studie, Risikoindikatoren zu definieren, die in der Lage sind Verkehrsleitstellen, aber auch Verkehrsteilnehmern Informationen über aktuelle oder mögliche Unfallgefahren zu liefern. Die Informationen, die von verschiedenen Risikoindikatoren abgeleitet werden können, werden in einem integrierten Modell zur effizienten Risikovorhersage zusammengefasst, das riskante Situationen in Echtzeit erkennen und vor aktuellen Verkehrsgefahren warnen kann.

Zunächst wird die Studie Risikoindikatoren besprechen, die in der Literatur bereits diskutiert wurden. Daran anschliessend werden neue Risikoindikatoren entwickelt. Beide Indikatorensets werden dann mit empirischen Verkehrsdaten aus der Schweiz abgeglichen. Sobald die Leistungsfähigkeit der einzelnen Risikoindikatoren quantifiziert werden kann, werden sie in einem integrierten Modell zusammengefasst, mit dem die Leistungsfähigkeit des Gesamtsystems verbessert werden soll. Abschliessend wird die Studie vorbeugende Massnahmen vorschlagen, mit denen das Unfallrisiko verringert werden kann.

Hintergrund

Untersuchungen zur Verkehrssicherheit können je nach Untersuchungsgegenstand grob

in drei Gruppen eingeteilt werden: Infrastruktur, Mensch und Fahrzeuge. In der Regel spielt der Faktor Mensch bei Verkehrsunfällen eine bedeutende Rolle. Dabei geht es jedoch nicht nur um Fahren unter Alkohol- oder Drogeneinfluss, sondern auch um die Grenzen, die dem Menschen generell gesetzt sind, wie etwa, die Tatsache, dass er nicht in der Lage ist, das zukünftige Verkehrsgeschehen vorherzusehen. Es ist daher die Aufgabe neuer Technologien, diese Unzulänglichkeiten auszugleichen.

Ein Trend im Rahmen der Verkehrssicherheit ist, Risiken mithilfe von Risikoindikatoren zu quantifizieren. Dazu können Daten herangezogen werden, die von Verkehrserfassungssystemen wie Überwachungskameras, Induktionsschleifendetektoren und ähnlichem stammen. Eine grosse Zahl von Risikoindikatoren wird eigens für diesen Zweck entwickelt. Zu den beliebtesten Indikatoren gehören Time-To-Collision, PICUD, TIT und TET. Bisher handelt es sich dabei um theoretische im Labor entwickelte Indikatoren. Zu einem späteren Zeitpunkt, wenn die Verkehrserfassungstechnologie weiter fortgeschritten ist, können diese Indikatoren auf reale Verkehrsdaten angewandt werden. Diese neue Technologie ermöglichte auch kleinräumige Untersuchungen, in denen einzelne Unfälle vor dem Hintergrund der erfassten Verkehrsdaten untersucht werden.

Andere Untersuchungen beschäftigten sich mit der Identifizierung von gefährlichen Verkehrssituationen auf Autobahnen in Echtzeit. Im Grunde geht es bei diesen Studien immer darum, das Verkehrsgeschehen, das Unfällen vorausgeht, so genannte PTS (Pre-Crash Traffic Situations), zu analysieren. PTS werden als gefährliche Verkehrssituationen eingestuft, da sie zu einem Unfall führten. Andere Verkehrssituationen, in denen kein Unfall geschah, werden als NTS (Non-Crash Traffic Situations) bezeichnet. Modelle, die es ermöglichen, gefährliche Verkehrssituationen vorherzusagen, basieren auf den Unterschieden zwischen NTS und PTS.

Methodik und Standortwahl

Methodik

Betrachtet man kleinräumige Untersuchungen näher, so weisen sie alle mehrere Nachteile auf. Die vorliegende Studie versucht diese zu vermeiden, indem sie folgende Probleme zu lösen versucht:

- Das Ungleichgewicht zwischen NTS und PTS. Da Unfälle selten auftreten, ist die Zahl von PTS wesentlich niedriger als die der NTS.
- Die Erhebung von NTS, die mit PTS verglichen werden, ist arbiträr. Um ein Modell entwickeln zu können, das in der Lage ist, zwischen NTS und PTS zu unterscheiden, werden NTS aus einer kleineren Stichprobe gewonnen, damit die Zahl der NTS der Zahl der PTS entspricht. Die Stichprobenerhebung bestehender Untersuchungen ist dadurch arbiträr.

Um den Zielen des vorliegenden Projektes zu genügen, umfasst unser Ansatz folgende Schritte:

- 1- Ermittlung von anwendbaren Risikoindikatoren. Die ausgewählten Indikatoren werden an realen Verkehrsdaten getestet, um festzustellen, ob sie tatsächlich in der Lage sind, zwischen Verkehrssituationen, die zu Unfällen führen (gefährlichen Situationen), und anderen Situationen zu unterscheiden.
- 2- Einsatz eines Verfahrens zur Stichprobenerhebung für Verkehrssituationen, in denen es nicht zu Unfällen kam. Das ist notwendig, da Unfälle nur selten auftretende Ereignisse sind und Auffahrunfälle oder seitliche Kollisionen zudem nur unter bestimmten Voraussetzungen auftreten.
- 3- Entwicklung von Modellen zur Risikoidentifizierung mit einer einzigen Variablen
- 4- Fusion verschiedener Modelle mit einer Variablen zu leistungsfähigeren Modellen
- 5- Vorhersagen von aktuell drohenden Unfallrisiken anhand der Ergebnisse des integrierten Modells in Echtzeit.

Schritt 1 ist wichtig, da die Zahl der Risikoindikatoren hoch ist und je nach Verfügbarkeit der Daten nicht alle Risikoindikatoren direkt angewandt werden können. Ziel des zweiten

Schritt 5 ist es, das Verkehrsgeschehen, das zu einem Unfall führt, mit relevanten Verkehrssituationen ohne Unfallfolgen zu korrelieren. Da Unfälle nur relativ selten auftreten, müssen bestimmte Bedingungen vorliegen, damit Auffahrunfälle oder seitliche Kollisionen überhaupt passieren. Durch die Erhebungsweise von NTS werden nur Verkehrsgeschehen erfasst, die mit PTS vergleichbar sind. Im Verlauf des Datenerfassungsprozesses werden Gruppen verschiedener Verkehrsgeschehen, so genannte Traffic-Regime gebildet, die jeweils eine Reihe von Bedingungen enthalten, unter denen entweder ein Unfall passiert oder auch nicht. Schritt 3 wird für jede Gruppe von Verkehrssituationen durchgeführt, bei denen PTS und NTS relevant sind. Bei den Variablen kann es sich um ausgewählte Risikoindikatoren oder Verkehrsparameter handeln. Jede Variable ermöglicht es, ein anderes Unfallrisiko zu ermitteln. Schritt 5 wird ebenfalls für jede Gruppe von Verkehrssituationen durchgeführt, um die Ergebnisse aller Variablen in ein einziges Modell zu integrieren. Schliesslich wird, um das Unfallrisiko einer Verkehrssituation vorherzusagen, die Risikoanalyse herangezogen, die das integrierte Modell für eine Reihe neuer Fälle ermöglichte. Daran wird gemessen, ob ein Verkehrsgeschehen unfallträchtig ist oder nicht.

Beobachtungsort

In der Schweiz gibt es mehr als 300 doppelte Messschleifen, die spezifische Fahrzeugdaten erfassen können. Da diese Messschleifen aber im grossen Abstand voneinander installiert sind, kann immer nur ein Beobachtungsort je Standort gewählt werden. Dieser wird nach folgendem Verfahren ausgewählt:

- Für die mehr als 300 Messschleifen werden jeweils die in einem Umkreis von einem Kilometer verzeichneten Unfälle gezählt. Diese werden als Considered Crashes bezeichnet. Die Standorte werden in eine Rangfolge entsprechend der Zahl der in Betracht gezogenen Unfälle gebracht.
- Da Verkehrsunfälle von der Kantonspolizei bearbeitet werden (die Schweiz verfügt über 26 Kantone) sind in verschiedenen Kantonen wie etwa dem Kanton Waadt oder dem Tessin nicht alle Unfälle im Detail erfasst. Diese Standorte werden daher ausgeschlossen.
- Danach werden alle Standorte eliminiert, die weiter als 15 km von einer Wetterstation entfernt sind. Die verbleibenden Standorte bilden die kondensierte Standortliste.
- Der erste Standort dieser Liste wird als Beobachtungsort gewählt.

Nun wird der Untersuchungsort ausgewählt. Er ist der 27. Standort auf der kondensierten Standortliste. Der Beobachtungsort erhält die Bezeichnung des Standortes der Messschleife: CH023. Standort CH023 befindet sich auf der Autobahn A1, die Bern mit Zürich verbindet. Standort CH023 ist zwar nicht der Ort mit der höchsten Unfallhäufigkeit, doch für ihn liegen sowohl Verkehrs-, als auch Wetter- und Unfalldaten vor. Diese sind für Risikoanalysemodelle, die der in der vorliegenden Studie gewählten Methodik zugrunde liegen, jedoch unbedingt notwendig.

Ergebnisse und Analysen

Grundlegende Unfallanalysen

Auf der Grundlage von Unfalldaten können folgende Beobachtungen zu Unfällen auf den Schweizer Autobahnen gemacht werden:

- Die Unfallhäufigkeit ist freitags und samstags am höchsten. Unfälle passieren am häufigsten am späten Freitagnachmittag (zwischen 16.00 und 19.00 Uhr). Zwar weist auch der Samstag eine hohe Unfallhäufigkeit auf, doch sind die Unfälle hier über den ganzen Tag verteilt. Unfälle ereignen sich auch häufiger am frühen Sonntagmorgen als frühmorgens an anderen Wochentage.
- Mehr als drei Viertel aller Unfälle passierten bei Regen- oder Schneefällen. Fast zwei Drittel aller Unfälle ereigneten sich bei trockener Fahrbahn.
- Unter dem Aspekt der Unfallschwere werden drei Parameter berücksichtigt, die Anzahl der Objekte in den Crash verwickelt, die Anzahl der Personen im Crash und die Anzahl der beteiligten Verletzten drei Vertretern der Schwere. Es wurde festgestellt, dass es vereinzelt Unfälle gibt, in die eine sehr grosse Zahl von Fahrzeugen und Personen verwickelt sind.

- Am Beobachtungsort treten häufig Auffahrunfälle und Unfälle auf, in die nur ein Fahrzeug verwickelt ist. Die meisten Unfälle am Beobachtungsort fanden im Berufsverkehr am Nachmittag bei hohem Verkehrsaufkommen statt.

Empfindlichkeit der Risikoindikatoren

Mögliche Risikoindikatoren

Aufgrund der Eigenschaften des Beobachtungsortes, von dem die realen Daten stammen, sind für die vorliegende Studie mehrere Risikoindikatoren geeignet. Von diesen wurden drei Risikoindikatoren ausgewählt: TTC, IBTR und PBTR.

Empfindlichkeitstests

Hauptziel des Empfindlichkeitstests ist es, die Leistungsfähigkeit von Risikoindikatoren bei der Analyse von existierenden Daten wie Verkehrs-, Wetter- und Unfalldaten zu testen. Der Empfindlichkeitstest sollte eine Bewertung der Risikoindikatoren ermöglichen und so als Richtschnur für die Verwendung der Risikoindikatoren in späteren Phasen der Studie dienen. Die Aussagekraft der ausgewählten Risikoindikatoren wird von zwei Faktoren beeinflusst: dem Verkehrsgeschehen und den Wetterbedingungen.

Es muss nun versucht werden, Risikoindikatoren in NTS und PTS (mit und ohne Niederschläge) zu testen, um so in jeder Situation die unterschiedliche Verteilung der Risikoindikatoren herauszufinden.

PBTR, IBTR und PICUD, deren Formel wetterabhängige Parameter (Bremsweg) aufweist, reagieren natürlich entsprechend auf die Wetterbedingungen. Die meisten Risikoindikatoren mit Ausnahme von TTC reagieren auch auf das Verkehrsaufkommen. Die Risikoindikatoren sind zudem abhängig von der jeweiligen Fahrspur. Der Empfindlichkeitstest zeigt, dass PICUD keine besonders verlässlichen Ergebnisse liefert, da dieser Risikoindikator suggeriert, dass der Verkehr unter guten Wetterbedingungen gefährlicher ist als bei Niederschlägen. Deshalb wird PICUD in dieser Studie nicht weiter benutzt.

Die beste Art und Weise die Leistungsfähigkeit von Risikoindikatoren zu testen, ist ihre Entwicklung im Verkehrsgeschehen kurz vor dem Unfall zu überprüfen, wenn feststeht, dass ein Unfallrisiko vor dem Unfall existiert. Sind die verschiedenen Unfallszenarien getestet, sollten bei der Einschätzung, ob eine Verkehrssituation gefährlich oder nicht gefährlich ist, nicht nur die Risikoindikatoren, sondern auch die Anzahl der Fahrzeuge berücksichtigt werden.

Traffic-Regime

Datenerfassung

Da Unfälle selten auftreten, ist die Zahl von PTS wesentlich niedriger als die der NTS. Die Unfälle, die in der vorliegenden Studie berücksichtigt werden, betreffen nur Auffahrunfälle und seitliche Kollisionen. Diese Unfälle ereignen sich nur unter bestimmten Bedingungen wie etwa im Berufsverkehr, wenn ein hohes Verkehrsaufkommen zu verzeichnen ist. Daher gibt es für eine gegebene PTS zahlreiche NTS, die nicht wirklich zum Vergleich herangezogen werden können, wie etwa mitternächtliche Verkehrssituationen, wenn nur wenige Fahrzeuge unterwegs sind.

Wird eine Stichprobe von NTS erstellt, so verringert dies nicht nur die Zahl der NTS, sondern erlaubt es auch, die PTS mit den NTS zu korrelieren, die sich am besten vergleichen lassen. Die Datenerfassung umfasst folgende Schritte:

- 1- Definition allgemeiner Verkehrssituationen, die durch Variablen charakterisiert sind
- 2- Definition von NTS und PTS
- 3- Anordnung von NTS in Clustern
- 4- Zuordnung der PTS zu den zuvor erhaltenen Clustern.

Am Ende des Prozesses erhält man Cluster verschiedener Verkehrssituationen. Jeder Cluster enthält eine gewisse Zahl von Verkehrssituationen, die nicht zu Unfällen führen,

und solchen, die zu Unfällen führen. Jeder Cluster entspricht einem Traffic-Regime, da die Verkehrssituationen jedes Clusters ähnliche Eigenschaften aufweisen.

Eine Verkehrssituation wird durch 22 Variablen bestimmt, etwa Unmittelbarkeit (2 Variable), Verkehr auf jeder Spur (14 Variable), Geschwindigkeitsunterschiede zwischen den Spuren (1 Variable), Veränderung des Verkehrsgeschehens zwischen zwei aufeinanderfolgenden Verkehrssituationen (4 Variable) und den Verkehrsbedingungen (1 Variable). Eine Verkehrssituation kann als NTS oder PTS eingestuft werden, je nachdem ob sie vor einem Unfall eintritt oder nicht. Im Fall eines Unfalls kann eine dreissigminütige Zeitspanne als Einlaufphase definiert werden. Ein weiterer Zeitpuffer vor der Einlaufphase dient als Trennung des normalen Verkehrsgeschehens von der Einlaufphase. Daneben gibt es auch eine Auslaufphase, die die Zeitspanne nach dem Unfall umfasst, bevor sich das Verkehrsgeschehen erneut normalisiert. Der Zeitpuffer vor der Einlaufphase und die Auslaufphase werden in der vorliegenden Studie nicht berücksichtigt. Alle Verkehrssituationen, bei denen es sich weder um die Einlaufphase, den Zeitpuffer noch die Auslaufphase handelt, werden als Non-Crash-Perioden betrachtet.

Die Clusterbildung erfolgt aufgrund des k-Means-Algorithmus, wobei die Zahl der Cluster K aufgrund des Clusterbildungsfehlers gewählt wird: Verringert sich bei einer Erhöhung der Clusterzahl der Clusterbildungsfehler nicht signifikant, wird die nächst kleinere Zahl von Clustern gewählt. Am Ende erhält man acht Cluster, die als acht Traffic-Regime betrachtet und mit den Buchstaben A bis H bezeichnet werden.

Durch die Clusterbildung der NTS werden PTS Traffic-Regime je nach Ähnlichkeit der PTS den acht Clustern zugeordnet.

Analyse der Traffic-Regime

Betrachtet man die 22 Variablen für jedes Traffic-Regime, so können die folgenden Feststellungen getroffen werden:

- Die Regime A, E und F sind durch flüssigen Verkehr gekennzeichnet. Bei diesen Regimen ist das Verkehrsaufkommen auf beiden Fahrstreifen nur gering.
- Das Regime H steht für hohes Verkehrsaufkommen und stockenden Verkehr. Dies gilt für beide Fahrspuren. Hohes Verkehrsaufkommen führt häufig zu beträchtlich verringerten Durchschnittsgeschwindigkeiten.
- Die Regime B und G weisen mittleres bis hohes Verkehrsaufkommen auf. Die Verkehrsdichte ist hier nur leicht geringer als beim Regime H. Aufgrund der Verkehrsänderungsvariablen nimmt der Verkehr im Regime B zu, während er im Regime G abnimmt. Daher kann das Regime B als ein Vorstau-Traffic-Regime betrachtet werden, wohingegen es sich beim Regime G um ein Nachstau-Traffic-Regime handelt.
- Die Regime C und D weisen ein niedriges bis mittleres Verkehrsaufkommen auf. Die Regime C und D unterscheiden sich vor allem durch das Vorhandensein von Niederschlägen. Bei den meisten Verkehrssituationen des Regimes C sind die Wetterbedingungen gut, bei Regime D hingegen liegen Niederschläge vor.

Die Zahl der PTS, die unter das Regime H fallen, ist höher als für alle anderen Regime. Die zweit- und dritthöchste Anzahl an PTS weisen Regime G und Regime B auf. Die Regime A und F mit ihrem flüssigen Verkehr enthalten keine einzige PTS, und das Regime E mit seinem zähen Verkehrsfluss enthält nur wenige. Verglichen mit Regime C hat D eine grosse Anzahl an PTS. Das mag daran liegen, dass Autofahrer bei Regen grössere Abstände einhalten, um Auffahrunfälle oder seitliche Kollisionen zu vermeiden.

Die Risikorate eines Traffic-Regime erhält man, indem man seine Anzahl an PTS durch seine Anzahl an NTS dividiert. Die Risikorate der Regime H und G ist höher als die anderer Regime. Das kann ein Anzeichen dafür sein, dass das Verkehrsgeschehen bei hohem Verkehrsaufkommen oder nach einem Stau höhere Risiken birgt.

Übergang zwischen Verkehrssituationen

Ein Übergang zwischen Verkehrssituationen ist ein Wechsel der Verkehrssituation von einem Regime zu einem anderen. Das Folgeregime kann dabei identisch oder verschieden

vom Ursprungsregime sein. Da PTS von fünf Minuten Länge benutzt werden, um das Verkehrsaufkommen vor einem Unfall zu quantifizieren, besteht eine Einlaufphase von 30 Minuten Dauer aus sechs PTS. Daher werden sechs aufeinanderfolgende PTS als ein Muster betrachtet. In diesem Abschnitt werden alle möglichen Muster untersucht, um diejenigen zu identifizieren, die häufig zu Unfällen führen.

Insgesamt gibt es $6^6 = 1'679'616$ mögliche Muster. Es gibt Übergänge, die unmöglich sind wie etwa von A nach H nach A (das heisst, ein Übergang von einer Strasse ohne Fahrzeuge zu einer Strasse mit hoher Verkehrsdichte und dann erneut wieder zu einer Strasse ohne Fahrzeuge und dies innerhalb von 15 Minuten). Daher beträgt die Gesamtzahl von Übergängen bei NTS-Mustern für die Zeitspanne, die die Studie abdeckt, 11'782 von 1'050'372 verfügbaren Mustern (Es müssen mindestens 30 Minuten aufeinanderfolgende Daten vorliegen, um ein Muster zu ermöglichen). Es gibt Muster, die häufiger auftreten als andere. So findet sich das Muster DDDDDD, das heisst Verkehr unter dem Regime D für 30 Minuten, 86'624 mal und macht damit 8,25 % aller verfügbaren Muster aus. Die nächst häufigsten Muster sind EEEEE, FFFFF, BBBBB und CCCCC mit jeweils 4,30 %, 2,40 %, 2,28 % und 1,54 % der verfügbaren Muster. Dieses Ergebnis fällt zusammen mit einer grossen Zahl von Übergängen vom und zum selben Muster in den Regimen B, C, D, E und F. Diese Muster können als stabil betrachtet werden.

Betrachtet man PTS-Muster, so findet man für 120 Unfälle nur 72 PTS-Muster. Dabei taucht das Muster HHHHHH achtmal und damit am häufigsten auf. Das heisst, ist das Verkehrsaufkommen unter Regime H stabil (mit anderen Worten, der Stau dauert 30 Minuten), ist das Unfallrisiko hoch. Bei 64 Unfällen (das sind 53,3 %) sind nur die beiden Regime G und H im Spiel. Dabei handelt es sich um den Verkehrsfluss zwischen den beiden Regimen H und G.

Vergleicht man PTS- mit NTS-Mustern, so finden sich die beiden PTS-Muster BBBBB und CCCCC am häufigsten bei den NTS-Mustern. Ihre Rate beträgt $1/23'913$ beziehungsweise $1/16'171$. Das bedeutet, dass das Risiko gering ist, wenn das Verkehrsgeschehen diesem Muster folgt. Es gibt jedoch auch Muster, die zu einer höheren Pre-Crash-Rate führen, wie etwa HGBCHH (ein PTS-Muster vs. vier NTS-Muster), HHHBHB (zwei PTS-Muster vs. zehn NTS-Muster), CGH-CHC (ein PTS-Muster vs. sieben NTS-Muster) und GDDGHB (ein PTS-Muster vs. acht NTS-Muster). Diese Muster treten nur selten auf, sind aber sehr unfallträchtig.

Fusion der Risikoindikatoren

Datenfusion

Das Data Fusion Technical Committee der Geosciences and Remote Sensing Society des IEEE definiert Datenfusion *als einen Prozess, durch den räumlich und zeitlich indizierte Daten, die durch verschiedene Instrumente gewonnen und aus unterschiedlichen Quellen stammen, miteinander verknüpft werden, um die Verarbeitung und Interpretation dieser Daten zu verbessern.*

Hauptziel der Datenfusion in FUSAIN sind eine Verbesserung der Risikoerkennung und eine Definition der Bedingungen, die die Strassen sicher machen. Das Besondere bei der Datenfusion für Sicherheitsindikatoren ist die Tatsache, dass die meisten Sicherheits-/Risikoindikatoren aus denselben Verkehrsdaten, die nur in unterschiedlicher Form vorliegen, abgeleitet wurden. Es werden daher dieselben Daten, die aus derselben Quelle stammen, miteinander verschmolzen. Dies entspricht der erst jüngst begründeten Praxis, Masszahlen zu gruppieren. Davon verspricht man sich nicht nur eine höhere Genauigkeit, sondern auch eine Stärkung des gesamten Risikoerkennungsprozesses.

Die verschiedenen Methoden der Datenfusion können in merkmalsbasierte Inferenzmethoden (wie Bayessche Wahrscheinlichkeitsinferenz, Evidenztheorie, Typparametrierung, künstliche neuronale Netze, Clusteranalyse oder Wahlverfahren) und Methoden auf kognitiver Grundlage (wie logische Muster oder wissensbasierte Expertensysteme) unterteilt werden. In der vorliegenden Studie wird die Evidenztheorie von Dempster und Shafer benutzt. Bei der Stichprobenerstellung für NTS ergab sich, dass Regime auftreten, denen

keine PTS entsprechen. Daher wird die Datenfusion nur für die Regime B, C, G und H eingesetzt, da hier die Anzahl der PTS hoch ist. Für jedes dieser Regime wird ein integriertes Modell entwickelt, das auf Modellen mit einer Variablen beruht und in der Lage ist, Verkehrsrisiken vorherzusagen. Für die Modelle mit nur einer Variablen wurden 22 Variablen benutzt, die Verkehrssituationen und drei Risikoindikatoren charakterisieren. Bei den letzteren handelt es sich dabei um TTC, PBTR und IBTR.

Ergebnisse

Es gibt für jedes der vier Regime B, C, G und H ein integriertes Modell der Risikoerkennung. Diese Modelle sind sehr leistungsfähig, was die Differenzierung von NTS und PTS betrifft. Daten (darunter NTS und PTS) aus jedem Regime werden unterteilt in Trainings- und Validierungsdatensätze. Die Trainingsdatensätze werden dazu verwendet, die Modelle zu entwickeln, und die Validierungsdatensätze, um diese Modelle zu validieren. In allen Regimen funktionieren die integrierten Modelle der Risikoerkennung mit Trainingsdatensätzen (alle PTS und mindestens 92 % aller NTS werden erkannt). Mit Validierungsdatensätzen, auch wenn die Leistungsfähigkeit der Modelle nicht so hoch ist wie die der Trainingsdatensätze, ist der Erkennungsgrad ebenfalls recht hoch (für NTS liegt er bei 85 % und für PTS bei 77 %). Bei vier Regimen werden 85 % der PTS richtig erkannt. Für NTS liegt der Prozentsatz bei 90 %. Werden die Verkehrssituationen von den vier Traffic-Regimen A, D, E und F ebenfalls hinzugezogen, ist die Erkennungsrate von NTS noch höher.

Mit den integrierten Modellen der Risikoerkennung erfolgt ein weiterer Schritt, um die zukünftige Verkehrssituation vorhersagen zu können. Grundlage dafür ist ein Konzept, das als *length of risk memory* bezeichnet wird. Damit ist gemeint, dass der in Echtzeit ermittelte Risikostatus der letzten riskanten Situation für die Einschätzung des Risikostatus der folgenden Verkehrssituation benutzt wird. Die Dauer der Risikospeicherung entspricht der Zahl der letzten Verkehrssituationen, deren ermittelter Risikostatus gespeichert wurde. Dadurch kann ein Mittel zwischen der Anzahl falscher Warnungen und verpasster Warnungen gefunden werden. Steigt die Dauer der Risikospeicherung an, verringert sich die Anzahl falscher Warnungen, während die Zahl der verpassten Warnungen ansteigt. Das Umgekehrte gilt ebenso: Verringert sich die Dauer der Risikospeicherung, steigt die Anzahl falscher Warnungen, während sich die Zahl der verpassten Warnungen verringert. Daher gibt es keine optimale Dauer für die Risikospeicherung, die gleichzeitig die Zahl falscher und verpasster Warnungen verringern würde. Es wird darauf hingewiesen, dass die Dauer der Risikospeicherung positiv und kleiner als 5 ist.

Schlussfolgerungen

Die Studie behandelt die Entwicklung von Modellen der Risikoerkennung auf der Grundlage von Datenfusionstechniken. Die Ergebnisse lassen folgende Schlussfolgerungen zu:

- Auffahrunfälle und seitliche Kollisionen finden häufiger während Stosszeiten und nach der Auflösung eines Staus statt sowie bei hoher Verkehrsdichte oder wenn das Verkehrsgeschehen zwischen verschiedenen Traffic-Regimen wechselt.
- Einzelne Risikoindikatoren sind nur beschränkt bei der Erkennung von Verkehrsrisiken hilfreich. Werden sie jedoch miteinander und anderen Verkehrsparametern fusioniert, sind die neuen Modelle (etwa integrierte Modelle der Risikoerkennung, die auf Traffic-Regimen basieren) sehr leistungsfähig.
- Integrierte Modelle der Risikoerkennung können dazu dienen, drohende Verkehrsrisiken in Echtzeit vorherzusagen. Ein Ausgleich zwischen falschen und verpassten Warnungen ermöglicht es nicht, die Anzahl beider Warntypen gleichzeitig zu verringern. Sollen die Modelle in der Praxis funktionieren, muss die Anzahl falscher Warnungen verringert werden, damit die Verkehrsteilnehmer nicht das Vertrauen in das Verkehrsleitsystem verlieren.
- Geschwindigkeitsunterschiede sind der Faktor, der den stärksten Einfluss auf das Unfallrisiko hat. Je nach Verkehrsdichte kann durch eine Verringerung des Geschwindigkeitsunterschieds zwischen zwei Fahrstreifen das Unfallrisiko verkleinert werden.
- Die Studie hat ebenfalls ergeben, dass viele Verkehrsunfälle der Vergangenheit mindestens fünf Minuten vor ihrem Eintritt hätten vorhergesagt werden können.

- Die entwickelten Modelle können Verkehrsleitstellen dabei helfen, den Verkehr entsprechend zu steuern. Die Informationen, die die Modelle liefern, dürften auch für Verkehrsteilnehmer interessant sein, da sie diesen gestatten, ihre Route entsprechend auszuwählen.

Summary

This study aims to contribute to improve motorway traffic safety by developing models that can identify real-time traffic risk. The approach is to classify potential risk indicators and test them with historical data including traffic, meteorological and crash data. First, risk indicators are short-listed based on the test of their sensitivity with real data. After that, two types of traffic conditions are distinguished: the conditions leading to crashes, called pre-crash cases and the conditions when no crash occurs, called non-crash cases. Pre-crash and non-crash cases are matched into groups such that under each group, pre-crash and non-crash cases are comparable. Each of such groups is considered as a traffic regime. Single-variable models are then developed under each traffic regime and provide inputs to the next step when a final model is developed under each traffic regime by fusing results of single-variable models.

A further step to predict short-term traffic crash risk is undertaken by combining the risk status of several most recent traffic cases in real-time. The test with historical data show that the developed prediction model can perform with high accuracy. Speed variation is detected to be the most important factor influencing the traffic crash risk. Depending on traffic density, the speed difference between two lanes can be reduced in order to diminish the crash risk. It is also found that a high percentage of historical crashes could have been identified at least 5 minutes before crash occurrences.

Although more test needs to be done, the system of short-term traffic crash risk prediction prove the potential of assisting traffic operators in a risk-oriented traffic management scheme.

Keywords: *risk indicator, risk identification model, preventive measure, crash prevention, safety indicator*

Introduction

Overview

Traffic safety is becoming one of the biggest concerns in many countries in the world. Although being ranked as one of the safest countries in term of traffic safety, there are still many deaths and injuries on Swiss roads. In order to contribute to improve Swiss traffic safety, the present study aims at defining risk indicators capable of providing information to traffic operators as well as road users about the traffic crash risk they are being involved or approaching. Information from several risk indicators is then fused to produce a unique effective risk identification model that can identify real-time risk information and predict short term traffic risk.

Firstly, the current study comes to identify risk indicators existing in the literature. Then, new risk indicators can be developed. After that, all the risk indicators are tested with field data from Swiss road network. Once the performance of individual risk indicators is quantified, they can be fused into unique models in order to optimize the overall performance. Finally, the study should suggest appropriate preventive measures to reduce crash risk.

Background

Studies in traffic safety can be broadly divided into three categories depending on the studied object: infrastructure, human, or vehicles. Usually, human factors play an important role in crash occurrences. Some factors do not limit to the state of human being right before crashes such as being drunk or influenced by drug, etc. but are caused by the general limit of human being such as the incapacity of foreseeing the coming traffic. Therefore, new technologies help to overcome those shortcomings of human beings.

To that extent, one of trends in traffic safety is to quantify risky traffic situation using risk indicators. The quantification can be based on real data collected from traffic sensors such

as traffic surveillance cameras, loop detectors, etc. An important number of risk indicators was developed for that purpose. Some of the widely used risk indicators include Time-To-Collision, PICUD, TIT, TET, etc. Previously, these were theoretical indicators created in laboratories. Later on, when the traffic sensing technology became well developed, those indicators could be applied using real traffic data. The new technology also opened a new trend of disaggregate studies where the individual crashes are considered with the help of collected traffic data.

Several studies have been developed attempting to identify risky traffic situations on motorways in real-time. In principal, these studies attempt to extract traffic situation right before crashes called pre-crash traffic situations – PTS. PTS are considered as risky traffic situations as they ended up with crashes. Other traffic situations are called non-crash traffic situations – NTS where there is no accident observed. Models for identifying risky traffic situations are developed by differentiating between NTS and PTS.

Methodology & Study Site

Methodology

By reviewing disaggregate studies, several drawbacks are identified. The methodology proposed within the present study aims to attain the study objectives by addressing the following issues:

- The imbalance between the populations of NTS and PTS. As crashes are rare events, the number of PTS is much lower than the number of NTS.
- The arbitrary nature of the process to sample NTS to be compared with PTS. To develop the model capable of differentiating NTS and PTS, NTS are usually sampled to a smaller population such that number of NTS is equivalent to the number of PTS. However, the sampling process within existing studies is arbitrary.

To conform to the objectives of the current project, the approach includes the following steps:

- 1-Identify applicable risk indicators. The selected indicators are tested with real data to verify their capacities in differentiating traffic conditions leading to crashes (risky conditions) and other conditions.
- 2-Apply a data sampling process to sample non-crash data. This is because crashes are rare events and rear-end and sideswipe crashes only occur under certain circumstance.
- 3-Develop single – variable risk identification models
- 4-Fuse results of single-variable models to obtain fused models with better performance
- 5-Provide prediction of short-term traffic risk using results of fused models in real-time framework.

Step 1 is important as the number of risk indicators is high and depending on the availability of data, not all risk indicators can be directly applied. In the second step, the objective is to match pre-crash traffic conditions with relevant non-crash conditions. This is because crashes are rare events and there must be certain context for rear-end and sideswipe crashes to occur. By sampling non-crash data, only non-crash conditions comparable with pre-crash conditions are kept for further processing. The data sampling process creates groups of traffic conditions called traffic regimes such that under each traffic regime, there is a list of non-crash conditions and a list of pre-crash conditions. Step 3 is undertaken under each traffic regime where pre-crash and non-crash cases are relevant. Variables can be risk indicators selected or traffic parameters. Each variable has certain performance in identifying traffic crash risk. Step 5 is also undertaken under each traffic regime aiming to combine the performance of all variables by fusing the single-variable models. Finally, to predict the traffic crash risk of the next traffic case, the risk assessment of several last traffic cases provided by fused identification models is used to judge whether the next traffic case is risky or not.

Study Site

In Switzerland there are more than 300 double traffic detectors who can provide individual vehicle data. As the spacing between any couple of detectors is high, the study site can only be chosen for one location. The study site is selected based on the following procedure:

- From the list of more than 300 traffic detector locations providing traffic data, count for each detector location the number of crashes satisfying conditions within the buffer of 1km from the location, called considered crashes. Re-arrange the list in descending order of the number of considered crashes.
- As road crashes are monitored by cantonal police (there are 26 cantons in Switzerland), crash positions can be undetailed in the crash records for several cantons such as Vaud or Ticino, etc. Therefore, locations in these cantons are removed from the list.
- Remove from the ordered list all traffic detector locations where is no meteorological station within a radius of 15km from the traffic detector locations. The remaining detector locations make up the reduced list of locations.
- Select the first location from the reduced list of locations.

Finally, the study site is selected and is at 27th position of the considered list. The study site is named according to the name of traffic detector stations: CH023. Site CH023 is on motorway A1 connecting two cities Bern and Zurich. The site CH023 is not the location where crash occurrences are the most numerous yet is selected because traffic, meteorological and crash data are altogether available, which is crucial in development of risk identification models by applying the methodology proposed in the current study.

Results and Analyses

Preliminary accident analysis

Based on accident data, the following observations about crashes on Swiss motorways are recognized:

- The number of crashes is the highest on Friday and then on Saturday. Crashes occur more often in late afternoon on Friday (from 16:00 to 19:00) than in any other time. Although the crash frequency is high on Saturday, crash occurrences did not concentrate on a particular hour of the day. Crashes occur more often in early Sunday mornings than in early morning of any other day of the week.
- More than 3/4 out of crashes occurred when the precipitation is positive (rain or snow). Nearly 2/3 out of crashes occurred when the pavement is dry.
- In term of crash severity, the number of objects involved in the crash, the number of people involved in the crash and the number of injuries are three representatives of the severity. It is observed that there are individual crashes with high numbers of vehicles and people involved in.
- At the study site, there are high number of rear-end crashes followed by the number of single-vehicle crashes. Most of crashes at the study site occurred during rush hour in the afternoon when the traffic flow is high.

Sensitivity of risk indicators

Potential risk indicators

Based on the characteristics of the study site that lead to the availability of real data, several risk indicators are considered as suitable for the current study. Finally, three main risk indicators are chosen to be used in the current study: TTC, IBTR and PBTR.

Sensitivity analyses

The main objective of the sensitivity analysis is to uncover the performance of risk indicators with data in the past, including traffic data, meteorological data, and crash data. The sensitivity test should provide the performance evaluation of risk indicators which serves as the guideline for the later usage of risk indicators in other stages of this study. With the

selected risk indicators, there are two factors that can influence their performance: the traffic flow itself and the meteorological conditions.

The approach is to test risk indicators under non-crash and pre-crash conditions with and without the presence of precipitation in order to figure out the difference in the distribution of risk indicators in each case.

PBTR, IBTR, and PICUD with meteorological elements in their formula (the deceleration rate) are sensitive to meteorological conditions. Most risk indicators except TTC are sensitive to flow ranges. The risk indicators are also sensitive to the lanes of the road. The sensitivity test shows that PICUD is less reliable as it indicates that the traffic under fine meteorological conditions is more unsafe than under rain conditions. Therefore, PICUD should not be used further in this study.

The best way to test the performance of risk indicators is to check the evolution of risk indicators during the traffic conditions just before the crash when it is sure that there is crash risk before the crash. With the crash cases tested, it is recommended that when considering if a traffic situation is risky or not, not only bound values of risk indicators but also the percentages of vehicles are essential.

Traffic Regimes

Data sampling

As crashes are rare events, the number of PTS is much lower than the number of NTS. The crashes considered in the current study involve only rear-end and sideswipe crashes. These are crashes that only occurred under certain conditions such as during dense traffic that is usually available during rush hour. Therefore, for a PTS, there are many NTS that are not really suitable for comparison such as NTS occurring at night when there are not many vehicles on the road.

Sampling NTS here does not only serve as a way to reduce the number of NTS but also serve as a process to match NTS with the most comparable PTS. Data sampling process includes the following steps:

- 1-Defining traffic situation in general that is characterized by variables.
- 2-Defining non-crash and pre-crash traffic situations
- 3-Clustering non-crash traffic situations into clusters
- 4-Classify pre-crash traffic situations into obtained clusters.

At the end of the process, clusters of traffic situations are obtained. Under each cluster, there is a list of pre-crash and non-crash traffic situations. Each cluster is a traffic regime as traffic situations under that cluster represent similar traffic condition.

A traffic situation is defined as characterized by 22 variables representing instantaneity (2 variables), state of traffic on each lane (14 variables), speed difference between lanes (1 variable), traffic evolution between consecutive traffic situations (4 variables) and traffic conditions (1 variable). A traffic situation can be classified as PTS or NTS depending on whether it occurs before a crash or not. For a crash, there is a crash time that allows identifying 30 minutes before the crash as pre-crash period. Pre-crash buffer period is right before pre-crash period and serves as the separator between the traffic periods before that with pre-crash period. There is also a post-crash period indicating the duration just after the crash when the traffic is getting back to normal state. Pre-crash buffer period and post-crash period are not considered in the present study. All traffic situations that are not within pre-crash, pre-crash buffer and post-crash periods are called non-crash periods.

The clustering process is undertaken based on K-mean method where the number of cluster K is chosen based on the clustering error: by increasing the number of clusters, when the clustering error is not getting significantly reduced, the previous number of clusters is chosen. Finally, 8 clusters are obtained and called as eight traffic regimes and named from A to H.

With the presence of eight traffic regimes obtained from NTS clustering process, PTS are classified into traffic regimes based on the similarity between the PTS and eight cluster centers.

Traffic regime analysis

Considering each of 22 variables under each traffic regime, the following observations are obtained:

- Regimes A, E, and F represent the free flow traffic conditions. These regimes are characterized by the very low flow on both lanes.
- Regime H represents the high flow and congested conditions. The flow on both lanes under this regime is high. Congestion is sometimes observed with significantly reduced average speed.
- Regimes B and G represents the high-medium traffic flow conditions. The traffic under these regimes is only less dense than the traffic under regime H. Based on traffic evolution variables, it is observed that under regime B, the traffic is increasing while it is decreasing under regime G. Therefore, regime B can be considered as pre-congestion traffic regime whereas, regime G can be post-congestion traffic regime.
- Regimes C and D represent the low-medium traffic flow conditions. The most significant difference between regime C and regime D is the presence of precipitation. While most of the time traffic situations under regime C occur under fine meteorological conditions, the precipitation is usually positive under regime D.

As with results of classifying PTS into traffic regimes, the number of PTS classified in regime H is greater than in any other regime. After that the numbers of PTS under regime G and under regime B are the second and the third greatest, respectively. Especially, there is no PTS classified under free flow regimes A and F. Only low number of PTS falling in low flow regime E. Between two regimes C and D, there are more PTS under regime D than under regime C. It is possible that drivers keep better distance under rainy conditions to avoid rear-end or sideswipe crashes.

Call risk rate of a traffic regime the number of PTS divided by the number of NTS under that regime. The risk rate under regimes H and G is higher than under other regimes. This might indicate the fact that traffic conditions during high flow or post-congestion conditions are more risky.

Traffic situation transitions

A traffic situation transition is a change of traffic situations from a regime to another regime. The next regime can be the same to or different from the current regime. As PTS of 5 minutes are used to quantify the traffic before crashes, a pre-crash period of 30 minutes is composed of six PTS. Therefore, a set of six consecutive PTS is considered as a pattern. This section is to consider all possible patterns and uncover the patterns who frequently lead to crashes.

There are totally $6^8 = 1'679'616$ possible patterns. There are transitions that can never occur such as from A to H to A (i.e. a change from no-vehicle on the road to very dense traffic and then to no-vehicle on the road – all this is within 15 minutes). For this reason, the total number of NTS transition patterns during the whole study period is 11'782 among 1'050'372 available patterns (there must be at least 30' consecutive data to produce a pattern). There are patterns that repeat more than other patterns. For example, pattern DDDDDD, i.e. the traffic remain under regime D for 30 minutes, repeats the most at 86'624 times, i.e. 8.25% of available patterns. After that, patterns EEEEE, FFFFF, BBBB, and CCCCC play 4.30%, 2.40%, 2.28%, and 1.54% of available patterns. This result conforms to high percentages of transitions from and to the same regime for regimes B, C, D, E, and F. These are stable patterns.

Considering PTS patterns, among 120 crashes, there are only 72 PTS patterns. Pattern HHHHHH repeats the most - eight times – which means that even when the traffic is stable under regime H (i.e. the congestion lasts for 30 minutes), the crash risk is getting high. There are 64 crashes (i.e. 53.3%) for which the traffic comes through only two regimes G and H and not any other regime. This is the traffic fluctuation between the two regimes H

and G.

Comparing PTS patterns with NTS patterns, two PTS patterns BBBBBB and CCCCCC are the most popular among NTS patterns with the rate of 1/23'913 and 1/16'171, respectively. This means that the risk is low when the traffic comes to this pattern. However, there are patterns that can lead to higher pre-crash rate such as HGBCHH (one PTS pattern against four NTS patterns), HHHBHB (two PTS patterns against ten NTS patterns), CGH-CHC (one PTS pattern against seven NTS patterns), and GDDGHB (one PTS against eight NTS patterns). These patterns are rare yet really risky.

Fusion of Risk Indicators

Data fusion

The IEEE Geoscience and Remote Sensing Society Data Fusion Technical Committee defines Data Fusion as *the process of combining spatially and temporally indexed data provided by different instruments and sources to improve the processing and interpretation of these data*.

The main goals of the data fusion process in FUSAIN are an improvement in risk detection and their ability to characterize road safety conditions. The specificity of the data fusion process for safety indicators is that most of the safety/risk indicators are derived from the same traffic data but in different forms; therefore the combination process is performed in fact for the same data provided by one source. This can be seen as the newly established practice of bagging estimators. The expected benefit is not only an increase of the accuracy but also a robustness of the whole risk detection process.

Among different available data fusion techniques such as Feature-Based Inference methods (such as Bayesian inference, Evidence theory, Parametric Template, Artificial neural networks, Cluster Algorithms, or Voting methods, etc.) and Cognitive-based models (such as Logical Templates, Knowledge-based expert systems, etc.), Dempster-Shafer method is used for the current study. As resulted from NTS sampling process, there are regimes where no PTS is classified into. Therefore, data fusion technique is only applied to regimes B, C, G and H – these are the regimes where the number of PTS is high. Under each of these regimes, a fused model capable of identifying traffic risk is developed based on single-variable models. Variables used for developing single-variable models include 22 variables characterizing traffic situations and three risk indicators: TTC, PBTR, and IBTR.

Results

There is a Fused Risk Identification Model under each of four regimes B, C, G, and H. In term of risk identification, these models show high performance in differentiating NTS and PTS. Data (including NTS and PTS) under each regime is divided into training and validation datasets. While training data is used for forming the models, validation data is used for validating the models. For all regimes, the fused risk identification models perform well with training data (all PTS are identified and at least 92% out of all NTS are identified). With validation data, although the performance of the models is not as good as with training data, the percentages of NTS and PTS identified remain very high (at least 85% for NTS and at least 77% for PTS). Overall for four regimes, 85% out of PTS are correctly identified. This percentage for NTS is 90%. If traffic situations of four other traffic regimes (A, D, E, and F) are considered, the overall percentage of NTS identified is higher.

With Fused Risk Identification Models, a further step is undertaken aiming to predict the risk status of the coming traffic situation. This is implemented based on a concept called *the length of risk memory*, i. e. in real-time context that is the identified risk status of the last risk situations that is used to estimate the risk status of the next traffic situation. The length of risk memory is the number of the last traffic situations whose identified risk status is memorized. That size makes up a trade-off between false alarm rate and the missed alarm rate. When the length of risk memory is increasing, the false alarm rate is reducing yet the missed alarm rate is increasing. And vice versa: if the length of risk memory is reducing, the false alarm rate is increasing while the missed alarm rate is reducing. Therefore, there is no optimal length of risk memory so that both false and missed alarm rates

are reducing. It is important to note that the length of risk memory is positive and smaller than 5.

Conclusions

The work presented here introduces the development of risk identification models using data fusion techniques. With the obtained results, the conclusions can be conducted:

- Rear-end and sideswipe crashes occurred more often during peak hours especially after the congestion or when there is high fluctuation of traffic making changing from/to different traffic regimes.
- Risk indicators separately provide moderate performance in identifying traffic risk. When they are fused together and with other traffic-flow variables, the resulted models (i.e. traffic regime-based fused risk identification models) give high performance.
- Fused Risk Identification Models can be used in real-time to predict short-term traffic risk. However, a trade-off between false and missed alarm rates to not allow minimizing the two rates at the same time. To implement the models in practice, it is important to reduce the false alarm rate in order not to lose the confidence of road users in the traffic management system.
- Speed variation is detected to be the most important factor influencing the traffic crash risk. Depending on traffic density, the speed difference between two lanes can be reduced in order to diminish the crash risk.
- It is also found that high percentage of historical crashes could have been identified at least 5 minutes before crash occurrences.
- The models developed can be used as an additional tool for traffic operators in order to help them to make traffic operating decisions. The information provided by the models can also be interesting for road users in their decision to make road choices.

1 Introduction

1.1 Motivation

Switzerland is one of the leading countries in road traffic safety. According to (BPU, 2011), in recent years, the number of deaths and heavy injuries by road traffic crashes is decreasing. This positive trend is obtained thanks to a series of preventive measures such as dedicated speed limit, active traffic management, improvement of pavement, etc. However, the total number of road traffic crashes is still very high with about 27 000 deaths and injuries in 2005, of which the number of motorway traffic crashes is about 320 and this number does not change much in comparison to previous years.

Many efforts have been taken to reduce the traffic crash risk and researches can contribute to improve the state of situation. The current study aims to define risk indicators capable of providing information to traffic operators as well as road users at a given location and at a concrete moment the level of traffic risk. Several parameters can be applied in order to develop a methodology allowing traffic operators to monitor risky traffic situations.

A methodology for the development of real-time motorway traffic risk identification models by fusing the risk indicators is proposed in the current study. The methodology is applied for a study site to develop a model capable of identifying real-time traffic crash risk. In parallel with risk identification, appropriate preventive measures are to be elaborated such that they can be taken to clear or reduce the risk before crashes occur.

1.2 Objective

The main objective of the project is to contribute to the improvement of road traffic safety, especially for motorways, by incorporating additional risk information provided by risk indicators.

To attain the main objective, the work reported here targets following objectives:

- Study the applicability of risk indicators reported in literature
- Develop and test risk indicators to identify traffic crash risk on Swiss motorways.
- Develop models for identifying real-time traffic crash risk on motorways using risk indicators with the help of data fusion techniques.
- Suggest appropriate preventive measures to reduce crash risk.

1.3 Project FUSAIN

FUSAIN project has been launched at LAVOC, EPFL. It is supported by FEDRO (Federal Roads Office), Switzerland, as an ARAMIS project for the period from January 01, 2007 to October 30, 2009.

IFSTTAR-LICIT contribution on FUSAIN is from Jan 2007 to Sept 2009. EPFL-LAVOC completes and submits this report in 2012.

The team working on FUSAIN is composed of:

From EPFL :

- Dr. Minh Hai Pham, LAVOC – EPFL
- Prof. André-Gilles Dumont
- Prof. Dr. Edward Chung, LAVOC – EPFL (until march 2009)

From IFSTTAR-LICIT:

- Dr. Nour-Eddin EL Faouzi, LICIT, IFSTTAR -ENTPE
- Dr. Olivier de Mouzon, LICIT, IFSTTAR -ENTPE (until October 2008)

- Solenne Durand, LICIT, IFSTTAR -ENTPE (until February 2009)
- Fouad Baouche LICIT, LICIT, IFSTTAR -ENTPE (from March to September 2009)

1.4 Report Structure

Chapter II discusses the overview of traffic safety and then concentrates on two related aspects: traffic crash risk identification and risk indicators. In chapter III, the methodology and study sites are discussed. As the study is based on Swiss traffic conditions, potential risk indicators applicable for Swiss roads are determined. Chapter IV presents the selected risk indicators and their test of sensibility with real Swiss traffic data. The methodology is applied to generate traffic regimes which are presented in chapter V. The development of traffic regime-based models and the results are presented in chapter VI. In chapter VII, the conclusions are discussed.

2 Traffic Safety – State of the Art

2.1 Introduction

Road traffic safety is a branch of traffic engineering aiming to reduce the number of deaths and injuries as well as property damages that are the consequence of collisions of vehicles traveling on public roads. The main road traffic danger is represented by collisions between a vehicle with other vehicles, pedestrians, and moving obstacles, such as animals or stationary obstructions (e.g. trees or utility poles). Therefore, the main objective of road traffic safety is to reduce the number and the severity of such crashes.

In term of terminology, it is worth noting that there are organizations such as the NHTSA (1997), RoadPeace (2011), “a CRASH is NO accident” (2011), cautioning not to use the term accident and using other names, such as crash or collision instead. The idea behind this renaming is to draw the attention to the fact that crashes are not acts of fate but are predictable and preventable (NHTSA, 1997), whereas accidents are events which are out of human control. Thereby, the term accident is not used in the current research specifically because we aim to reduce the number or at least the severity of road traffic collisions.

Different efforts were dedicated to uncover the factors that cause crashes. Rumar (1985) using British and American crash reports, found that 57% of car crashes were due solely to driver factors, 3% solely to roadway factors, 2% solely to vehicle factors, 27% to combined roadway and driver factors, 6% to combined vehicle and driver factors, 1% to combined roadway and vehicle factors, and 3% to combined roadway, driver, and vehicle factors. These data suggests that, except collisions due solely to vehicle factors, roadway factors or to both, driver factors are wholly or partly involved in 94% out of the total crashes analyzed.

Crash risk interventions can be categorized in many forms, depending on the driver, the vehicle, and roadway factors. For example, drivers can be divided into different groups depending on age, type of vehicles, etc. such that the specific or typical problems of each group can be better addressed. In Switzerland for instance, young drivers from 18 to 24 years of age are recognized by the BPU (2011) as the group of drivers highly influenced by alcohol and having the highest rate of fatalities and severe injuries (10.5 every 100'000 inhabitants during weekdays and 17.9 every 100'000 inhabitants during weekends). Therefore, legal measures adopted against young drivers are usually strict. One of such measures is related to application of the “trial driving license” for a period of three years after the success of driving examination, called probationary period. The permanent driving license can only be obtained after three years if no traffic code violation by the driver is observed (SAN, 2011).

The interventions related to roadway or vehicle factors mainly aims to improve the design and maintenance. In Europe, there has been a Campaign for Safe Road Design since 2009, calling to make safe road design a Europe transport priority (EuroRAP, 2009). This European campaign is an expansion of the UK Campaign for Safe Road Design and targets to cut Europe's toll of road deaths and serious injuries by a third by improving road design (EuroRAP, 2009). For better addressing roadway factors, roads are classified into three main categories: built-up areas, non-built-up areas, and motorways. In many countries, such as in the UK (DfT, 2009) or in Switzerland (BPU, 2011), the highest number of casualties is observed in built-up areas, which is much more than on motorways. Therefore, a higher number of preventive measures are applied in built-up areas, especially on shared space locations, where vulnerable road users, such as pedestrians and bicyclists, can be found. Vehicle safety is also an effective approach to improve traffic safety by reducing the occurrences of driver's errors (called active safety) or by improving safety equipment and vehicle design in order to reduce crash severity once crash occurs (called passive safety). Examples of active safety are the in-board intelligent systems designed to assist drivers such as: automatic breaking systems, adaptive cruise control, pre-crash system, etc. Examples of passive safety equipment are: seat belts and air bags.

2.2 Human Factor

Although various efforts undertaken prove that traffic safety is improved, the concept of problem solving does not exist yet. Even when safe roads and safe vehicles are designed and implemented, the number of crashes can only be reduced, but does not disappear. The reason being is that driver factors are critical, and it is exactly the driver who decides and performs all driving tasks. As early as in 1939, Farmer and Chambers (1939) introduced the term accident proneness to indicate a personal idiosyncrasy predisposing the individual who possesses it in a marked degree to a relatively high crash rate. This means that a number of drivers have to be responsible for causing crashes because they have certain harmful personal characteristics. Later on, Rumar (1985) found that drivers are responsible for 94% of crashes. Yet even considering such a result, it is always possible to claim that drivers could have better reacted or could have been more prepared to avoid collisions due to road or vehicle problems. In order to improve safety, driver factor approaches emphasize safety education and motivation of persons aiming to modify unsafe driving behaviors.

However, driver blaming is not supported by many research groups. People following systems theory consider that collisions arise from interaction among humans, machines, and environment. Collisions are seen as a failure of the whole traffic system rather than a failure of the driver. Under normal conditions, it is assumed that there is the harmony between humans and environment, such that the chance of collision is naturally low. For this reason, drivers are victims of crashes, since the traffic system is too complex for the driver's capacity to process information.

To participate in road traffic, drivers are required to have sufficient driving skills including both physical and mental capabilities. Therefore, lapses and errors are influenced by perceptual, intentional, or judgmental processes. To avoid lapses and errors, the drivers can improve their limitations in observation as well as technical mastery. Vehicles and roads can also be improved to facilitate driver's behaviors. On the contrary, violations are believed to be based on motivational and/or social factors. General solutions to diminish violations deal with changing people's beliefs and/or motives for avoiding violations such as attitude campaigns, police surveillance, speed cameras, or influencing the driver subconsciously through smart-design of the road environment.

2.3 Risk Identification

2.3.1 Risk Indicators

Parameters used to characterize traffic risks linked to particular traffic situations have been studied since 1960s. They are called risk indicators and can be used in real-time. Risk indicators aim to predict risky traffic situations and allow identifying factors leading to crashes – based on which specific preventive measures can be applied to improve traffic safety.

Introduction

Motorway system has a vital role in transport system in many countries. It links cities and allows fast movements that can only be undertaken by high speed supporting vehicles such as automobiles, HGVs, and motorcycles. On the one hand, the fast movements on motorways allow reducing travel time. On the other hand, a series of motorway traffic safety regulations have to be applied to reduce traffic accident risks against such high speed traffic conditions where an accident on freeway once happens would be very serious. In general, motorways are safest roads by design.

Many studies have been undertaken to better understand and improve the motorway traffic safety. One of the research directions is to use traffic measures to characterize traffic situations for the purpose of identifying crash-prone traffic conditions and taking necessary operational actions to reduce the crash risk. These measures are called "risk indicators", which are traffic parameters that allow describing partially crash risk status of a road section with its unchanged infrastructural configuration. The indicators can characterize traffic flow

properties before the occurrences of events such as crashes, congestions, etc. that happened in the past. The indicators also allow identifying potential risky traffic conditions that may occur as the progress of the current traffic flow. It means in the both cases that traffic data have to be collected to be used in the analysis.

In this study, the term “risk indicator” may be replaced by the term “safety indicator”. For some indicators, their high values indicate high risk situations. They are called “risk indicators”. There are also indicators that their high values indicate safe situations. They are called “safety indicators”. If their values are low, they also indicate the crash risk.

Traffic Conflict Techniques

For many years, traffic engineers observed the traffic to find out the association between unsafe situations and operational problems. However, the procedures and conclusions were given based on opinion and judgment. The first traffic measure dated back in 1960s when a set of formal definitions and procedures for observing traffic conflicts at intersections was developed by two engineers working for General Motors [Perkins S. R., 1969]. This was then developed to Traffic Conflict Technique that is used in many countries (for instant, in Sweden since 1970, in England and in United States since 1979). The risk indicators proposed by these literatures are used to quantify the severity of hypothesized traffic crashes on many road types including motorways. Due to the lack of field data, the measures are principally based on traffic theory with the traffic assumptions and the manual count of vehicles by observers.

For a long time, Time-To-Collision (TTC) has been an important indicator that represents the time required for two vehicles to collide if they continue at their present speed and on the same path (Hayward, 1971). TTC plays an important role in Traffic Conflict Techniques in many countries. Horst and Hogema (1993) introduced TTC as a core parameter in the Collision Avoidance system. The TTC value of a vehicle implies the remaining time for that vehicle to collide with its front vehicle if both of them still keep their speed and travel direction. The smaller the TTC value, the more dangerous the situation is. Another condition for the collision to happen is that the speed of the considered vehicle has to be greater than the speed of its front vehicle.

$$TTC_i = \frac{v_{i-1} * Gap_i}{v_i - v_{i-1}}, \text{ with } v_i > v_{i-1}$$

Where TTC_i , V_i , Gap_i are respectively TTC value, speed, and time gap of the i -th vehicle – the considered vehicle. V_{i-1} is the speed of its front vehicle.

Other indicators similar to TTC were then presented within the Traffic Conflict Techniques in different countries such as Time-Exposed-TTC (TET) and Time-Integrated-TTC (TIT) (Minderhoud and Bovy, 2001).

Recently, Aron et al., (2007) presented two indicators: *IBTR* (Individual Braking Time Risk) or G-value and *PBTR* (Platoon Braking Time Risk) or J-values. The G-value represents the rear-end crash risk of a vehicle when its front vehicle stops. G-value is characterized by the relationship between the speed of that vehicle and the time gap between that vehicle and its front vehicle. If the speed is high while the time gap is low, the risk becomes higher.

$$G(i) = Max \left[0, \log_2 \left(\frac{Tb}{2 * Gap_i} \right) \right] = Max \left[0, \log_2 \left(\frac{1}{2} * \frac{v_i}{|\gamma_{max}|} \frac{1}{Gap_i} \right) \right]$$

Where γ_{max} , V_i , Tb , Gap_i are the maximum deceleration rate, the speed, the minimum braking time until the vehicle stops and the time gap of the considered vehicle. The maximum acceleration rate is defined based on pavement conditions and should reflect meteorological conditions on the road. The definition of *IBTR* also comply with the standard values of time gap and speed limit: if a vehicle keeps standard time gap and travels at speed limit, its crash risk will be zero.

The risk of collision with the preceding vehicle increases when vehicles are considered in a platoon: the number of “events” which can affect the current vehicle is proportional to the number of events which may occur on the preceding vehicles, so to the number of these vehicles.

That is why the Platoon Braking Time Risk (PBTR) or J-value for the vehicle “*i*” of the platoon is defined by the following formula:

- for the leader of the platoon: $J(1) = 0$
 - for the vehicle *i*, PBTR value is:
 - $J(i) = G(i) + J(i-1)$ if $G(i) > 0$
 - $J(i) = 0$ if $G(i) = 0$;

For the *i*-th vehicle, *PBTR* is additive, but reset to zero as soon as one vehicle is *safe*. *PBTR* is a cumulative risk indicator. The values of this parameter can be obtained based on individual vehicle data, e.g. time gap between two consecutive vehicles, individual vehicle speed which can be collected from motorway traffic loop detectors.

2.3.2 Risk Identification Models

Introduction

Researches in motorway traffic safety can be broadly divided into groups of studies based on the objects of the improvement, such as infrastructure, vehicles, drivers, etc. Improving traffic safety by improving traffic flow is also part of these studies, and aims at predicting the future trend of traffic risks by analyzing past crashes and suggesting countermeasures when risks are identified, in order to prevent potential future ones. As the traffic flow is the main object of our study, we classify our study into the group of motorway traffic safety studies using traffic flow data.

According to Golob et al., (2004), consecutively by Pande, (2005), studies on motorway traffic safety using traffic flow data can be divided into two groups, depending on the units of analyses undertaken: aggregate studies and disaggregate studies. Aggregation studies consider counts of crashes or crash rates for specific time periods (typically months or years), as well as for specific spaces (specific roads or networks), as units of analysis and use statistical distribution parameters of traffic flow for that specific time and space. For disaggregate studies, the units of analysis are the crashes themselves and traffic flow is represented by parameters of at the time and place of each crash. As our study examines each crash to verify the difference between traffic conditions leading to the crash in crash case and non-crash traffic conditions, it can be classified into the group of disaggregate studies. This classification is illustrated in Fig. 2.1, where traffic safety studies using traffic flow data (Traffic Flow related studies) are partitioned into the two smaller groups: aggregation studies and disaggregation studies.

Infrastructure related studies focus on improving traffic safety by improving the infrastructure. Stine et al. (2010), for example, investigated the safety of highway medians through iterative simulations of off-road median encroachments. Another example is given by the study of Donnell and Mason (2006), who investigate median design policies for high-speed divided highways to assess existing median barrier warrant criteria.

Vehicle and driver related studies investigate traffic safety issues that are applicable both for motorways and non-motorways. For vehicle related studies, Blum and Eskandarian (2006), reviewed the research that has been conducted on intelligent speed adaptation, and presented possible strategies to maximize both effectiveness and acceptability to mitigate deleterious effects on roadway safety. Umedu et al. (2010) focused on the inter-vehicular communication, namely on the distributed detection of dangerous vehicles on roads and highways, and proposed a dangerous-vehicle-detection protocol to detect drivers who violate speed limits. Klingender et al. (2009) concentrated on the different effects of adapting the maximum weight and dimensions of heavy commercial vehicles allowed by the European Commission by providing an overview of the in-depth safety analysis of heavy commercial vehicles on European roads. Regarding driver related studies, a large effort has been devoted to discover the influence of driver state on driving (Arnedt et al., 2001;

Rzepecki-Smith et al., 2010); (Mulder et al., 2008).

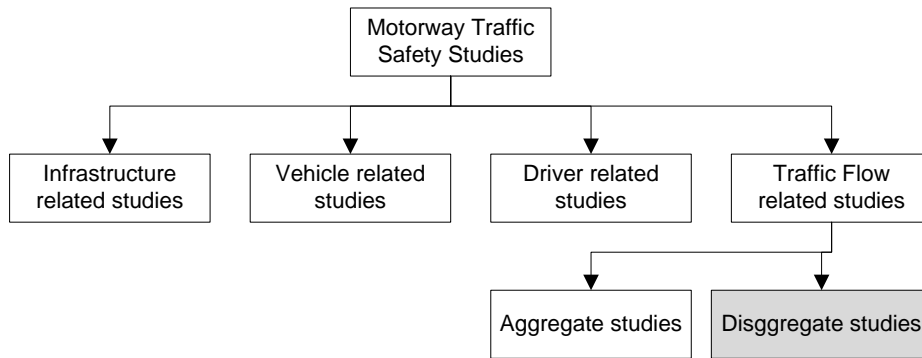


Fig 2.1: Research area of the current study

It is worth noting that the classification presented in Fig 2.1 is rather relative, as traffic is a complex system. To improve its safety the interaction between different groups is necessary and has to be addressed in a holistic way.

Aggregate Studies

During the past decades a large effort has been devoted to aggregate studies on motorways traffic safety. To date, aggregate studies are still widely used to find and to analyze the relationship between crash rate/frequency and other factors such as the congestions, infrastructure, weather effect, etc.

Wang et al. (2009) investigated the relationship between traffic congestion and road accidents to clarify the speculation that there may be an inverse relationship between the latter and road accidents. The study aimed to explore the impact of traffic congestion on the frequency of road accidents using a spatial analysis approach, while controlling for other relevant factors that may affect road accidents. The results suggest that traffic congestion has little or no impact on their frequency on the M25 motorway in the UK.

Aguero-Valverde and Jovanis (2006; 2008) focused on spatial analyses to produce spatial models. The authors tried to explore the effect of spatial correlation in models of road crash frequency at the segment level. Different segment neighboring structures were tested to establish the most appropriate one in the context of modeling crash frequency in road networks. A full Bayesian hierarchical approach was used with conditional autoregressive effects for the spatial correlation terms. Analyses of crash, traffic, and roadway inventory data from a rural county in Pennsylvania indicate the importance of including spatial correlation in road crash models. The models with spatial correlation show a significantly better fit than the Poisson lognormal model with heterogeneity only. Parameters significantly different from zero included annual average daily traffic (AADT) and shoulder widths of less than 4 ft, and between 6 and 10. One important result relates to the potential of spatial correlation to reduce the bias associated with model misspecifications.

Recently, Lord and Mannering (2010) provided a detailed review of the key issues associated with crash-frequency data, as well as reporting the strengths and weaknesses of the various methodological approaches that researchers have used to address these problems. The authors concluded that while the steady march of methodological innovation (including recent applications of random parameter and finite mixture models) has substantially improved our understanding of the factors that affect crash-frequencies, it is the prospect of combining evolving methodologies with far more detailed vehicle crash data that holds the greatest promise for the future.

In general, aggregate studies, although they bring more insights of crash frequency, do not propose any real-time models for identifying instant traffic crash risks.

Disaggregate Studies

Disaggregate studies are relatively new, and are made possible by the availability of data

being collected in support of intelligent transportation systems developments. Motorway traffic management centers routinely archive traffic flow data from sensor devices such as inductive loop detectors. Different disaggregate data types can be obtained from such sensor devices depending on the types of sensors and on the aggregation level of collected traffic data. For example, single inductive loop detectors collect traffic data for intervals of several seconds, and provide the vehicle count and the occupancy during the intervals. Double inductive loop detectors can provide more detailed data than vehicle counts and occupancies. The average speed during the time interval is the extra information. Moreover, double loop detectors can also provide individual vehicle data including: speed, headway, and time gap, length of each vehicle. It is worth noting that there is no time interval in collecting individual vehicle data (i.e. data are extracted whenever there is a vehicle passing by the detectors)

There are a small number of research groups who have undertaken disaggregate studies and have reported positive results in the literature. Most of the studies aim to develop real-time models for identifying traffic risk using disaggregate traffic flow data. The original idea of these studies is to compare non-crash traffic conditions (i.e. traffic conditions when no crash is recorded) and traffic conditions where traffic crash risk is high. In reality, it is difficult to determine which traffic conditions are highly risky. Therefore, crash databases are used to extract traffic conditions leading to crashes (*pre-crash traffic conditions*). Pre-crash traffic conditions are highly risky because they end up in crashes. By developing models capable of differentiating pre-crash and non-crash conditions, it is expected that a new real-time traffic condition can be collected and classified into non-crash or pre-crash conditions. Therefore, preventive actions can be implemented, if the traffic condition is in its “pre-crash” state, to avoid potential crashes.

3 Methodology, Study Sites and Data

3.1 Methodology

To conform to the objectives of the current project, the approach includes the following steps:

- 1-Identify applicable risk indicators. The selected indicators are tested with real data to verify their capacities in differentiating traffic conditions leading to crashes (risky conditions) and other conditions.
- 2-Apply a data sampling process to sample non-crash data. This is because crashes are rare events and rear-end and sideswipe crashes only occur under certain circumstance.
- 3-Develop single – variable risk identification models
- 4-Fuse results of single-variable models to obtain fused models with better performance
- 5-Provide prediction of short-term traffic risk using results of fused models in real-time framework.

Step 1 is important as the number of risk indicators is high and depending on the availability of data, not all risk indicators can be directly applied. In the second step, the objective is to match pre-crash traffic conditions with relevant non-crash conditions. This is because crashes are rare events and there must be certain context for rear-end and sideswipe crashes to occur. By sampling non-crash data, only non-crash conditions comparable with pre-crash conditions are kept for further processing. The data sampling process creates groups of traffic conditions called traffic regimes such that under each traffic regime, there is a list of non-crash conditions and a list of pre-crash conditions. Step 3 is undertaken under each traffic regime where pre-crash and non-crash cases are relevant. Variables can be risk indicators selected or traffic parameters. Each variable has certain performance in identifying traffic crash risk. Step 5 is also undertaken under each traffic regime aiming to combine the performance of all variables by fusing the single-variable models. Finally, to predict the traffic crash risk of the next traffic case, the risk assessment of several last traffic cases provided by fused identification models is used to provide a judge whether the next traffic case is risky or not.

3.2 Study Sites and Data

3.2.1 Swiss Motorways Network

In Switzerland, motorways are called autobahnen in German, autoroutes in French or autostrade in Italian. Two of the most important motorways are the A1, connecting St. Margrethen in eastern canton of St. Gallen to Geneva in western part of the country, and the A2, connecting Basel in the northern part to Chiasso in canton of Ticino in the south. The common speed limit on Swiss motorways is 120 km/h.

The Federal Roads Office (FEDRO) is the Swiss authority responsible for the country's road infrastructure and private road transport. As of the January 1st, 2008, its range of duties increased significantly. With the entry into effect of the redistribution of financial responsibility and the accompanying division of duties between the federal government and the cantons, it assumed the functions of developer and operator of the motorway network. It belongs to the Federal Department of the Environment, Transport, Energy and Communications (DETEC), and focuses on securing sustainable and safe mobility on the country's roads.

According to FEDRO, (2009), a total of 1'765.6 kilometers of motorway are currently in operation. The network is planned to comprise 1'892.5 kilometers. The remaining 126.9 kilometers are expected to be completed within the next 15 years. Until the end of 2008 a total of 1'765.6 kilometers of motorways were in operation including:

- 7-lane sections 1.2 km
- 6-lane sections 80.7 km
- 4-lane sections 1'300.8 km
- 3-lane sections 1.9 km
- 2-lane sections 269.5 km
- Mixed sections 111.5 km

This corresponds to 93.3 percent of the planned network presented in Fig 3.1.

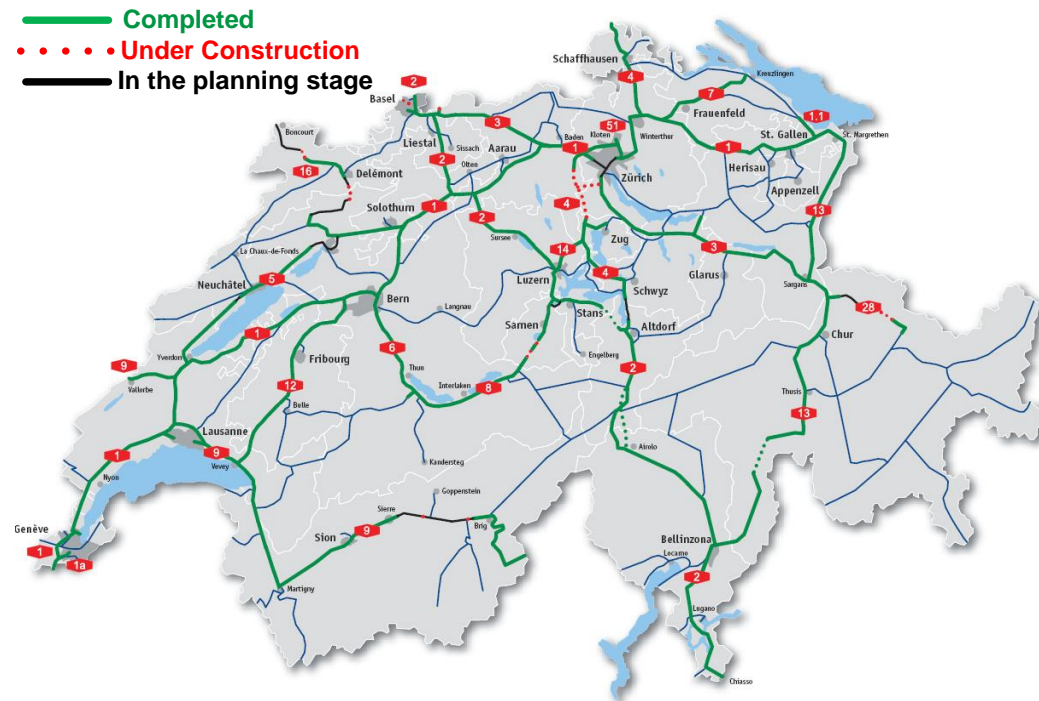


Fig 3.1 Swiss motorway networks. Source: FEDRO

3.2.2 Traffic Detectors

Until Jan, 2010, about 300 traffic detector stations are installed on Swiss motorways and provide online downloadable data. The map of the stations is presented in Fig 3.2. According to FEDRO, (2009), there are criteria to choose the location on the road for installing traffic detectors. The criteria serve for two main purposes: traffic statistics and better traffic management. The traffic data collected are used by engineering companies or traffic operators.

Most of traffic detectors on Swiss motorways are double inductive loop detectors. Each loop detector records the time of passage and the duration of presence of each vehicle on the detector. Two loop detectors installed at a distance of 4 meters make up a double loop detector on one road lane collecting individual vehicle information when vehicles pass by the detectors.

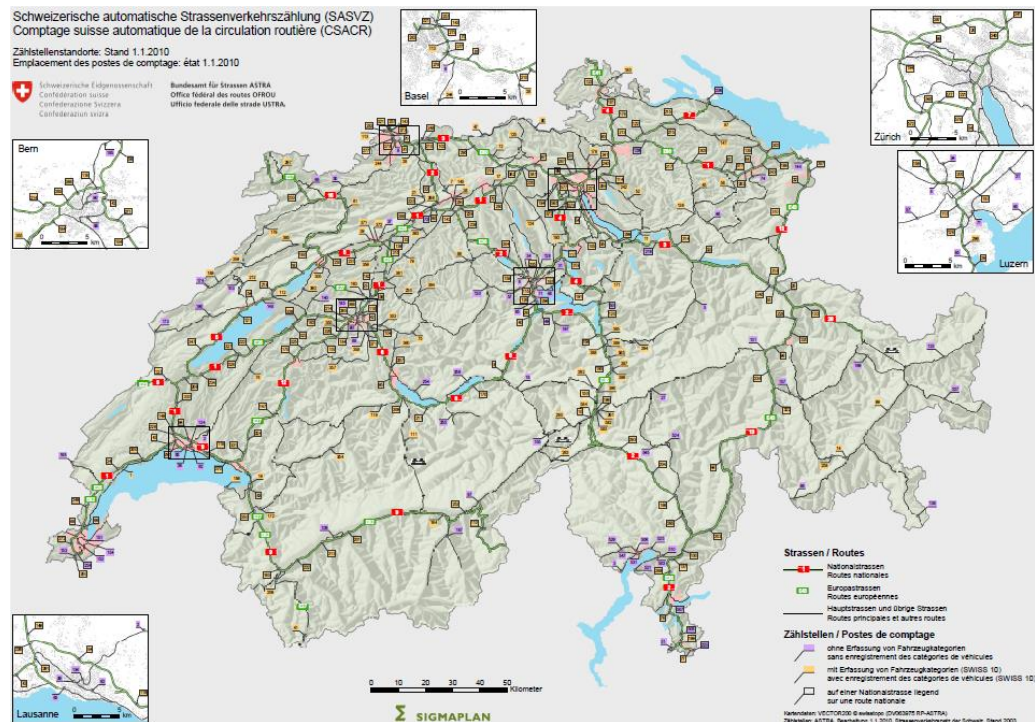


Fig 3.2 Swiss automatic road traffic counts

Together with traffic data detectors, there are also meteorological sensors for measuring weather information. There are two meteorological station systems exist in parallel in Switzerland: MeteoSwiss system and Boschung system.

The Boschung stations are installed along Swiss roads, including motorways for collecting road weather information aiming to improve road maintenance and safety. According to Boschung, (2010), the information collected depends on particular stations. Unfortunately, due to technical problems, only data from Boschung stations on motorways in Vaud canton are accessible and used in this study.

The other meteorological system in Switzerland comprises MeteoSwiss stations providing weather information for general purposes including daily weather forecast. MeteoSwiss stations are installed to collect information for the whole area around the station.

3.2.3 Meteorological Data

Boschung Stations

Depending on locations, Boschung stations provide meteorological data with the following data fields:

- i. Station code
- ii. Station name
- iii. Date time
- iv. Temperature
- v. Soil Temperature
- vi. Relative of Humidity
- vii. Dew Point
- viii. Type of precipitation (Three types: rain, snow, or no precipitation)
- ix. Quantity of precipitation (Five levels: nothing, weak, normal, strong, very strong)
- x. Number of spraying (this is based on the deicing program with salt solution)

Each data line represents a query for 5-minute intervals. While data fields iv, v, vi, vii, and x contain quantitative values, data fields viii and ix are categorical.

MeteoSwiss Stations

Fig 3.3 presents the layout of a program allowing extracting MeteoSwiss data. MeteoSwiss data from all MeteoSwiss stations are managed and stored on a central server. Data extraction is undertaken via Java-based software Clipmap. The software can provide meteorological data at many aggregation levels. The most detailed data represent information for 10-minute intervals. Some example meteorological parameters are instant temperature at two meters above ground (*tre200s0*), instant relative humidity at two meter above the ground (*ure200s0*), the total precipitation (in mm) of the last 10 minutes (*rre150z0*).

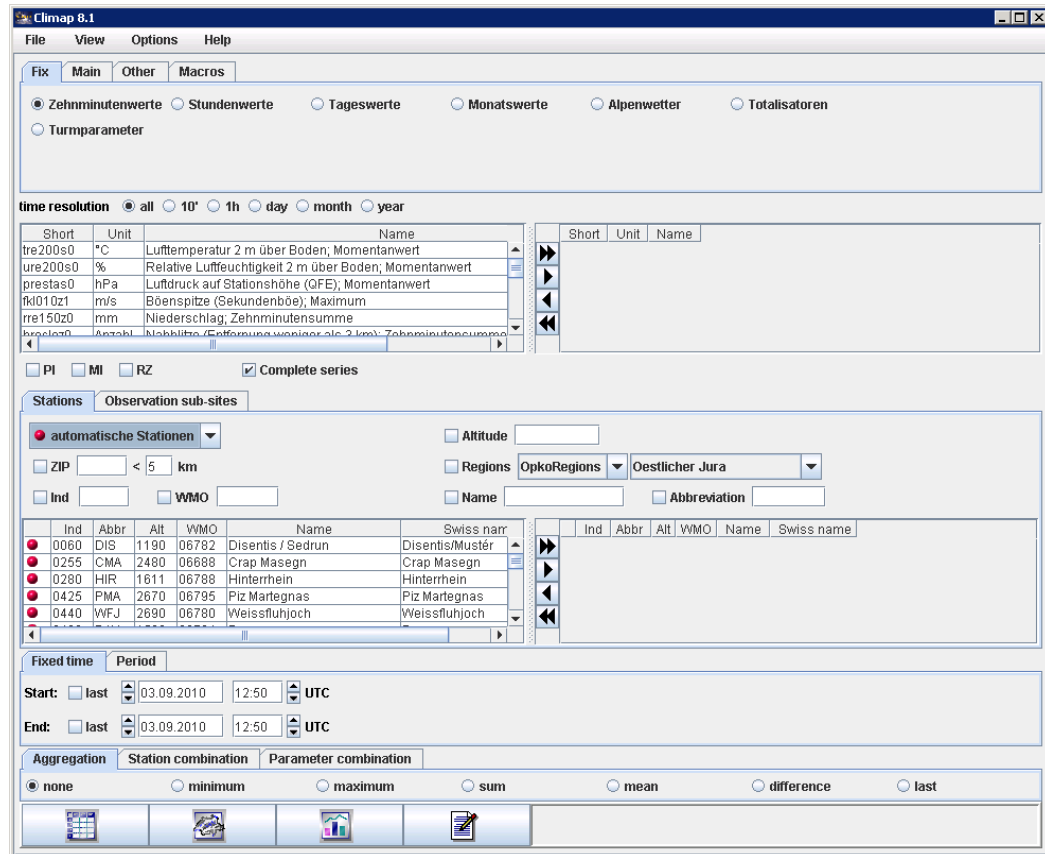


Fig 3.3 Clipmap interface (in English and German)

3.2.4 Summary

MeteoSuisse and Boschung stations are alternatives that provide meteorological data. Whereas Boschung data is specific to road weather conditions, MeteoSuisse data is more general for localities. The common characteristics of these two types of stations include the lack of sight distance, the presence of fog or the information on ambient conditions.

However, data from these stations can still be useful for the current study with information such as precipitation, temperature, etc.

Crash Database

In Switzerland, all traffic crashes recorded by the police are informed to FSO who stores the data for statistics purposes. According to FSO, (2010), the concept of crash depends on the mode of transport. Road traffic crashes since 1992 are all crashes occurring on public roads. Yet before that until 1991, only crashes causing damage estimated at more than 500 Swiss francs (until 1975 the limit was set at 200 francs) were taken into account. Since 2002, a crash is considered as road traffic crash if it occurs on public roads and causes injuries. Anyone who has suffered injuries, regardless of gravity, falls into the category of injured.

According to FSO, (2010), since 1992, the road traffic police collect anonymous individual data on traffic crashes, vehicles and people involved with features registered including:

- Circumstances of the crash (date, time, type of crash, type of road, location, conditions, etc.)
- Type of vehicles involved, information on drivers (age and driving license) and the people involved (position in the vehicle security system, consequences of the crash, sex, age)

The submission of crash data to the FSO needs to be transmitted using informatics or statistical survey forms which is listed in (FSO, 2005). The statistical survey forms are considered as the statement of crashes or a guideline for the census of the crashes. The FSO also provides the guideline on how to fill the survey form (FSO, 2005).

The survey form can be used for any road traffic crash type and for any road type. Ten crash categories are used to distinguish different crash types as listed in Fig 3.4.

Motorways are safe by design such that the probability for some accident types such as accidents of categories C, G, and H to occur is minimized whereas; some other crash types such as crashes of categories B, D, E, and F are more common. Fig 3.5 presents the percentages of four most common crash types plus the percentage of all other crashes for six years from 2002 to 2007 with a total of 46'641 crashes on Swiss motorways.

Fig 3.4 Ten crash types for all road types

Category	Name
A	Crashes related to pedestrian
B	Skidding or losing control
C	Crossing crashes
D	Crashes while overtaking
E	Rear-end crashes
F	Crashes while changing lanes (for pre-selection) or bypassing
G	Crashes while changing directions
H	Crashes while turning (without changing the direction)
I	Crashes caused by an animal
K	Other crashes

Note that crashes under category B are considered as single vehicle crashes. According to FSO, (2005), a crash is classified into category B when the driver tries to avoid a collision or deviates from the carriageway by his own fault. There should have been no prior collision with another road user; otherwise it is another type of accident. The crash due to evasive action, for example triggered by an overtaking maneuver of a vehicle on other direction, is classified into this category. When a driver overtake or being overtaken loses control of his vehicle, the correct type is however an accident when overtaking.

The category D of crashes includes only crashes related to overtaking. The collision occurs between the overtaking vehicle and a vehicle overtaken, a vehicle traveling in the opposite direction or a vehicle traveling behind the first vehicle and already conducting an overtaking.

Rear-end crashes under category E indicates crashes by a vehicle hitting another vehicle, whether moving or momentarily stopped, which borrows the same way. If the first vehicle hits a parked or permanently stopped vehicle, it is an accident by slipping or loss of control (of category B), or other type of crash.

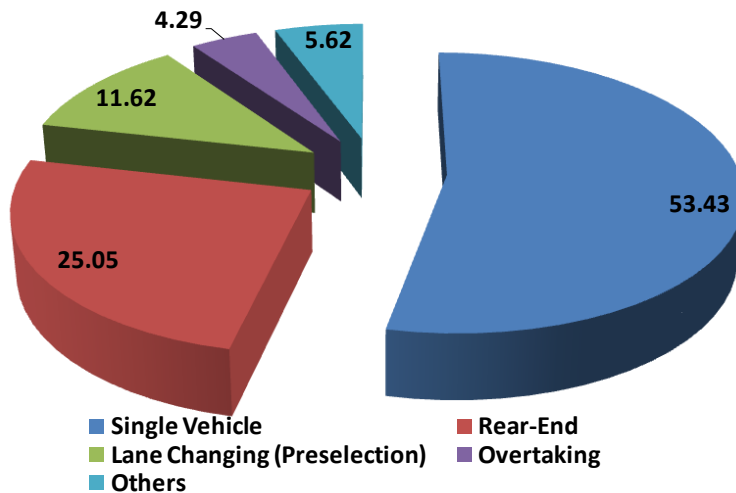


Fig 3.5 Percentages of most common crash types on Swiss motorways

Crashes under category F can be divided into two cases: bypassing and pre-selecting. The bypassing collision between two vehicles traveling in the same direction applies when none of the two vehicles intend to change lanes. A pre-selecting collision occurs when a vehicle collides before, during, or after a lane change, with another vehicle traveling in the same direction on a different lane. The lane change is planned or executed in order to pre-select before a junction, an on/off-ramp of motorways, but not to overtake or attempt to overtake.

The details of crashes are presented in the statistics survey form (FSO, 2005). There are three parts to be filled in for each crash: general about the crash; the object causing the crash; and the persons involved in the crash. The general information about a crash includes time, location, type, and external conditions (such as weather, pavement status, lighting conditions, etc.). Most of the general information is important for the present study. In the second part of a crash record, the information about the cause of the crash is important as it allows identifying whether the crash was caused by the driver's status, the vehicle's technical problems, or external factors. The third part of the crash record is not of interest of the current study.

Study Sites

A study site is selected and used in this study. The procedure of study site selection includes the following steps:

- i. From the list of more than 300 traffic detector locations providing traffic, count for each detector location the number of crashes satisfying conditions within the buffer of 1km from the location, called *considered crashes*. Re-arrange the list in descending order of the number of considered crashes.
- ii. Remove from the ordered list all traffic detector locations that there is no meteorological station within a radius of 15km from the traffic detector locations. The remaining detector locations make up the reduced list of locations.
- iii. Select the first location from the reduced list of locations.

As road crashes are monitored by cantonal police (there are 26 cantons in Switzerland), crash positions can be undetailed in the crash records for several cantons such as Vaud or Ticino, etc. Therefore, study site cannot be selected in those cantons.

The selected study site is at 27th position of the considered list. The study site is named according to the name of traffic detector stations: CH023. The location of site CH023 is illustrated in Fig 3.6. Site CH023 lies on motorway A1 connecting two cities Bern and Zurich. The site CH023 is not the location where crash occurrences are the most numerous yet is selected because traffic, meteorological and crash data are altogether available, which is crucial in development of risk identification models by applying the methodology proposed in section II.1.

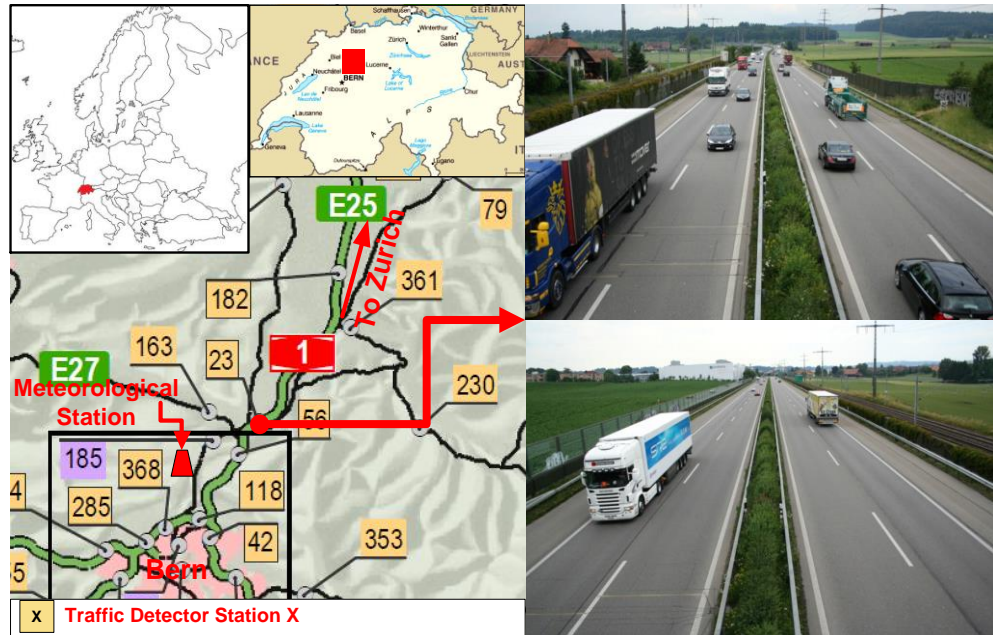


Fig 3.6 Study site in Bern canton

The road section at site CH023 includes two traffic directions: Bern - Zurich and Zurich - Bern. On each of directions, the road pavement is divided into two lanes. The lane on the left is called *normal lane* whereas; the lane on the right is called *overtaking lane*. Within 1km from the location of traffic detectors, road sections are straight and the inclination of the section is almost zero.

3.3 Preliminary analysis of accidents

This section provides different statistics on crashes based solely on crash data. The statistics in this section can be referenced in the next chapters.

3.3.1 Crash Distribution by Time of the Day and Day of the Week

Fig 3.7 presents the crash distribution. Here the crash time in the crash records is used. In Fig 3.7, the last row represents the distribution of crashes by time of day averaged by all the days of the week. The last column represents the crash distribution by day averaged by all time of the day. The number of crashes is the highest on Friday and then on Saturday. Crashes occur more often in late afternoon on Friday (from 16:00 to 19:00) than in any other time. Although the crash frequency is high on Saturday, crash occurrences did not concentrate on a particular hour of the day. Crashes occur more often in early Sunday mornings than in early morning of any other day of the week.

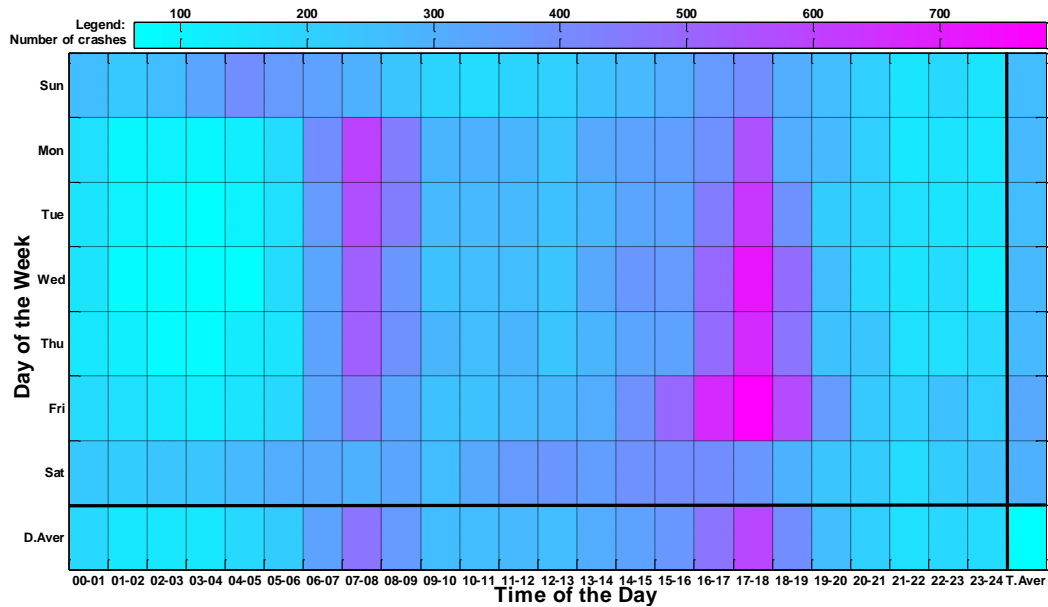


Fig 3.7 Crash distribution by day of the week and time of the day

The statistics in Fig 3.7 indicates that time of the day and day of the week contribute to crash occurrences.

3.3.2 Crash Distributions by Weather Conditions and Pavement Conditions

The crash distribution by weather conditions is presented in Fig 3.8. The crash distribution by pavement conditions is presented in Fig 3.9. The two distributions show that even when there is no precipitation in the air, the pavement can still be non-dry, i.e. there can be other sources for the non-dry state of the pavement.

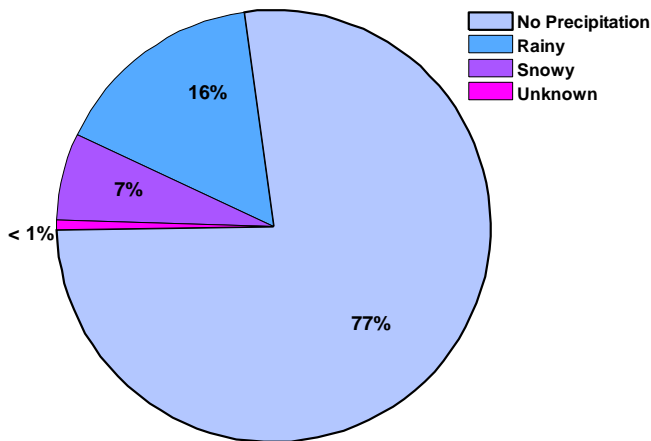


Fig 3.8 Crash distribution by weather conditions

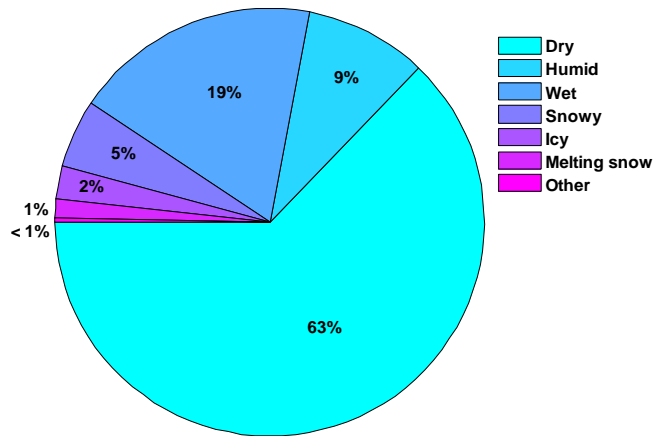


Fig 3.9 Crash distribution by pavement conditions

3.3.3 Crash Distributions by Crash Severity

There are many criteria to evaluate the severity of crashes. Fig 3.10, Fig 3.11, and Fig 3.12 present three different views on the crash severity.

In Fig 3.10, the number of objects involved in a crash can be as high as 17. As the crashes are on motorways, the objects are vehicles. Due to the high speed practiced on the motorways, it is not abnormal to observe many vehicles involved in a crash.

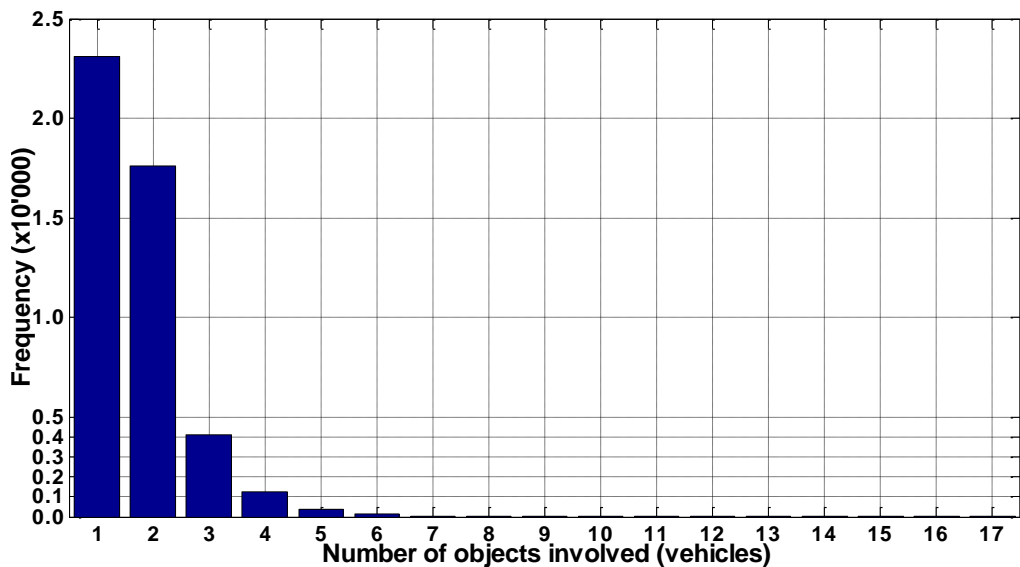


Fig 3.10 Crash distribution by the number of objects involved

In Fig 3.11, the crash severity is viewed according to the number of persons involved in the crash. Depending on the capacity of vehicles, there can be more than one person occupying a vehicle. The maximum number of persons involved in a crash is 86.

Fig 3.12 presents the crash distribution by the number of injuries. The number of crashes without any injury is high, contributing up to about 75% of all crashes. However, there are also crashes with the number of injuries increasing up to 27 or 57, which is high.

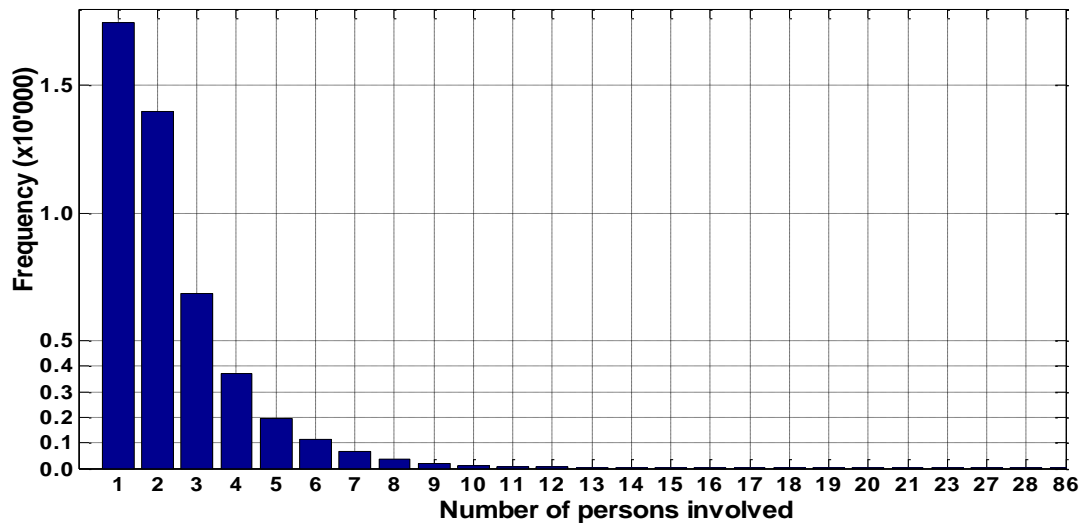


Fig 3.11 Crash distribution by the number of persons involved

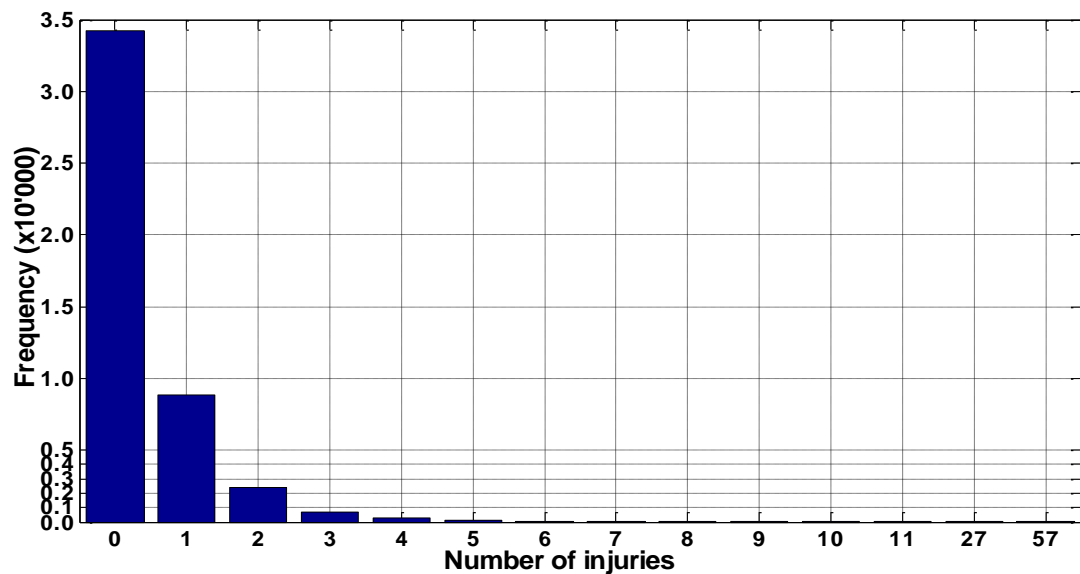


Fig 3.12 Crash distribution by the number of injuries

3.3.4 Crash Statistics at Study Site CH023

The statistics presented in Fig 3.12 and Fig 3.14 are based on all original crash data at study site CH023, i.e. the crash time was not corrected and crashes other than traffic-induced crashes are still counted. On a road section of 2 km (i.e. one km each direction from the location of traffic detectors) and during 7 years (i.e. from 2002 to 2007), there are 85'841'219 vehicles passing by traffic detectors. On the same road section and during the same period, 170 crashes are reported by the police. As result, the crash rate for the road section during the period is 99.02 million of vehicle kilometers traveled.

Fig 3.13 suggests that most of crashes (about 67.6% of all crashes) at study site CH023 are rear-end whereas the proportion of single vehicle crashes is about 26.5%.

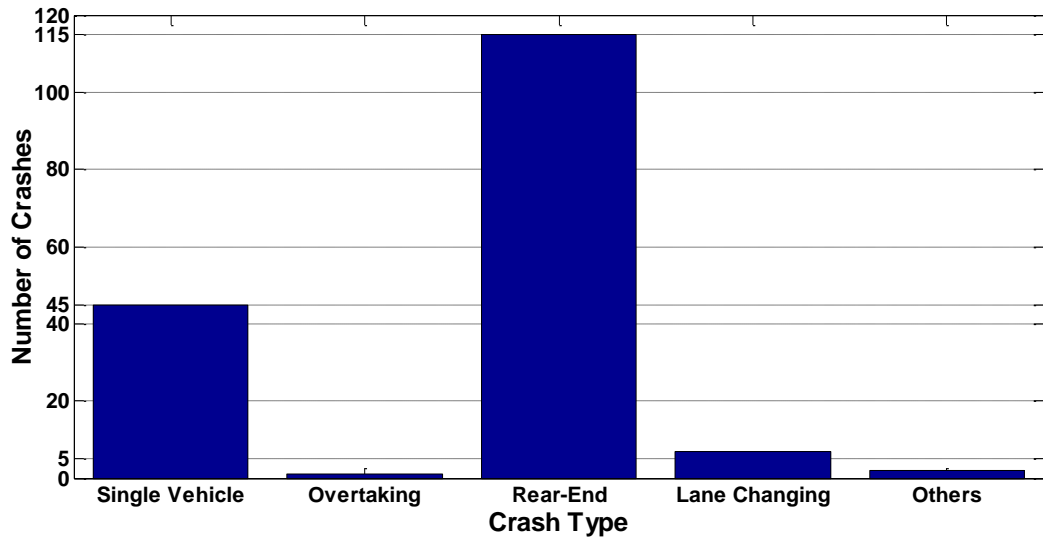


Fig 3.13 Crash distribution by crash types (Site CH023)

Fig 3.14 suggests that most of the crashes recorded at CH023 (about 51.2% of all crashes) occurred from 16:00 to 18:00 (i.e. period of two hours in the late afternoon). Besides, two crash peaks in Fig 3.14 also correspond to morning and evening peaks when the traffic flow is high.

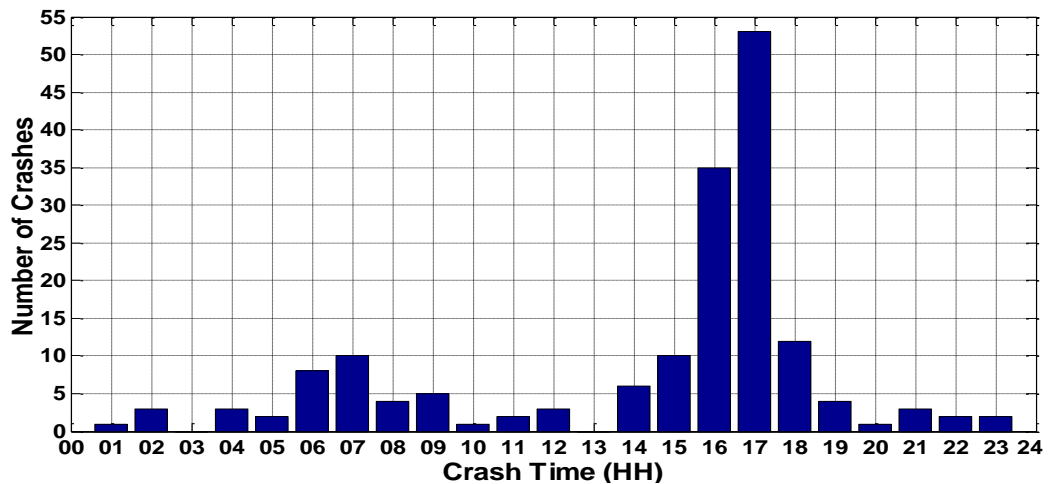


Fig 3.14 Crash distribution by crash time (Site CH023)

Speed/flow relationship of traffic conditions on the right lane before crashes is presented in Fig 3.15. Here, most of crashes occurred under high flow conditions. Several crashes occurring under low flow – high speed conditions are single-vehicle crashes. Crashes occurring under low flow – high speed conditions are rear-end crashes under congestions.

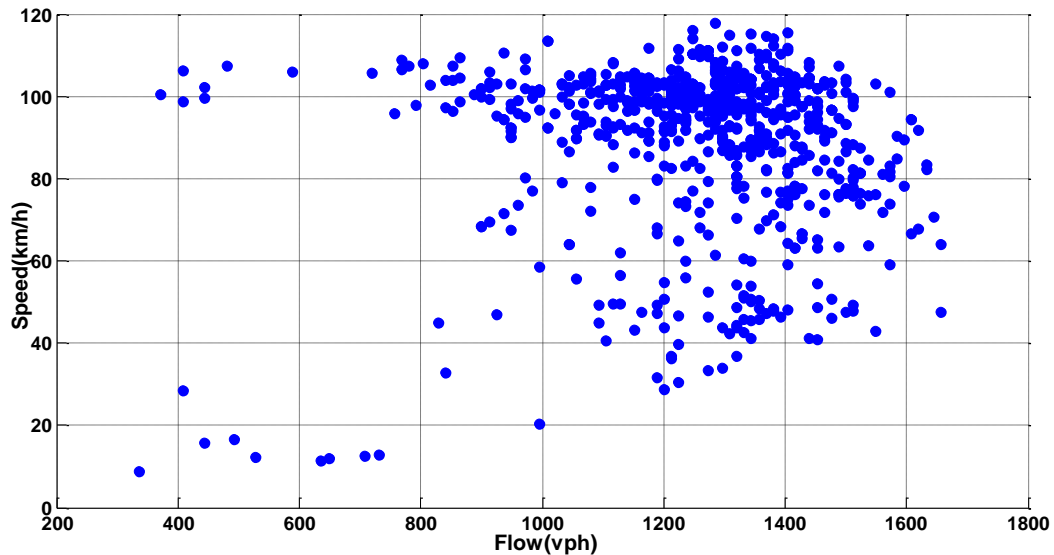


Fig 3.15 Speed/Flow diagram for traffic conditions preceding crashes on the right lane. Dots representing 5-minutes aggregated speed and flow

The speed profile based on the time of the day is presented in Fig 3.16. The clear speed drop can be observed from 11:00 on weekend and from 16:00 on weekdays which starts the afternoon peaks. This means that the congestions did occur although without high frequency. There is almost no congestion during the morning peak at the study site.

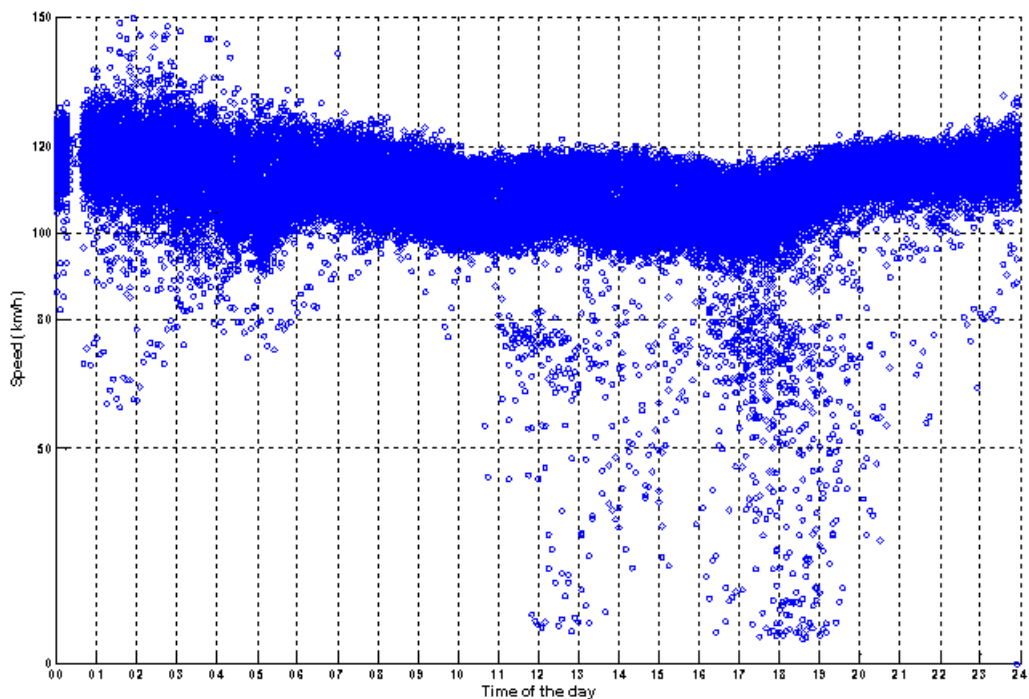


Fig 3.16 Speed profile according to time of the day (weekdays and weekend included)

3.4 Potential Risk Indicators

In section 3.1, different types of risk indicators are presented. The selection of risk indicators for this study is undertaken. As this study focuses on motorway traffic safety in Switzerland and due to the particularity of individual traffic data collected from traffic detectors, the selected risk indicators should match to the following criteria:

- Can be obtained based on a single traffic detector. Some of risk indicators in section 3.1 can be calculated based on a single traffic detector but their usage requires the presence of many adjacent detectors. The risk indicators like that are also excluded from this study.
- Can be obtained based on individual traffic data.

The potential risk indicators are listed in Fig 3.17. These are the risk indicators that satisfy the aforementioned criteria.

Fig 3.17 List of potential risk indicators

Risk Indicators	Description
TTC	Time To Collision
IBTR/G-value	Individual Braking Time Risk
PBTR/J-Value	Platoon Braking Time Risk

4 Sensitivity analyses of risk indicators using field data

4.1 Objective

This study looks at the status of traffic safety through the patterns of risk indicators. The understanding about the performance of risk indicators under traffic conditions in the past is important to determine if the risk indicators are applicable to monitor road traffic risk and if they can be integrated into a pro-active risk monitoring system.

The sensitivity test of risk indicators is a way to find out the performance of risk indicators with data in the past, including traffic data, meteorological data, and crash data. The sensitivity test composes of two following steps:

- Find out the distributions of risk indicators from which, extreme values as well as bound of risk indicators can be identified.
- Check the evolution of risk indicators' values just before crashes. As risk indicators bear information about crash risk on the road, their values should reflect the crash risk. However, it is not always evident to observe a traffic conditions with high crash risk. Therefore the traffic conditions just before crashes are good moments to check the risk indicators because the crash risk at that time turned into real crash later.

The sensitivity test should provide the performance evaluation of risk indicators which serves as the guideline for the later usage of risk indicators in other stages of this study.

4.2 Factors Influencing Traffic Safety

Traffic safety can be influenced by many factors such as geometry of the road section, meteorological conditions, traffic flow as well as drivers' behaviors. This research aims at improving traffic safety by reducing accident risk that can be caused by the meteorological conditions and by traffic flow. Hence, the selections of a straight and flat ATC-equipped section and of non-DUI crashes help minimize the effect of other factors.

The factors below are considered in this research:

- Traffic Flow: Motorway traffic flow is self-adapted with the interaction of vehicles with various operational regulations such as speed limit, minimum time gap, etc. Traffic flow can affect traffic safety in many ways. In this paper, different ranges of traffic volume will be considered to test the sensitivity of safety indicators.
- Meteorological Conditions: There are three weather conditions according to precipitation types from weather data: no precipitation (fine weather), rain, and snow. With the selected safety indicators, the effect of each weather condition can be quantified by a parameter such as the maximum deceleration rate (see [Chung (Ed.), 2007]) as shown in Fig 4.1.

Fig 4.1 Maximum deceleration rate is a function of weather conditions

Weather conditions	Maximum deceleration rate (m/s ²) - γ_{max}
Fine weather	6.87
Rain	4.81

4.3 Sensitivity analyses

4.3.1 Non-crash Cases

Processing steps

Traffic data and meteorological data for the whole year of 2005 were used to observe the distribution of selected risk indicators. The algorithm to process the data includes following steps:

- Calculate risk indicator values for each vehicle. This calculation involves traffic data and meteorological data. For each vehicle the selected data set, there is a set of values calculated for it such as speed, headway, time gap, vehicle type, and risk indicators' values.
- Aggregate the data by lanes with the aggregation interval of 5 minutes. The aggregation process should provide the following information for each 5-minute interval:
 - The traffic volume of the lane during that 5-minute interval
 - The average speed
 - The distribution of vehicle types
 - The distributions of risk indicators.
 - Determine the flow range of that 5-minute interval.

Flow ranges are used because traffic properties under each traffic flow range are very different. To understand the performance of risk indicators, they should be considered under separated flow ranges. There are five flow ranges used in this study representing the traffic flow conditions from free flow to dense flow. The five flow ranges are: 0-500vph, 500-800vph, 800-1100vph, 1100-1500vph, and greater than 1500vph.

In the next sub-sections, the distributions of selected risk indicators are presented. Some of the risk indicators may be influenced by the meteorological conditions and as consequence, their distributions under two meteorological conditions are essential. The two meteorological conditions are fine meteorological conditions and rain conditions. At the study site, road sections contain two traffic directions. For each direction, there are two lanes. The left lane is called the slow lane while the right lane is called the fast lane.

As the goal of this study is to focus on risky traffic situations, the safety domains of risk indicators are not presented in the figures of section 3.3.3. Depending on risk indicators, the risk domains of risk indicators can be their high or low values.

IBTR or G-value distribution

In Fig 4.2, Fig 4.3, Fig 4.4, and Fig 4.5, the IBTR distributions are presented for the slow lane and the fast lane, under fine conditions or rain conditions.

On the same lane, IBTR tends to be higher under rain conditions. This trend is expressed in two ways:

- Upper bound values under rain conditions are higher than upper bound values under fine meteorological conditions. For example, on the slow lane, the maximum IBTR value can reach to 9 under fine meteorological conditions (see Fig 4.2, flow range greater than 1500vph) and 10 under the rain conditions (see Fig 4.4, flow range greater than 1500vph).
- Higher percentage of vehicles having near IBTR upper bound value under rain conditions than under fine meteorological conditions. This means that there are more vehicles seen as risky by IBTR.

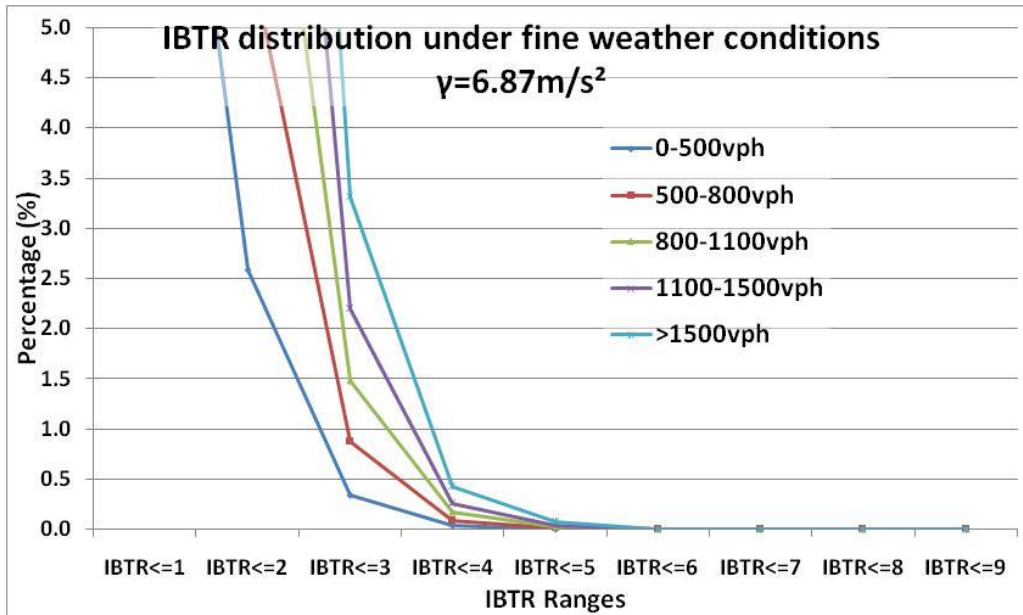


Fig 4.2 IBTR distribution under fine meteorological conditions on the slow lane

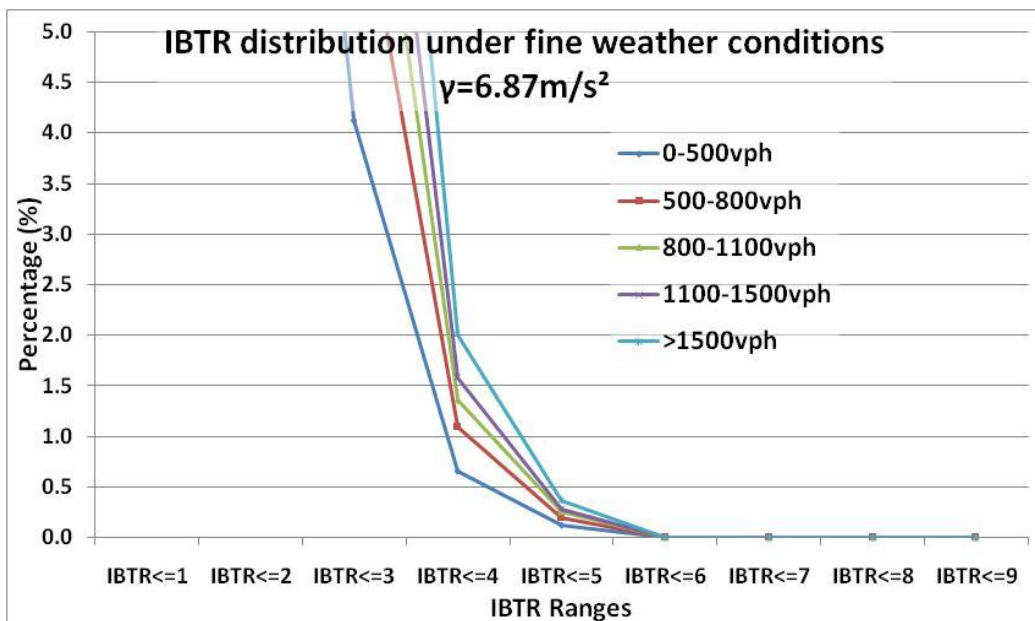


Fig 4.3 IBTR distribution under fine meteorological conditions on the fast lane

Under the same meteorological conditions, IBTR values tend to be higher on the fast lane than on the slow lane. Comparing Fig 4.2 with Fig 4.3 (fine meteorological conditions) or Fig 4.4 with Fig 4.5 (rain conditions), for the same IBTR range, higher percentage of vehicles having that IBTR range on the fast lane than on the slow lane. This is due to the higher average speed on the fast lane.

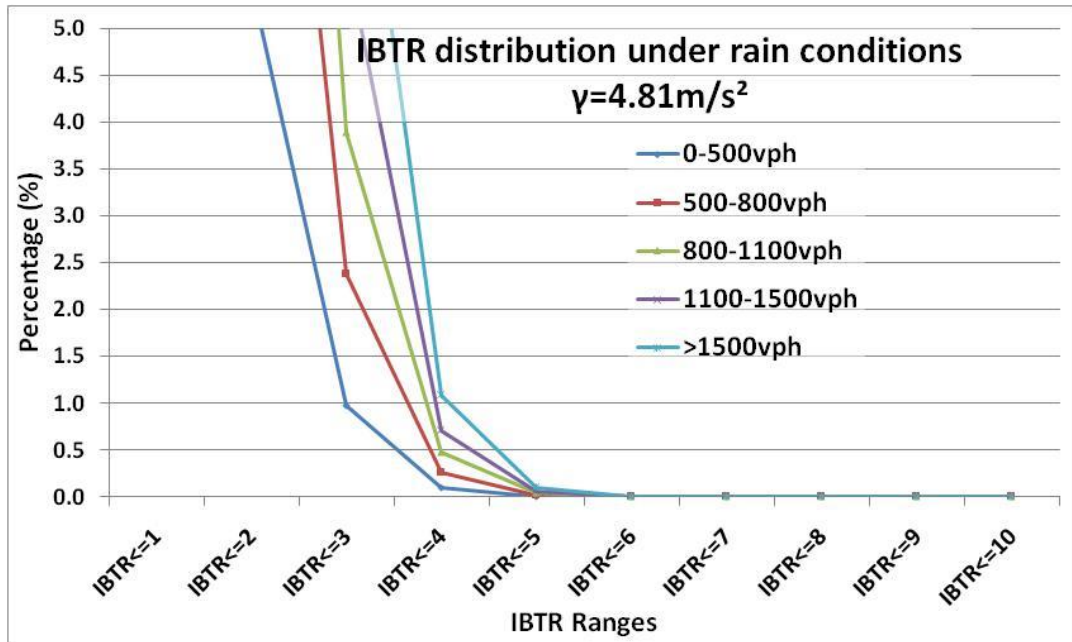


Fig 4.4 IBTR distribution under rain conditions on the slow lane

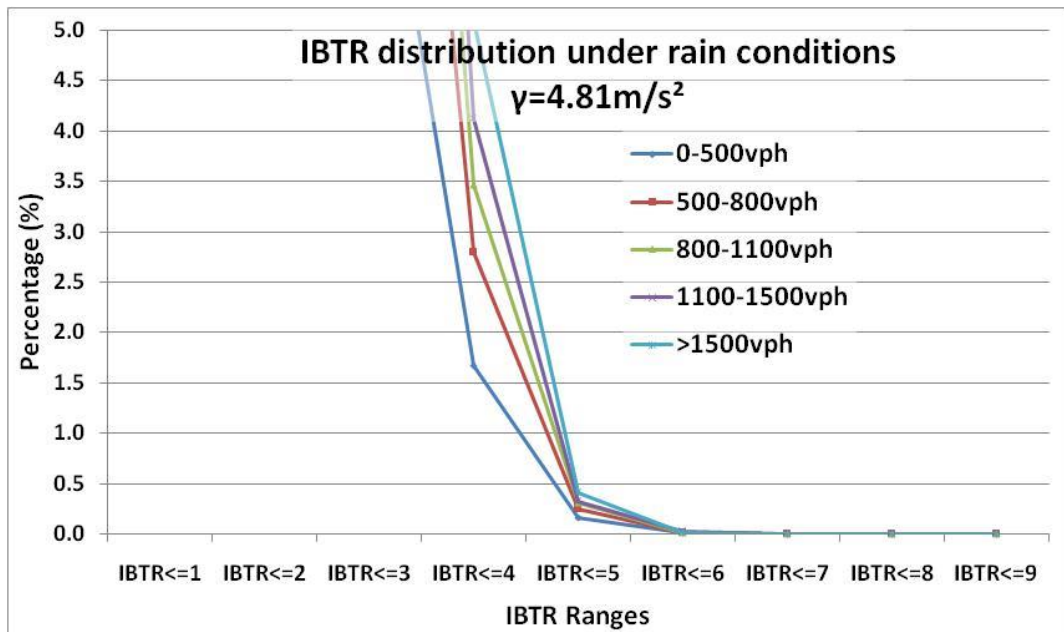


Fig 4.5 IBTR distribution under rain conditions on the fast lane

PBTR or J-value distribution

PBTR is an accumulation of IBTR over vehicles in a platoon. Therefore, value ranges of PBTR are extended and the trends that are observed with IBTR become bolder. Under fine meteorological conditions, PBTR values can raise up to 37 on the slow lane (see Fig 4.6) and more than 49 on the fast lane (see Fig 4.7).

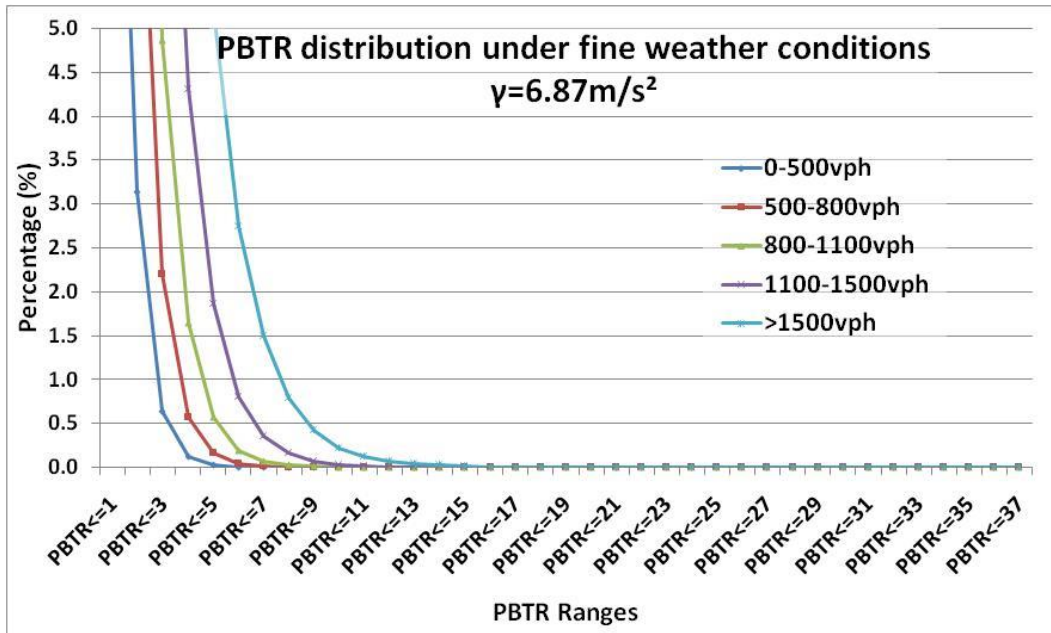


Fig 4.6 PBTR distribution under fine meteorological conditions on the slow lane

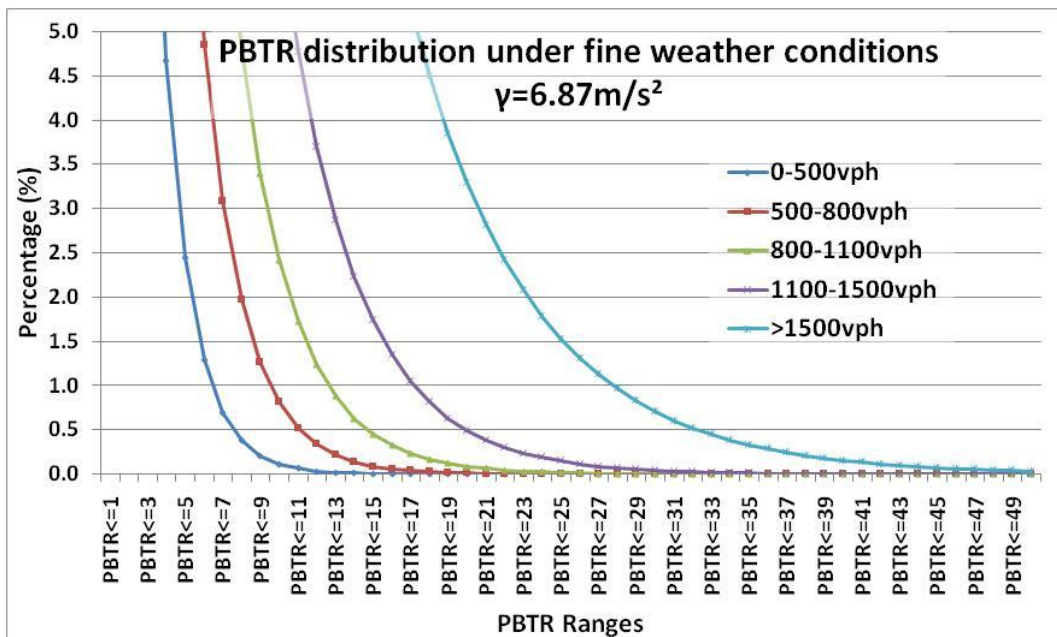


Fig 4.7 PBTR distribution under fine meteorological conditions on the fast lane

PBTR is more sensitive to meteorological conditions. For example, on the slow lane, the upper bound of PBTR is 37 for high flow range when the meteorological conditions are fine and can rise up to 43 when it rains.

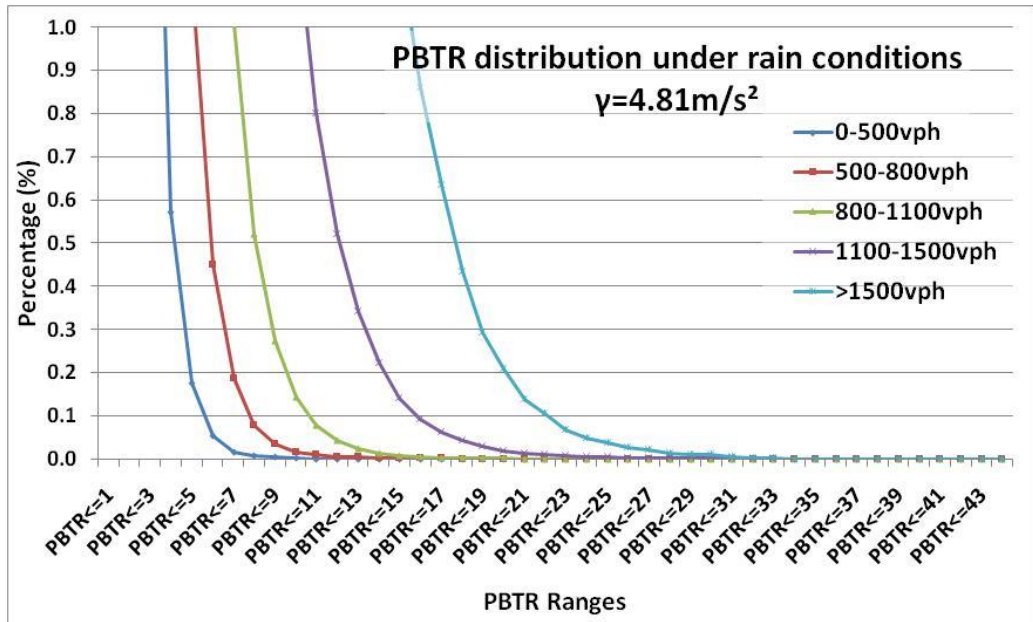


Fig 4.8 PBTR distribution under rain conditions on the fast lane

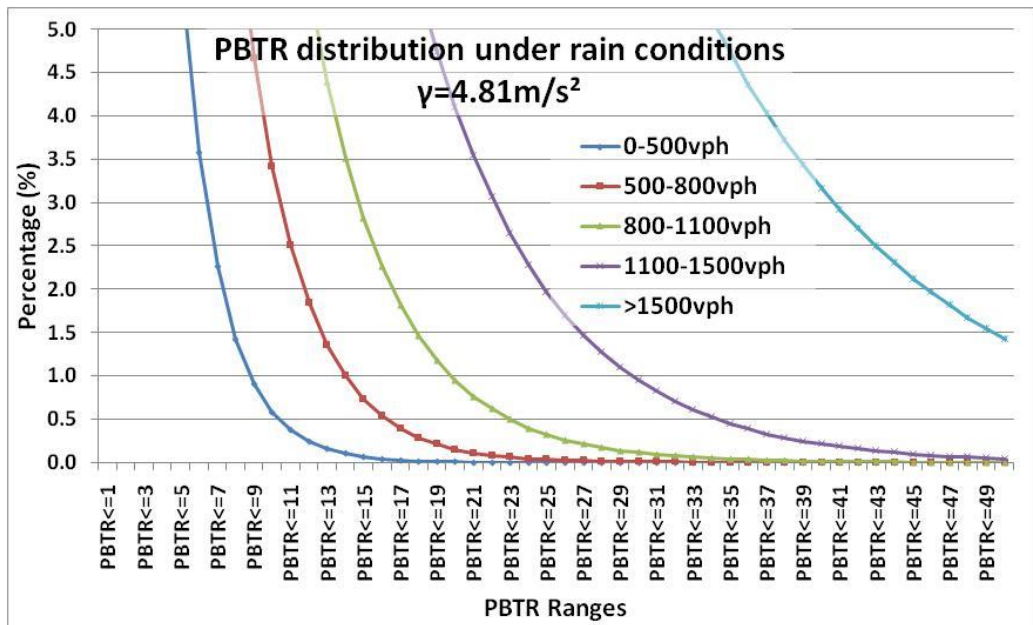


Fig 4.9 PBTR distribution under rain conditions on the fast lane

PBTR is also more sensitive to flow ranges with the clearer separation of curves in Fig 4.6, Fig 4.7, Fig 4.8, and Fig 4.9. Under higher flow ranges, vehicles move closely while maintaining high speed which causes high percentage of vehicles having high PBTR values.

TTC distribution

On the contrary to IBTR and PBTR, TTC indicates higher crash risk when its value decreases. Therefore, low TTC values are more interesting in this study than its high values. Fig 4.10 and Fig 4.11 show the TTC distributions on the slow lane and on the fast lane respectively.

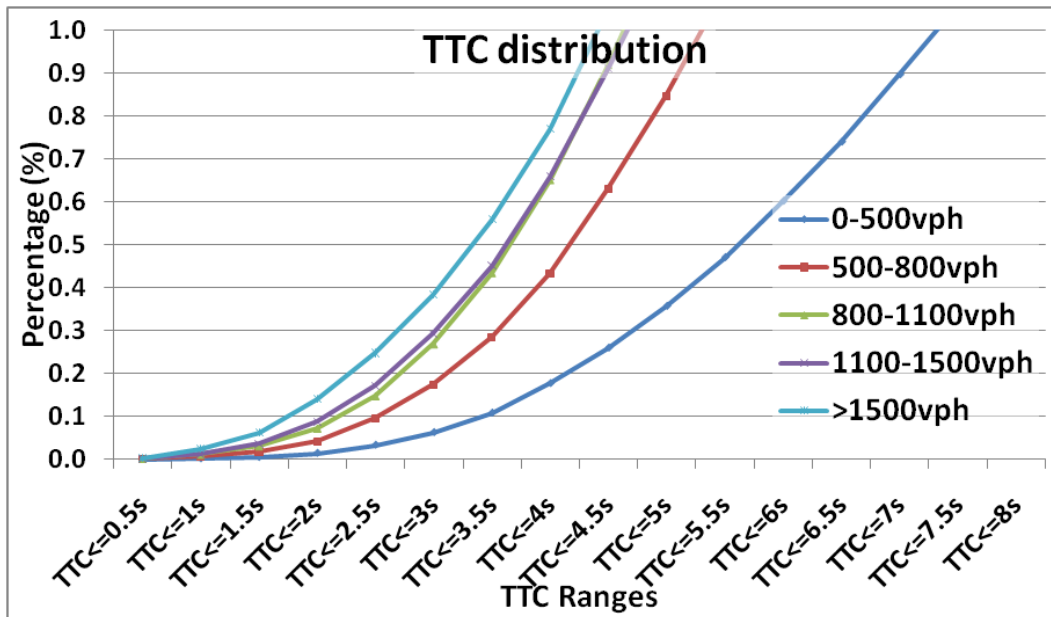


Fig 4.10 TTC distribution on the slow lane

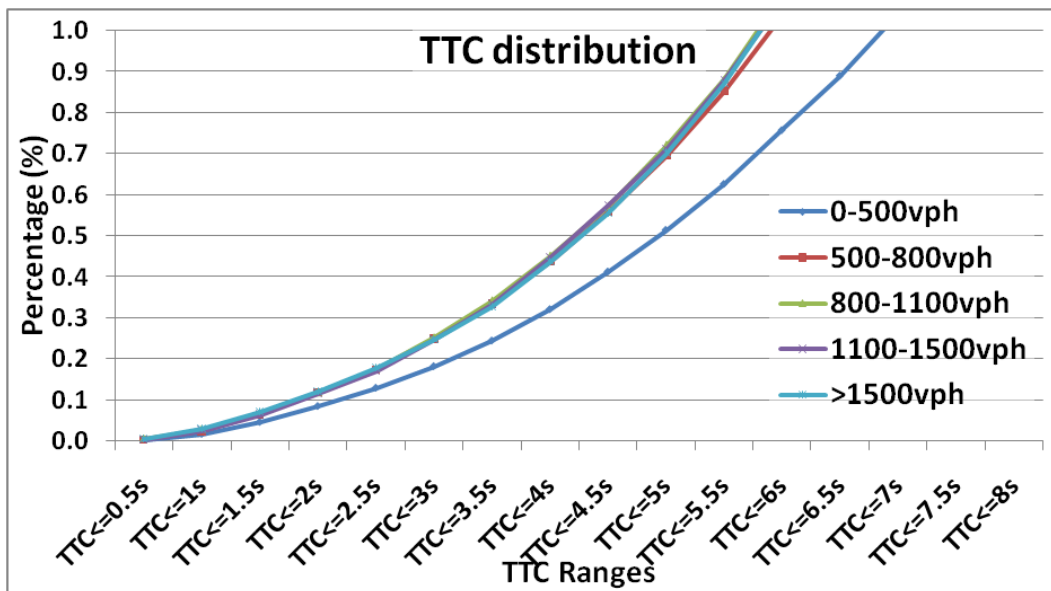


Fig 4.11 TTC distribution on the fast lane

In Fig 4.10, TTC distribution under free flow is very different from other flow ranges while for two flow ranges, 800-1100vph and 1100-1500vph, TTC distributions are almost overlapped. This means that TTC distribution is relatively stable under a wider flow range.

In Fig 4.11, TTC distribution under free flow is separated from TTC distributions under other traffic flows ranges. Apart from the free flow range, there are always 1% of vehicles having TTC smaller than 5.5 seconds.

It can be seen from Fig 4.10 and Fig 4.11 that TTC is less sensitive to traffic flow than IBTR or PBTR.

SOSL distribution

Motorway speed limits are speed levels that are recognized by law and any violation to the speed limits is considered as risky. Fig 4.12 and Fig 4.13 shows the distributions of SOSL on the slow lane and on the fast lane.

For the same flow range, the percentages of vehicles running faster than speed limits on the slow lane are lower than on the fast lane. This is normal because vehicles usually come to the fast lane to overtake and then come back to the slow lane.

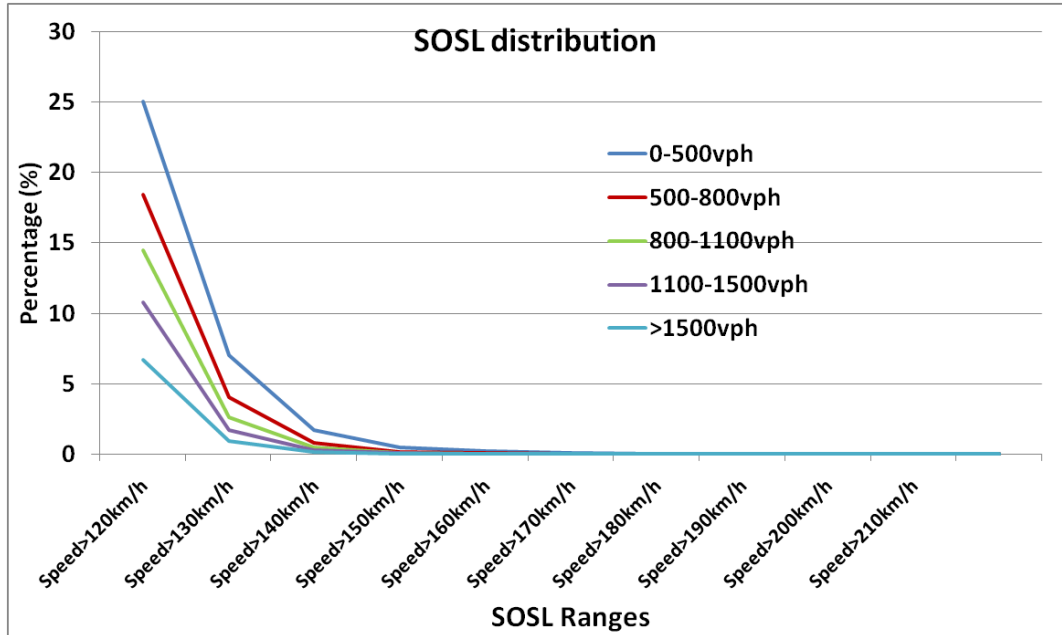


Fig 4.12 SOSL distribution on the slow lane

SOSL is also sensitive to flow conditions: under higher flows, it is difficult for vehicles to accelerate because the traffic is dense. On the contrary, the percentage of vehicles with speed greater than the speed limit is the highest under the free flow when drivers can find front space to accelerate their vehicles.

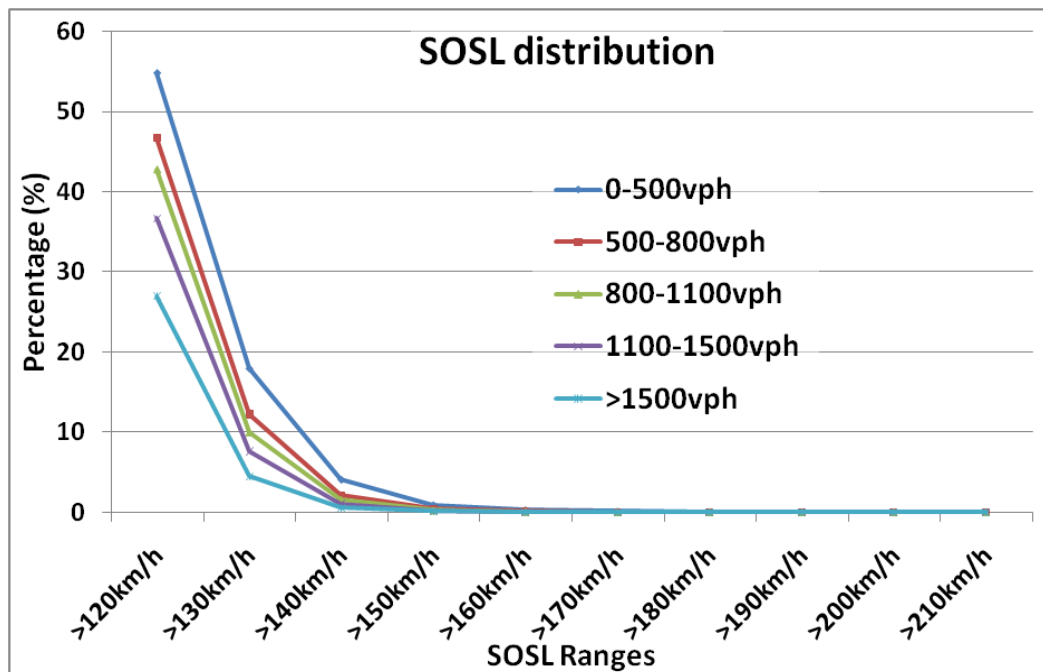


Fig 4.13 SOSL distribution on the fast lane

PICUD distribution

PICUD represents the remaining distance between two vehicles when the front vehicle brakes. If that distance is smaller, the situation becomes more risky. Therefore, low PICUD

values are more interesting to this study.

PICUD distributions are shown in Fig 4.14, Fig 4.15, Fig 4.16, and Fig 4.17. In these figures, the common trend is that PICUD distribution curves under different traffic flow ranges are clearly separated. This means that traffic flow has a sharp influence to PICUD distribution. When the traffic flow increases, the distant gap between vehicles is reduced and the vehicle speed becomes more homogenous which causes smaller PICUD values. That is why there are more vehicles having low PICUD under higher flow ranges.

Under the same meteorological conditions, more vehicles on the fast lane are indicated as risky by PICUD. For example, under fine meteorological conditions and under high flow range (greater than 1500vph), there are 3.5% out of vehicles having PICUD smaller than 3m on the slow lane while this percentage is 5% on the fast lane.

About the influence of meteorological conditions, PICUD tends to be lower under fine meteorological conditions than under rain conditions. For example, considering the slow lane, in Fig 4.14, there are 5% out of vehicles under high flow range (greater than 1500vph) and under fine meteorological conditions having PICUD smaller than 4m while this percentage is about 4.3% under the same flow range and under rain conditions (see Fig 4.16). This result is unexpected because rain conditions are supposed to have bad effect on traffic safety. For this reason, PICUD will not be considered in the later stage of this study.

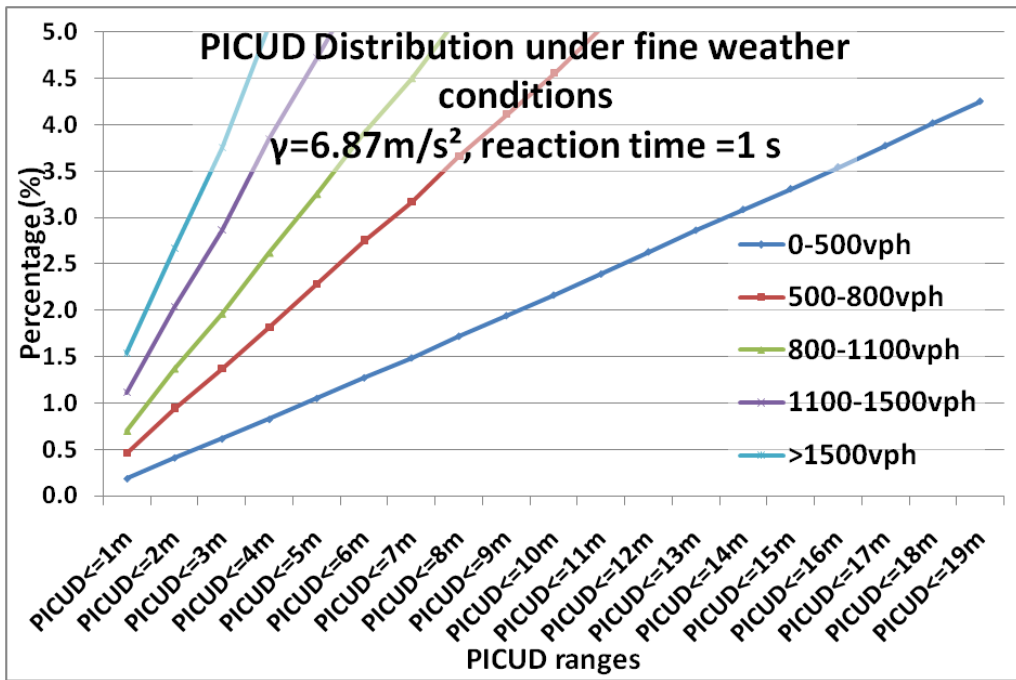


Fig 4.14 PICUD distribution under fine meteorological conditions on the slow lane

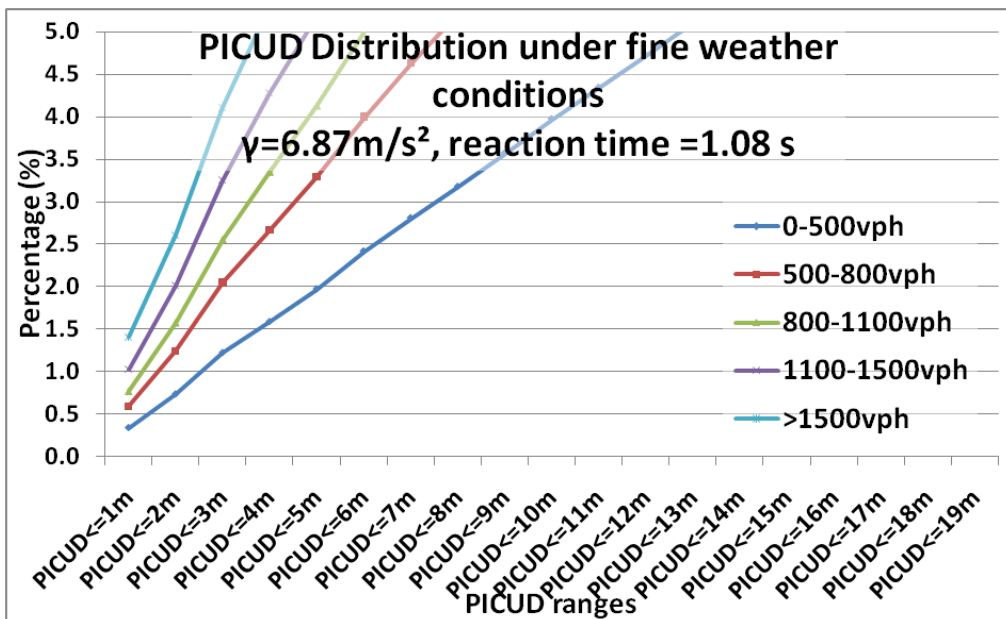


Fig 4.15 PICUD distribution under fine meteorological conditions on the fast lane

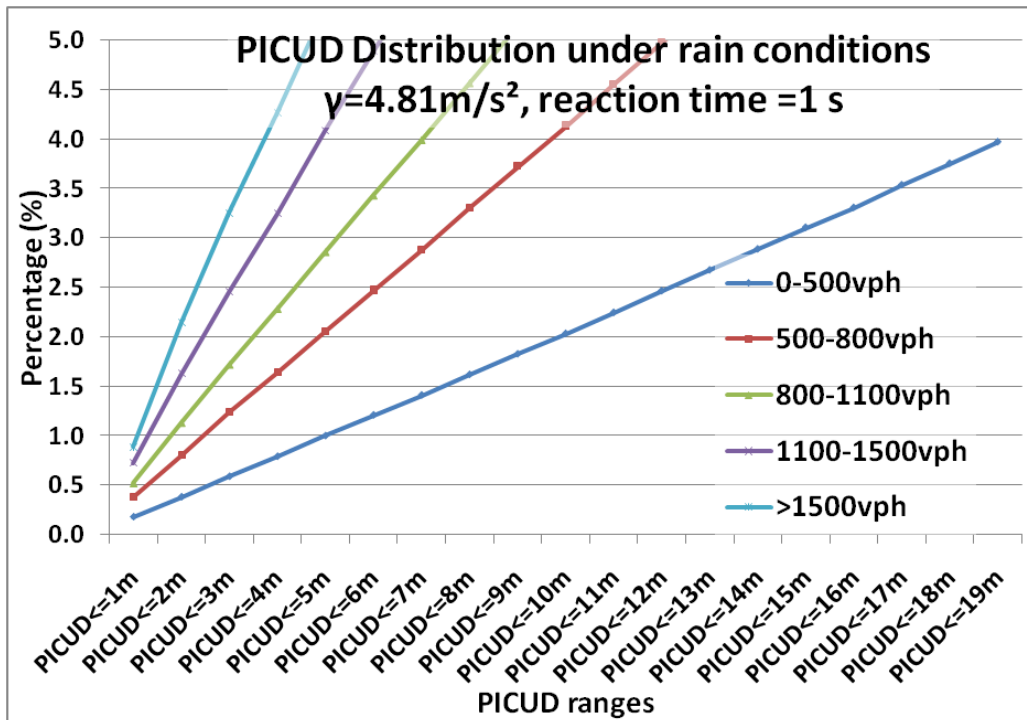


Fig 4.16 PICUD distribution under rain conditions on the slow lane

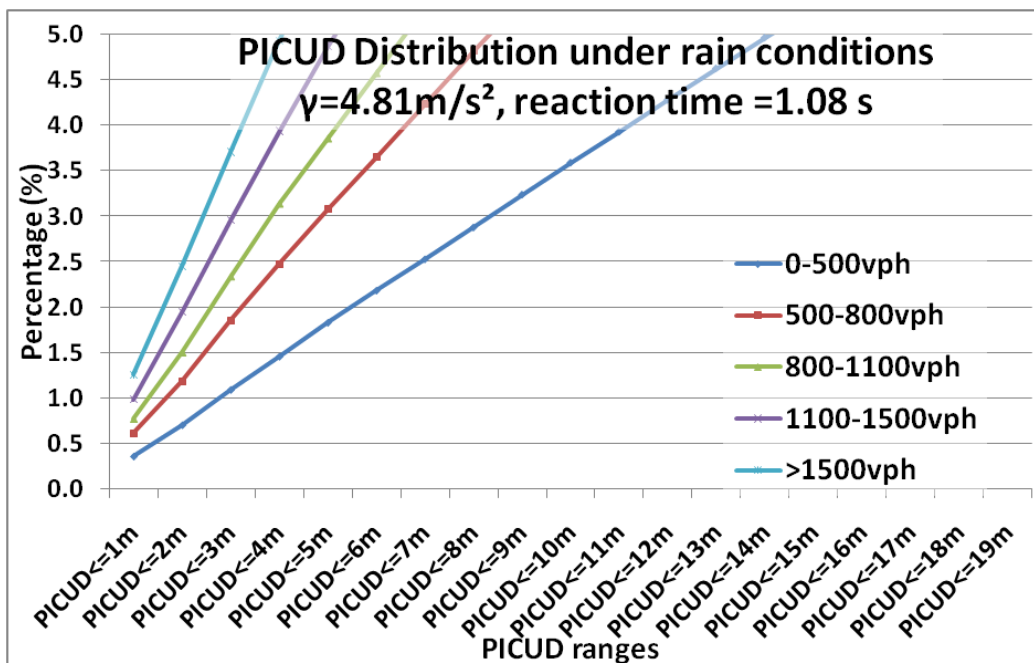


Fig 4.17 PICUD distribution under rain conditions on the fast lane

4.3.2 Crash cases

Crash cases are important in testing the performance of risk indicators because the traffic situations just before occurrences of crashes contain crash risk which ended in crashes. These traffic situations are the only known risky situations so far.

In this sub-section, the details about the evolution of risk indicators before several crashes are described and the results with all processed crashes are summarized.

Motorway crash data at the study site (see section III.2.), contain information about crash positions, date and time of crashes, weather conditions, pavement conditions, lighting conditions, types of crashes, and drivers' status during the crashes. Data in the files are coded according to OFS code (Office Fédéral de la Statistique). The time of a crash is stored under the form of a one-hour period where the crash happened: e.g., if a crash happened between 02:00PM and 03:00PM, then the time of the crash recorded is 14:00-15:00. During 4 years (2002-2005), there are totally 3693 crashes on motorway sections in Vaud canton, among which 470 (12.7% of all crashes) are DUI crashes.

All DUI-crashes are not interested in this study and they are excluded. The non-DUI crash database contains 3223 crashes. The study focuses only on study site including two motorways N1 (the section from Lausanne to Bern – N1LB), and N9 (the section from Lausanne to Sion, N9LS) where traffic data are available from the two ATCs 116 and 149. Eliminating crashes on other road sections, the number of crashes is 1940.

To test the evolution of traffic before and during a crash, the traffic patterns leading to the crash need to be analyzed. However, the constraint of distance between a crash and the location of the nearest ATC causes the strong reduction of number of crashes that can be used in this study. The locations of the ATCs are fixed points. The crashes are selected if they are inside a buffer of 500m around an ATC. Fig 4.18 shows the number of Non-DUI crashes for each ATC.

Fig 4.18 Crash distribution by ATCs' locations

ATC	Upstream crashes	Downstream crashes
116	15	14
149	24	18

Many tests were done to validate different risk indicators and are reported in different publications. In [Pham, M.-H. et al., 2007], TTC and PBTR are tested with a crash described in Fig 4.19. A method was presented to find the exact time of the crash and to analyze the evolution of TTC and PBTR.

Fig 4.19 Crash information

Date & Time	Sunday, 2005-01-23, between 20:00-21:00
Position	A9, direction Sion-Lausanne: 15km500. This crash happened at upstream from the ATC 116.
DUI-crash?	No
Type of accident	Single vehicle accident
Weather	Fine
Pavement	Dry
Lighting condition	dark

Although this is a single vehicle crash, it was caused by the involved vehicle while trying to avoid a collision with another vehicle.

In Fig 4.20, Fig 4.21, Fig 4.22, and Fig 4.23 the individual values of TTC and PBTR are presented together with vehicle speeds and traffic flow at each 5-minute interval. The individual speed and the flow help identifying the exact time of the crash and they are repeated in every figure to become the background for the evolution of TTC or PBTR.

In Fig 4.20, there is a speed drop at about 20:55. At that moment the traffic flow also drop down. 5-minute intervals before 21:00 were considered and TTC and PBTR distributions

during these intervals are tested.

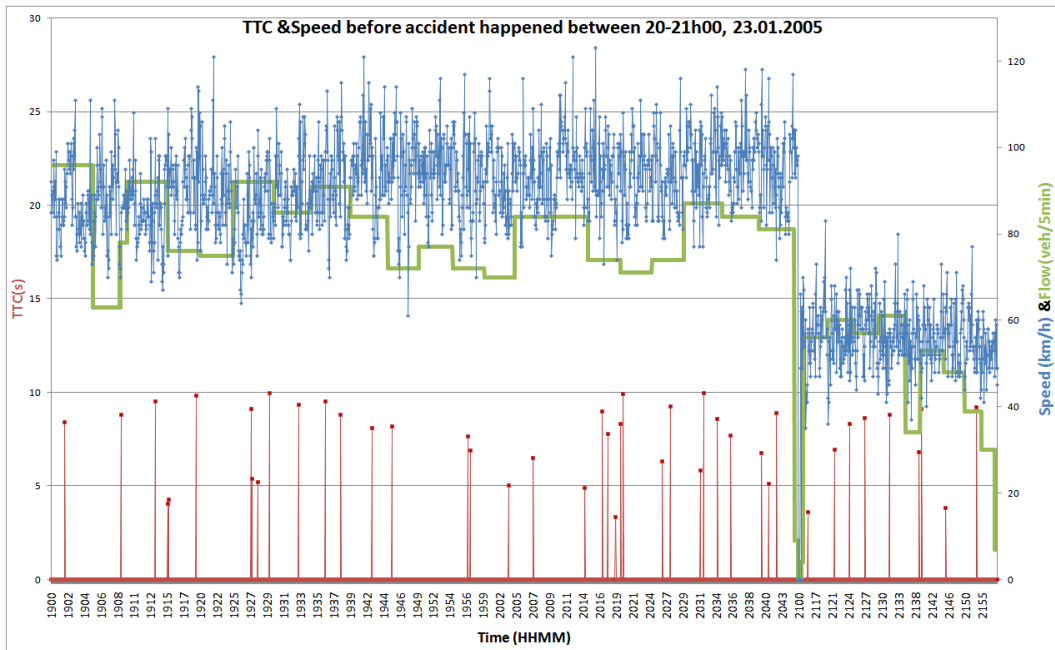


Fig 4.20 Flow and Individual TTC, speed on the normal lane when the crash happened. TTC values greater than 10s are made equal to zero.

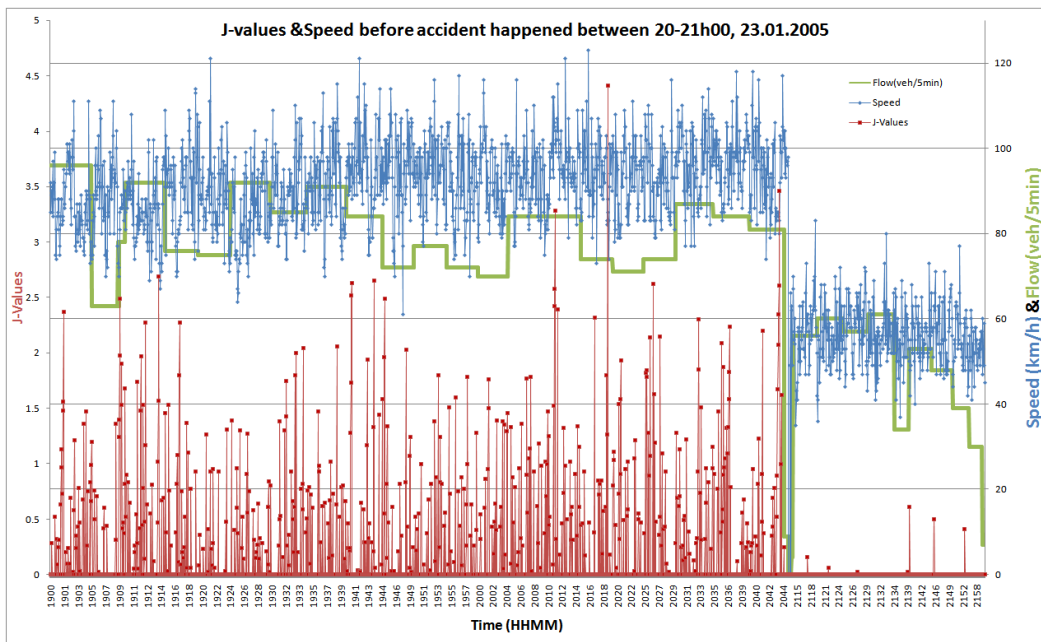


Fig 4.21 PBTR evolution on the slow lane

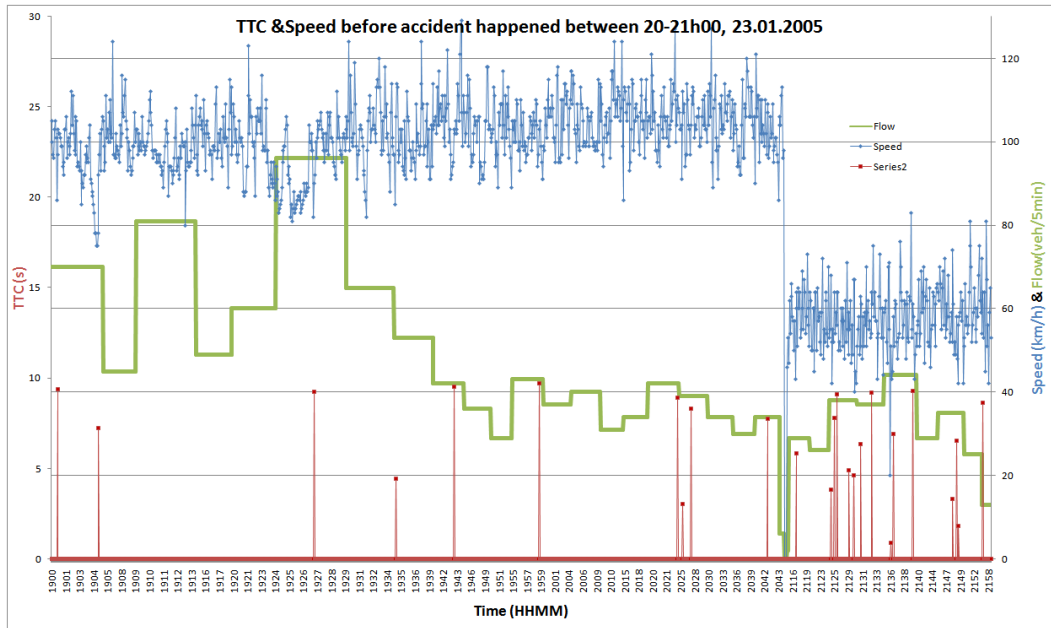


Fig 4.22 TTC evolution on the fast lane

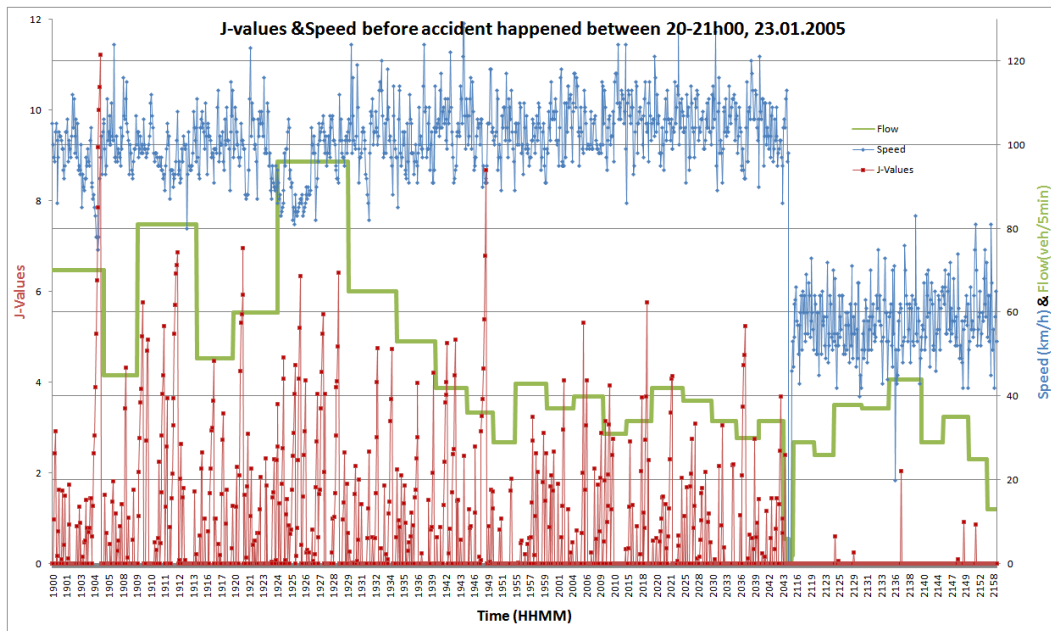


Fig 4.23 PBTR evolution on the fast lane

In Fig 4.24 and Fig 4.26, TTC distributions on the slow lane and on the fast lane are presented. The two right most columns contain information about the index of 5-minute interval and the flow during that interval. One day contains 24 hours which are 1440 minutes, i.e. 288 5-minute intervals. If the period from 00:00 to 00:05 is the first 5-minute interval of a day with an index of 0, the 5-minute interval from 20:55-21:00 has the index of 251. The other columns on the left represent the percentages of vehicles during each 5-minute interval having TTC values represented by the columns.

The traffic situation just before the crash is the 5-minute interval having index of 248. On the slow lane, the lowest TTC value during this interval is from 5.0 to 5.5seconds which is not as low as the lowest TTC value during the interval 243. Besides, by comparing the percentage of vehicles having TTC smaller than 7seconds during these two 5-minute intervals, the percentage is higher during the interval 248. This suggests that risk indicators may have better performance if not only their bound values are considered but also the

percentages of vehicles that can approach those bound values.

The same conclusion can be taken from PBTR distributions in Fig 4.25 and Fig 4.27 where the red color means the risk is high.

Fig 4.24 TTC distribution on the slow lane

0<TTC<2.5	0<TTC<3.0	0<TTC<3.5	0<TTC<4.0	0<TTC<4.5	0<TTC<5.0	0<TTC<5.5	0<TTC<6.0	0<TTC<6.5	0<TTC<7.0	0<TTC<7.5	0<TTC<8.0	0<TTC<8.5	0<TTC<9.0	0<TTC<9.5	0<TTC	Flow(veh/h)	Time (5Min)
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.03	1.03	1.03	45.36	1164	228
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.28	1.28	48.72	936	229
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.09	47.83	1104	230
0.00	0.00	0.00	0.00	2.63	2.63	2.63	2.63	2.63	2.63	2.63	2.63	2.63	2.63	2.63	46.05	912	231
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	50.67	900	232
0.00	0.00	0.00	0.00	0.00	0.00	2.17	2.17	2.17	2.17	2.17	2.17	2.17	2.17	3.26	48.91	1104	233
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.18	52.94	1020	234
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.10	2.20	47.25	1092	235
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.19	1.19	1.19	44.05	1008	236
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.39	1.39	1.39	44.44	864	237
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	49.35	924	238
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.39	1.39	2.78	2.78	2.78	2.78	51.39	864	239
0.00	0.00	0.00	0.00	0.00	0.00	1.43	1.43	1.43	1.43	1.43	1.43	1.43	1.43	1.43	55.71	840	240
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.19	1.19	1.19	1.19	1.19	1.19	1.19	50.00	1008	241
0.00	0.00	0.00	0.00	0.00	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	39.29	1008	242
0.00	0.00	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	2.70	2.70	4.05	4.05	45.95	888	243
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.41	1.41	1.41	47.89	852	244
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.35	1.35	1.35	1.35	1.35	1.35	2.70	40.54	888	245
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.15	1.15	1.15	1.15	1.15	1.15	2.30	42.53	1044	246
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.19	1.19	1.19	1.19	40.48	1008	247
0.00	0.00	0.00	0.00	0.00	0.00	1.23	1.23	1.23	2.47	2.47	2.47	2.47	2.47	3.70	50.62	972	248
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	44.44	108	249
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	250
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	251
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	252
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	50.00	24	253
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	50.00	48	254
0.00	0.00	0.00	1.79	1.79	1.79	1.79	1.79	1.79	1.79	1.79	1.79	1.79	1.79	1.79	53.57	672	255
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.67	1.67	1.67	3.33	3.33	3.33	51.67	720	256
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.75	1.75	45.61	684	257
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.64	1.64	50.82	732	258
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.94	2.94	2.94	2.94	2.94	2.94	47.06	408	259
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.89	50.94	636	260
0.00	0.00	0.00	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	39.58	576	261
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.56	48.72	468	262
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	50.00	360	263
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.17	2.17	2.17	52.17	552	264
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	46.00	600	265
1.69	1.69	1.69	1.69	1.69	3.39	3.39	3.39	3.39	3.39	3.39	3.39	5.08	5.08	6.78	44.07	708	266

Fig 4.25 PBTR distribution on the slow lane

Time(5min)	Flow(veh/h)	J>=0	J>0	J>1	J>2	J>3	J>4
228	1164	100.00	36.08	7.22	1.03	0.00	0.00
229	936	100.00	26.92	11.54	1.28	0.00	0.00
230	1104	100.00	33.70	9.78	2.17	0.00	0.00
231	912	100.00	31.58	9.21	1.32	0.00	0.00
232	900	100.00	21.33	2.67	0.00	0.00	0.00
233	1104	100.00	29.35	3.26	0.00	0.00	0.00
234	1020	100.00	30.59	8.24	1.18	0.00	0.00
235	1092	100.00	27.47	3.30	1.10	0.00	0.00
236	1008	100.00	33.33	11.90	3.57	0.00	0.00
237	864	100.00	29.17	8.33	2.78	0.00	0.00
238	924	100.00	27.27	3.90	0.00	0.00	0.00
239	864	100.00	27.78	5.56	0.00	0.00	0.00
240	840	100.00	30.00	10.00	0.00	0.00	0.00
241	1008	100.00	29.76	8.33	0.00	0.00	0.00
242	1008	100.00	36.90	14.29	4.76	1.19	0.00
243	888	100.00	22.97	5.41	2.70	1.35	1.35
244	852	100.00	30.99	11.27	0.00	0.00	0.00
245	888	100.00	29.73	14.86	4.05	0.00	0.00
246	1044	100.00	36.78	8.05	1.15	0.00	0.00
247	1008	100.00	30.95	11.90	2.38	0.00	0.00
248	972	100.00	34.57	9.88	6.17	1.23	0.00
249	108	100.00	0.00	0.00	0.00	0.00	0.00
250	-	-	-	-	-	-	-
251	-	-	-	-	-	-	-
252	-	-	-	-	-	-	-
253	24	100.00	0.00	0.00	0.00	0.00	0.00
254	48	100.00	0.00	0.00	0.00	0.00	0.00
255	672	100.00	1.79	0.00	0.00	0.00	0.00
256	720	100.00	1.67	0.00	0.00	0.00	0.00
257	684	100.00	1.75	0.00	0.00	0.00	0.00
258	732	100.00	0.00	0.00	0.00	0.00	0.00
259	408	100.00	2.94	0.00	0.00	0.00	0.00
260	636	100.00	1.89	0.00	0.00	0.00	0.00
261	576	100.00	2.08	0.00	0.00	0.00	0.00
262	468	100.00	2.56	0.00	0.00	0.00	0.00
263	360	100.00	0.00	0.00	0.00	0.00	0.00
264	552	100.00	0.00	0.00	0.00	0.00	0.00
265	600	100.00	0.00	0.00	0.00	0.00	0.00
266	708	100.00	1.69	0.00	0.00	0.00	0.00

Fig 4.26 TTC distribution on the fast lane.

0<TTC<1.0	0<TTC<1.5	0<TTC<2.0	0<TTC<2.5	0<TTC<3.0	0<TTC<3.5	0<TTC<4.0	0<TTC<4.5	0<TTC<5.0	0<TTC<5.5	0<TTC<6.0	0<TTC<6.5	0<TTC<7.0	0<TTC<7.5	0<TTC<8.0	0<TTC<8.5	0<TTC<9.0	0<TTC<9.5	0<TTC	Flow(veh/h)	Time (5Min)	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.41	1.41	1.41	1.41	2.82	42.25	852	228	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	40.00	540	229	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	48.15	972	230	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	40.82	588	231	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	36.67	720	232	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.04	50.00	1152	233	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	43.08	780	234	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	43.40	636	235	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.38	52.38	504	236	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	50.00	432	237	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	37.93	348	238	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	41.86	516	239	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	48.65	444	240	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	37.50	480	241	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	51.61	372	242	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	47.06	408	243	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.38	2.38	42.86	504	244	
0.00	0.00	0.00	0.00	0.00	2.56	2.56	2.56	2.56	2.56	2.56	2.56	2.56	2.56	2.56	5.13	5.13	5.13	38.46	468	245	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	50.00	408	246	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	36.67	360	247	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.94	2.94	2.94	2.94	47.06	408	248	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	66.67	72	249	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	250	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	251
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	252
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	253
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	50.00	24	254	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.45	3.45	3.45	3.45	3.45	3.45	3.45	3.45	44.83	348	255	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	42.31	312	256	
0.00	0.00	0.00	0.00	0.00	0.00	2.63	2.63	7.89	7.89	7.89	7.89	7.89	7.89	10.53	10.53	10.53	13.16	42.11	456	257	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.70	2.70	2.70	2.70	2.70	2.70	5.41	54.05	444	258	
2.27	2.27	2.27	2.27	2.27	2.27	2.27	2.27	2.27	2.27	2.27	2.27	2.27	4.55	4.55	4.55	4.55	6.82	50.00	528	259	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	51.72	348	260	
0.00	0.00	2.86	2.86	2.86	5.71	5.71	5.71	5.71	5.71	5.71	5.71	5.71	8.57	8.57	8.57	8.57	8.57	48.57	420	261	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	64.00	300	262	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.69	46.15	156	263	
0.00	0.00	3.45	3.45	6.90	6.90	6.90	6.90	6.90	10.34	13.79	17.24	17.24	17.24	17.24	17.24	17.24	17.24	51.72	348	264	
0.00	0.00	2.63	2.63	2.63	2.63	2.63	2.63	2.63	2.63	2.63	2.63	2.63	5.26	5.26	5.26	5.26	5.26	42.11	456	265	
0.00	2.13	2.13	2.13	2.13	2.13	2.13	2.13	2.13	2.13	2.13	2.13	2.13	4.26	4.26	4.26	4.26	4.26	40.43	564	266	

Fig 4.27 PBTR distribution on the fast lane.

Time(5min)	Flow(veh/h)	J>=0	J>0	J>1	J>2	J>3	J>4	J>5	J>6	J>7	J>8	J>9	J>10	J>11
228	852	100.00	52.11	29.58	16.90	11.27	9.86	9.86	8.45	7.04	5.63	5.63	2.82	1.41
229	540	100.00	35.56	20.00	4.44	4.44	2.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00
230	972	100.00	50.62	40.74	29.63	20.99	13.58	9.88	3.70	0.00	0.00	0.00	0.00	0.00
231	588	100.00	44.90	26.53	18.37	6.12	2.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
232	720	100.00	48.33	36.67	20.00	8.33	8.33	6.67	1.67	0.00	0.00	0.00	0.00	0.00
233	1152	100.00	62.50	43.75	30.21	19.79	13.54	5.21	2.08	0.00	0.00	0.00	0.00	0.00
234	780	100.00	41.54	24.62	12.31	7.69	3.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00
235	636	100.00	43.40	24.53	11.32	3.77	1.89	0.00	0.00	0.00	0.00	0.00	0.00	0.00
236	504	100.00	45.24	33.33	23.81	14.29	7.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00
237	432	100.00	52.78	36.11	22.22	16.67	11.11	8.33	5.56	2.78	2.78	0.00	0.00	0.00
238	348	100.00	13.79	6.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
239	516	100.00	41.86	23.26	13.95	2.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
240	444	100.00	35.14	24.32	8.11	2.70	2.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00
241	480	100.00	35.00	22.50	12.50	7.50	5.00	2.50	0.00	0.00	0.00	0.00	0.00	0.00
242	372	100.00	38.71	29.03	16.13	9.68	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
243	408	100.00	41.18	29.41	20.59	8.82	2.94	2.94	0.00	0.00	0.00	0.00	0.00	0.00
244	504	100.00	45.24	21.43	9.52	7.14	4.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00
245	468	100.00	46.15	28.21	7.69	2.56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
246	408	100.00	20.59	11.76	11.76	2.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
247	360	100.00	46.67	26.67	16.67	13.33	10.00	3.33	0.00	0.00	0.00	0.00	0.00	0.00
248	408	100.00	38.24	17.65	5.88	2.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
249	72	100.00	16.67	16.67	16.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
250	-	-	-	-	-	-	-	-	-	-	-	-	-	-
251	-	-	-	-	-	-	-	-	-	-	-	-	-	-
252	-	-	-	-	-	-	-	-	-	-	-	-	-	-
253	-	-	-	-	-	-	-	-	-	-	-	-	-	-
254	24	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
255	348	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
256	312	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
257	456	100.00	7.89	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
258	444	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
259	528	100.00	2.27	2.27	2.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
260	348	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
261	420	100.00	5.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
262	300	100.00	4.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
263	156	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
264	348	100.00	3.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
265	456	100.00	2.63	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
266	564	100.00	6.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

In (Pham et al., 2008), four crashes are used for testing six indicators. The four crashes are near traffic detector 149 and described in Fig 4.28. The risk criteria are defined based on the desired alarm rates and then applied to traffic conditions before the four crashes.

A guideline about how to define the risk criteria is also presented in (Pham, Mouzon, Chung and Faouzi, 2008). This guideline can be applied in both cases where the crash database is available or not. If the crash database is available and crash time can be observable in traffic data, pre-crash conditions are used to validate the risk criteria. If the crash database is not available or the number of usable crashes is limited, the desired alarm rate can be used to propose the risk criteria. In that case, the risk criteria are not validated.

Fig 4.28 Description of four crashes

ID + Site number (detector code)	Date (yyyymmdd)	Hour period (hhmm)	Light (<u>D</u> ay / <u>N</u> ight)	Weather (<u>D</u> ry / <u>R</u> ain)	Road surface (<u>D</u> ry / <u>W</u> et)	Distance from traffic detector (m)	Position (<u>U</u> p- / <u>D</u> own-stream) from detector	Direction (+ for lanes 1 and 2 / -for lanes 3 and 4)
1 149	20030308	1600-1700	D	D	D	696	D	+
2 149	20030310	0900-1000	D	D	D	244	D	-
3 149	20040116	1900-2000	N	R	W	359	U	+
4 149	20040922	1700-1800	D	D	D		U	+

4.4 Conclusions

In this section, the sensitivity of safety indicators is tested against the traffic flow, the meteorological conditions and the lanes of the road.

PBTR, IBTR, and PICUD with meteorological elements in their formula are sensitive to meteorological conditions. Most risk indicators except TTC are sensitive to flow ranges. The risk indicators are also sensitive to the lanes of the road. The sensitivity test shows that PICUD is not reliable because it indicates that the traffic under fine meteorological conditions is more unsafe than under rain conditions. Therefore, PICUD should not be used further in this study.

The best way to test the performance of risk indicators is to check the evolution of risk indicators during the traffic conditions just before the crash when it is sure that there is crash risk before the crash. With the crash cases tested, it is recommended that when considering if a traffic situation is risky or not, not only bound values of risk indicators but also the percentages of vehicles are essential.

5 Data Sampling & Traffic Regimes

5.1 Introduction

5.1.1 Motivation & Objective

Crashes are rare events and therefore pre-crash traffic conditions are rare in comparison to traffic conditions where no crash is recorded. Preliminary analyses presented in section III. 3 also indicate that rear-end and sideswipe crashes only occur during certain traffic conditions, especially during rush hours when the traffic density is high. Therefore, for better capturing the risky traffic conditions (which are similar to pre-crash traffic conditions), similar traffic conditions need to be extracted.

In the current study, the pre-crash and non-crash traffic conditions can be determined based on traffic and crash data. However, as non-crash conditions are abundant, it is a challenge to decide which non-crash conditions are relevant to the pre-crash conditions. Therefore the objective of this part is to present the methodology to sample non-crash data and to provide analyses on obtained results called traffic regimes.

5.1.2 Data Sampling Framework

Data sampling process includes the following steps:

- 1-Defining traffic situation in general that is characterized by variables.
- 2-Defining non-crash and pre-crash traffic situations
- 3-Clustering non-crash traffic situations into clusters
- 4-Classify pre-crash traffic situations into obtained clusters.

At the end of the process, clusters of traffic situations are obtained. Under each cluster, there is a list of pre-crash and non-crash traffic situations. Each cluster is a traffic regime as traffic situations under that cluster represent similar traffic condition.

5.1.3 Traffic Situation Characterization

A traffic situation – TS is defined as a piece of traffic information aggregated for an interval of 5 minutes. There are 5 types of variables that represent instantaneity, state of traffic on each lane, speed difference between lanes, traffic evolution between consecutive traffic situations and traffic conditions. Fig 5.1 presents the list of variables. The instantaneity is represented by time of the day and day of the week. Seven variables representing the traffic status on each lane are traffic flow, average speed, average headway, occupancy, standard deviation of headway, standard deviation of speed, and percentage of heavy vehicles. As there are two lanes per direction at the study site, the number of variables representing traffic status on each lane is 14. The speed difference between two lanes represents the third type of variables specifying the difference between two lanes. Traffic evolution is represented by four variables: flow and speed changes on each of two lanes. Finally, whether condition is represented by the discretized precipitation X22, i.e. when the precipitation is positive, X22 is set to 1; otherwise it is set to 0.

Fig 5.1 List of variables characterizing traffic situations

Variable	Alias	Meaning	Numeric/ Categorical	Unit	Value range	Specification
X1	<i>Day</i>	Time of the day	Categorical	-	0 - 277	Instantaneity
X2	<i>Day</i>	Day of the week	Categorical	-	1, 2, 3, 4, 5, 6, and 7	
X3	<i>LFlow</i>	Right lane's flow	Numeric	Vehicle per hour (vph)	0 - 4000	Right lane's status
X4	<i>LASpd</i>	Right lane's average speed	Numeric	Kilometers per hour (km/h)	0.0 - 200.0	
X5	<i>LAHw</i>	Right lane's average headway	Numeric	Second	0.0 - 100.0	
X6	<i>LOcc</i>	Right lane's occupancy	Numeric	Percentage (%)	0.0 - 100.0	
X7	<i>LVHw</i>	Right lane's standard deviation of headway	Numeric	Kilometers per hour (km/h)	0.0 - 100.0	
X8	<i>LVSpd</i>	Right lane's standard deviation of speed	Numeric	Second	0.0 - 50.0	
X9	<i>L%HV</i>	Right lane's percentage of heavy vehicles	Numeric	Percentage (%)	0.0 - 100.0	
X10	<i>HFlow</i>	Left lane's flow	Numeric	Vehicle per hour (vph)	0 - 4000	Left lane's status
X11	<i>HASpd</i>	Left lane's average speed	Numeric	Kilometers per hour (km/h)	0.0 - 200.0	
X12	<i>HAHw</i>	Left lane's average headway	Numeric	Second	0.0 - 100.0	
X13	<i>HOcc</i>	Left lane's occupancy	Numeric	Percentage (%)	0.0 - 100.0	
X14	<i>HVHw</i>	Left lane's standard deviation of headway	Numeric	Kilometers per hour (km/h)	0.0 - 100.0	
X15	<i>HVSpd</i>	Left lane's standard deviation of speed	Numeric	Second	0.0 - 50.0	
X16	<i>H%HV</i>	Left lane's percentage of heavy vehicles	Numeric	Percentage (%)	0.0 - 100.0	
X17	<i>Spd#</i>	Speed difference between two lanes	Numeric	Kilometers per hour (km/h)	-200.0 - 200.0	Discrepancy between lanes
X18	<i>LFCg</i>	Right lane's flow difference compared to previous TS	Numeric	Vehicle per hour (vph)	-4000 - 4000	Traffic evolution
X19	<i>LSCg</i>	Right lane's speed difference compared to previous TS	Numeric	Kilometers per hour (km/h)	-200.0 - 200.0	
X20	<i>HFCg</i>	Left lane's flow difference compared to previous TS	Numeric	Vehicle per hour (vph)	-4000 - 4000	
X21	<i>HSCg</i>	Left lane's speed difference compared to previous TS	Numeric	Kilometers per hour (km/h)	-200.0 - 200.0	
X22	<i>Prec</i>	Precipitation	Numeric	-	0 and 1	Weather conditions

5.1.4 Pre-crash and Non-crash Traffic Situations

Here, pre-crash and non-crash TS are defined based on crash and traffic data. Given a crash, the traffic evolution before, during and after the crash makes up the crash period as illustrated in Fig 5.2. The traffic conditions outside of crash periods of all crashes make up non-crash traffic situations - NTS.

Within crash periods, post-crash periods represent traffic evolution after crash occurrences. Traffic evolution before crashes can be divided into two periods: pre-crash buffer and pre-crash periods. Post-crash periods are out of interest in the current study as traffic nature is changed after crashes. Pre-crash buffer period is also unused as it can contain a mixture of non-crash and pre-crash traffic situations.

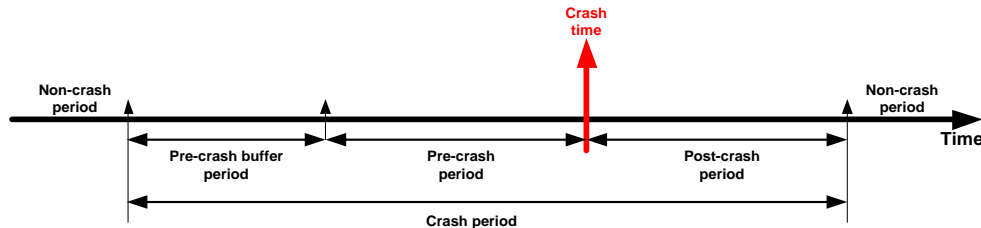


Fig 5.2 Two types of traffic situations

Pre-crash traffic situations – PTS are defined as traffic situations within pre-crash periods of all rear-end and sideswipe crashes.

5.1.5 Traffic Regimes

With the definition of NTS and PTS, the next step is to cluster NTS into groups and then classify PTS into obtained groups. This process is illustrated in Fig 5.3.

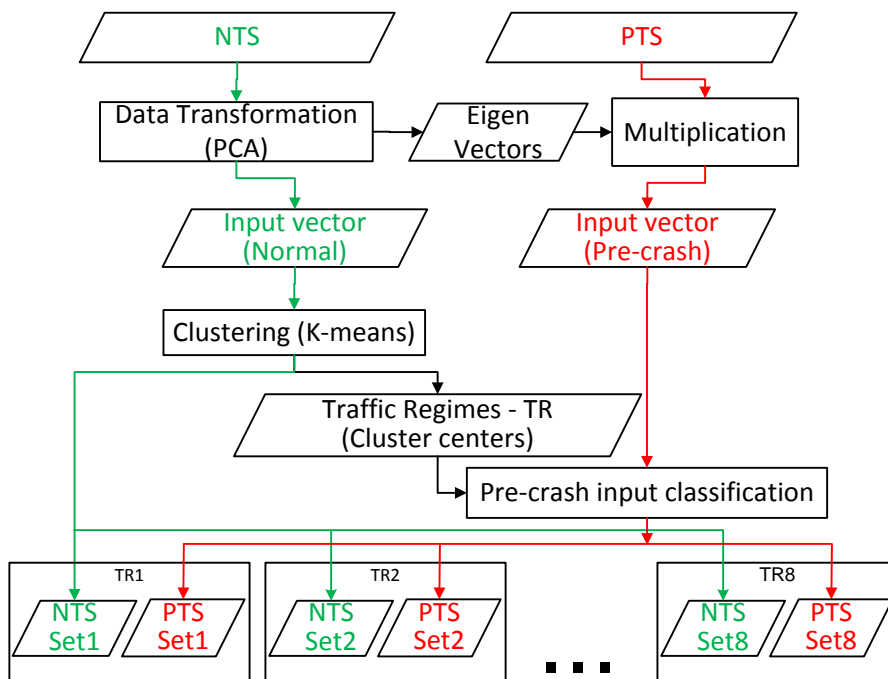


Fig 5.3 Clustering-Classification process

Before clustering, NTS matrix is pre-processed using PCA to reduce the number of dimension and each NTS become an input vector with number of dimensions reduced from 22 variables to 6 principal components. Then the input vectors are clustered using K-means to obtain clusters. Based on the clustering error, the number of clusters is set to 8, i.e. the

clustering error does not significantly reduce when the number of clusters increases from 8 to 9. Each cluster is called a Traffic Regime and named from A to H. With the obtained traffic regimes, PTS are pre-processed to reduce the number of dimensions to 6 and then classified into one of eight traffic regimes.

Fig 5.4 presents the results of the clustering-classification process. There are regimes (for instant regimes A and E) where there is no PTS observed. The objective of data sampling in the methodology presented in Section V.1.1 is to match PTS data with relevant NTS data. Therefore, the absence of PTS in several regimes indicates that NTS under those regimes should not be used to compare with any PTS.

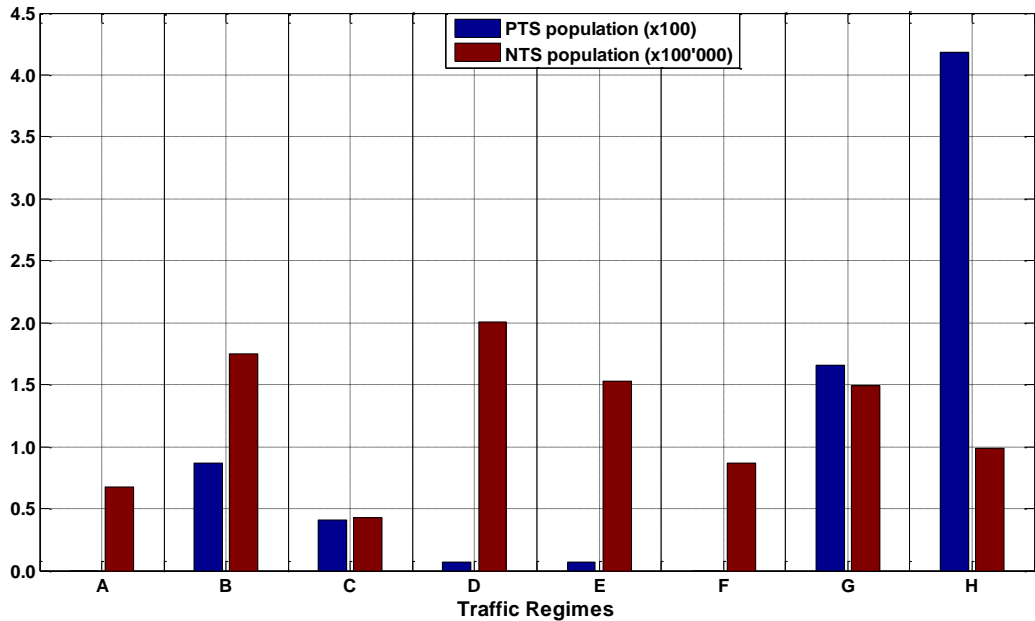


Fig 5.4 Population of NTS and PTS under each traffic regime

According to Fig 5.4, almost PTS occur during day time, especially during peak hours (6:00-8:00AM and 4:00-6:00PM, weekday) whereas; there is low number of PTS occurring during the night (from 8:00PM to 6:00AM) when the traffic demand is low. This distribution of PTS is reasonable as the traffic must be sufficiently dense for traffic – induced crashes (i.e. rear-end and sideswipe crashes in the present study) to occur.

In Fig 5.5, the locations of traffic regimes represented by cluster centers are characterized by several variables pointing in the same direction. For example, regime F is characterized by high values of X5, X7, X8, and X22 (i.e. lying in the fourth quarter of the first two PC space) and low values of X17 (speed difference becomes negative), X11 (left lane’s speed becomes zero), and X15 (speed variation becomes zero). It means that regime F represent traffic conditions where there are vehicles only on one lane due to low traffic or due to some other reasons.

Two regimes A and E lie on the first quarter of the first two PC space and are most underlined by high values of X4, X12, and X14 and low values of X3, X10, X6, and X13. It means that regimes A and E represent traffic conditions where there are vehicles on both lanes yet the flow is low. More principal components are needed to explain the difference between regimes A and E.

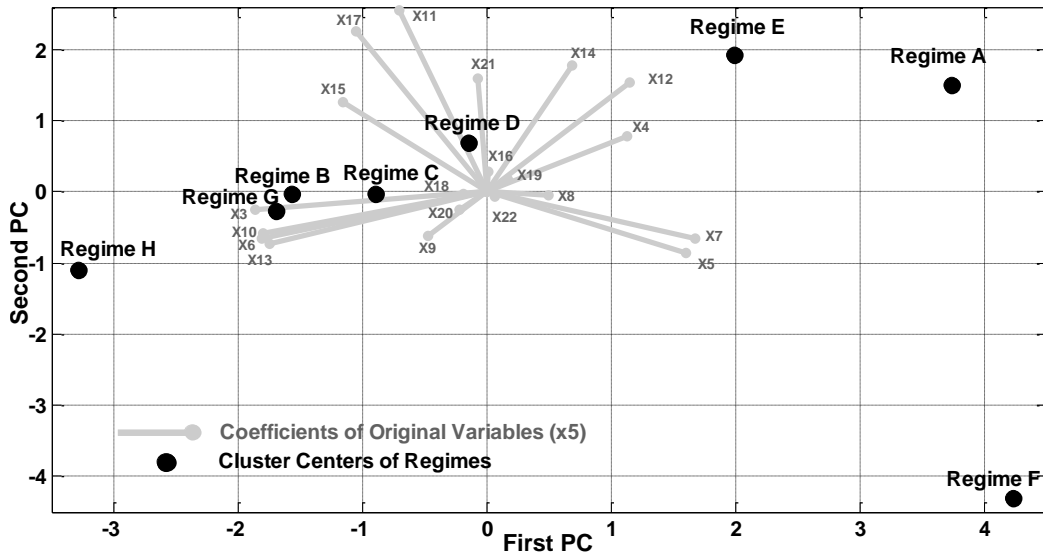


Fig 5.5 Location of cluster centers the first two PC space

Regime D is the only regime lying in the second quarter of the first two PC space and is strongly supported by X_{15} , X_{17} , X_{11} , and X_{21} . Regimes B, C, G, and H lie on the third quarter of the first two PC space and are indicated by high X_3 , X_6 , X_{10} , and X_{13} . B, C, G, and H are the traffic regimes where the population of PTS is high and the PTS/NTS rate is also high.

According to Fig 5.5, traffic data are divisible into three big groups:

- group of traffic active on one lane: regime F
- group of low traffic active on two lanes: regime A and E
- group of medium or high traffic on two lanes: regime B, C, D, G, and H

To explain the difference between each regime, more variables need to be used. In the subsequent sections, the detailed characterization of each regime is presented.

5.2 Traffic Regime Analyses

5.2.1 Overview

Fig 5.6 and Fig 5.7 present traffic regimes in the fundamental speed/flow diagram for the right and left lanes, respectively. It can be seen that data points from one regime overlap with data points from some other regimes. However, the overlapping is apparent as there are other variables used for separating different traffic regimes, not only speed and flow on two lanes.

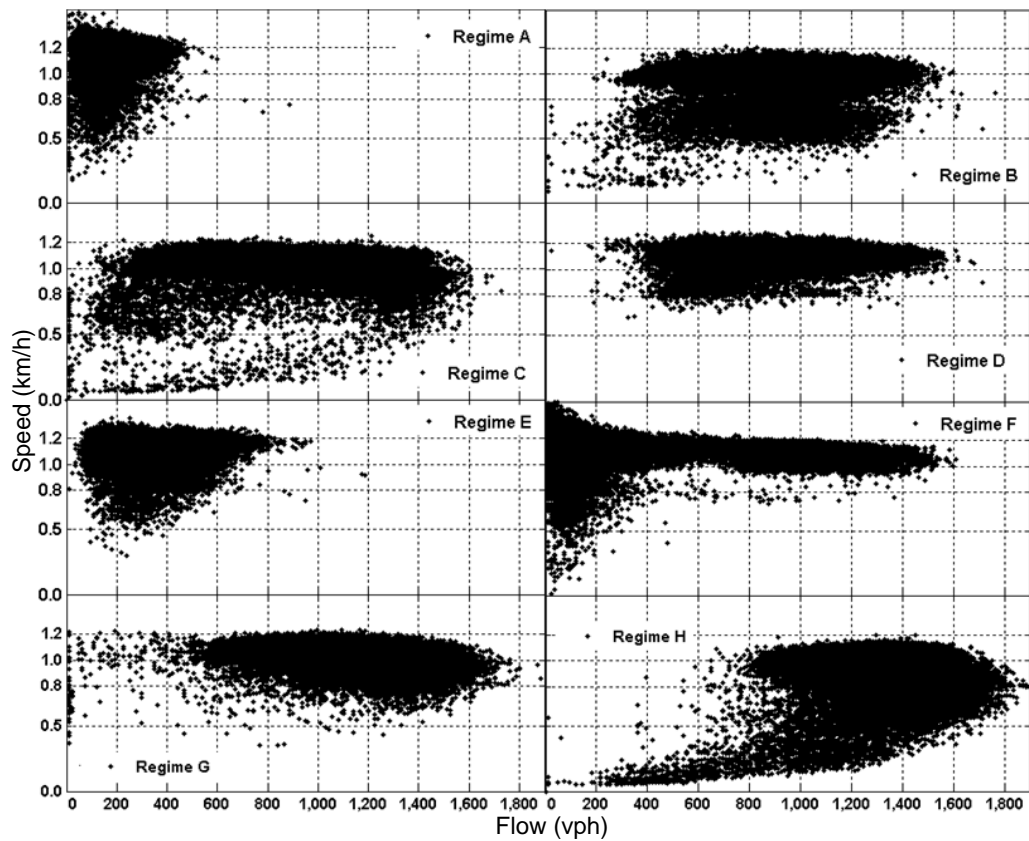


Fig 5.6 Traffic regimes under speed-flow diagram for the right lane

It can be easily recognized that the traffic under regime F is only active on one lane, mostly on the right lane and some Traffic Situations on the left lane. According to Fig 5.6, traffic flow on the right lane can be high up to 1400vph. This means that regime F represent traffic conditions where the left lane is closed due to some unknown reason.

Two regimes A and E also represent low flow conditions. The difference between regimes A and E is that the traffic flow is lower under regime A than under regime E, especially on the left lane.

For other regimes (i.e. regimes B, C, D, G, and H), speed and flow on two lanes cannot explain the differences.

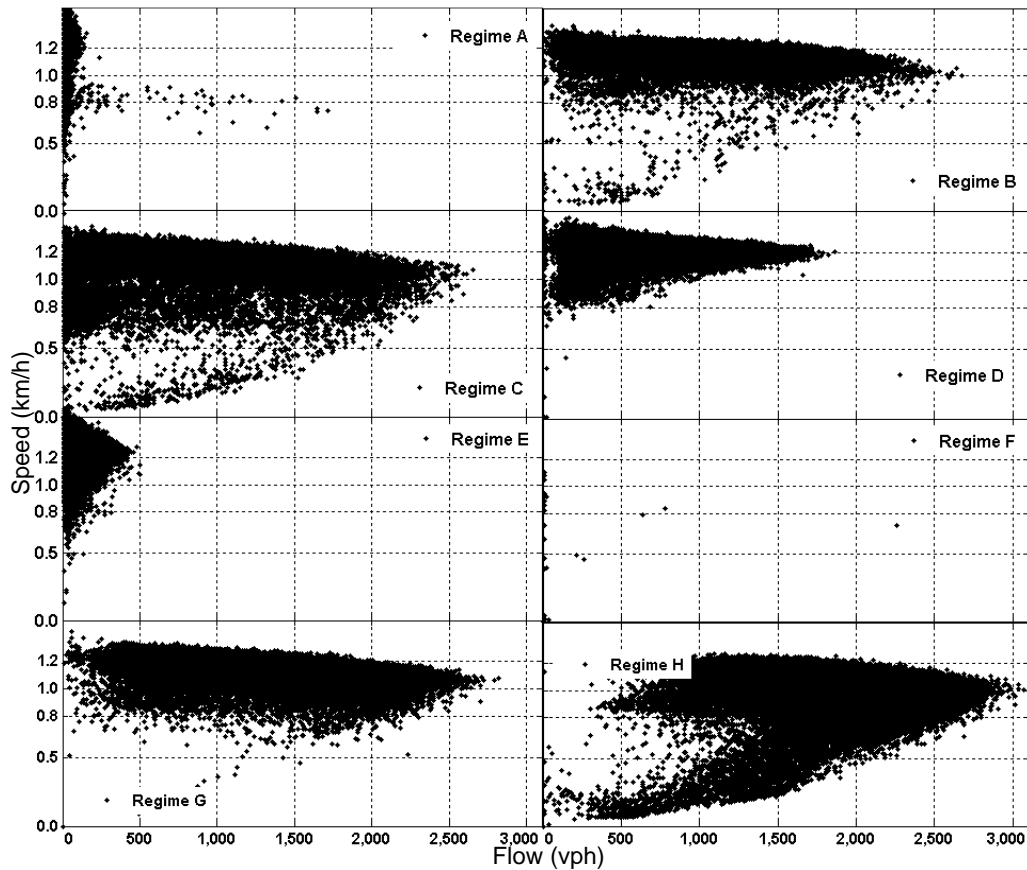


Fig 5.7 Traffic regimes under speed-flow diagram for the left lane

5.2.2 Traffic States on Two Lanes

Fig 5.8 presents statistics on variables indicating traffic states on both of the lanes, i.e. variables from X3 to X9 for the right lane and from X10 to X16 for the left lane. Using variable X3 (flow on the right lane), Traffic Situations can be partitioned into three groups:

- 1-The group of low flow ($X3 < 500\text{vph}$) including regimes A, E, and F
- 2-The group of medium traffic ($500\text{vph} \leq X3 \leq 1100\text{vph}$) including regimes B, C, and D.
- 3-The group of high traffic ($X3 > 1100\text{vph}$) including two regimes G and H.

Comparing variable X10 (flow on the left lane) between two regimes G and H, it can be stated that regime H represents high flow and congested conditions. Considering X4 and X11 (the average speeds on the right and left lanes), it is shown that the speed is lower under regime H than under any other regime. The high flow traffic under regime H is also reflected by the outstanding values of variables X6 and X13 (occupancy on the left and right lanes, respectively) compared to other regimes.

Regime G is the regime where traffic flows on both lanes (i.e. X3 and X10) are the second highest (flows under regime H are the highest). According to Fig 5.6 and Fig 5.7, there is no congestion under regime G. Therefore, the average speeds under regime G are higher than under regime H. Regime B is also very similar to regime G by comparing variables from X3 to X8, and X10 to X15. The clearest difference between regimes G and B is the percentage of heavy vehicles on both lanes – X9 and X16.

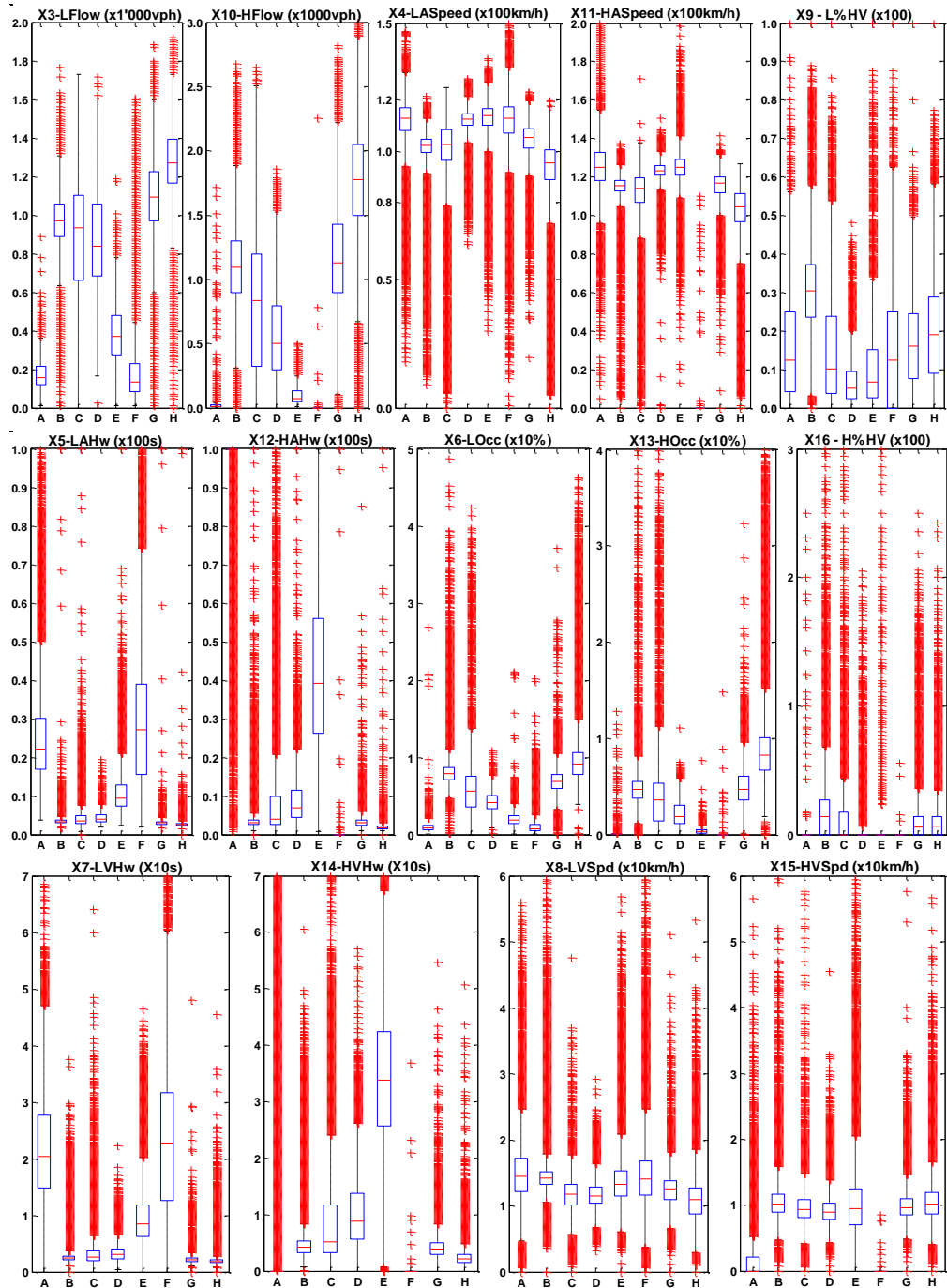


Fig 5.8 Statistics on variables from X3 to X16 representing lane states

Regimes C and D are also similar by comparing most of variables representing lane states. Some significant difference between regimes C and D are:

- The average speeds on both lanes (X4 and X11) are higher under regime D than under regime C.
- The occupancies on both lanes (X6 and X13) are lower under regime D than under regime C.
- The percentages of heavy vehicles on both (X9 and X16) are slightly lower under regime D than under regime C

Finally, using variables representing lane states allow differentiating the following traffic regimes or groups of traffic regimes:

- The free flow traffic conditions represented by regimes A, E, and F.

- The high flow and congested conditions represented by regime H.
- The high-medium traffic flow conditions represented by regimes B and G
- The low-medium traffic flow conditions represented by regimes C and D.

5.2.3 Traffic State Variations and Non-traffic Characteristics

Fig 5.9 presents variables from X17 to X21 indicating the variations of traffic state across lanes and over two continuous traffic situations. Fig 5.9 also presents non-traffic variables X1, X2, and X22. Based on those variables, the difference between regimes B and G and between regimes C and D can be clarified.

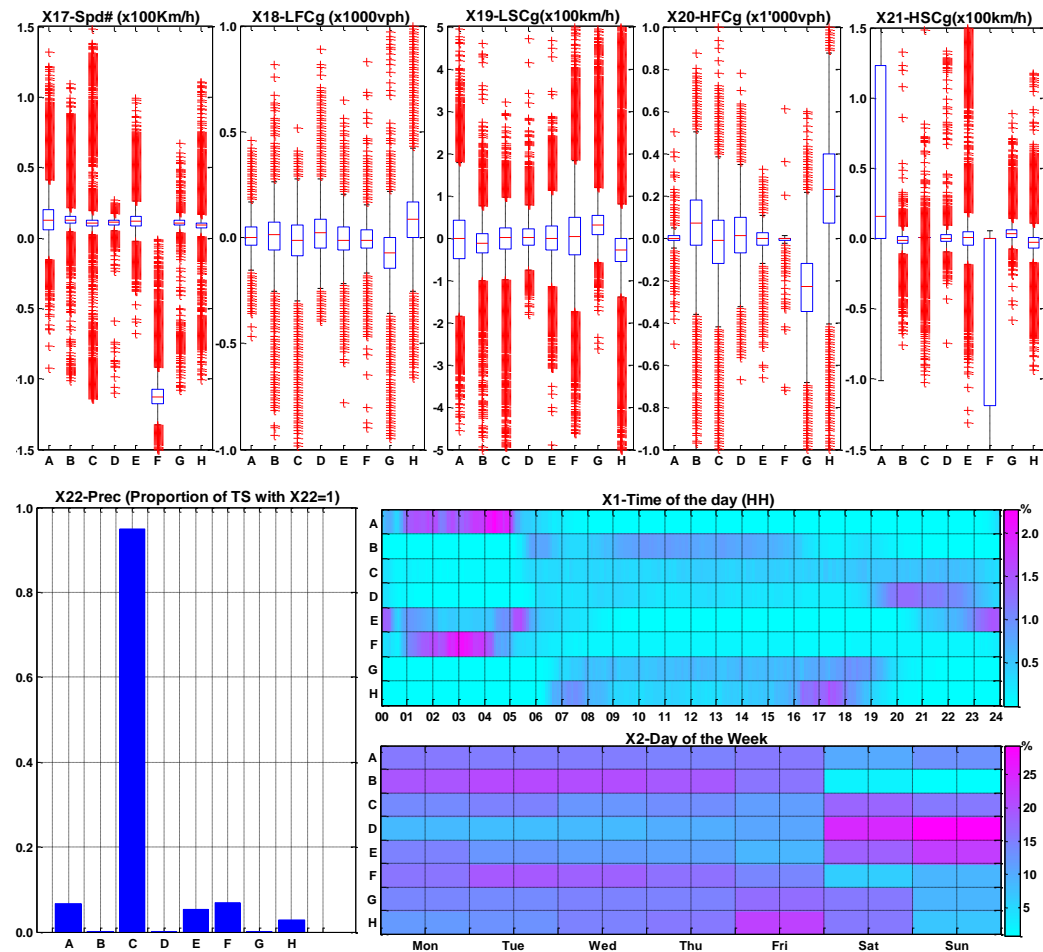


Fig 5.9 Statistics on variables from X17 to X21 and non-traffic variables X1, X2, and X22

By comparing X18 and X20 (the changes of flow on the right and left lanes, respectively) of two regimes B and G, X18 and X20 under regime G are mostly negative which means that there are less vehicles for Traffic Situations under regime G compared to the traffic situations that precede. This reflects the reduction of traffic flow under regime G. That can be the reason explaining the increases of average speeds on both lanes (represented by X19 and X21). For regime B, the inversed tendency occurs: the traffic flow is increasing and the average speed is reducing.

However, that tendency does not happen to regimes C and D. There is a slight increase of traffic flow under regime D compared to regime C. However, this change is not significant. It means that traffic characteristics under regimes C and D are similar.

Consider three non-traffic variables X1, X2, and X22, the most significant difference between regimes C and D is indicated by variable X22 (the type of precipitation). More than 95% out of NTS belonging to regime C are under rainy conditions whereas; only less than

1% of NTS under regime D are under rainy conditions. As X_{22} is one of variables involved in clustering process, it contributes actively to the separation of traffic into regimes C and D. According to Fig 5.8, the average speed is lower under regime C than under regime D. As most traffic characteristics are similar under regimes C and D, the type of precipitation is the reason for the higher PTS/NTS rate under regime C.

Two categorical variables X_1 and X_2 are not used in data sampling process and therefore do not contribute to the formation of Traffic Regimes. However, X_1 and X_2 also characterizes the difference between Traffic Regimes. The following can be observed from Fig 5.9:

- Regimes A, E, and F occur mostly during early morning before 6:00AM. Regime E occurs also after 9:00PM. Regimes A and F occur more often on weekday whereas; regime F occurs mostly during weekend.
- Regime H occurs more often during rush hours. Regime H occurs rarely on Sunday yet very often on Friday.
- Regime G is more frequent from 8AM to 7:00PM every day except Sunday when regime G is less frequent.
- Regime B is most frequent from 5:30AM to 6:30AM and from 9:00AM to 4:00PM and from Monday to Friday. During weekends, regime B is rare.
- Regime D can occur every day yet with high frequency on Sunday. High frequency of regime D's occurrences is also observed from 7:00PM to midnight.
- Regime C is very similar to regime D by considering X_1 and X_2 .

5.2.4 Examples of Traffic Regimes

This section presents several examples of traffic regime evolution in one day traffic. Three dates are selected: a normal day (working, non-rain, and uncongested day), a rainy day, a congested day, and a weekend. In Fig 5.10, Fig 5.11, and Fig 5.12, variables X_3 (flow on the right lane) and X_4 (average speed on the right lane) for the whole day are presented and referenced on the left vertical axis. Traffic Regimes are referenced on the right vertical axis. There is also a period in each figure where there is no data due traffic detector reset. During this period, speed and flow are set to zero.

In general, on weekdays, starting from midnight the traffic volume reduces to its lowest level from 1:00AM to 4:00AM before increasing to morning peak at about 8:00AM. Thereafter, the traffic reduces and stays at high volume level before increasing from 3:00PM to afternoon peak at about 5:30PM. After that, traffic volume reduces until midnight.

Fig 5.10 presents traffic evolution for a normal working day with no precipitation ($X_2=2$ - Tuesday and $X_{22}=0$) and uncongested traffic. As it is non-rainy day the traffic did not come to regime C.

Fig 5.11 presents traffic evolution for a weekend with some rainy period on Saturday, June 17, 2006. In early morning (from 1:00AM to 5:00AM), the traffic on Saturday is not as low as on working days. However, the traffic increases gradually to high flow at around 10:00AM. In the afternoon and evening, there are periods where the precipitation is positive: from 4:00PM to 4:30PM and some periods within 7:00PM to 8:00PM. During these periods, the traffic comes to regime C.

Fig 5.12 presents a working day traffic without precipitation and with congestion during morning peak (at around 7:00AM to 7:30AM). The congestion is indicated by the speed drop to below 40km/h.

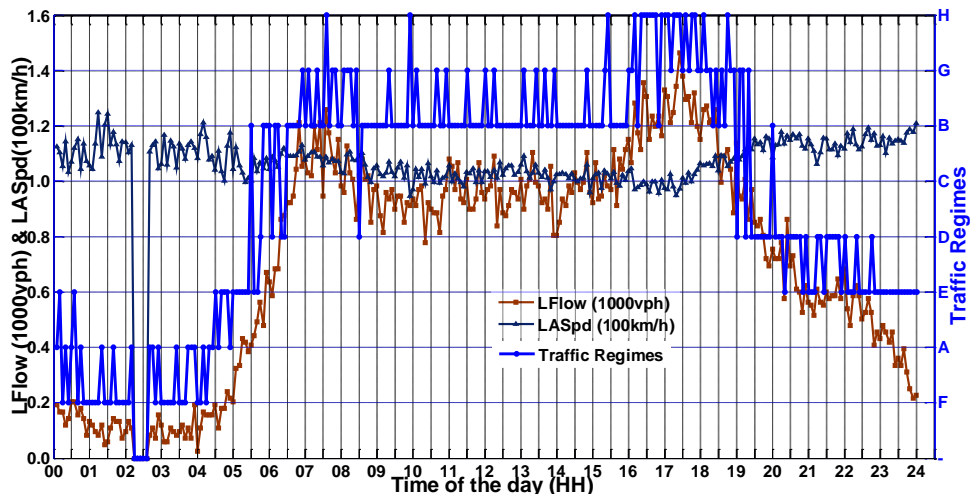


Fig 5.10 Traffic Regime evolution on Tuesday, April 25, 2006

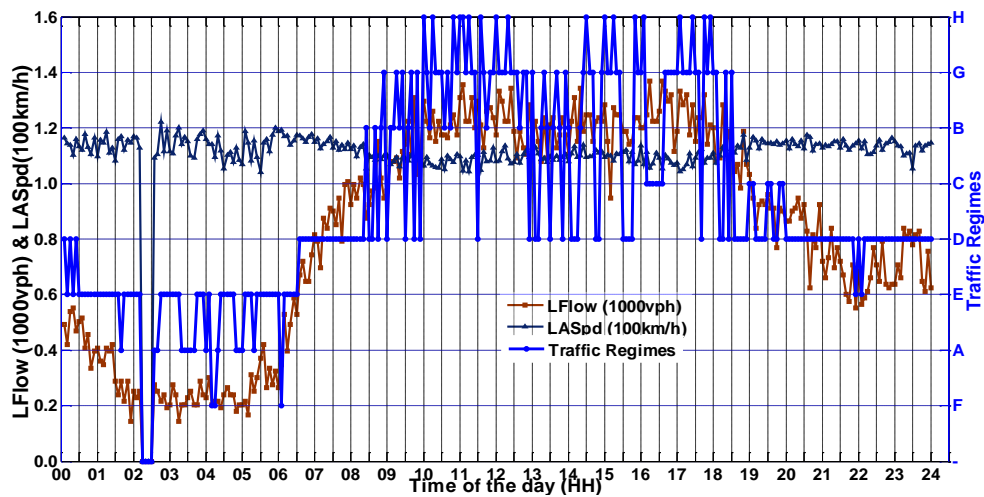


Fig 5.11 Traffic Regime evolution on Saturday, June 17, 2006

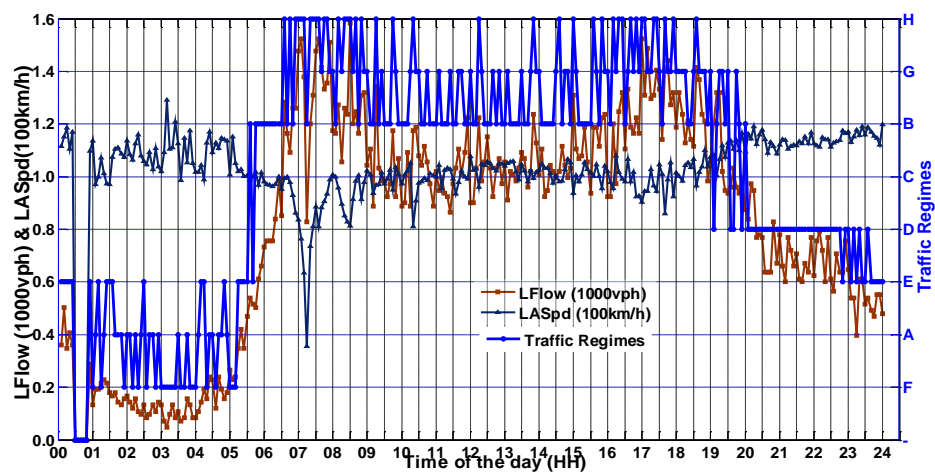


Fig 5.12 Traffic Regime evolution on Thursday, August 21, 2003

5.2.5 Summary

Three examples presented in Fig 5.10, Fig 5.11, and Fig 5.12 enforce the observations from 1) to 6) in section 5.2.3. Moreover, the following conclusions can be conducted:

- Regime B usually occurs when the traffic is increasing. Therefore, it represents increasing traffic and it usually precedes high flow and congested traffic regime H.

- Regime G usually occurs when the traffic is decreasing. Therefore, it represents decreasing traffic and usually follows regime H.
- When the traffic is already at regime B and does not significantly increase, it will come to regime G.

5.3 TS Transitions

5.3.1 Introduction

TS transitions are the movements of the TS from one TR to another TR. The typical transitions of TR for one day via three examples presented in Fig 5.10, Fig 5.11, and Fig 5.12 is that the traffic starts from midnight in free flow regimes F, A, and E then stays in high density regime B, G, and H during daytime before coming back to regimes E, A, and F in the evening. Regimes C and D play intermediate role when the traffic changes from free flow to high density regimes.

This section aims to provide more insight about TS transitions for NTS and eventually for PTS. Because six PTS are used for each crash, the pattern of six TR transitions is considered here.

5.3.2 NTS Patterns

NTS Transitions

An NTS pattern is a TR transition pattern of NTS, i.e. the traffic conditions where there are no crashes recorded. Fig 5.13a and Fig 5.13b presents the proportions of NTS transitions among TR. In Fig 5.13a and Fig 5.13b, Traffic Regimes are grouped in three groups: group of low traffic (regimes A, E, and F), group of intermediate traffic (regimes D and C), and group of high traffic (regimes B, G, and H).

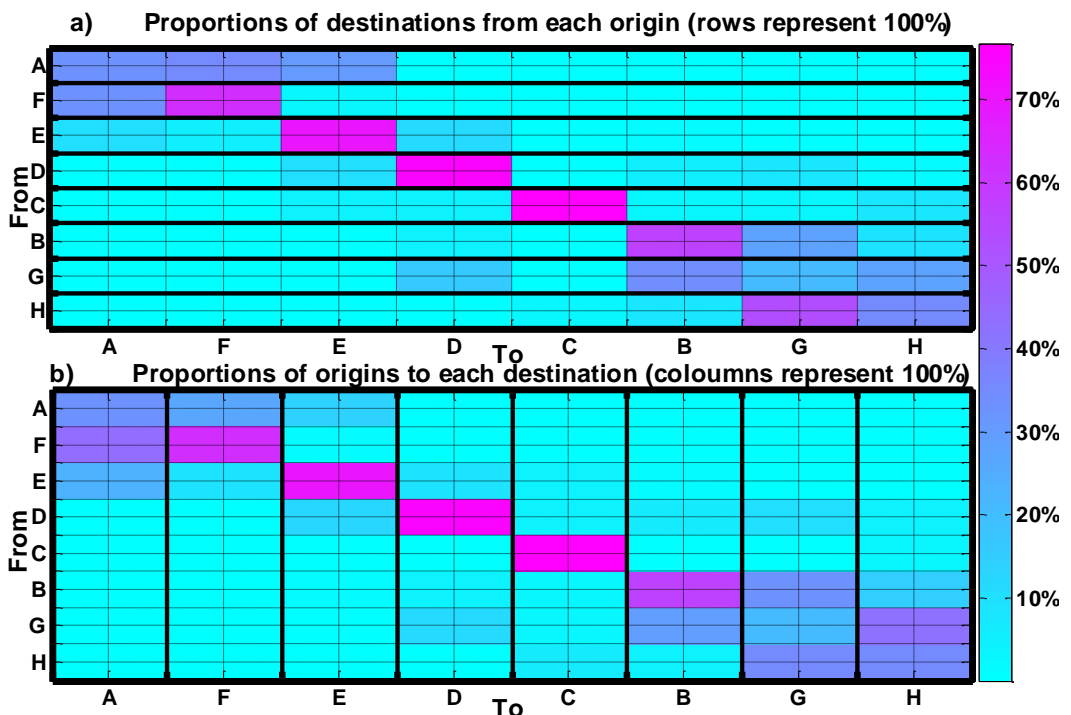


Fig 5.13 Proportions of NTS transitions from origins to destinations

Consider two consecutive traffic situations TS_{t-1} and TS_t whose traffic regimes are α and β , respectively. Two regimes α and β can be the same or different and are among eight regimes from A to H. Call $trans(\alpha, \beta)$ the number of transitions from regime α to β in NTS data. Call $P_t(\alpha, \beta)$ is the proportion of NTS transitions from regime α to regime β compared

to all NTS transitions from regime α to all regimes. $P_r(\alpha, \beta)$ is calculated as follows:

$$P_r(\alpha, \beta) = \frac{trans(\alpha, \beta)}{\sum_{\forall \beta} trans(\alpha, \beta)}$$

Similarly, $P_c(\alpha, \beta)$ is the proportion of NTS transitions from regime α to regime β compared to all NTS transitions from all regimes to regime β . $P_c(\alpha, \beta)$ is calculated as follows:

$$P_c(\alpha, \beta) = \frac{trans(\alpha, \beta)}{\sum_{\forall \alpha} trans(\alpha, \beta)}$$

Each cell in rows in Fig 5.13a represents $P_r(\alpha, \beta)$. For instance, as shown in Fig 5.12, more than 70% of transitions from regime D are destined for regime D, i.e. traffic situations remain in regime D and $\alpha = \beta = D$. Each cell in columns in Fig 5.13b represents $P_c(\alpha, \beta)$. Among NTS transitions from (and to) regimes F, E, D, C, and B, more than 50% remain in (or originate) the same regimes.

Regime G is the most unstable regime because once the traffic is under regime G, it remains under regime G only 21% of the time. From regime G, about 34% of NTS transitions have regime B as destination and about 28% have regime H as destination. Among NTS transitions to regime G, transitions from H and B contribute 35% and 33%, respectively.

The high flow or congested regime H is also unstable with about 35% of NTS transitions from regime H to regime H. Regime G exchanges the most with regime H: 53% of transitions from H is to G and 43% of transition to H is from G. Here high percentage of transitions from G to H among the transitions to H means that there is a fluctuation between G and H.

Regime B is rather stable with about 57% of NTS transitions from B remain in B. From B, about 28% of transitions are to G and 8% are to H. To B, about 29% are from G and 5% are from H. It means that the percentage of transitions from H to B is low.

The direct transitions between regimes B, G, and H and regimes A, E, and F are rare. The percentages of transitions between regimes B, G, and H or regimes A, E, and F with regimes C and D are low because regimes C and D are stable.

NTS Pattern Statistics

A TS pattern represents 30' evolution of TS under six traffic regimes. There are eight TR candidates (from A to H) for each of six traffic regimes. Therefore, there are totally $6^8 = 1'679'616$ possible patterns. There are transitions that can never occur such as from A to H to A. For this reason, the total number of NTS transition patterns during the whole study period is 11'782 among 1'050'372 available patterns (there must be at least 30' consecutive data to produce a pattern). There are patterns that repeat more than other patterns.

The pattern DDDDDD, i.e. the traffic remain under regime D for 30 minutes, repeats the most at 86'624 times, i.e. 8.25% of available patterns. After that, patterns EEEEE, FFFFFFF, BBBB, and CCCCC contribute 4.30%, 2.40%, 2.28%, and 1.54% of available patterns. This result conforms to high percentages of transitions from and to one regime for regimes B, C, D, E, and F.

5.3.3 PTS Patterns

Fig 5.14 presents the evolution of traffic under traffic regimes before crashes used in this study. Each cell column represents a crash, each cell row represents the moment of the PTS before crashes. Each cell is a PTS and the color within each cell represents the traffic regime of that PTS. The top row represents the frequencies of corresponding patterns.

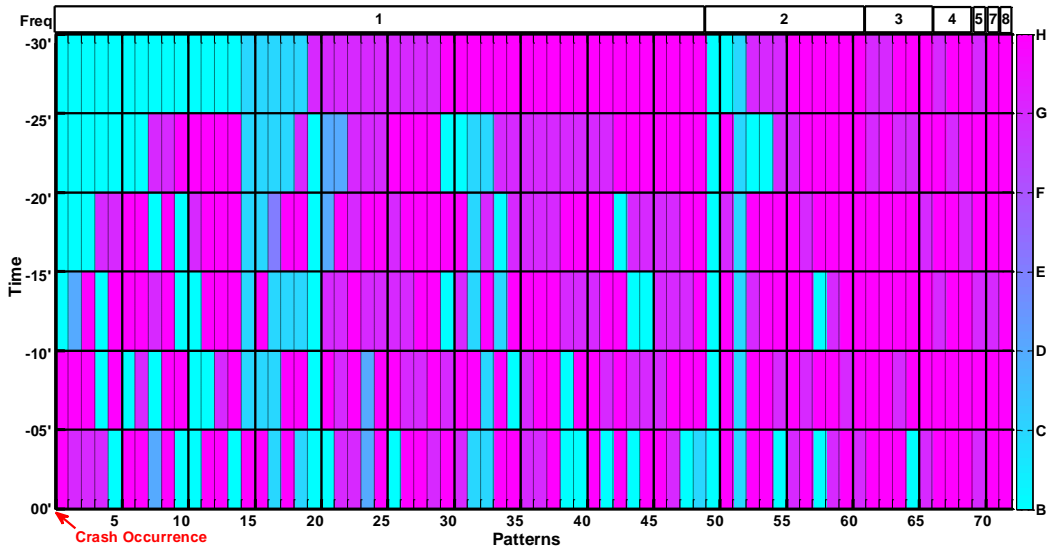


Fig 5.14 PTS Pattern repetitions

The PTS pattern HHHHHH repeats the most - eight times – which means that even the traffic is stable under regime H, the crash risk is still high. There are 64 crashes (i.e. 53.3%) for which the traffic comes only to regimes G and H such as the pattern 23, 25, 27-29, 36-38, 41, 46, 47, 56, 57, 58-64, and 66-72.

PTS patterns are also observed within NTS data. Fig 5.15 presents the repetition frequencies of PTS patterns and the pre-crash rate which is the ratio between the frequencies of PTS patterns in pre-crash data and PTS pattern in NTS data. The pattern indices in Fig 5.15 are the same to pattern indices in Fig 5.14. The PTS patterns BBBBBB and CCCCCC are the most popular in NTS data. After that, the frequencies of PTS patterns GBBBBB and HHHHHH are also high. It means that the chance for the traffic to end up with a crash after these patterns is not high.

There are patterns that can lead to higher pre-crash rate such as HGBCHH (one of four cases ended up with a crash), HHHBHB (two of ten cases ended up with a crash), CGH-CHC (one of seven cases ended up with a crash), and GDDGHB (one of eight cases ended up with a crash).

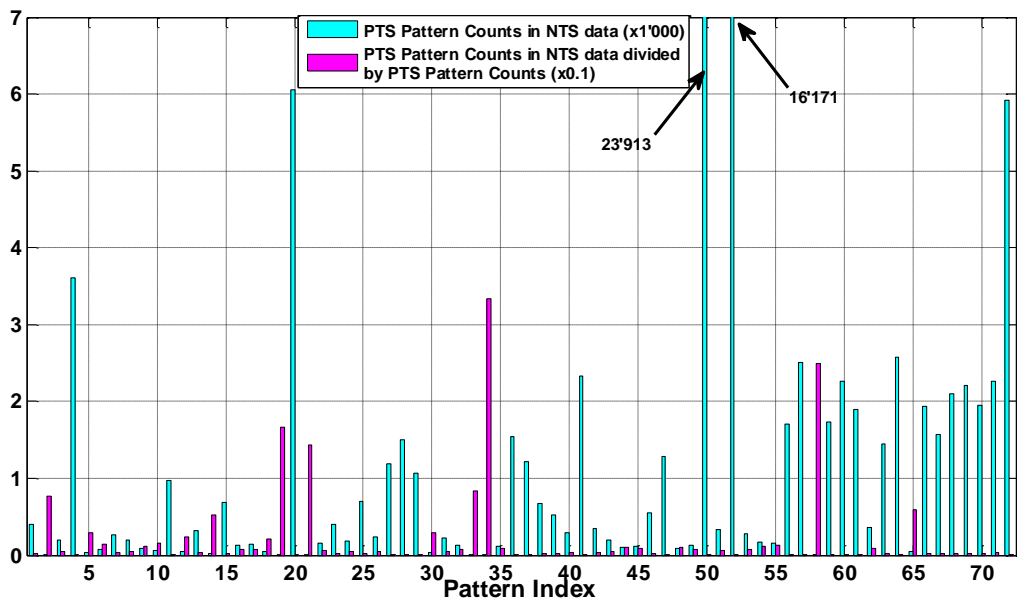


Fig 5.15 PTS patterns observed in NTS data

5.4 Summary

This chapter concentrates on uncovering the characteristics of eight traffic regimes. The fundamental Speed/Flow diagrams on two lanes cannot properly characterize each of traffic regimes as only four variables can be presented and different regimes overlap in the diagrams.

Three groups of traffic regimes are observed: group of low traffic, group of medium traffic and group of high traffic. The typical one-weekday traffic evolution starts from midnight with the first group, then changes to the second group at around 5:00 to 6:00 AM before reaching the third group and remain in the third group during day time. The traffic in the evening changes to the second group and then the first group until midnight. Weekend traffic evolution is slightly different with the traffic coming at the third groups at some time near noon.

Within the first group, traffic under regime E is denser than under regimes A and F. Regime F represents special traffic conditions when there is only traffic on one lane.

Regimes C and D are classified into the second group and are similar in traffic characteristics. The only difference between C and D is that regime C occurs under rainy conditions whereas regime D occurs under normal weather conditions.

In the third group, regime H represents high flow or congested traffic conditions whereas regimes B and G represents traffic conditions leading to and following regime H, respectively. There is high fluctuation between regimes H and G and between regimes G and B.

Traffic evolution can be presented by the transitions of TS from one regime to another regime. It is observed that before almost all crashes, the traffic comes to regime H at least once.

6 Fusion of Risk Indicators

During our study we have seen that no single risk indicator will achieve highly accurate risk characterization of traffic situation except in some special situations (for some network configuration/traffic situation and/or with high detector coverage). This is induced to some extent by traffic/safety dynamic which cannot be formalized by a single procedure. Therefore, in this section we will combine a set of selected safety indicators. Often, the approach used is to find the 'best' (single) indicator in some sense (showing best performance over validation set) among the available safety indicators and for each traffic situation obtained from the clustering, then combining these individual safety indicators. Since the best combination of a set of safety indicators depends on the application and on the safety indicators to be combined, there is no single, best combination scheme or any unequivocal relationship between the accuracy of a multiple risk indicator system and the individual constituent safety indicator.

6.1 Introduction of data fusion

Data fusion is defined as a framework for the combination of data originating from different sources in order to achieve higher quality information for decision making than a single source.

The IEEE Geosciences and Remote Sensing Society Data Fusion Technical Society define the Data Fusion as *the process of combining spatially and temporally indexed data provided by different instruments and sources to order to improve the processing and interpretation of these data.*

Data fusion is used on many applications like weather forecasting, vehicular traffic control, military operation, etc. Regarding risk indicators and depending of available input information, several data fusion frameworks and approaches are of potential help to combine in real time information provided by those safety indicators. The methods considered were expert systems, fuzzy logic, statistical techniques, Bayesian inference, Dempster-Shafer evidence theory, and several voting logic approaches. Of these, statistical aggregation, voting, and Dempster-Shafer inference were selected as candidates for further evaluation, based on the types of data and information that are easily extracted from the risk indicators. Hence, the information available from each risk indicator determines the data fusion approach to be applied. All of the approaches share a common requirement to be successful, namely that risk indicators need to supply some measures of the quality, reliability, or likelihood that its quantitative or qualitative outputs are correct in addition to the risk indicator value itself.

The main goals of the data fusion process in FUSAIN are an improvement in risk detection and their ability to characterize road safety conditions.

The specificity of the data fusion process for safety indicators is that most of the safety/risk indicators are derived from the same traffic data but in different forms; so the combination process is performed in fact for the same data provided by one source. This can be seen as the newly established practice of bagging estimators. The expected benefit is not only an increase of the accuracy but also a robustness of the whole risk detection process.

The architecture of such a process is illustrated in Fig 6.1. The benefit expected from the data fusion process is to improve the quality of the risk indicator measures and particularly the conclusions drawn. With respect to the risk associated to traffic situations.

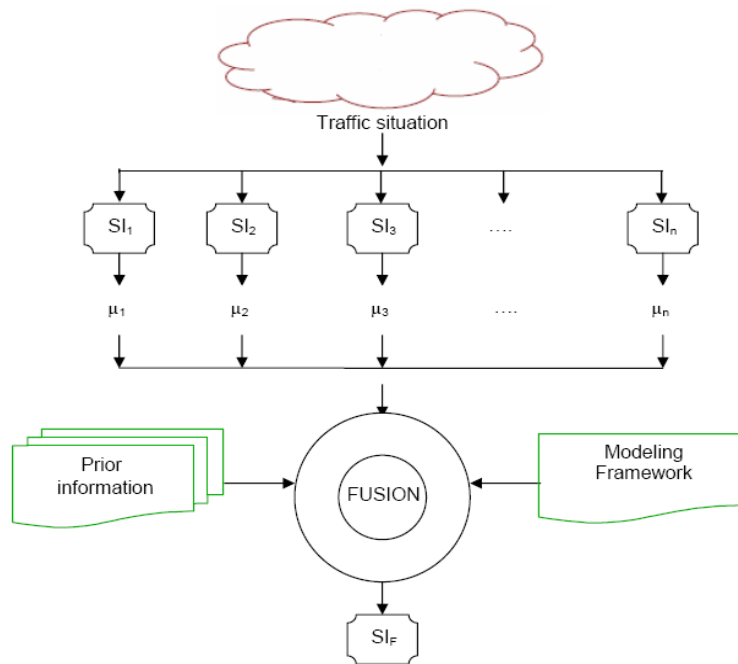


Fig 6.1 Architecture of safety indicators fusion

Sensor fusion is the combining of sensory data or data derived from sensory data from disparate sources such that the resulting information is in some sense better than would be possible when these sources were used individually. It is a combination of information from multiple sensors/sources to achieve inferences that are not feasible from a single sensor or source.

The concept of multi-sources fusion is hardly new. The most widely used fusion model, developed in 1985 by the U.S Joint Directors of Laboratories (JDL) Data Fusion Group, aims at facilitating understanding and communication among acquisition managers, theoreticians, designers, evaluators, and users of data fusion techniques to permit cost effect system design, development, and operation (White, 1991). This model divides the DF into low and high level processes. The low level support pre-processing of data (classification, identification, target tracking...). High level process support situation and threat assessment and fusion process refinement.

A new revision of the JDL Data Fusion was proposed by Steinberg et al., (1999) and it's structured as a five-hierarchical level processing described as follows:

- Level 0: Data Assessment: source pre-processing (normalizing, formatting, ordering, compressing...).
- Level 1: Object Assessment: estimation and prediction of entity states on the basis of inferences from observations.
- Level 2: Situation assessment: involves the relations among elements being aggregated.
- Level 3: Impact Assessment: estimation and prediction of effects on situations of planned or predicted actions by participants.
- Level 4: Process Refinement: achieves improved results by continuously refining estimates and assessments through planning and control.

6.2 Fusion Schemes

After defining the 5 global levels of the data fusion process, we focus on the methods used for aggregate (fuse) the data from the sources. Generally the DF isn't an easy task and it can be structured into 4 important steps (Bloche, 2002)

- 1-Modelling: in this step the formalism (fusion framework) and the expression of the data that is needed for the fusion are selected. For example if every source provide an information about the decision . So the form of depend on the chosen formalism. (Can be a statistical distribution, mathematic formula...).
- 2-Estimation: here model parameters and likelihood measure need to be estimated and calibrated these two tasks are performed in this step.
- 3-Combination: this is actually the step at which data are combined according to the selected formalism.
- 4-Decision: the final step of the fusion where the decision is taken in depends on available data.

In the literature there are three main categories of the fusion process/methodologies know as: Physical Models, Feature-Based Inference Techniques and Cognitive-Based Models. In our study we will focus on two families which are the Feature-Based Inference Techniques and the Cognitive-Based Models because of the inconsistencies of the Physical Models with the nature of the safety indicator.

6.2.1 Feature-Based Inference

Feature Based technique are divided into probabilistic and information theoretic techniques. Statistical parametric mapping entails the use of the general linear model and classical statistics. The parametric techniques include Bayesian inference, and Dempster-Shafer evidential theory.

Machine learning model transform parametric data into an identity declaration and are known by the fact that they share similar concept: similarity in identity is reflected in observable parameters. This category include methods such as: Artificial neural networks, cluster algorithms, voting methods, entropy measuring techniques...

Bayesian inference

Bayesian inference is a probability-based approach grounded in Bayes rule. The information is modelled with a conditional probability. For example the information M_j^i can be expressed $M_j^i = P(D_i / S_j)$ where $P(D_i / S_j)$ is the conditional probability to take de decision D_i from the data source S_j .

The major advantage of the probabilistic methods is based on a theoretical and mathematical ground and was the object of numerous works. They thus propose a very rich range of tools allowing as well as modelling tool (law) and the learning approaches.

In the Bayesian Inference we distinguish to type of fusion framework Simultaneous and Sequential. In the simultaneous model the data to fuse are available at the same time, so we have to calculate Probability of this form $P\langle D_i | S_1, \dots, S_j \rangle$ By means of the Bayesian rule:

$$P\langle D_i | S_1, \dots, S_j \rangle = \frac{P\langle S_1, \dots, S_j | D_i \rangle P(D_i)}{P(S_1, \dots, S_j)},$$

where all the terms are learnt by apprenticeship. With the Sequential framework we use Bayes rule directly where new information came to update the estimated old one (used when the data are updated frequently):

$$P(D_i | S_1, \dots, S_j) = \frac{P(S_1 | D_i) P(S_2 | S_1, D_i) \dots P(S_j | S_1, \dots, S_{j-1}, D_i) P(D_i)}{P(S_1, \dots, S_j)}$$

Note that this equation can be simplified under the independence assumption.

The best way to take a decision in the Bayesian inference is to use maximum a posteriori estimate (MAP), so we choose D_i when:

$$P(D_i) = \text{Max} \left\{ P(D_i | S_1, \dots, S_j), 1 \leq i \leq l \right\}.$$

In his book, Klein compared the two probabilistic approaches and gave a table of advantages and weakness; we are going to reproduce some of the arguments provided by him in Fig 6.2 and Fig 6.3.

Fig 6.2 Advantages of Classical and Bayesian inferences

Classical	Bayesian
Probability model links observed data and population.	Probability of a hypothesis being true is determined from known evidence.
Probability model is usually empirically based on parameters calculated from large number of samples.	Likelihood of hypothesis is updated using a previous likelihood estimate and additional evidence.
A number of decision rules may be used to decide between the null hypothesis H_0 and an alternative hypothesis H_1 , including maximum likelihood, Newman-Pearson and min-max.	Subjective probabilities are inferred from experience and vary from person to person Support more than two hypotheses at a time.

Fig 6.3 Weakness of Classical and Bayesian inference

Classical	Bayesian
When generalized to include multi-dimensional data from multiple sensors, a priori knowledge and multidimensional probability density function are required.	A priori probabilities and likelihoods must be defined.
Generally only two hypotheses can be assessed at a time, namely H_0 and H_1 .	Complexities are introduced when multiple conditional dependent events are present.
Multivariate data produce evaluation complexities.	Competing hypotheses must be mutually exclusive.
A priori assessments are not utilized.	Cannot support an uncertainty class.

Evidence theory

Evidential theory or Dempster-Shafer theory is a data fusion approach that been developed in the 70's and based on seminal works of Dumpster.

This approach is known to be more flexible then previous one as it allows to model at the

same time the indistinctness (ignorance) and uncertainty with the help of probability mass m , plausibility and belief functions.

The mass functions are defined in all the sub sets of the discernment space $\Omega = \{d_1, d_2, \dots, d_j\}$ not only the singleton.

Mass functions is a mapping for $2^\Omega \rightarrow [0, 1]$ that satisfy: $m(\emptyset) = 0$ and $\sum_{A \subset \Omega} m(A) = 1$, and a proposition is called a focal element if its mass is greater than zero.

The Dempster-Shafer theory is useful when the sensors (data source) contributing information cannot associate a 100% probability of certainty to the decision, so the information (knowledge) are combined using Dempster's rule to find the intersection or conjunction, and when the intersection is empty the conflicting probability are redistributed to the non-empty set element.

The mass functions can be expressed as a probability estimated from the singleton but this expression (model) is reducing and don't take advantage of the Evidential theory, but there are others techniques which takes this initial model then distribute the masses on all the compound hypotheses.

Parametric Template

Here multi sources information's are compared with priori conditions to determine if the observation contains certitude about entity.

An example of parametric template applied to multispectral or hyper spectral sensor data, the sensor detect the value of the radiance R_i emitted by object over many spectral bands $\Delta\lambda_i$.

Objects are defined by templates consisting of radiance values for each spectral band in the sensor. The measured radiance values are overlaid on the templates. Identification is made when the measured radiance values over the ensemble of spectral bands correspond to or are best represented by those of a known object.

Artificial neural networks

An Artificial Neural Network (ANN) (Stergiou and Siganos, 1996) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like humans, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons.

In the artificial neural networks, data fusion process proceeds like the clustering with artificial networks except that the neuron are de decision of each data source.

Cluster Algorithms

In the various branches of mathematics that fall under the heading of abstract algebra, the kernel of a homomorphism measures the degree to which the homomorphism fails to be injective. An important special case is the kernel of a matrix, also called the null space. The definition of kernel takes various forms in various contexts. But in all of them, the kernel of a homomorphism is trivial (in a sense relevant to that context) if and only if the homomorphism is injective. The fundamental theorem on homomorphism (or first isomorphism theorem) is a theorem, again taking various forms, that applies to the quotient algebra defined by the kernel. In this article, we first survey kernels for some important types of algebraic structures; then we give general definitions from universal algebra for generic algebraic structures.

Clustering algorithms for data fusion can be gathered into two main categories (Yu et al., 2008).

The first approach is clustering ensemble, also known as clustering aggregation or consensus clustering that combines multiple partitions into a consolidated partition by consensus function. Different ensemble algorithms have the same conceptual framework as shown in Fig 6.4 they mainly vary on the choice of consensus functions. Clustering ensemble is originally applied on single source where various partitions are generated by different representations, dimensionality reduction techniques and clustering algorithms.

The strategy of clustering ensemble can be extended to data fusion, where the main difference is that partitions now are varied by sources. An underlying assumption is: If the information contained in multiple sources is highly correlated, their partitions should also still contain "common agreement" thus a consolidated partition can also be obtained.

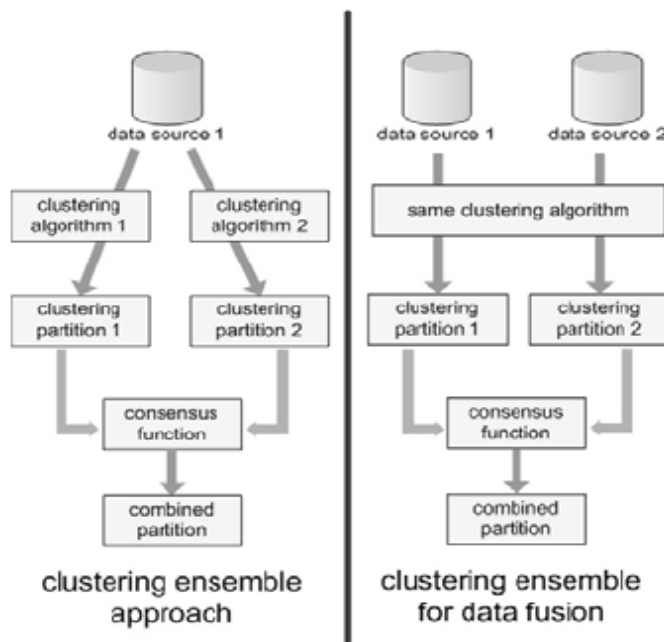


Fig 6.4 Clustering approach for data fusion (Yu, de Moor and Moreau, 2008)

An alternative approach of clustering by data fusion is achieved by fusing similarity matrices. If the similarity matrices are positive semi-definite, the data integration problem can be formulated as a kernel fusion problem. The main difference of kernel fusion approach illustrated in Fig 6.5 is that the integration is carried in kernel space before clustering algorithm is applied (early integration) while clustering ensemble aggregates partitions after clustering (late integration). Kernel method elegantly resolves the heterogeneities of data sources by representing them as same-size kernel matrices.

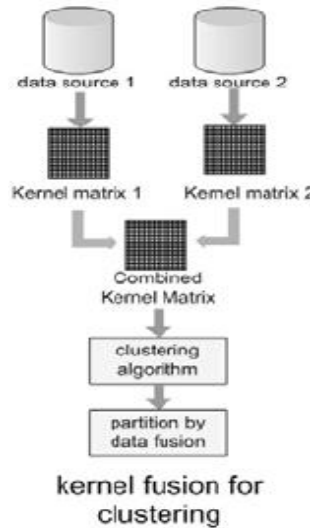


Fig 6.5 Kernel fusion Approach (Yu, de Moor and Moreau, 2008)

Voting methods

This approach considers each source’s (Sensor) declaration as a vote in which majority, plurality, or decision-tree rules are used. Also discrimination can be introduced via weighting of the source declaration.

There is another technique called Figures of merit that have similar aspect with the weighted decision formulas, it is metrics derived from heuristic arguments that aid in establishing a degree of association between observations and object identity.

There are three types of configurations in the voting methods: The parallel, series and series/parallel in Fig 6.6 relate those three configurations.

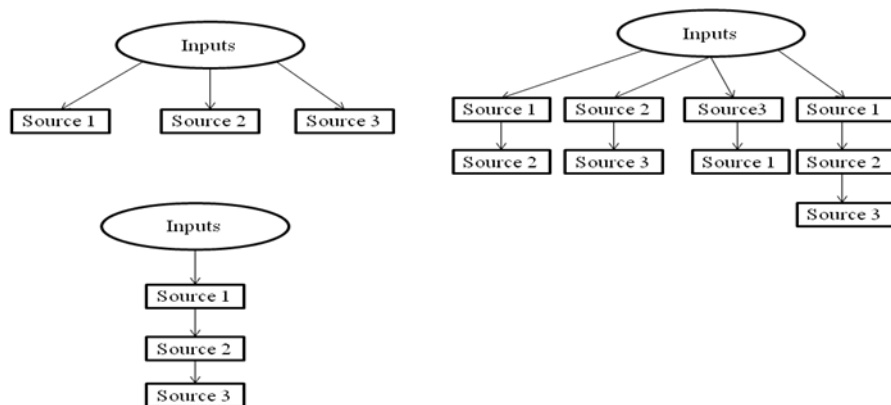


Fig 6.6 Voting Methods configuration (Klein, 2004)

The series source represent the “and” operator of the logic and parallel the “or” operator.

6.2.2 Cognitive-based models

This class of methods includes Logical template, Knowledge-based systems and fuzzy set theory; it tries to reproduce the decision taken by human analysts.

Logical Templates

Template is a comparison between stored pattern and an observed data to infer the classification or to predict a situation. Parametric templates that compare real-time patterns

with stored ones can be combined with logical templates derived from Boolean relationships.

Knowledge-based expert systems

Knowledge-based expert systems, or simply expert systems, use human knowledge to solve problems that normally would require human intelligence. These expert systems represent the expertise knowledge as data or rules within the computer. These rules and data can be called upon when needed to solve problems.

Computed-based expert systems frequently consist of four components:

- Knowledge base
- Global data base that contain input
- Control structure
- Human/machine interface

Knowledge base contain rule of the type "IF-THEN", IF: represent the condition, THEN: the conclusion or decision. This rule can be used for the data fusion by apprenticeship over the data base like we have seen on the Bayesian inference.

Fuzzy set theory

Fuzzy sets are sets whose elements have degrees of membership. Fuzzy sets are an extension of the classical notion of set. In classical set theory, the membership of elements in a set is assessed in binary terms according to a bivalent condition an element either belongs or does not belong to the set. By contrast, fuzzy set theory permits the gradual assessment of the membership of elements in a set this is described with the aid of a membership function valued in the real unit interval [0, 1]. Fuzzy sets generalize classical sets, since the indicator functions of classical sets are special cases of the membership functions of fuzzy sets, if the latter only take values 0 or 1. Classical bivalent sets are in fuzzy set theory usually called crisp sets. Fuzzy set theory data fusion operate by aggregating decision from data sources (risk indicators) decision using the membership function concept and the result is a decision with a certain membership (the result membership is an aggregation of the memberships of each data source)

This part has focused on the definition of the Data Fusion concept and the description of the works and approaches associated to the Data Fusion domain.

6.3 Risk Indicators Fusion

The main aim of this section is to present a framework for data fusion using the Dempster Shafer Evidential Theory for the risk indicators.

First we discuss a general approach for the multi-source fusion using the evidential theory and then, we propose a specific framework for the safety/risk indicators fusion of the investigated in the FUSAIN project.

6.3.1 Dempster -Shafer technique of data fusion

Dempster-Shafer technique is based on the theory of evidence developed by Dempster in 1969 and popularised by Shaffer in 1976. This theory allows dealing with uncertainty in a more general setting than the use of the classical probabilistic approach.

Knowledge from multiple data sources is combined with the use of the Dempster's rule to define a conjunction of each proposition and their likelihood. On the case where the intersection of the propositions reported by data sources is an empty set, the weight of the conflicting probability is attributed to the non-empty sets.

This approach allows to model at the same time the indistinctness and uncertainty with the help of probability mass m , support and plausibility.

The mass functions unlike probabilities are quite defined in all the sub sets of the discernment space Ω , $\Omega = \{d_1, d_2, \dots, d_j\}$ not only the singleton.

The mass functions fulfil the following condition:

$$m(\emptyset) = 0 \text{ and } \sum_{A \subset \Omega} m(A) = 1, .$$

A proposition is called a focal element if its mass is greater than zero.

Support

The support is the sum of all masses assigned directly by the data source to a given proposition and its subsets; the subsets are focal if their masses are greater than zero. Set $S(a) = m(a)$, $S(a \cup b) = m(a) + m(b) + m(a \cap b)$. (6)

Plausibility

The plausibility of a proposition is the sum of all mass not assigned to its negation. The plausibility of a proposition can be written as:

$$Pl(a) = 1 - S(\bar{a}), \text{ (7)}$$

Plausibility is also the sum of all masses of subsets a_j that have non-null intersection with **a**:

$$Pl(a) = \sum_{a_j \cap a \neq \emptyset} m(a_j). \text{ (8)}$$

An uncertainty interval is defined by [S (a), Pl (a)], where S(a) < Pl(a). In the probabilistic setting S (a) =Pl (a) = Pr (a).

The lower bound (support) is equal to the minimal commitment for the proposition based on direct evidence of the data source. The upper bound (plausibility) is equal to the support plus any potential commitment. Fig 6.7 summarizes interpretation of uncertainty intervals.

Fig 6.7 Interpretation of uncertainty intervals

Uncertainty interval	Interpretation
[0,1]	Ignorance
[0.6,0.6]	A definite probability 0.6.
[1,1]	Certain proposition.
[0,0]	Null proposition.
[0.3,0.75]	Probability of a is between 0.3 and 0.75.
[0,0.9]	Evidence provides partial support for \bar{a} .
[0.27,1]	Evidence provides partial support for a .

Dempster’s rule for combination of multiple data source

Given two mass functions, m_1 and m_2 we define the combined mass function, denoted by m_{12} as:

$$m_{12}(E_i) = K \sum_{E_j \cap E_l = E_i} m_1(E_j)m_2(E_l), \quad (9)$$

Where:

$$K = \frac{1}{1 - \sum_{E_j \cap E_l = \emptyset} m_1(E_j)m_2(E_l)}. \quad (10)$$

The probability mass calculated on (6) is called the orthogonal sum and is denoted:

$$m_1(E_j) \oplus m_2(E_l) \quad (11)$$

6.3.2 Implementation of the Dempster Shaffer technique

Dempster's Shaffer rule is implemented by forming a matrix with the probability masses to combine along the first column and last row (the first column and last row represent the detection of each data source) as presented in Fig 6.8.

Element of the matrix are computed as the product of the probability mass. Matrix elements represent the propositions of the intersection of the propositions that are multiplied.

Fig 6.8 Example of possible outputs of a data fusion process

	A	B	C	$B \cup C$	Ω
A	A	\emptyset	\emptyset	\emptyset	A
B	\emptyset	B	\emptyset	$B \cap (B \cup C)$	B
C	\emptyset	\emptyset	C	$(B \cup C) \cap C$	C
$A \cup C$	$(A \cup C) \cap A$	$(A \cup C) \cap B$	$(A \cup C) \cap C$	$(A \cup C) \cap B \cup C$	$A \cup C$
$B \cup C$	$A \cap (B \cup C)$	$(B \cup C) \cap B$	$(B \cup C) \cap C$	$B \cup C$	$B \cup C$
Ω	A	B	C	$B \cup C$	Ω

When the intersection of the propositions that define the matrix elements is an empty set, the mass of the empty set is fixed to zero and the probability mass assigned to the nonempty set elements is increased by the K factor defined in equation (10).

6.3.3 Fusion of risk indicators with Dempster-Shaffer Framework

In this subsection we will introduce two strategies for the FUSAIN risk indicators fusion using the Dempster modelling. The two proposed strategies are named “static fusion” versus “online fusion” respectively. Generally speaking the fusion process is the same in the two strategies and the difference lies on the attribution of the mass associated to each risk indicator. More specifically the definition of the mass function is the first step of the fusion, we will set apart the two mass attribution approaches method and once attributed their fusion operate with the same technique.

Mass function modelling

The preliminary step of the risk indicator (RI) fusion is to define the mass function. The mass function must represent the risk likelihood associated to a given traffic situation (derived by traffic data). Thus for a given real time situation the value of each risk indicator must be computed together with the type of traffic situation it belongs to. Then an apprenticeship can be operated within the stored crash data for the same traffic situation. More specifically we consider here that outputs in terms of decision of a given risk indicator R_i are:

H_1 =”risk” and H_2 =”no decision” so the mass function values are associated to H_1 and to frame of discernment.

The static fusion strategy

For the static fusion, the mass attribution is computed once, so for a given traffic situation (TS_j) and for each risk indicator (RI_i) a same value of the mass ($m_{i,j}$) is attributed to H_1 and it obtained from apprenticeship over the available data, a threshold t_i is defined for each RI_i value regarding to pre-crash data, hence the mass is computed as follows:

$$m_{i,j}(\text{"risk"}) = \frac{r_{i,j}}{n_j} \quad (12)$$

Where:

$r_{i,j}$: represent the number of traffic situations with values of risk indicator (RI_i) higher than the threshold t_i .

n_j : Total number of traffic situation in the cluster j (TS_j).

And $m_{i,j}(\text{"no decision"}) = 1 - m_{i,j}(\text{"risk"})$ (13)

Online fusion strategy

For the online fusion the mass is computed online for every new traffic situations. So for a given real time situation the value of the risk indicator is calculated and the situation is assigned to a one of the cluster achieved by clustering. The apprenticeship process operates as follows:

$$m_j(\text{"risk"}) = \frac{1}{n_j} \sum_{l=1}^{n_j} \mathbf{1}\{v_i \geq v_{ij}^{pc}\}, \quad (14)$$

v_i : Real time value of the risk indicator (l),

n_j : The total number of traffic situations (safe and pre-crash) in the cluster (j),

v_{ij}^{pc} : The risk indicator (l) Value for the pre-crash situation (l) from the traffic situation (cluster j).

v_{ij}^{nc} : The risk indicator (l) Value for a non-crash situation (l) from the traffic situation (cluster j).

Thus, the mass attribution for the online fusion proceeds by comparing the risk indicators values of the real time situation with each pre-crash available data. The mass associated to "no decision", is derived by normalization so that the sum of mass equals to 1. To account for a non-detection rate defined is a number of times the value of risk indicator is greater than those obtained for a non-crash situation which could be risky, we can introduce it as an amplification (reliability) factor. In this case, the mass function given in is modified as:

$$m_j(\text{"risk"}) = \frac{1}{n_j} \sum_{l=1}^{n_j} \mathbf{1}\{v_i \geq v_{ij}^{pc}\} \alpha_{ij}, \quad (15)$$

$$\text{where } \alpha_{ij} = 1 - \frac{1}{n_j} \sum_{l=1}^{n_j} \mathbf{1}\{v_i \geq v_{ij}^{nc}\} \quad (16)$$

And the mass functions are normalized to sum up to one.

Fusion of the risk indicators mass

In this step we can perform mass function contribution. This formula carries out the fusion of the weights of the risk indicators (mass function) calculated on the confusion vector.

The fusion of risk indicators operate by fusing RI two by two. For instance, for the RI_1 and RI_2 which outputs are $(O_l)_{l=1\dots}$ Dempster's rule is defined as follows:

$$m_{12/j}(O_i) = K \sum_{O_s \cap O_l = O_i} m_{1/j}(O_s) m_{2/j}(O_l),$$

$$\text{where } K = \frac{1}{1 - \sum_{O_s \cap O_l = \emptyset} m_{1/j}(O_s) m_{2/j}(O_l)}.$$

This means generating a 2*2 matrix structured as shown in Fig 6.9:

Fig 6..9 Decision matrix

	Risk	No decision
Risk	$m_{1/j}m_{2/j}$	$m_{1/j}(1-m_{2/j})$
No decision	$(1-m_{1/j}) m_{2/j}$	$(1-m_{1/j})(1-m_{2/j})$

m_1 and m_2 represent respectively the mass associated to the risky situation calculated for each risk indicators RI_1 and RI_2 .

$(1-m_{1/j})$ and $(1-m_{2/j})$ represent respectively the mass associated to the uncertainty relative to each risk indicators RI_1 and RI_2 .

In this first matrix the only commensurate element are those along the diagonal. All other are empty set members. The redistribution of the mass of the empty set members to the non-empty set propositions is calculated via the K factor as presented in Fig 6.10

Fig 6.10 calculation of non-empty set propositions

	Risk	No decision
Risk	$K(m_{1/j}m_{2/j})$	0
No decision	0	$K(1-m_{1/j})(1-m_{2/j})$

$$K = \frac{1}{1 - ((1 - m_{1/j})m_{2/j} + m_{1/j}(1 - m_{2/j}))}$$

The resulting mass function $m_{12/j}$ represent the mass of the new fused RI_{12} (composed of RI_1 and RI_2) and the value " $K(m_{1/j}m_{2/j})$ " is the mass associated to the risk of the new fused indicator as presented in Fig 6.11.

Fig 6.11 mass associated to the risk of the new fused indicator

	$m_{12/j}$
Risk	$K(m_{1/j}m_{2/j})$
No decision	$K(1-m_{1/j})(1-m_{2/j})$

Now with new fused RI the next step is to fuse with the same approach with all the rest indicators RI_{12} with $RI_3...$ and the last fusion will give the aggregated decision with the final value of risk. Fig 6.12 presents the online fusion framework.

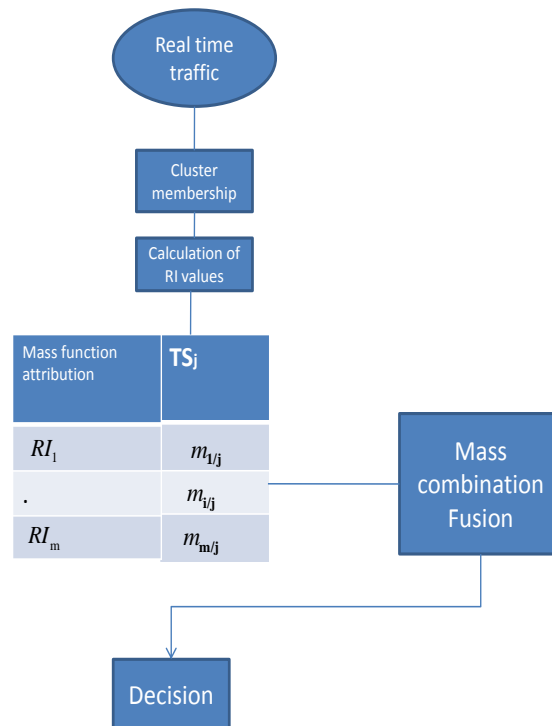


Fig 6.12 Online fusion framework.

Fig 6.13 provides the static fusion results and Fig 6.14 shows an online fusion example.

Fig 6.13: Static Fusion result

Traffic regimes	Size	iTTC	TF	IBTR	PBTR	IBTRRT	PBTRT	PBTRbis	PBTRRT bis	MAGIK	CIBTR	Decision
		Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	
A												
B												
C	564	0,071	0,098	0,018	0,16	0,072	0,028	0,157	0,177	0,014	0,174	7.992686e-12
D	33	0,606	0,455	0,364	0,515	0,545	0,03	0,272	0,060	0,181	0	0.0001
E												
F												

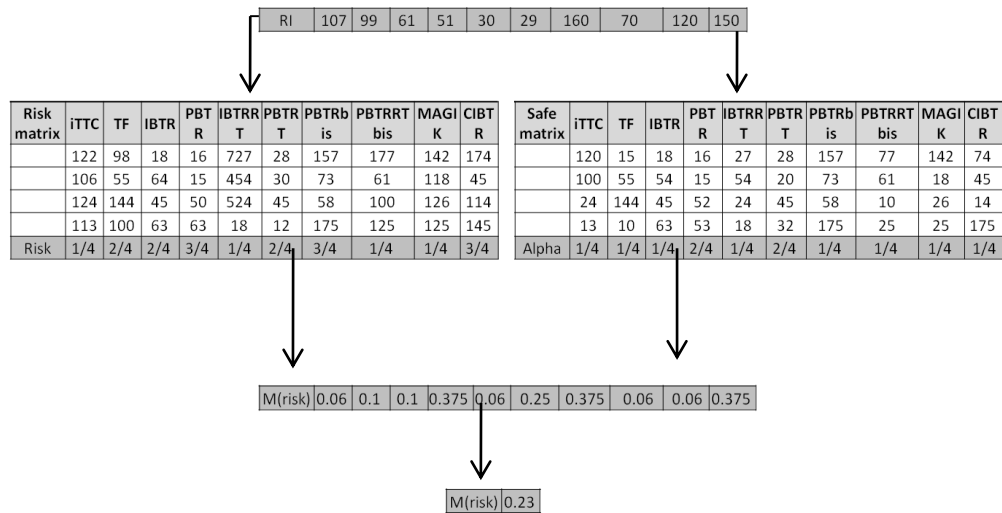


Fig 6.14 Online Fusion Example

We have addressed to the problem of the fusion of risk indicators. We have shown in our study how to perform the Dempster Shaffer method and define two frameworks for the data fusion problem especially for the fusion of the traffic risk indicators studied for the FUSAIN project.

6.4 Results

6.4.1 Regime-based Risk Identification Models (RIM)

Single-variable models are developed using Random Forest (Breiman, 2001) to classify a traffic case into one of two class NTS or PTS. For a traffic case, there are 25 results from 25 single-variable models. The Fused RIM result is decided by voting single-variable results.

Fig 6.15 presents the results obtained with Fused RIM for four traffic regimes B, C, G, and H. With data including NTS and PTS divided into three smaller data sets: training and validation the performance of the models is reduced for the validation data set yet still at high level.

Fig 6.15 Fused RIM results

Traffic Regime	Training (%)		Validation (%)	
	NTS	PTS	NTS	PTS
B	99.58	100.00	96.13	85.33
C	95.73	100.00	91.81	90.00
G	92.89	100.00	89.25	89.02
H	96.45	100.00	85.47	77.05
All four regimes	96.41	100.00	90.73	85.24

6.4.2 Crash Risk Prediction Model

Fused RIMs provide the affirmation of traffic crash risk for a traffic case that has occurred. In a real-time framework, the affirmation of traffic crash risk for several last traffic cases can be employed to predict the coming traffic case. Call the false alarm rate the percentage of NTS that are wrongly predicted as PTS and missed alarm rate the percentage of PTS identified as NTS. The sensitivity test with historical PTS shows that there is a tradeoff between the false alarm rate and missed alarm rate: it is impossible to minimize false alarm rate and miss alarm rate at the same time.

As illustrated in Fig 6.16, the number of the last traffic cases considered to predict the coming traffic case, called *the length of risk memory*, ranges from 1 to 5. The optimum length of risk memory can be 2 or 3 as with this length, the false alarm rate and the missed alarm rate are although not the lowest yet acceptably low compared to other length of risk memory.

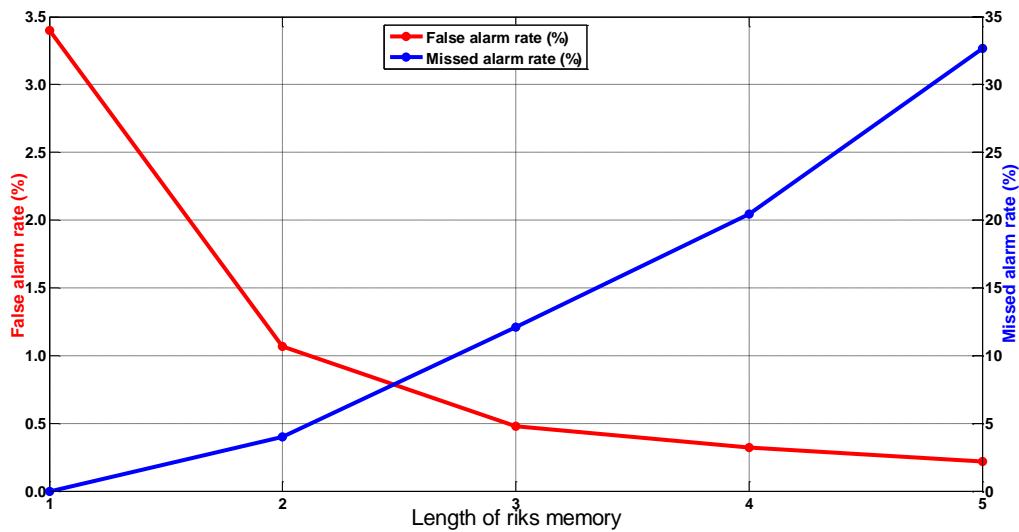


Fig 6.16 Tradeoff between false alarm rate and miss alarm rate

If the length of risk memory is fixed at 2, nearly 99 NTS are correctly predicted and more than 95% PTS are correctly predicted.

7 Conclusions

7.1 Summary

This work focuses on the development of risk indicators in real time road their test and assessment on motorway-like infrastructure (urban expressways or highways) based on individual data collected by stationary inductive loop detectors (ILDs). The initial goal was to capture real-time risk induced by potential front-rear collision. A set of indicators has been proposed and tested on real data; some of those indicators also take into account exogenous parameters such us weather conditions. This initial goal was supplemented by (i) a framework and procedures for fusion in real-time the set of indicators previously selected in order to derive a unique meta-indicator to be used as an activation variable of traffic management strategies and (ii) a search for risk indicators based on individual data type of vehicle trajectories (see annex 2) which becomes more and more available thanks to probe and floating car data (FCD). Based on trajectory-based data, the derived indicators as sought to be able to capture both rear-end collision side-wipe collisions.

A first extension of the proposed tools is to extend the coverage of risk assessment and its propagation. A piece of work towards this objective is performed (see annex 2) but need to be deepened. Thus, it is possible to highlight/visualise the risk propagation, similar to what is actually done by traffic simulation tools but from risk prospective. Furthermore, knowing the risk level associated with a traffic situation in real time, i.e. at each time period can distinguish risky situations vs. safer one. Note here, that is a situation is considered risky, this does not necessarily mean that there will be a crash, but the combination of ingredients making the probability of occurrence of a crash is very high and its impact on traffic conditions (level of service) may be dramatically high.

The ultimate goal of this work is the implementation of appropriate management and control actions, i.e. risk-oriented or risk-sensitive management strategies on a black spots or locations with high recurrent congestion, to lower the level of risk, and thus avoid some predictable crashes due to traffic configurations, or at least reduce their severity. This is in the line with the sustainable traffic management requirements. However, the use of such indicators has some limitations. First, they allow understanding the risk of collision between two vehicles, and not fail in capturing the risk associated with other type of crashes (e.g. crash involving a single vehicle). On the other hand, we have seen that these indicators and their thresholds are site-dependant. To insure the portability and universality of the proposed tools, an online calibration and tuning procedure based on the site specificities, traffic configurations, level of crashes and their severities is necessary. However, some field data may not be available, a solution then could be the use of microscopic simulation tool, first to calculate the indicator values, and secondly to test potentially different control strategies and find those that actually reduce road safety risk. However, at present, simulation models cannot be used for such purposes because they are not able to accurately simulate the risk due to the car following models which intentionally not allow collision between vehicles.

Finally, we would like here to insist that the risk-oriented control strategies should be limited to high risk zones in terms of geometry, traffic configurations and traffic inducted crashes. Indeed, such strategies would aim primarily at reducing the risk and traffic-dependent crashes, i.e. reducing the number of front-rear collisions or/and side-wipe collisions due to the lane changing, and we can assert that despite strategy implementation for road traffic control, crashes can still occur because of the strong correlation of the human (driver errors, lack of caution and non-respect of road regulation) or vehicle failure in the causes of crashes.

7.2 Applications

The developed models can be used to provide additional information to traffic operators.

The models can provide binary or multiple level outputs. Besides, with the current version, the following immediate applications can be considered in the context of the selected study site:

- 1-To provide directly to drivers the real-time traffic crash risk status (warning, alarm or both) via a real-time traffic information system, such as Variable Message Signs. This helps drivers to take the necessary precautions when they are within risky traffic conditions. Drivers have though no obligation to change their driving behavior.
- 2-To be used as a tool for evaluating traffic crash risks at the same road location before and after a change at the corresponding road section. The change can be, for instance, the opening/closure of a new on-ramp upstream or of a new off-ramp downstream. It is advisable that there should not be any infrastructural change at the location of the traffic detectors where required data is available the developed model can be applied for the evaluation.

7.3 Perspectives

Firstly, the obtained models can be improved by optimizing the design choices at each step of the methodology. Fig 7.1 presents the potential areas of improvement to optimize the obtained results. With the developed models, different design choices in methodology can be selected aiming to optimize performance.

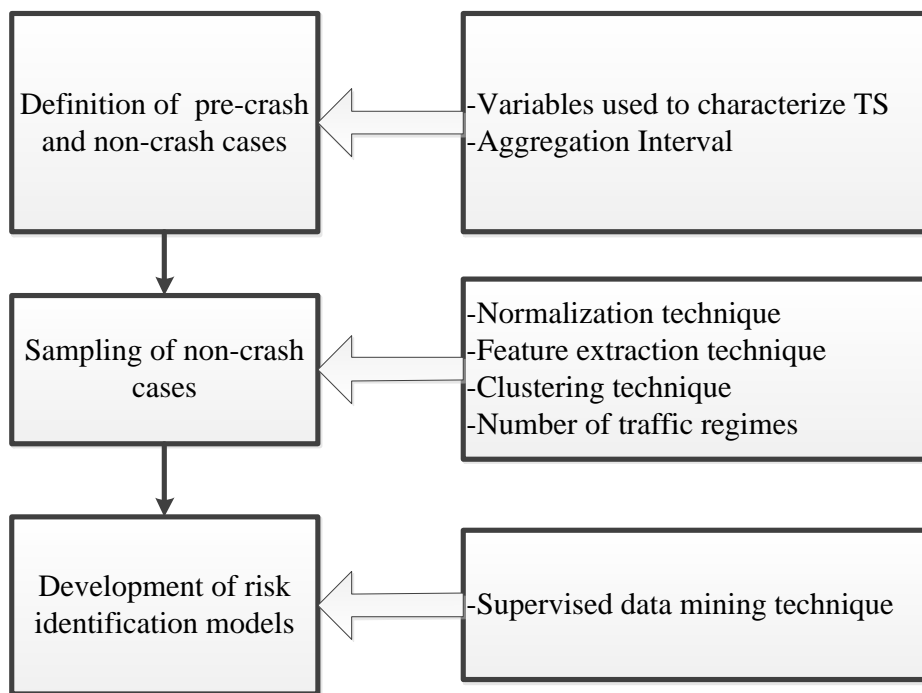


Fig 7.1 Potential improvement for optimizing performance

Several improvements can be implemented when defining NTS and PTS. Most of traffic related variables characterizing traffic situations in the current research are related to speed and flow. These variable choices facilitate the result interpretation as the variables are fundamental. The use of other variables such as risk indicators (for instance, PICUD, MARS, etc.) can be tested. The duration of the aggregation can also be optimized. Here, 5-minute intervals are used. However, this duration can be optimized with respect to the prediction performance. Similarly, in sampling non-crash cases and in developing risk identification models, the choices of normalization, feature extraction, clustering, and supervised data mining techniques can also be optimized.

Moreover, other study developments can be conducted based on the current study. The directions include:

- Extending to a larger study site/road network
 - Use of a series of detectors
 - Use of floating car data.
 - In both cases, additional data can be used to define traffic situations.
- Extending to other data types such as Vehicle-To-Vehicle - V2V data/Vehicle-To-Infrastructure - V2I data
- Developing algorithms to manage motorways with objective based on Risk Indicators

Acknowledgement

We acknowledge:

- 1-This study is financed by the Fund for Roads-Bridge-Tunnel Research, Federal Road Office (FEDRO).
- 2-Crash data in Switzerland used in this project is provided by the Federal Statistics Office (FSO).
- 3-Traffic data collected from Swiss motorways used in this project is provided by the Federal Road Office (FEDRO).
- 4-Meteorological data used in the current study are provided by MeteoSuisse and Boschung Systems.

Annexes

I	Généralités sur les indicateurs et sur le risque	109
I.1	La notion d'indicateur	109
I.2	Les indicateurs de trafic	110
I.3	Paramètres influençant la sécurité.....	112
I.4	Risque associé à une situation de trafic	115
II	Les indicateurs de risque en temps réel	118
II.1	Historique et état de l'art	118
II.2	Typologie des indicateurs de risque	125
II.3	Indicateur global de sécurité : I_{crash}	127
II.4	Homogénéité des vitesses et des densités	128
II.5	Différence relative de vitesse et de distance inter-véhiculaire.....	128
III	Propriétés des indicateurs de risque en temps réel	130
III.1	Relations entre indicateurs et paramètres relatifs au trafic.....	130
III.2	Comparaison des différents indicateurs.....	131
IV	Etablissement de seuils	133
IV.1	Tracés des distributions des valeurs des indicateurs	133
IV.2	Utilisation des algorithmes de détection automatique d'incidents	133
V	Estimation du risque à partir des données trajectoires	137
V.1	Présentation des données NGSIM	137
V.2	Etude du risque de collision avant-arrière	143
V.3	Etude du risque de collision latérale	150
	Références de l'annexe	153

I Généralités sur les indicateurs et sur le risque

I.1 La notion d'indicateur

I.1.1 Définition

Diverses définitions de la notion d'indicateur se trouvent dans la littérature. Selon l'Institut Français de l'Environnement, un indicateur est "quelque chose qui simplifie l'information en provenance de phénomènes complexes et qui la quantifie de manière à la rendre significative à l'échelle désirée" (Barraud, 2008). D'après la norme ISO 8402, c'est une "information choisie, associée à un phénomène, destinée à en observer périodiquement les évolutions au regard d'objectifs périodiquement définis" (Lucas, 2002). Enfin, selon Gras et al., (1989), "les indicateurs sont des variables [...] qui fournissent des renseignements sur d'autres variables plus difficiles d'accès".

Ainsi, un indicateur est une indication qui tente de refléter une situation donnée. Il peut s'agir d'une variable qualitative ou quantitative, selon la nature du phénomène à décrire. Certains indicateurs qualitatifs peuvent également s'exprimer sous forme de proportions ou de pourcentages, ce qui facilite la perception en ordre de grandeur des phénomènes observés. D'une manière générale, plus la réalité que l'on cherche à évaluer est complexe, plus on a besoin d'indicateurs différents. En effet, un indicateur donne une information utile mais incomplète.

L'existence d'indicateurs suppose l'existence d'une question ou de préoccupations qu'ils cherchent à éclairer. Ils sont censés fournir à la personne qui en dispose une information significative relativement à ses propres préoccupations. En effet, les indicateurs sont souvent des éléments importants dans l'aide à la décision. Ils permettent de faciliter la compréhension d'un système complexe, en aidant les utilisateurs à prendre des décisions appropriées au vu de l'atteinte de leurs objectifs. Ainsi, un indicateur donne une vision synthétique d'un aspect d'un système et permet de simplifier l'information. C'est un compromis entre les résultats scientifiques et la demande d'information simple et concise. (Mitchell et al., 1995)

Un indicateur répond à plusieurs besoins, et la mise en place d'indicateurs répond à plusieurs attentes. Un indicateur est donc à la fois :

- un instrument de mesure,
- un outil de communication, qui permet de simplifier l'information et de la transmettre à un auditoire particulier,
- un outil de suivi et d'évaluation, de prévision, qui peut par exemple permettre de mesurer l'efficacité d'un dispositif mis en place,
- un outil de comparaison et d'aide à la décision, qui peut permettre de comparer plusieurs solutions à un problème.

I.1.2 Propriétés

Chaque phénomène ou situation peut être décrit par un jeu d'indicateurs. Cet ensemble se doit de former une famille cohérente d'indicateurs, c'est-à-dire qu'elle possède les propriétés suivantes :

- exhaustivité : ils permettent de caractériser la situation dans son ensemble,
- non-redondance : il n'y a pas deux indicateurs qui expriment exactement la même chose,
- cohérence.

Par ailleurs, chaque indicateur doit être de bonne qualité. Il existe différentes manières d'évaluer la qualité d'un indicateur. Selon Maurizi et Verrel (2002), les critères de sélection d'un indicateur sont les suivants:

- qualité des données d'après lesquelles il est calculé, qui doivent être sûres et précises,
- éléments de faisabilité : les données qu'il nécessite doivent être accessibles, au niveau disponibilité et coût,
- pertinence vis-à-vis de l'objectif visé : il doit permettre une représentation spatiale et temporelle de la situation, et répondre effectivement à l'objectif visé,
- pertinence vis-à-vis des acteurs :
 - vis-à-vis des décideurs et des utilisateurs de l'indicateur, il doit définir un objet quantifié et rendre possible la comparaison entre différents scénarios,
- vis-à-vis du public :
 - simplicité,
 - lisibilité : il doit être compréhensible de façon immédiate par le lecteur,
 - correspondance avec les centres d'intérêt du public.

Labouze et al. (1995) donnent une autre liste de qualités qu'un bon indicateur doit posséder selon eux :

- accessibilité : capacité de l'indicateur à être calculé assez rapidement à un coût acceptable,
- fidélité : conservation lors de l'évaluation d'un biais éventuel à un niveau constant,
- objectivité : capacité à être évalué sans ambiguïté à partir de données observables,
- pertinence : capacité à refléter tous les aspects d'un phénomène et à garder sa signification dans le temps,
- précision/robustesse : définition de l'indicateur avec une marge d'erreur acceptable, en fonction de la précision des mesures sur les grandeurs observables,
- sensibilité : variations significatives de l'indicateur pour des variations assez faibles du phénomène,
- univocité : variation de l'indicateur de façon monotone par rapport au phénomène décrit pour pouvoir interpréter ses variations sans équivoque.

En résumé, un bon indicateur est un indicateur qui est sensible aux changements attendus, qui se calcule à partir de données fiables, précises et accessibles, et qui est facilement compréhensible par ses utilisateurs.

I.2 Les indicateurs de trafic

Le long des réseaux routiers se trouvent des capteurs de trafic permettant de recueillir les données de trafic. A partir de ces données sont ensuite construits des indicateurs de trafic, qui caractérisent la circulation sur chaque tronçon de route. La combinaison de plusieurs indicateurs permet de caractériser le mieux possible une situation de trafic.

De façon générale, l'état du trafic se caractérise principalement par le débit, qui représente le nombre de véhicules par unité de temps passant en un point donné, et par la densité, qui représente le nombre de véhicules par unité de longueur se trouvant sur une section donnée. Ces deux variables sont des indicateurs de trafic, mais de nombreux paramètres entrent en compte dans leur calcul, ainsi que dans le calcul d'autres indicateurs de trafic.

Suivant la classification proposée par Chung (2006), les indicateurs de trafic sont déterminés par des paramètres qui concernent aussi bien les conditions météorologiques, la demande, l'offre et l'interaction entre l'offre et la demande.

I.2.1 Influence des conditions météorologiques

Les paramètres relatifs aux conditions météorologiques sont directement observables ou mesurables, notamment à l'aide d'une station météorologique. Ils sont par exemple :

- température de l'air,
- température de surface de la chaussée,

- état de la chaussée (mouillée, sèche, glacée...),
- vitesse et direction du vent,
- taux d'humidité,
- type et hauteur des précipitations.

Les indicateurs de trafic liés aux conditions météorologiques sont ceux qui dépendent essentiellement de ces paramètres, comme par exemple la distance de freinage. Cependant, ces paramètres peuvent influencer d'autres indicateurs, comme la vitesse moyenne, qui est plus ou moins élevée suivant l'état de la chaussée.

La distance de freinage est un indicateur de trafic caractérisant l'état de circulation quand on la compare avec la distance intervéhiculaire. En effet, c'est la relation entre ces deux grandeurs qui permet d'analyser l'état de trafic. Ainsi, certains indicateurs de trafic sont complémentaires et ne peuvent s'interpréter les uns sans les autres.

I.2.2 Influence de la demande

Les paramètres relatifs à la demande sont principalement la composition du trafic et la proportion de poids lourds. Ils influent notamment sur la capacité, qui est le nombre maximal de véhicules pouvant s'écouler en un point par unité de temps, soit le débit maximal. Cependant, il n'existe pas d'indicateur de trafic exclusivement lié à la demande.

I.2.3 Influence de l'offre

Les paramètres liés à l'offre sont ceux qui se rapportent aux propriétés de la route :

- géométrie de la route,
- largeur,
- rayon de courbure,
- nombre de voies,
- type de revêtement,
- vitesse limite,
- type et durée des incidents ou restrictions.

L'indicateur relatif à l'offre est la capacité. Tous les paramètres cités ci-dessus ont une influence sur la capacité.

I.2.4 Relation entre l'offre et la demande

De nombreux indicateurs de trafic sont liés à l'interaction entre l'offre et la demande. Ce sont notamment ceux qui se calculent à partir des données fournies par les dispositifs de mesure situés sur les routes. Ces dispositifs peuvent être par exemple des boucles électromagnétiques, des caméras vidéo ou des capteurs à ultra-sons. Le système le plus utilisé sur les voies rapides est celui des boucles électromagnétiques. Ces boucles donnent, à chaque fois qu'un véhicule passe dessus, l'heure exacte, la vitesse du véhicule et sa longueur. A partir de ces données, on peut connaître différents indicateurs de trafic :

- débit,
- densité,
- temps intervéhiculaire,
- vitesse instantanée.

D'autres indicateurs peuvent également être calculés par le biais de différentes méthodes, comme le temps de parcours moyen ou la vitesse moyenne.

I.2.5 Synthèse

Les paramètres liés au trafic sont de différents types, et ils permettent le calcul d'indicateurs

de trafic, qui, ensemble, caractérisent la circulation. Chacun de ces indicateurs est déterminé par plusieurs paramètres. Ainsi, comme le montre le schéma de la figure 1, les interactions entre les paramètres et les indicateurs de trafic sont multiples et permettent une caractérisation complète de la circulation.

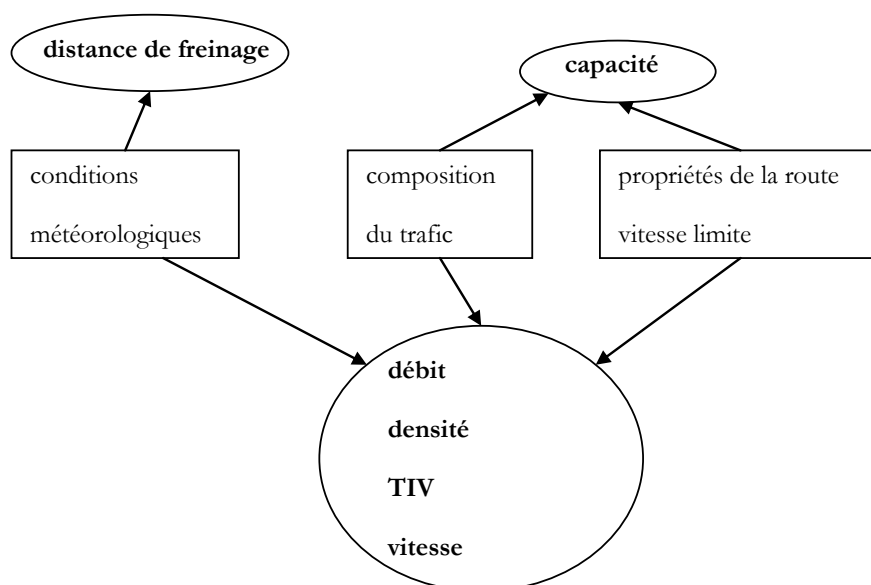


Fig. I.1 Relations entre paramètres et indicateurs de trafic

I.3 Paramètres influençant la sécurité

Certains indicateurs de trafic, ainsi que certains paramètres caractérisant le trafic, sont supposés ou connus pour avoir un impact direct sur la sécurité routière. Plusieurs recherches ont permis d'aboutir à une relation entre ces facteurs et la fréquence des accidents. Ces recherches consistent en une étude statistique des bases de données d'accidents, et sont donc une étude de risque a posteriori. Ce sont les relations entre les facteurs dont on cherche à connaître l'influence sur la sécurité et les accidents ayant eu lieu qui sont recherchées. Par exemple, dans le cas d'études de risque sur des voies rapides, les facteurs suivants peuvent être pris en compte en tant que paramètres dont les variations sont susceptibles d'augmenter ou de réduire le risque d'accident :

I.3.1 Localisation dans le temps

L'analyse de données accentogènes montre souvent une différence entre les taux d'accidents correspondant à la journée et ceux correspondant à la nuit. Ainsi, dans le projet européen INTRO (Chung et al., 2007), l'analyse des bases de données d'accidents correspondant aux autoroutes suisses du canton de Vaud durant quatre ans a permis d'aboutir à différentes conclusions relatives à la relation entre les accidents et le moment auquel ils ont lieu. Le taux d'accidents est ici exprimé en accidents par véhicules*kilomètres. Les résultats sont les suivants :

- Les mois les plus accidentogènes sont les mois de janvier, mars, avril, août et octobre. Ce sont des mois durant lesquels il y a des vacances scolaires, et donc beaucoup de véhicules particuliers sur les routes.
- Le taux d'accidents est plus élevé le week end, en particulier le dimanche.
- Le taux d'accidents est 3 à 5 fois plus élevé entre 1 heure et 5 heures du matin que le reste de la journée.

I.3.2 Débit, densité et taux de congestion

Différents chercheurs ont démontré l'existence d'une relation entre la fréquence des accidents et le débit et la densité du trafic. Par exemple, Lord et al. (2005) ont étudié le rapport entre le type de l'accident (selon qu'il implique un ou plusieurs véhicules) et le débit horaire de la voie. Leurs recherches montrent la nécessité de prendre en compte non seulement le débit et la densité, mais aussi le rapport débit/capacité. Ils ont obtenu les conclusions suivantes :

- Le nombre total d'accidents augmente quand le débit augmente, mais cette augmentation se fait à un taux décroissant.
- Le nombre d'accidents en fonction de la densité suit une courbe de la forme de celle de la figure 2.
- Le graphique représentant le nombre d'accidents total en fonction du ratio débit/capacité a la même forme que celui en fonction de la densité.
- Si on ne s'intéresse qu'aux accidents impliquant plusieurs véhicules, leur nombre augmente quand la densité augmente ou quand le ratio débit/capacité augmente.

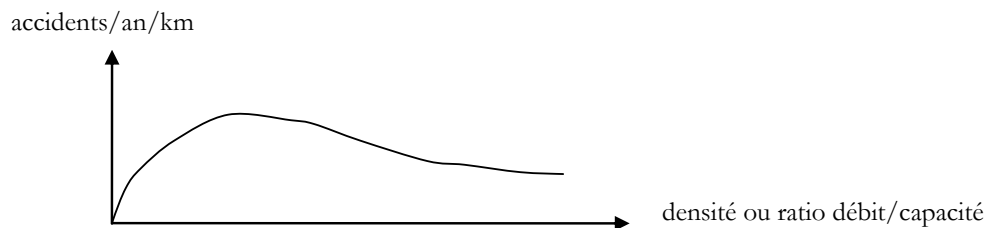


Fig. 1.2 Taux d'accidents en fonction de la densité ou du ratio débit/capacité

I.3.3 Vitesse

De nombreuses recherches ont mis en avant les effets négatifs de l'augmentation de la vitesse sur la sécurité routière. La relation entre vitesse moyenne, accidents corporels graves et nombre de tués a été modélisée par de nombreux chercheurs. Par exemple, Nilsson (2004) a mis en place un modèle appelé modèle puissance, qui estime les effets du changement de la vitesse moyenne sur le nombre d'accidents mortels, le nombre d'accidents graves et mortels, et le nombre total d'accidents corporels (figure 3).

D'après ce modèle, une augmentation de 5% de la vitesse moyenne entraîne une hausse d'environ 10% du nombre total d'accidents corporels et de 20% du nombre d'accidents mortels. A l'inverse, une diminution de 5% de la vitesse moyenne entraîne une baisse d'environ 10% du nombre total d'accidents et de 20% du nombre d'accidents mortels.

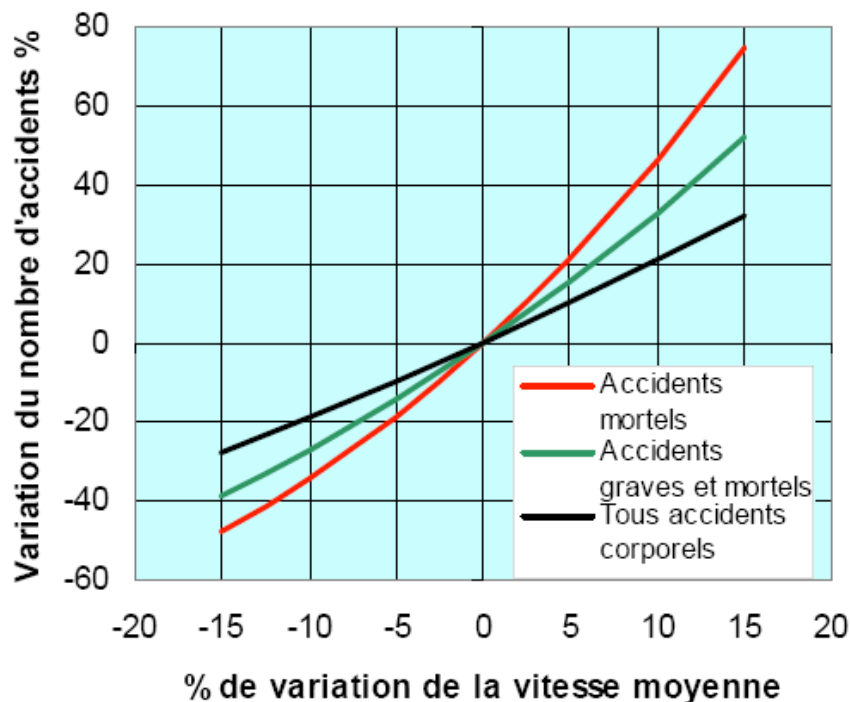


Fig. 1.3 Modèle de Nilsson (source : OCDE, 2007)

Cette représentation, comme tout modèle, est une représentation simplifiée de la réalité. Ainsi, certains paramètres ne sont pas pris en compte. Par exemple, les conséquences d'une variation des vitesses moyennes sur le taux d'accidents sont beaucoup plus importantes en agglomération que sur autoroute, et ceci n'est pas visible dans le modèle de Nilsson. Ce modèle montre néanmoins l'existence d'un lien entre vitesse moyenne et taux d'accidents.

En revanche, d'autres études comme celle de Garber et Gadiraju (1992) arrivent à la conclusion qu'une augmentation de la vitesse limite des véhicules légers ne conduit pas à une variation significative du taux d'accidents. Dans une autre étude qu'ils publient en 1989, ils montrent que le taux d'accidents augmente quand la variance des vitesses pratiquées augmente, mais qu'il ne varie pas en fonction de la vitesse moyenne.

D'autres modèles établissent une relation entre la variance des vitesses (ou l'écart-type) et le taux d'accidents. Par exemple, le modèle de Salusjärvi (cf Janssen, 2008) se base sur des observations empiriques pour établir un lien entre la variation de l'écart-type des vitesses et la variation du taux d'accidents (figure 4). D'après ce modèle, une réduction de l'écart-type des vitesses pratiquées de 10 km/h entraîne une baisse de plus de 50% du nombre d'accidents. Ainsi, plus les vitesses sont homogènes, plus le risque d'accident est faible.

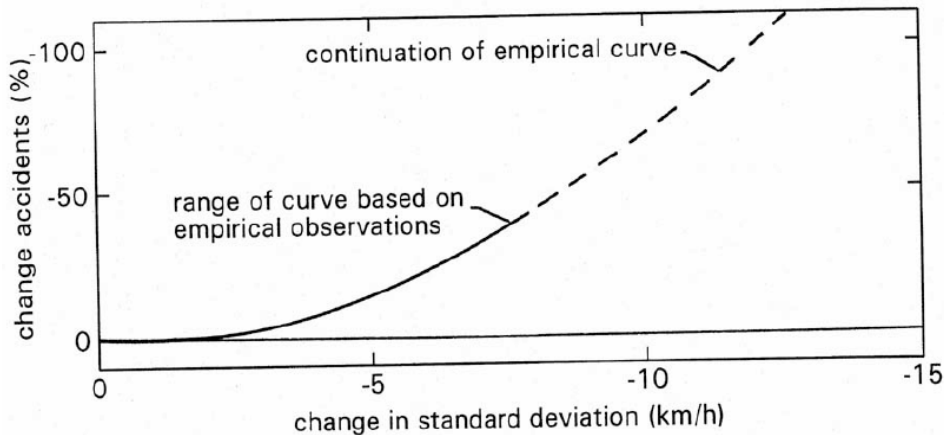


Fig. I.4 Modèle de Salusjärvi (source : Janssen, 2008)

I.3.4 Conditions météorologiques

Dans plusieurs études, l'effet de la pluie ou de mauvaises conditions météorologiques sur le risque d'accident est mis en avant. Ainsi, les études de Keay et Simmonds (2005) sur la ville de Melbourne (Australie) ont montré qu'un temps pluvieux impliquait un risque d'accident plus élevé qu'un temps sec, et que plus il y a de pluie, plus ce risque augmente, même si l'effet de la pluie est différent suivant la période de la journée considérée.

Par ailleurs, d'après les résultats du projet INTRO, sur les autoroutes suisses, les accidents sont trois fois plus nombreuses par temps de pluie que par temps sec, et deux fois plus nombreux sur route enneigée que sur route sèche. Ceci peut venir du fait qu'en Suisse, les limites de vitesses sont les mêmes quel que soit le temps.

I.3.5 Géométrie de la route

Les statistiques montrent que la présence de géométries particulières sur la route, telles que des tunnels ou des virages, peut augmenter le risque d'accidents. Par exemple, des études du CETE sur les accidents en courbe (Dupré, 2005) utilisent des données tirées de la base BAAC selon lesquelles 18% du total des accidents et 33% des accidents mortels ont lieu dans un virage.

Dans le projet INTRO, trois facteurs principaux sont identifiés comme entraînant un taux d'accidents plus élevé que la moyenne : la nuit, les virages et la pluie. La présence d'un de ces trois facteurs fait augmenter le taux d'accidents, et la combinaison de plusieurs de ces facteurs entraîne un taux d'accidents encore plus élevé.

I.4 Risque associé à une situation de trafic

I.4.1 La notion de risque

D'après le dictionnaire Petit Larousse (1997), le risque est un "danger, inconvénient plus ou moins probable auquel on est exposé". D'après le Petit Robert (1996), c'est un "danger éventuel plus ou moins prévisible". Il y a une idée de potentialité et d'incertitude dans ces définitions. Ainsi, le risque d'accident caractérise la probabilité qu'il y ait un accident. Plus cette probabilité est élevée, plus le risque est fort.

La définition scientifique du risque est qu'il résulte de la superposition d'un aléa et d'un enjeu. L'aléa se rapporte à la probabilité d'occurrence d'un événement, d'origine naturelle ou humaine, ainsi que par l'intensité de sa manifestation. L'enjeu correspond aux dommages directs dus à l'événement si celui-ci se produisait, ces dommages pouvant être matériels ou humains, voire économiques et sociaux. Dans le cas du risque d'un accident

de la route, l'enjeu dépend essentiellement de la gravité de l'accident, et donc de l'intensité de l'aléa. Le risque correspond donc à une éventualité, et une situation peut être jugée comme étant très risquée sans nécessairement donner lieu à un accident.

I.4.2 Risque en temps réel

Selon le moment auquel il est évalué, le risque peut être de deux types. Il peut être calculé a posteriori, en analysant les bases de données d'accidents en relation avec les conditions de trafic. Ce type d'analyse permet d'identifier des facteurs augmentant le risque d'accidents, et donc certaines situations de trafic durant lesquelles la probabilité d'avoir un accident est plus élevée que dans d'autres situations. C'est notamment de cette façon qu'ont été mis en évidence les facteurs influençant la sécurité dont nous avons parlé précédemment.

Mais le risque peut également être calculé en temps réel, c'est-à-dire a priori. Pour cela, des indicateurs appelés indicateurs de risque en temps réel sont définis. Ils se calculent à partir des données de trafic recueillies en temps réel. Ainsi, à chaque instant, les données de trafic caractérisent une situation de trafic particulière, à laquelle peut être associé un risque routier particulier. La façon d'évaluer le risque et les indicateurs utilisés pour cela diffèrent suivant le type de risque que l'on cherche à appréhender et le type de données dont on dispose.

Le fait d'évaluer le risque en temps réel, et donc a priori, permet de pouvoir mettre en œuvre des actions permettant de minimiser le risque avant que l'accident qui risquait de se produire ne puisse se produire. Ainsi, l'utilisation d'indicateurs de risque en temps réel pourrait permettre d'éviter certains accidents ou au moins de diminuer leur gravité, et donc d'améliorer la sécurité routière.

Un indicateur de risque en temps réel est donc une information chiffrée calculée à partir des données de trafic recueillies en temps réel. Il varie significativement en fonction du risque, ce qui signifie qu'une valeur élevée de l'indicateur traduit une situation de trafic à fort risque d'accident, même si ce risque ne se traduit pas nécessairement par un accident.

I.4.3 Différents types de risque routier

L'existence de différents types d'accidents entraîne l'existence de différents types de risque routier. D'abord, un accident peut résulter de l'interaction entre un véhicule et l'environnement (un obstacle fixe par exemple), ou de l'interaction entre plusieurs véhicules. Les indicateurs de risque que nous allons présenter dans la suite de ce document permettent uniquement de détecter le risque dans le cas d'accidents impliquant plusieurs véhicules. Ils ne s'appliquent pas au risque de collision entre un véhicule unique et un obstacle.

Il existe deux principaux types d'accidents mettant en cause plusieurs véhicules, ce qui implique l'existence de deux types de risque :

- risque de collision avant-arrière entre deux véhicules qui se suivent,
- risque de collision latérale entre deux véhicules roulant l'un à côté de l'autre.

La suite de ce document s'intéresse à l'étude de ces deux types de risque sur voie rapide. En France, 14% des accidents corporels sur routes nationales et départementales, et 40% des accidents sur autoroute sont des collisions avant-arrière (Aron et al., 2004). Ainsi, au vu du nombre des accidents de ce type, il est important de s'intéresser au risque de collision avant-arrière sur voie rapide. Un autre exemple est celui des autoroutes suisses, dont la répartition des accidents suivant leur type est montrée sur la figure 5. Les collisions avant-arrière et latérales représentent 37% des accidents sur les autoroutes suisses du canton de Vaud (Chung et al., 2007).

Répartition des accidents par type

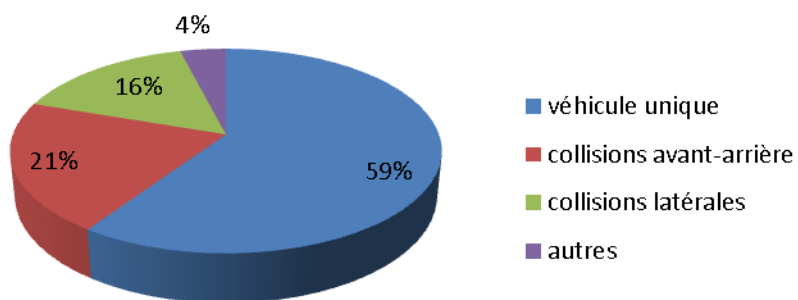


Fig. I.5 Répartition des types d'accidents sur les autoroutes suisses

II Les indicateurs de risque en temps réel

Depuis les années 60, des chercheurs de différents pays ont cherché à créer des indicateurs qui caractérisent le risque routier associé à une situation de trafic particulière à chaque instant. Ce sont les indicateurs de risque en temps réel. Le but de leur mise en œuvre est de pouvoir prévoir les situations à risque, et donc éventuellement de mettre en œuvre des actions spécifiques adaptées au risque encouru afin d'améliorer la sécurité routière.

Selon de Mouzon et al. (2008), un indicateur de risque est de bonne qualité si l'on peut lui attribuer un seuil de déclenchement d'une alarme caractérisé par :

- une bonne sensibilité, ce qui signifie que le seuil n'est atteint qu'un faible pourcentage du temps, car sinon l'alarme perdrait en crédibilité,
- une bonne précision, c'est-à-dire que, pour le plus d'accidents possibles, le seuil était effectivement dépassé peu de temps avant l'accident.

Ainsi, les indicateurs de risque sont liés statistiquement avec l'occurrence des accidents. Mais il faut également que le minimum de situations possible donne lieu à une valeur de l'indicateur assez élevée pour caractériser la situation comme étant risquée.

II.1 Historique et état de l'art

Les indicateurs de risque sont apparus pour la première fois dans le cadre d'une technique appelée "Traffic Conflict Technique" (TCT) initiée en 1967 par Perkins et Harris, du laboratoire General Motors à Detroit, cités par Archer (2005). Ils ont introduit un premier indicateur de risque nommé Time-to-Accident (TA), qui représente le temps au bout duquel deux véhicules donnés qui se suivent entreraient en collision s'ils maintenaient leur vitesse et leur trajectoire.

Depuis la fin des années 60, cet indicateur a été amélioré et d'autres indicateurs permettant de mesurer le risque ont pu être définis.

II.1.1 Time-to-Collision (TTC) et ses dérivés

Time-to-Collision (TTC)

Le TTC est l'indicateur de risque le plus utilisé, depuis son introduction en 1972 par Hayward, cité par Archer (2005). Comme le TA, il calcule le temps au bout duquel il y aurait un accident si les véhicules considérés gardaient la même vitesse et la même direction. Il existe différentes façons de calculer le TTC, dont certaines prennent en compte le temps de réaction des usagers de la route.

Cet indicateur est adapté à une estimation du niveau de risque en milieu urbain, notamment aux intersections, mais il peut aussi être utilisé pour caractériser le risque correspondant aux collisions avant-arrière sur voie rapide. Dans ce cas, il suffit de connaître la distance entre les deux véhicules et la vitesse de chacun pour mesurer le TTC. Ainsi, une boucle électromagnétique peut permettre le calcul du TTC.

La figure 6 illustre le calcul classique du TTC dans le cas d'un risque de collision avant-arrière. On cherche à calculer le TTC correspondant au véhicule i , à l'instant auquel ce véhicule passe sur le capteur. Sa vitesse est notée v_i , et la vitesse du véhicule le précédant v_{i-1} (elle est inférieure à v_i sinon le TTC n'existe pas). Enfin, g_{i-1} représente la différence entre les instants de passage sur le capteur des véhicules i et $(i-1)$. Le choc a lieu entre l'arrière du véhicule $(i-1)$ et l'avant du véhicule i . Leurs longueurs sont notées respectivement l_{i-1} et l_i .

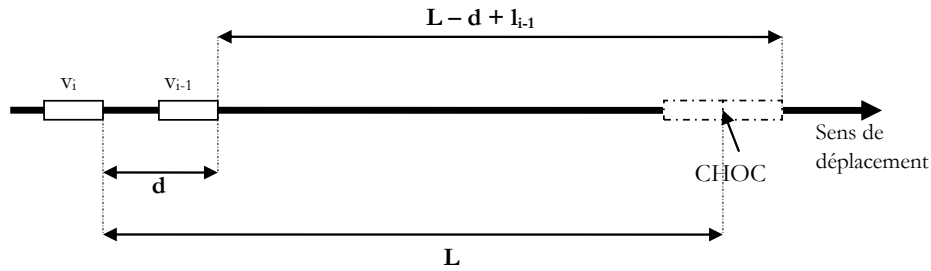


Fig. II.6 Calcul du TTC

$$d = gap_i \times v_{i-1} ; L = TTC_i \times v_i ; L - d + l_{i-1} = TTC_i \times v_{i-1}$$

Ces relations donnent la formule de calcul du TTC :

$$TTC_i = \frac{gap_i \times v_{i-1} - l_{i-1}}{v_i - v_{i-1}}$$

Par exemple, si deux véhicules se suivent, celui de devant se déplaçant à 120 km/h, celui de derrière à 130 km/h, que la différence entre leurs instants de passage sur le capteur est de 1 seconde, et que la longueur du premier véhicule est de 4 mètres, le TTC est de 10.56 secondes. Plus le TTC est faible, plus la situation est risquée.

Post-Encroachment Time (PET)

Le PET est une variante du TTC. Il a été introduit en 1983 par Cooper, et la façon de le calculer est détaillée par exemple dans Archer (2005). La principale différence entre le PET et le TTC est que le calcul du TTC suppose que si les véhicules considérés gardent la même vitesse et la même trajectoire, alors ils vont entrer en collision, alors que le calcul du PET suppose qu'il existe une zone de conflit mais que les véhicules ne s'y trouveront pas au même moment s'ils maintiennent leurs vitesse et trajectoire. Ainsi, dans les situations où on peut calculer un TTC, on ne peut pas calculer un PET, et vice versa.

Le PET est un indicateur particulièrement adapté pour les intersections, et donc pour le milieu urbain. On définit une zone de conflit (en l'occurrence le lieu de l'intersection), et on considère deux véhicules allant traverser cette zone de conflit et risquer d'entrer en collision. Le PET est la différence entre le temps qu'il y aura entre le moment où le premier véhicule quittera la zone de conflit et le moment où le second véhicule y entrera. Plus le PET est faible, plus la situation est risquée.

Fig. II.3 illustre le calcul du PET dans le cas d'une intersection, vers laquelle se dirigent deux véhicules, de vitesses v_1 et v_2 , de longueurs l_1 et l_2 et dont les distances à l'intersection sont d_1 et d_2 .

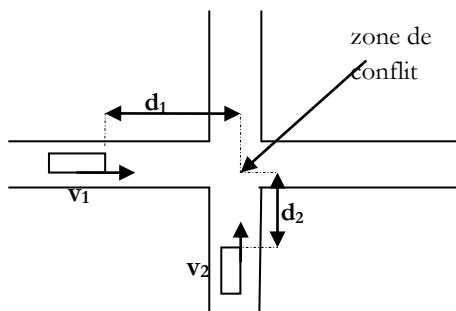


Fig. II.7 Calcul du PET

Supposons que le véhicule 1 arrive au niveau de la zone de conflit en premier. Il la quitte au bout d'un temps $(d_1 + l_1)/v_1$. Le véhicule 2 y arrive au bout d'un temps d_2/v_2 . On a donc

:

$$PET = \frac{d_2}{v_2} - \frac{d_1 + l_1}{v_1}$$

L'hypothèse selon laquelle le véhicule 1 quitte la zone de conflit avant que le véhicule 2 y arrive revient à dire que $\frac{d_2}{v_2} > \frac{d_1 + l_1}{v_1}$.

Time Exposed Time-to-Collision (TET) et Time Integrated Time-to-Collision (TIT)

Minderhoud et Bovy (2001) ont introduit deux nouveaux indicateurs dérivés du TTC. Ces indicateurs se basent sur la définition préalable d'une valeur critique du TTC, notée TTC^* , au-dessous de laquelle la situation est considérée comme risquée.

Le TET et le TIT se calculent dans une période donnée, entre t_1 et t_2 , et dans un espace donné, entre x_1 et x_2 . Pour chaque véhicule i se trouvant entre x_1 et x_2 entre t_1 et t_2 , on calcule son TTC_i à chaque instant, puis on le trace en fonction du temps, sachant que le TTC n'existe que lorsqu'il y a un véhicule devant le véhicule i et que celui-ci a une vitesse moins élevée que celle du véhicule i .

Le TET_i est le temps durant lequel TTC_i est inférieur à la valeur critique TTC^* . On peut ensuite faire la somme des TET_i sur tous les véhicules présent pour caractériser le risque global de la situation. Plus le TET est faible, moins la situation est risquée.

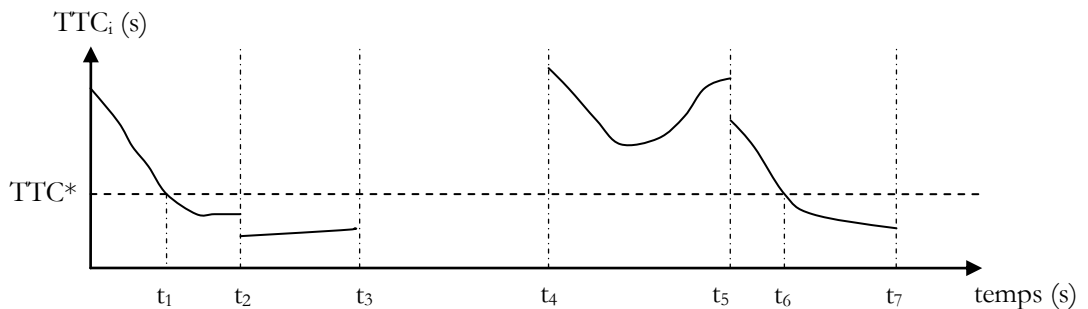


Fig. II.8 Calcul du TET

Fig. II.4 montre un exemple de variation du TTC en fonction du temps pour un véhicule particulier. Les moments où le TTC n'est pas calculé (comme entre t_3 et t_4) correspondent aux instants durant lesquels le véhicule considéré se déplace moins rapidement que celui qui le précède. Les discontinuités dans la courbe représentant le TTC en fonction du temps correspondent à une des situations suivantes :

- Le véhicule considéré change de file.
- Le véhicule le précédant change de file.
- Un véhicule s'insère devant le véhicule considéré.

Les changements de pente dans les parties continues de la courbe correspondent à des variations de vitesse de la part du véhicule considéré ou de celui le précédant.

Dans l'exemple de Fig. II.4, le TET vaut :

$$TET_i = t_3 - t_1 + t_7 - t_6$$

L'inconvénient du TET est qu'il ne tient pas compte de la valeur du TTC mais seulement du fait que celle-ci soit au-dessous ou au-dessus d'une valeur critique particulière. Ainsi, si par exemple le TTC est très faible mais durant un temps très court, la valeur du TET ne sera pas représentative du risque relatif à la situation.

Le TIT sert à remédier à ce problème, car pour chaque véhicule i , TIT_i est la somme des différences entre le TTC_i et TTC^* pour les moments où TTC_i est inférieur à TTC^* . Dans le cas d'une représentation continue,

$$TIT_i = \int_{t_1}^{t_2} (TTC^* - TTC_i(t)) \delta(t) dt$$

où $\delta(t) = 1$ si $TTC_i < TTC^*$,

0 sinon.

Le TIT est un indicateur qui s'exprime en secondes carrées, et plus il est faible, moins la situation est risquée.

De nombreux chercheurs ont défini d'autres indicateurs dérivés du TTC, comme par exemple le DST (Deceleration to Safety Time) introduit par Hupfer (1997), qui caractérise l'importance de l'action que doit faire un conducteur pour éviter une collision, c'est-à-dire que la valeur du DST sera élevée et donc correspondra à une situation risquée si un freinage d'urgence est nécessaire, alors qu'elle sera faible s'il suffit que le conducteur lève le pied de la pédale d'accélérateur.

Bien que tous ces indicateurs puissent être adaptés à l'étude du risque sur voie rapide, ils s'utilisent principalement en milieu urbain, et particulièrement aux intersections.

II.1.2 Indicateurs caractéristiques du risque de collision latérale

Chang et Kao (1991) définissent un indicateur de trafic lié aux changements de voie : l'AFR (Average Flow Ratio). Pour chaque voie i , $AFR_i(t)$ caractérise les changements de voie en provenance de la voie i durant une période de durée t , en générale prise égale à cinq minutes. Il se calcule de la manière suivante :

$$AFR_i(t) = \frac{\frac{q_i(t)}{q_{i-1}(t)} \times NL_{i,i-1}(t) + \frac{q_i(t)}{q_{i+1}(t)} \times NL_{i,i+1}(t)}{NL_{i,i-1}(t) + NL_{i,i+1}(t)}$$

où : les voies $i-1$ et $i+1$ sont les voies adjacentes à i si elles existent

$q_j(t)$ est le débit moyen sur la voie j durant t

$NL_{i,j}(t)$ est le nombre de changements de voie de la voie i vers la voie j durant t

Chang et Kao observent que lorsque l'AFR augmente, la fréquence des changements de voie augmente également.

Lee et al. (2006) s'intéressent à l'étude des collisions latérales. Ils considèrent des données recueillies sur une autoroute à Orlando (Floride, Etats-Unis) pendant quatre ans, durant lesquels ils distinguent en particulier 725 collisions avant-arrière et 220 collisions latérales. Pour expliquer ces dernières, ils tentent de transformer l'indicateur AFR défini précédemment. Tout d'abord, ils ont ressenti le besoin de modifier sa définition, car ils ont remarqué que les collisions latérales ayant lieu sur la voie i étaient généralement dues à un changement de voie vers la voie i , et non en provenance de cette voie. La nouvelle définition de $AFR_i(t)$ prend donc en compte les changements de voie à partir des voies $(i-1)$ et $(i+1)$, ce qui fait qu'il convient d'introduire également les changements de voie à partir de la voie $(i-1)$ vers la voie $(i-2)$ et à partir de la voie $(i+1)$ vers la voie $(i+2)$, si celles-ci existent. On a alors :

$$AFR_i(t) = \frac{q_{i-1}(t)}{q_i(t)} \times \frac{NL_{i-1,i}(t)}{NL_{i-1,i}(t) + NL_{i-1,i-2}(t)} + \frac{q_{i+1}(t)}{q_i(t)} \times \frac{NL_{i+1,i}(t)}{NL_{i+1,i}(t) + NL_{i+1,i+2}(t)}$$

Sur une route à n voies, Overall Average Flow Ratio (OAFR) est la moyenne géométrique

des AFR_i.

$$OAFR(t) = \left[\prod_{i=1}^n AFR_i(t) \right]^{\frac{1}{n}}$$

L'étude des collisions ayant eu lieu sur l'autoroute considérée à Orlando commence par le tracé du diagramme fondamental expérimental. Les points correspondants à des collisions sont alors placés dans le graphique, et on remarque que la plupart des collisions latérales se trouvent dans la partie fluide du graphique, alors que les collisions avant arrière sont réparties assez uniformément dans la partie fluide et dans la partie congestionnée. Ainsi, la fluidité du trafic peut être un indicateur caractéristique d'un risque de collision latérale.

Les collisions avant-arrière et latérales sont ensuite comparées par le biais de sept indicateurs : OAFR, vitesse moyenne, coefficient de variation en vitesse (CVS), débit moyen, coefficient de variation en débit (CVF), taux d'occupation moyen, et un indicateur valant 1 en période de pointe et 0 sinon. Trois de ces facteurs sont trouvés significativement représentatifs des collisions latérales :

- OAFR : plus il est élevé, plus le risque de collision latérale est élevé,
- CVF : plus il est élevé, plus le risque de collision latérale est élevé,
- période de pointe ou pas : le risque de collision latérale est moins élevé en période de pointe.

II.1.3 Distance de freinage et Potential Index for Collision with Urgent Deceleration (PICUD)

Définition du PICUD

Uno et al. (2005) étudient un croisement d'autoroutes au Japon, où certains véhicules sont obligés de changer de voie sur une courte distance pour rejoindre leur destination, comme les véhicules venant de la zone D et se dirigeant vers la zone A sur la figure 9.

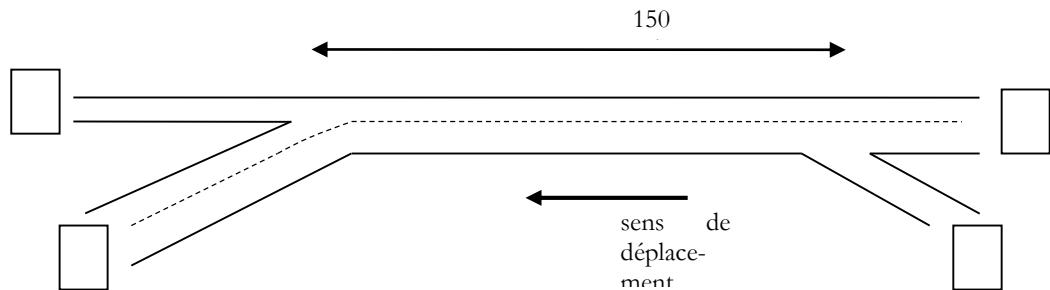


Fig. II.9 Croisement d'autoroutes avec changement de voie obligatoire

Lorsqu'un véhicule s'insère entre deux autres, il y a risque de collision. Si la vitesse du véhicule de devant est plus élevée que celle du véhicule suiveur mais que la distance entre eux est très courte, le TTC n'est pas calculable, et pourtant le danger existe. Pour pallier cet inconvénient du TTC, un nouvel indicateur de risque est défini : le PICUD.

Le PICUD représente la distance finale qu'il y aurait entre les deux véhicules considérés si celui de devant freinait brusquement, et que celui de derrière l'imitait après un certain temps de réaction. La mesure du PICUD nécessite la connaissance de la décélération γ (en m/s²) en tout instant et du temps de réaction t_r .

Calcul du PICUD à décélération constante

On suppose dans un premier temps que la décélération est constante, et on note γ sa valeur absolue.

On considère un véhicule de vitesse v_0 , commençant à freiner à $t=0$. Sa vitesse décroît

alors à la vitesse γ ($v(t)=v_0-\gamma \times t$), et il s'immobilise à . La distance qu'il a parcourue durant ce temps t_0 s'écrit alors $\int_0^{t_0}(v_0 - \gamma \times t)dt = v_0 \times t_0 - \gamma \times \frac{t_0^2}{2} = \frac{v_0^2}{2 \times \gamma}$

En notant $(x_{i-1}-x_i)$ la distance initiale entre les deux véhicules, v_{i-1} la vitesse du premier et v_i celle du deuxième, on peut calculer le PICUD correspondant au véhicule i :

$$PICUD_i = x_{i-1} - x_i + \frac{v_{i-1}^2}{2\gamma} - t_r \times v_i - \frac{v_i^2}{2\gamma}$$

Le PICUD peut prendre des valeurs positives ou négatives, et plus il est faible, plus il y a risque de collision. On considère qu'une valeur du PICUD positive entraîne une absence de risque.

Calcul du PICUD à décélération variable

Dans le projet européen INTRO (Chung et al., 2007), la décélération au moment d'un freinage d'urgence n'est pas considérée comme constante, mais elle augmente d'abord linéairement durant un temps de transition t_s , de 0 à sa valeur maximale. Des études du LCPC ont montré que le temps t_s est égal à $0.0022 \times v$, où v est la vitesse en km/h, et t_s est exprimée en secondes.

On note γ_{max} la valeur absolue de la décélération maximale, et $t=0$ est l'instant de début du freinage d'urgence.

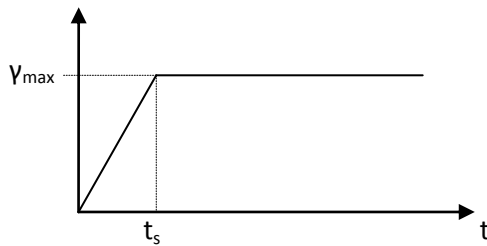


Fig. II.10 Evolution de la décélération lors d'un freinage d'urgence

A un instant t , la vitesse s'écrit $v(t) = v_0 - \int_0^t \gamma(u)du$, où v_0 est la vitesse à $t=0$.

$$\text{Si } 0 \leq t \leq t_s \quad v(t) = v_0 - \frac{\gamma_{max}}{t_s} \times \frac{t^2}{2}$$

$$\text{Si } t \geq t_s \quad v(t) = v_0 + \frac{\gamma_{max} \times t_s}{2} - \gamma_{max} \times t$$

Le temps de freinage $t_{freinage}$ est tel que $v(t_{freinage})=0$, donc $t_{freinage} = \frac{v_0}{\gamma_{max}} + \frac{t_s}{2}$

La distance de freinage vaut $\int_0^{t_{freinage}} v(t)dt$, ce qui donne :

$$d_{freinage} = \frac{v_0 \times t_s}{2} + \frac{v_0^2}{2 \times \gamma_{max}} - \frac{\gamma_{max} \times t_s^2}{24}$$

Nous avons vu que la valeur de t_s dépendait linéairement de la valeur de v_i . Nous pouvons donc introduire la constante $a_s = \frac{t_s(i)}{v_i} = \frac{t_s(i+1)}{v_{i+1}}$.

La distance d'arrêt du véhicule i est alors $d_{freinage} = v_i^2 \left(\frac{1}{2 \times \gamma_{max}} + \frac{a_s}{2} - \frac{\gamma_{max} \times a_s^2}{24} \right)$.

En tenant compte du temps t_s mis pour atteindre la décélération maximale, le PICUD s'exprime de la manière suivante :

$$PICUD_i = x_{i-1} - x_i + (v_{i-1}^2 - v_i^2) \left(\frac{1}{2 \times \gamma_{max}} + \frac{a_s}{2} - \frac{\gamma_{max} \times a_s^2}{24} \right) - v_i \times t_r$$

Enfin, on peut introduire la longueur des véhicules dans le calcul du PICUD, car la distance qui restera entre les deux véhicules après les deux freinages d'urgence est celle entre l'arrière du véhicule (i-1) et l'avant du véhicule i. Si le dispositif de comptage relève la position de l'avant des véhicules, il faut alors remplacer $(x_{i-1} - x_i)$ par $(x_{i-1} - x_i - l_{i-1})$, où l_{i-1} est la longueur du véhicule (i-1).

Taux de friction

Les études du LCPC et du projet INTRO ont montré que la valeur de γ_{max} dépend essentiellement de l'état (mouillé ou sec) de la chaussée. Les valeurs utilisées dans les projets INTRO et FUSAIN, et déduites d'études du LCPC, sont $\gamma_{max}^{sec} = 6.87 \text{ m/s}^2$ et $\gamma_{max}^{mouillé} = 4.81 \text{ m/s}^2$.

La distinction entre les situations pour lesquelles la chaussée est sèche et celles pour lesquelles elle est mouillée peut se réduire à l'introduction d'un taux de friction α défini de la manière suivante :

$$\alpha = \gamma / \gamma_{max}$$

où γ_{max} est la décélération maximale absolue, c'est-à-dire sur chaussée sèche. Il suffit ensuite d'utiliser $(\alpha \times \gamma_{max})$ à la place de γ_{max} , et de fixer $\gamma_{max} = 6.87 \text{ m/s}^2$ quelles que soient les conditions météorologiques.

L'utilisation du taux de friction permet une meilleure lisibilité des formules, mais également de tenir compte de toutes les variations possibles de la décélération. En effet, celle-ci peut dépendre non seulement des conditions météorologiques, mais aussi du type de revêtement ou encore de la géométrie de la route.

Dans le cas simple où seul l'état de la chaussée influe sur la décélération, et où on distingue uniquement deux états possibles (sec ou mouillé), on aura :

$$\alpha = \begin{cases} 1 & \text{si la chaussée est mouillée,} \\ 0.7 & \text{si la chaussée est sèche.} \end{cases}$$

II.1.4 Individual Braking Time Risk (IBTR) et ses dérivés

Individual Braking Time Risk (IBTR)

L'IBTR et ses dérivés sont des indicateurs de risque utilisés dans le projet INTRO. Chung et al. (2007) les décrit, en rappelant la définition initiale de Aron en 2003 :

Pour un véhicule i circulant avec une vitesse v_i et un temps intervéhiculaire gap_i sous des conditions caractérisées par un taux de friction α ,

$$IBTR_i = \max(0, \log_2 \left(\frac{v_i / \gamma_{max}}{2 \times gap_i \times \alpha} \right))$$

Un IBTR nul correspond à une situation sûre, alors qu'un IBTR positif correspond à une situation risquée, qui est d'autant plus risquée que l'IBTR est élevé.

Nous avons vu que $\frac{v_i}{\gamma_{max} \times \alpha}$ est le temps minimum nécessaire au véhicule i pour s'arrêter.

Si ce ratio est supérieur à gap_i , le véhicule i n'a pas assez de temps pour s'arrêter sans risquer d'entrer en collision avec le véhicule situé devant lui. Cependant, comme le véhicule i-1 circule à une certaine vitesse, on peut se permettre d'utiliser $2 \times gap_i$ au lieu de gap_i , ce qui permet de minimiser le nombre de situations considérées comme risquées.

En réalité, le facteur 2 introduit ici est adapté à une route où la vitesse maximale autorisée est de $v_{lim}=90$ km/h, où le temps intervéhiculaire minimum légal imposé est de $gap_{min}=2$ secondes, et où la décélération maximale est $\gamma_{max}=6.25$ m/s². La situation limite, pour laquelle $\log_2((v_i/\gamma_{max})/(2*gap_i))=0$, est la situation dans laquelle le véhicule i roule à 90 km/h et se trouve à 2 secondes du véhicule qui le précède. Cette situation est considérée comme sûre, mais si le temps intervéhiculaire devient inférieur à 2 secondes ou si la vitesse devient supérieure à 90 km/h, elle sera alors risquée.

De Mouzon et al. (2008) ont adapté cet indicateur à tout type de route en introduisant un facteur k tel que $k = \frac{v_{lim}}{gap_{min} \times \gamma_{max}}$. Ils nomment Compensated Individual Braking Time Risk (CIBTR) l'indicateur suivant :

$$CIBTR_i = \log_2 \left(\frac{v_i/\gamma_{max}}{k \times gap_i \times \alpha} \right)$$

On a alors :

$$IBTR_i = \max(0, CIBTR_i)$$

PBTR (Platoon Braking Time Risk)

Si un véhicule i se trouve dans un peloton de véhicules, le risque de collision avec le véhicule qui le précède augmente avec son rang dans le peloton. Le Platoon Braking Time Risk (PBTR) est un indicateur de risque qui illustre ce phénomène. Il est défini de la manière suivante :

$$PBTR_i = 0 \text{ si } IBTR_i = 0,$$

$$\left. \begin{array}{l} \\ \\ \\ \end{array} \right\} PBTR_{i-1} + IBTR_i \text{ sinon.}$$

PBTR est donc un indicateur qui s'additionne au fur et à mesure que le véhicule i est loin dans le peloton. Les analyses effectuées sur les données de trafic et accidentogènes du canton de Vaud par de Mouzon et al. (2007) ont montré que la valeur du PBTR pouvait augmenter jusqu'à prendre des valeurs très élevées, puis redescendre d'un coup à 0. Un des inconvénients du PBTR est qu'il n'est pas toujours représentatif de la réalité. Par exemple, considérons un véhicule i roulant derrière un véhicule ($i-1$) se trouvant très près du véhicule ($i-2$), le véhicule ($i-1$) ayant donc un PBTR positif. S'il y a collision entre les véhicules ($i-2$) et ($i-1$), le véhicule i est sûr d'être en sécurité si le temps qu'il mettra à s'arrêter est inférieur au temps intervéhiculaire. Si nous avons $k=1$, cela serait équivalent à $IBTR_i=0$. Mais nous avons vu que $k>1$, donc il est possible que $IBTR_i=0$ alors que $\frac{v_i}{\gamma_{max} \times \alpha} > gap_i$, et donc que le véhicule i ne soit pas réellement en sécurité.

Compensated Platoon Braking Time Risk (CPBTR)

Pour remédier à cet inconvénient du PBTR, de Mouzon et al. (2008) proposent un nouvel indicateur de risque : le Compensated Platoon Braking Time Risk (CPBTR), qui se calcule de la manière suivante :

$$CPBTR_i = \max(0, CPBTR_{i-1} + CIBTR_i)$$

Le principe additif du PBTR est conservé, mais le CPBTR n'est pas remis à 0 à chaque fois qu'un véhicule a un comportement non risqué. Cependant, ce comportement compense en partie les attitudes risquées des véhicules le précédant, car le CIBTR, alors négatif, permettrait une diminution du CPBTR.

II.2 Typologie des indicateurs de risque

II.2.1 Différents types d'indicateurs de risque

Les indicateurs de risque peuvent être classés dans différentes catégories. D'abord, ils

peuvent être regroupés par finalité : Ce sont soit des indicateurs de risque en temps réel, calculés a priori, soit des indicateurs de risque en temps différé, calculé a posteriori. Par exemple, le taux d'accidents est un indicateur de risque en temps différé.

Nous nous intéresserons ici uniquement aux indicateurs de risque en temps réel, mais ceux-là aussi sont de différents types. Il est d'abord possible de les classer selon la nature des données auxquelles ils s'appliquent. En effet, certains se calculent sur des données individuelles alors que d'autres se calculent sur des données agrégées. Enfin, chaque indicateur permet d'évaluer le risque d'un type d'accident particulier : collision avant-arrière ou collision latérale.

La figure 11 présente ces différentes catégories d'indicateurs de risque. Pour chaque catégorie, des exemples sont donnés, et ceux-ci sont détaillés dans la suite.

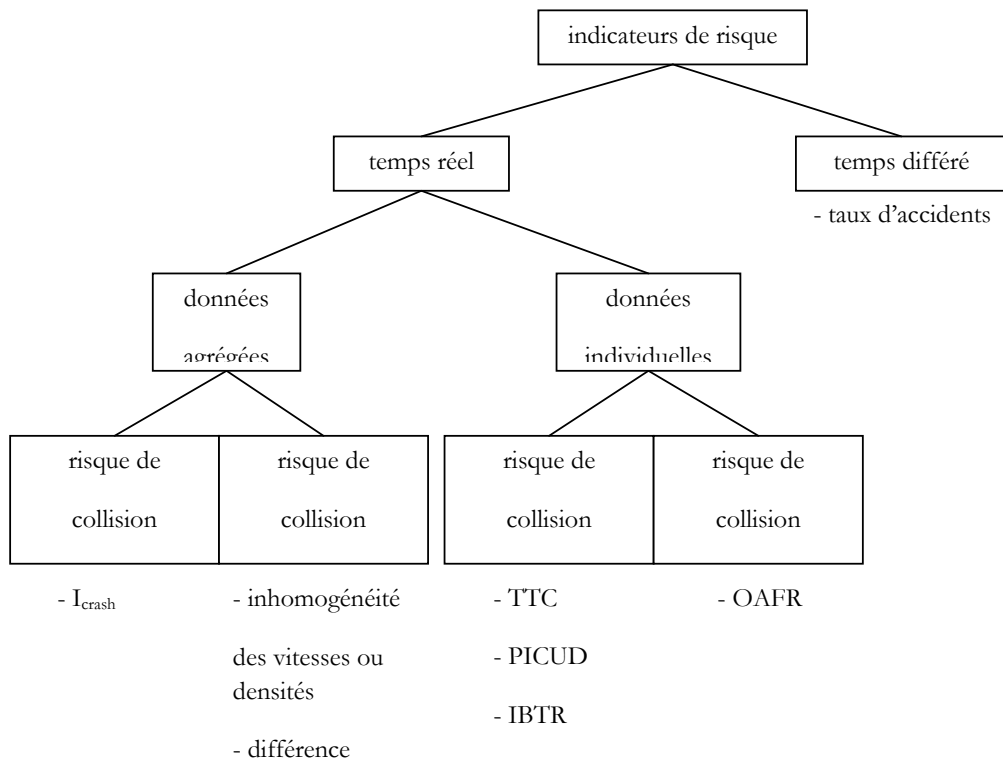


Fig. II.11 Typologie des indicateurs de risque

II.2.2 Exemples d'indicateurs de risque en temps réel se calculant sur des données individuelles

Tous les indicateurs de risque en temps réel présentés dans la partie précédente se calculent sur des données individuelles. Comme nous l'avons explicité, le TTC, le PICUD, l'IBTR, le PBTR et le CPBTR sont des indicateurs de risque en temps réel caractéristiques du risque de collision avant-arrière, alors que l'OAFR est un indicateur de risque en temps réel caractérisant le risque de collision latérale.

Cependant, il est possible d'utiliser d'autres données individuelles de trafic en tant qu'indicateurs de risque en temps réel. Par exemple, Vogel (2003) utilise le temps intervéhiculaire comme un indicateur de risque et le compare au TTC. Ses conclusions sont que, même si ces deux indicateurs sont indépendants l'un de l'autre, seul un véhicule au faible temps intervéhiculaire peut avoir un faible TTC. Ainsi, un temps intervéhiculaire élevé entraîne nécessairement un TTC élevé, et donc un faible risque d'accident. A l'instar du TTC, le temps intervéhiculaire peut donc être considéré comme un indicateur de risque en temps réel caractéristique du risque de collision avant-arrière sur voie rapide. De plus, il a l'avantage d'être facile à mesurer, et de nécessiter moins de données que le calcul du TTC par

exemple.

D'autre part, la vitesse peut également être utilisée comme indicateur de risque en temps réel. En effet, une vitesse individuelle élevée entraîne un risque d'accident supérieur à une vitesse individuelle plus faible. Ce lien entre vitesse et accidents a été établi a posteriori, après analyse des bases de données d'accidents, mais il est quand même possible de s'en servir dans l'analyse du risque en temps réel, car les vitesses peuvent être des données individuelles recueillies en temps réel.

II.2.3 Exemples d'indicateurs de risque en temps réel se calculant sur des données agrégées

Dans le cadre du projet européen INTRO, les données disponibles sont des données agrégées sur des périodes de 6 minutes. Il a donc fallu créer des indicateurs de risque en temps réel se calculant à base de données agrégées. Les exemples suivants présentent des indicateurs ayant été mis en œuvre dans ce projet.

II.3 Indicateur global de sécurité : I_{crash}

I_{crash} est un indicateur de risque de collision avant-arrière s'appliquant aux situations où la vitesse est élevée et la densité est élevée. Il se calcule sur une période de 30 minutes partagée en 5 périodes de 6 minutes. Il se base sur deux indicateurs :

- I_1 : indicateur de "surfusion". I_1 est le nombre de périodes de 6 minutes durant lesquelles la vitesse moyenne excède un seuil donné correspondant à un seuil de sécurité et dépendant du taux d'occupation de la période considérée. I_1 déclenche une alerte quand sa valeur est supérieure à un seuil N_1 . Dans le projet INTRO, le seuil utilisé est $N_1=4$.
- I_2 : indicateur d'"onde de choc établie"

L'indice "u" est l'indice représentant une période de 6 minutes. Il varie de 1 à 5.

I_2 est la somme :

- du nombre de périodes u durant lesquelles :
 - $TO(u)-TO(u-1) < \text{seuil_TO_accélération}$,
 - et $V(u)-V(u-1) > \text{seuil_vitesse_accélération}$,
 - et $V(u) > 80$ km/h,
 - et $TO(u) \in [5,12]$.
- et du nombre de périodes u durant lesquelles :
 - $TO(u)-TO(u-1) > \text{seuil_TO_décélération}$,
 - et $V(u)-V(u-1) < \text{seuil_vitesse_décélération}$,
 - et $V(u) > 80$ km/h,
 - et $TO(u) \in [5,12]$.

où TO est le taux d'occupation et V la vitesse.

Ces périodes correspondent à des accélérations ou décélérations rapides entraînant un risque. Le seuil de l'indicateur I_2 est noté N_2 .

Les valeurs utilisées dans le projet INTRO sont :

- $N_2=1$,
- $\text{seuil_TO_accélération} = \text{seuil_TO_décélération} = 3\%$,
- $\text{seuil_vitesse_accélération} = \text{seuil_vitesse_décélération} = 10$ km/h.

L'indicateur I_{crash} est un indicateur binaire : Il ne peut valoir que 0 ou 1. Lorsqu'il est égal à 1, cela signifie que la situation est risquée, et lorsqu'il est nul, la situation est sûre. I_{crash} déclenche une alerte, c'est-à-dire vaut 1, si au moins l'un des deux indicateurs I_1 et I_2 prend

une valeur supérieure ou égale à sa valeur seuil N_1 ou N_2 . Dans le cas contraire, il est égal à 0.

Dans le projet INTRO, cet indicateur a été calculé sur différentes situations, certaines ayant donné lieu à un accident et d'autres non. Les résultats sont les suivants :

- 64.5% des accidents ont été précédés d'une alerte.
- La durée moyenne entre le début de l'alerte et l'accident quand il y en a est de 24 minutes.
- Une alerte est déclenchée 24.6% du temps durant lequel le taux d'occupation est entre 5 et 12% et que la vitesse est supérieure à 80 km/h.
- 99.8% des fois où l'alerte est déclenchée, il n'y a pas d'accident.

II.4 Homogénéité des vitesses et des densités

Le projet INTRO a étudié l'influence de l'homogénéité des vitesses et des densités entre les différentes voies d'une route sur le risque. Les indicateurs associés se calculent en un point particulier (en général la localisation d'un détecteur), sur une route à n voies, où l'on note, pour une période j donnée, V_i^j la vitesse pratiquée sur la voie i , et V_{fluide} la vitesse en situation fluide (par exemple 130 km/h). V_i^j est la vitesse agrégée sur la période j considérée.

L'inhomogénéité de la vitesse est représentée par :

$$E_{speed}^j = \frac{1}{V_{fluide}} \sum_{i=1}^{n-1} |V_{i+1}^j - V_i^j|$$

De même, en notant K_{max} la densité maximale et K_i^j la densité de la voie i sur la période j , l'inhomogénéité de la densité est représentée par :

$$E_{density}^j = \frac{1}{K_{max}} \sum_{i=1}^{n-1} |K_{i+1}^j - K_i^j|$$

Les valeurs de ces deux indicateurs sont sans unité et peuvent donc s'additionner pour former un troisième indicateur : $E_{density}^j + E_{speed}^j$. Cet indicateur caractérise l'inhomogénéité des vitesses et des densités entre les différentes voies. Plus sa valeur est élevée, plus les différences entre les voies en termes de vitesses et de densités sont importantes. A l'inverse, lorsque cet indicateur est nul, les vitesses et les densités sont parfaitement homogènes.

Les études réalisées dans le cadre du projet INTRO montrent qu'une forte inhomogénéité des vitesses et/ou des densités entre les différentes voies, soit des valeurs élevées d'au moins un de ces indicateurs, entraîne un risque d'accident plus élevé que si les vitesses et les densités étaient homogènes. Cela est principalement dû au fait que les changements de voie sont plus nombreux lorsqu'il y a de l'inhomogénéité entre les voies, et un fort taux de changement de voie entraîne un fort risque d'accident.

II.5 Différence relative de vitesse et de distance inter-véhiculaire

Cet indicateur s'intéresse aux changements de file gênants. L'idée est de chercher une relation entre les vitesses et les débits des différentes voies et la fréquence des changements de voie. On se place sur des périodes de durée 6 minutes. Pour chaque couple de voies adjacentes i et j , on cherche à prévoir le nombre de changements de voies durant une période en fonction des deux paramètres suivants :

- différence relative de vitesse : $\sigma_{i-j} = \frac{|V_i - V_j|}{(V_i + V_j)/2}$

- différence relative de distance intervéhiculaire : $\delta_{i-j} = \frac{|D_i - D_j|}{(D_i + D_j)/2}$

où V_n est la vitesse moyenne sur la voie n et D_n est la distance intervéhiculaire moyenne sur la voie n , soit la vitesse divisée par le débit.

Les études menées dans le projet INTRO ont utilisé des données d'une section de l'auto-route A8 en France. L'analyse de vidéos a permis de prendre en compte uniquement les changements de voie dangereux. Les conclusions sont que la fréquence de ces changements de voie ne dépend pas de la différence relative de distance intervéhiculaire.

On construit alors un indicateur basé uniquement sur les différences de vitesses :

$$\Delta = \frac{1}{2} \sum_{i=1}^{n-1} |V_i - V_j| \quad \text{où } n \text{ est le nombre total de voies}$$

L'étude des données de l'A8 disponibles montre une relation parabolique entre la fréquence des changements de voie gênants et l'indicateur Δ défini précédemment, comme le montre la figure 12. La définition de seuils T_1 et T_2 correspondant respectivement à une fréquence de changements de voie transitoire et à une fréquence de changements de voie dangereuse donne les seuils Δ_1 , Δ_2 , Δ_3 et Δ_4 pour l'indicateur Δ tels qu'une valeur de Δ entre Δ_2 et Δ_3 correspond à une situation dangereuse, et une valeur de Δ inférieure à Δ_1 ou supérieure à Δ_4 correspond à une situation sûre. La fréquence de changements de voies gênants se trouve pour une valeur de Δ d'environ 13 km/h.

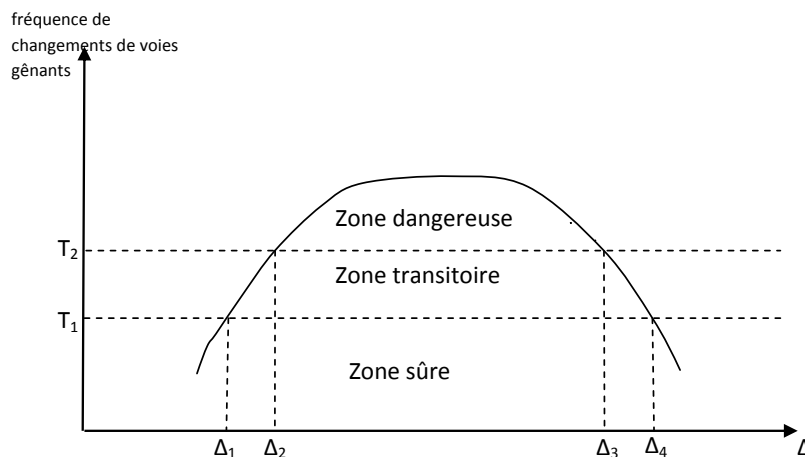


Fig. II.12 Fréquence de changements de voie gênants en fonction de la différence de vitesse

III Propriétés des indicateurs de risque en temps réel

III.1 Relations entre indicateurs et paramètres relatifs au trafic

Nous avons vu qu'il existe différents facteurs de risque, comme la géométrie de la route ou les conditions météorologiques. Le seul de ces facteurs qui est pris en compte dans le calcul de certains indicateurs de risque en temps réel est la présence ou l'absence de pluie. En effet, l'effet accidentogène de la pluie est notamment du à la perte d'adhérence, ce qui est pris en compte dans plusieurs indicateurs de risque en temps réel, car ceux-ci nécessitent la connaissance de la décélération ou du taux de friction, qui diffère suivant que la chaussée est sèche ou mouillée.

Dans le cadre du projet FUSAIN, de Mouzon et al. (2008) analysent des données provenant d'autoroutes suisses et établissent que, pour chaque site considéré, le seuil du TTC ou du PBTR correspondant à une situation risquée doit être différent suivant le niveau de trafic, caractérisé par les valeurs du débit. En effet, les valeurs moyennes du TTC et du PBTR varient de façon significative selon les valeurs du débit. Il convient donc d'étudier séparément les indicateurs pour des valeurs de débit élevées et pour des valeurs de débit faible. Ainsi, le débit est un paramètre dont le calcul des indicateurs de risque en temps réel ne tient pas compte, mais qui doit quand même être pris en considération dans l'analyse des données de trafic et de ces indicateurs.

Cette dépendance au débit est vraie pour tous les indicateurs testés dans le projet FUSAIN :

- TTC,
- IBTR,
- PBTR,
- CPBTR, indicateur créé dans le cadre du projet,
- temps intervéhiculaire,
- PICUD,
- vitesse.

De même, la géométrie de la route n'a pas d'incidence sur la valeur des indicateurs, mais elle est quand même prise en compte dans leur étude, car ils sont étudiés séparément selon le site sur lequel les données sont recueillies. Ainsi, les valeurs d'indicateurs provenant de sites de géométries différentes ne sont pas comparables.

A l'inverse, d'autres paramètres tels que le taux d'alcoolémie sont totalement indépendants de la valeur des indicateurs. C'est pourquoi dans le projet FUSAIN les accidents lors desquels un conducteur impliqué avait une alcoolémie positive ne sont pas pris en compte.

Les indicateurs de risque en temps réel existants sont nombreux, et ceci montre qu'il existe de nombreuses façons de caractériser le risque associé à une situation. La méthode et les indicateurs utilisés dépendent d'abord :

- de l'environnement dans lequel on se trouve (urbain ou voie rapide),
- du type de risque que l'on veut caractériser (risque de collision avant-arrière ou risque de collision latérale),
- des données dont on dispose (véhicule traceur, trajectoires de véhicules, données individuelles recueillies par des détecteurs de trafic, données simulées, données agrégées...).

Il faut ensuite prendre en compte d'autres facteurs tels que le débit ou les conditions météorologiques dans le calcul des indicateurs retenus.

Les indicateurs de risque en temps réel donnent une information chiffrée caractérisant le risque associé à une situation de trafic à tout instant. Ces indicateurs ne sont pas destinés à être utilisés par les usagers de la route. En effet, ils ne sont pas compréhensibles par le public, qui n'est pas familier avec les notions abordées, et donc qui n'a aucune idée de l'ordre de grandeur des valeurs des indicateurs qui rendrait leur situation sûre ou risquée. Par contre, ils sont utilisables par les gestionnaires des routes, pour qui le fait de savoir en temps réel en quel lieu de leur réseau il y a un risque d'accident est important.

III.2 Comparaison des différents indicateurs

Chacun des indicateurs définis précédemment a des propriétés qui lui sont propres. Il est possible de comparer ces différentes propriétés dans le but d'évaluer les performances de chaque indicateur. Ils sont comparables au niveau des aspects suivants :

- simplicité : accessibilité des données nécessaires, qui ne doivent pas être trop nombreuses,
- lisibilité : compréhensibilité de l'indicateur par le public,
- exhaustivité : capacité à refléter tous les aspects du risque routier, et donc à prendre en compte tous les paramètres influant le risque,
- calculabilité : possibilité d'être calculé à tout instant, sous toutes conditions.

Fig. III.13 compare les indicateurs de risque de collision avant-arrière se calculant sur des données individuelles locales, et pouvant s'utiliser sur voie rapide.

Fig. III.13 Comparaison des indicateurs

	TTC	PICUD	IBTR	PBTR
Données nécessaires	-longueur du véhicule, - vitesse, -temps intervéhiculaire.	-longueur du véhicule, - vitesse, -distance intervéhiculaire, - état de la route.	- vitesse -temps intervéhiculaire - état de la route - vitesse limite.	- vitesse - temps intervéhiculaire - état de la route - vitesse limite.
lisibilité	bonne lisibilité : temps au bout duquel il y aurait collision.	bonne lisibilité : distance qui resterait après collision	pas de signification concrète	pas de signification concrète
Paramètres pris en compte	- vitesse, -temps intervéhiculaire	- vitesse, -temps intervéhiculaire, -temps de réaction, -conditions météorologiques.	- vitesse, -temps intervéhiculaire, -conditions météorologiques, - type de route.	- vitesse, -temps intervéhiculaire, -conditions météorologiques, - type de route, - place du véhicule dans le peloton
calculabilité	ne se calcule que si la vitesse du véhicule de devant est la plus faible.	se calcule dans toutes les situations.	se calcule dans toutes les situations.	se calcule dans toutes les situations.

IV Etablissement de seuils

L'utilisation d'indicateurs de risque routier en temps réel sert à déceler, à tout instant, si une situation de trafic est risquée ou si elle est sûre. Pour cela, il convient d'établir, pour chaque indicateur, une valeur seuil permettant de distinguer les situations sûres des situations risquées. Les performances d'un système de détection du risque dépendent fortement du choix des seuils de déclenchement des alarmes : des seuils trop faibles génèrent des fausses alarmes, alors que des seuils trop grands passent à côté de vrais risques. Dans la littérature, il existe diverses méthodes pour évaluer ces seuils.

IV.1 Tracés des distributions des valeurs des indicateurs

Une première méthode est mise en œuvre par De Mouzon et al. (2007). Elle consiste d'abord à tracer, pour chaque indicateur et chaque condition dont il dépend (débit, conditions météorologiques...), la distribution cumulée des valeurs de l'indicateur sur une période d'une durée assez longue (en général au moins un an). Ainsi, pour chaque indicateur, pour un lieu donné, une classe de débit donnée et un état de la chaussée donné, les courbes tracées permettent de connaître la proportion des valeurs de l'indicateur qui dépasse chaque valeur. Il suffit ensuite de fixer un taux défini comme étant le pourcentage de situations qui seront considérées comme risquées. Le seuil de chaque indicateur correspond ensuite à la valeur de l'indicateur dépassée par ce pourcentage de véhicules.

Par exemple, dans leur étude, De Mouzon et al. cherchent à établir des seuils pour les indicateurs caractéristiques des collisions avant-arrière à partir de données provenant d'un capteur sur une autoroute suisse. Ils choisissent 0.1% comme proportion de véhicules risqués. Pour valider les seuils qu'ils trouvent à partir de cette proportion, ils s'intéressent ensuite à certains accidents de type collision avant-arrière ayant eu lieu à moins d'un kilomètre du capteur considéré, et dans lesquels les conducteurs ont un taux d'alcoolémie nul. Ils vérifient alors, pour chaque indicateur, que le seuil qu'ils ont trouvé est effectivement dépassé durant la demi-heure précédant chaque accident, et donc que l'alarme mise en place aurait été tirée avant l'arrivée de ces accidents. Si c'est le cas, ils déclarent leurs seuils comme étant admissibles, et les indicateurs comme étant effectivement caractéristiques du risque de collision avant-arrière.

L'utilisation de cette méthode nécessite la connaissance des données de trafic et des conditions de trafic durant une longue période au lieu où l'on veut mettre en place un tel système de détection du risque routier en temps réel. Elle nécessite également la disponibilité de données complètes relatives à certains accidents ayant eu lieu au lieu considéré.

IV.2 Utilisation des algorithmes de détection automatique d'incidents

Une autre approche est de se servir des algorithmes utilisés dans la détection automatique d'incidents. Les performances des systèmes de détection sont alors caractérisées par trois indicateurs (Cohen, 2005) :

- le taux de détection (TD), défini par la fraction du nombre d'incidents détectés sur le nombre d'incidents effectifs,
- le taux de fausses alarmes (TFA), défini par le rapport du nombre de fausses alarmes sur le nombre de tests effectués au cours d'une période de temps donnée. Il y a fausse alarme lorsqu'à l'issue d'un test, l'indicateur diagnostique un incident qui, en réalité, ne s'est pas produit.
- le délai moyen de détection (DMD), correspondant au temps moyen entre l'arrivée d'un incident et sa détection par le système.

Dans le cadre de l'établissement de seuils pour un indicateur de risque en temps réel, il convient de rechercher un compromis optimal entre les critères taux de détection et taux de fausses alarmes. Cela revient à maximiser le taux de détection tout en minimisant le taux de fausses alarmes, sachant que les valeurs des mesures du taux de détection et du taux de fausses alarmes dépendent l'une de l'autre. De façon générale, si l'un augmente, l'autre augmente également.

Dans le cas de la détection automatique d'incidents, de nombreux algorithmes ont été mis en place. Ils se basent sur des variables de trafic recueillies par des capteurs, telles que le débit, le taux d'occupation et la vitesse. Différents algorithmes compilant ces variables ont été mis en place afin de détecter les situations correspondant à un incident, puis ils ont été testés par différents chercheurs sur des données de trafic sur différentes routes. La description d'un bon nombre de ces algorithmes et de leurs résultats peut être trouvée dans Martin et al. (2001). Les taux de détection varient entre 67% et 100%, les taux de fausses alarmes entre 0% et 1.9%, et les délais moyens de détection entre 0.4 et 4 minutes.

Cohen (2005) remarque que ces approches révèlent une certaine variabilité de performance suivant l'état de paramètres extérieurs tels que l'intensité du trafic, les conditions météorologiques ou les caractéristiques géométriques des aménagements. Par exemple, lors d'une expérimentation sur des données provenant de l'A43 près de Lyon, le passage d'un temps sec à une pluie soutenue a fait diminuer le taux de détection de 84% à 58%, tandis que le taux de fausses alarmes était multiplié par 6.

Il propose alors de résoudre le problème en fixant préalablement un taux de détection α (en général compris entre 80% et 90%), et en déterminant alors le seuil de détection conduisant à un taux de fausses alarmes minimal. Ainsi, le problème revient à un problème d'optimisation sans contrainte, portant le nom de calibrage des systèmes.

Cette façon de raisonner peut facilement s'appliquer à l'établissement de seuils pour des indicateurs de risque en temps réels. Si toutes les valeurs d'un indicateur donné peuvent être calculées sur une certaine période, et que tous les accidents ayant eu lieu durant cette période sont complètement connus, alors il est possible d'associer à chaque valeur possible de l'indicateur un taux de détection et un taux de fausses alarmes correspondant au taux effectifs si cette valeur était le seuil choisi. Il faut ensuite retenir toutes les valeurs de l'indicateur entraînant un taux de détection supérieur à α comme étant des seuils admissibles. Parmi ces seuils admissibles, celui qui est choisi est celui qui minimise le taux de fausses alarmes.

Cela revient à trouver $\min_s \{TFA / (TD > \alpha)\}$.

Il est clair que ce genre de calibrage requiert l'élaboration d'une base de données associant mesures de trafic et base exhaustive des incidents et accidents survenus pendant les mêmes périodes. La qualité de cette base de données conditionne directement celle du calibrage effectué.

Considérons un indicateur de risque dont nous calculons la valeur pour N situations, dont n résultent en un incident ou un accident. Nous considérons que plus la valeur de cet indicateur est élevée, plus la situation est risquée, et nous notons $[a_1 a_2 a_3 \dots a_n]$ les valeurs de l'indicateur pour les n situations résultant en un accident, classées dans l'ordre croissant, c'est-à-dire $a_1 \leq a_2 \leq \dots \leq a_n$. Connaissant les N valeurs de l'indicateur, il est possible de tracer la distribution cumulée de ces valeurs, notée $F(x)$. Comme le montre Fig. IV.14, $F(x)$ est la proportion de valeurs de l'indicateur étant supérieures ou égales à

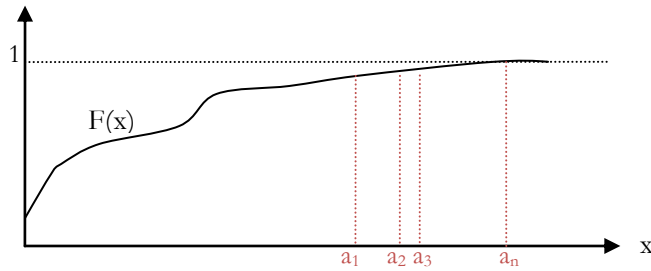


Fig. IV.14 Distribution cumulée des valeurs d'un indicateur

La distribution cumulée $F(x)$ des valeurs de l'indicateur se trace empiriquement, à partir de N "situations". Considérons que les données de trafic sont recueillies par un capteur. L'indicateur prend une valeur différente à chaque passage d'un véhicule sur le capteur, mais il ne paraît pas pertinent d'assimiler chaque passage d'un véhicule à une des N situations précédemment définies. Il faudrait plutôt considérer qu'une situation correspond à tous les véhicules passant sur le capteur durant une certaine période (par exemple une demi-heure), et que la valeur de l'indicateur associée vaut le maximum des valeurs de l'indicateur durant la période. Une situation résultant en un accident correspondrait alors à une période telle qu'un incident survient durant la période suivante.

A partir des données de $F(x)$ et de la suite $(a_1, a_2 \dots a_n)$, il est possible de calculer, pour chaque valeur possible x de seuil de l'indicateur, les taux de détection et de fausses alarmes associés.

1^{er} cas : $x \leq a_1$

Tous les accidents se passent dans une situation ayant une valeur de l'indicateur supérieure à x , donc le taux de détection vaut $TD(x) = 1$.

Le nombre de situations détectées comme risquées, c'est-à-dire dont la valeur de l'indicateur est supérieure à x , est égal à $N \times (1 - F(x))$. Parmi celles-ci, n correspondent à des accidents ; il y a donc $(N \times (1 - F(x)) - n)$ qui sont des fausses alarmes.

Ainsi, le taux de fausses alarmes vaut $TFA(x) = \frac{N \times (1 - F(x)) - n}{N}$.

2^{ème} cas : $a_i < x \leq a_{i+1}$, avec i compris entre 1 et $(n-1)$

Un raisonnement analogue au précédent donne :

$$TD(x) = \frac{n-i}{n} \quad \text{et} \quad TFA(x) = \frac{N \times (1 - F(x)) - (n-i)}{N}$$

3^{ème} cas : $x > a_n$

$$TD(x) = 0 \quad \text{et} \quad TFA(x) = 1 - F(x)$$

Le taux de détection est fixé : il doit être supérieur à α .

Ceci correspond à $x \leq a_j$, tel que $\frac{n-(j-1)}{n} \geq \alpha$ et $\frac{n-j}{n} < \alpha$.

Autrement dit, j est la plus grande valeur entière telle que $j \leq (1-\alpha) \times n + 1$.

Notons $j = E((1-\alpha) \times n + 1)$, où E représente la partie entière.

La fonction $TFA(x)$, que nous cherchons à minimiser, est continue par morceaux. Elle est

continue sur chacun des segments $]a_i, a_{i+1}]$. La fonction F étant croissante, si $a_i < x < y \leq a_{i+1}$, alors $TFA(x) > TFA(y)$.

Le minimum de la fonction $TFA(x)$ sur un segment $]a_i, a_{i+1}]$ est donc $\frac{N \times (1 - F(a_{i+1})) - (n - i)}{N}$.

Le seuil qui est solution de l'algorithme est donc la valeur a_s telle que :

$$\min\{ [N \times (1 - F(a_i)) - (n - (i - 1))], 1 \leq i \leq E((1 - \alpha) \times n + 1) \} \\ = N \times (1 - F(a_s)) - (n - (s - 1))$$

Au lieu de fixer le taux de détection et de chercher la valeur seuil à partir de ce taux, une autre stratégie utilisée en détection automatique d'incidents consiste à minimiser un critère composite compilant le taux de détection et le taux de fausses alarmes (Cohen, 2005) :

$$C = TFA \times TE \times DMD$$

où TE est le taux d'erreur, défini par $TE = 1 - TD$.

Dans le cas de l'établissement de seuils pour des indicateurs de risque en temps réel, la notion de délai moyen de détection n'a pas de sens ; il suffit donc de minimiser $(TFA \times TE)$.

En reprenant les notations précédentes, $TFA \times TE(x)$ est une fonction continue par morceaux, telle que, si $x \in]a_i, a_{i+1}]$, alors $TFA \times TE(x) = \frac{N(1 - F(x)) - (n - i)}{N} \times \frac{i}{n}$.

Nous ne tenons pas compte du segment $]a_n, +\infty]$, sur lequel $TD=0$, et donc ne contient pas de valeur intéressante en tant que valeur seuil.

Le minimum de $TFA \times TE(x)$ sur $]a_i, a_{i+1}]$ est atteint pour $x=a_{i+1}$.

Le seuil solution de l'algorithme est donc la valeur a_s telle que

$$\min\{ [(N \times (1 - F(a_i)) - (n - (i - 1))) \times (i - 1)], 1 \leq i \leq n \} \\ = (N \times (1 - F(a_s)) - (n - (s - 1))) \times (s - 1)$$

V Estimation du risque à partir des données trajectoires

Les données trajectoires (e.g. données NGSIM) sont des données fines et individuelles, c'est-à-dire qu'elles regroupent, sur une section de route particulière et dans un intervalle de temps particulier, toutes les informations relatives à chaque véhicule ayant circulé sur cette section. A chaque instant, la vitesse et la position de chaque véhicule sont connues, ce qui permet de pouvoir tracer des courbes appelées "trajectoires de véhicules" dans un diagramme espace-temps.

L'utilisation de ce type de données sur une section de voie rapide dans un contexte d'élaboration d'indicateurs de risque en temps réel sur voie rapide peut se faire selon deux axes :

- une élaboration du risque de collision longitudinale (collision avant-arrière),
- une élaboration du risque de collision latérale.

Nous avons vu que le risque de collision latérale est lié à la fréquence des changements de voies, et ceux-ci ne sont décelables que dans le cas de données de type trajectoires de véhicules.

V.1 Présentation des données NGSIM

V.1.1 Présentation du site et des données disponibles

Les données utilisées ont été recueillies sur une autoroute, la "Highway 101", situé à Los Angeles (Californie, Etats-Unis). La section étudiée mesure 640 mètres, est à sens unique et comporte cinq voies principales ainsi qu'une voie auxiliaire reliant une entrée et une sortie comme montrée dans Fig V.15.

Le recueil de ces données s'inscrit dans le cadre du projet NGSIM (*Next Generation Simulation*), ayant pour objectif de promouvoir l'utilisation des outils de simulation et de favoriser l'émergence d'outils fiables de microsimulation (Federal Highway Administration, 2007).

Le recueil a été effectué le 15 juin 2005, en utilisant huit caméras placées au sommet d'un bâtiment situé en bordure de la route considérée, et filmant chacune un morceau de la section de route. Les vidéos sont ensuite analysées automatiquement, en discrétisant le temps en intervalles d'une durée d'un dixième de seconde, ce qui permet de connaître les informations relatives à chaque véhicule durant chacun de ces intervalles de temps.

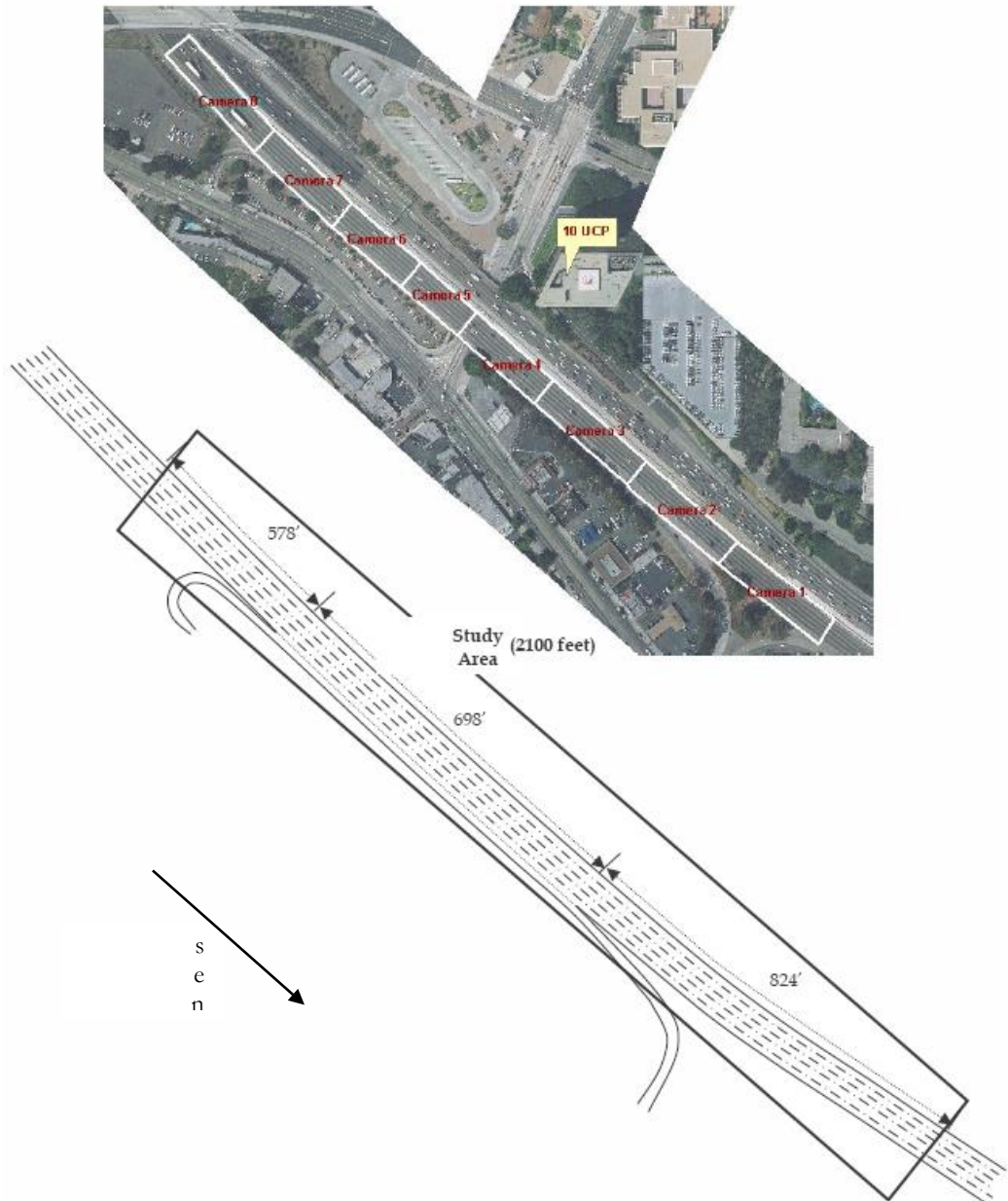


Fig. V.15 Section de la "Highway 101" étudiée
(source : <http://ngsim.fhwa.dot.gov>)

Nous allons nous intéresser ici à une période d'un quart d'heure, allant de 8h05 à 8h20.

Les données sont sous la forme d'un fichier composé de plusieurs lignes et de plusieurs colonnes. Nous convertissons les données dans le système métrique, car les mesures sont données en pieds et les vitesses en pieds par seconde, et nous retenons uniquement les informations utiles au calcul d'indicateurs de risque en temps réel. Ceci nous donne une matrice composée des 11 lignes suivantes comme illustrées dans Fig V.16:

- *Identifiant Véhicule* : Numéro caractéristique du véhicule considéré,
- *Temps* : Temps en dixièmes de seconde, avec comme origine 8h05,
- *X* : Coordonnée latérale du centre de l'avant du véhicule en mètres, avec comme origine le point le plus à gauche de la section dans le sens de la circulation,
- *Y* : Coordonnée longitudinale du centre de l'avant du véhicule en mètres, avec comme origine l'entrée de la section dans le sens de la circulation,
- *Longueur* : Longueur du véhicule en mètres,

- **Classe** : Catégorie du véhicule selon le code suivant :
 - 1 = Deux-roues motorisé,
 - 2 = Véhicule léger,
 - 3 = Poids lourd.
- **Vitesse** : Vitesse du véhicule en mètres par seconde
- **Voie** : Indice variant de 1 à 5 caractérisant la voie sur laquelle se trouve le véhicule considéré ; 1 caractérisant la voie la plus à gauche (voie rapide) et 5 la voie la plus à droite (voie lente),
- **Véhicule Devant** : Identifiant du véhicule précédant le véhicule considéré,
- **Véhicule Derrière** : Identifiant du véhicule suivant le véhicule considéré,
- **Distance Intervéhiculaire (DIV)** : Distance entre le centre de l'avant du véhicule considéré et le centre de l'avant du véhicule le précédant en mètres.

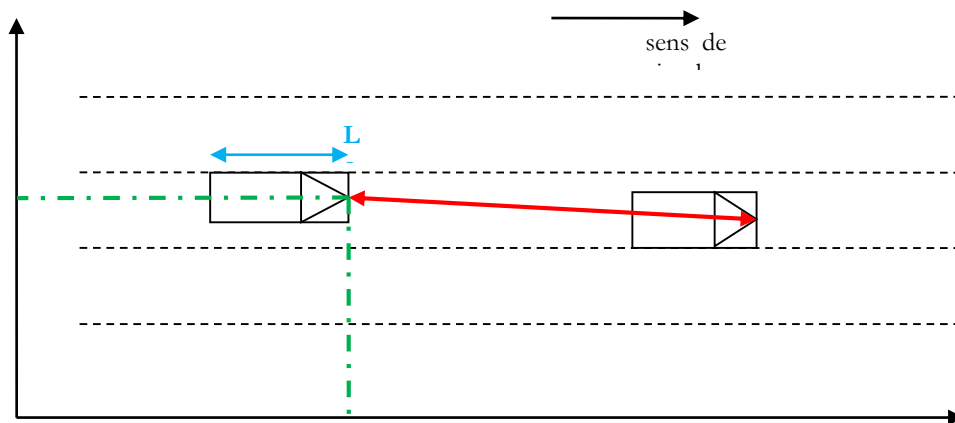


Fig. V.16 Principales informations disponibles pour un véhicule

Chaque ligne représente un véhicule particulier en un instant particulier. La matrice se compose de plusieurs blocs mis les uns à la suite des autres et représentant chacun un véhicule particulier. Ainsi, la structure ligne par ligne est du type suivant :

- *ligne 1* : véhicule 1 à l'instant 0,
- *ligne 2* : véhicule 1 à l'instant 1,
- *ligne 3* : véhicule 1 à l'instant 2,
- *ligne (n+1)* : véhicule 1 à l'instant n, n étant le nombre d'intervalles de temps d'un dixième de seconde durant lesquels le véhicule 1 apparaît dans la section,
- *ligne (n+2)* : véhicule 2 à l'instant 0,
- *ligne (n+3)* : véhicule 2 à l'instant 1...

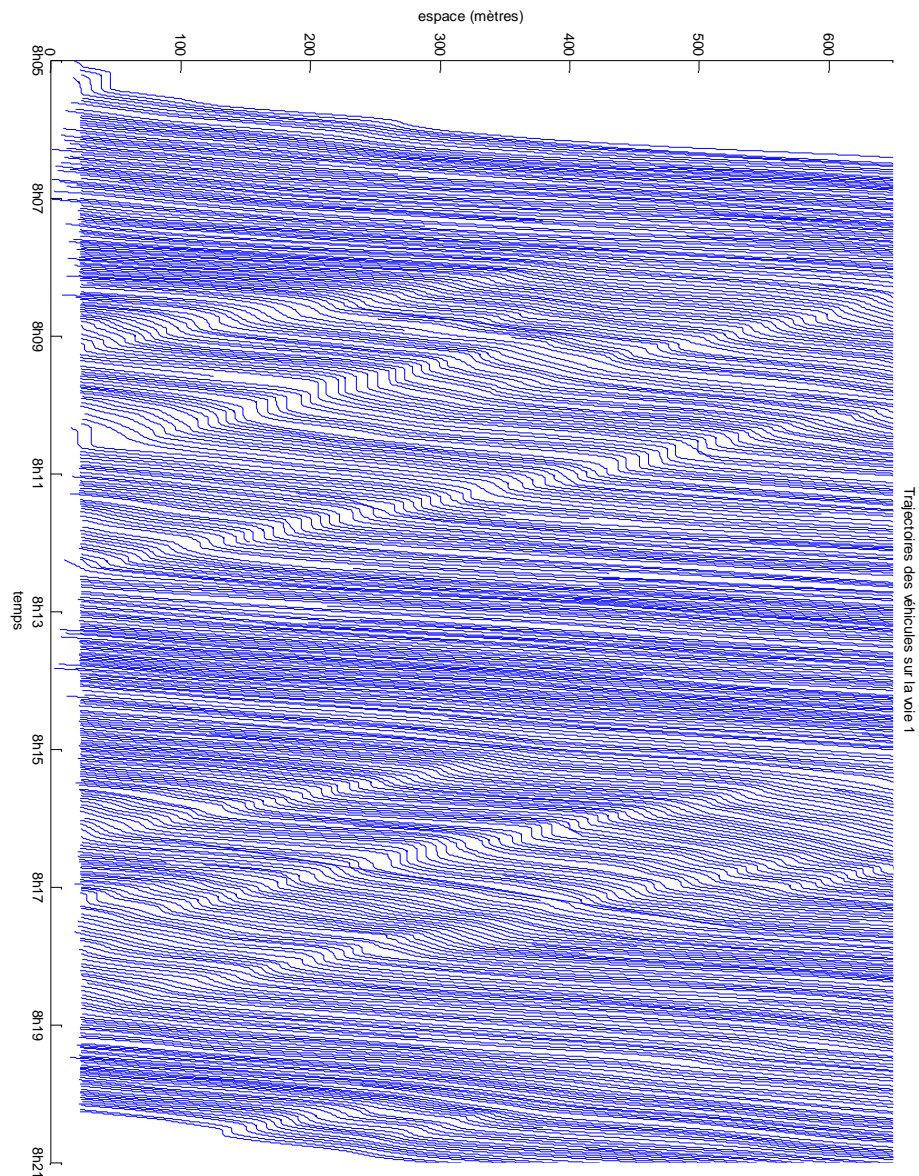


Fig. V.17 Trajectoires de véhicules sur la voie 1

Ces informations permettent de tracer les trajectoires de véhicules dans le diagramme espace-temps pour chaque voie. Fig. V.17 montre ces trajectoires sur la voie 1. Les diagrammes espace-temps correspondant aux voies 2, 3, 4 et 5 se trouvent en annexe F (figures 122 à 125).

Les trajectoires de véhicules tracées laissent apparaître quelques zones de congestion. Nous pouvons calculer à partir de ces courbes la vitesse "fluide", égale à la pente des trajectoires situées entièrement à l'extérieur des zones de congestion. Cette vitesse vaut environ 40 km/h.

V.1.2 Vérification de la qualité des données

Dans le projet NGSIM, chacune des informations relatives aux véhicules est calculée séparément, alors qu'il existe un lien entre certaines d'entre elles. Par exemple, la voie sur laquelle se trouve le véhicule dépend de la valeur de sa position latérale, notée ici X. Il est possible de vérifier la cohérence de ces deux informations comme montrées dans Fig. V.18 et Fig. V.19.

Fig. V.18 Caractéristiques des distributions des positions latérales suivant la voie

	moyenne	minimum	maximum	médiane	écart type
voie 1	2.12 m	0.27 m	4.07 m	2.11 m	0.43 m
voie 2	5.68 m	3.46 m	7.51 m	5.70 m	0.44 m
voie 3	9.11 m	7.04 m	10.85 m	9.12 m	0.41 m
voie 4	12.49 m	10.18 m	14.21 m	12.49 m	0.41 m
voie 5	16.16 m	13.56 m	21.06 m	16.13 m	0.53 m

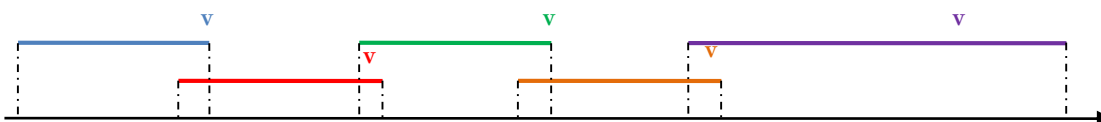


Fig. V.19 Répartition des positions latérales selon la voie

La moyenne des positions latérales par voie suit bien l'évolution attendue. Par contre, il existe un chevauchement au niveau des extrémités des plages de valeurs de X relatives à chaque voie : La valeur de X maximum rencontrée sur chaque voie i est plus grande que la valeur de X minimum pour la voie (i+1). Il existe donc certaines valeurs de X qui peuvent correspondre à plusieurs voies. Ceci est dû en général soit aux véhicules en train de changer de voie, pour lesquels il est difficile de savoir sur quelle voie ils se trouvent, soit aux véhicules chevauchant deux voies, notamment des deux-roues motorisés roulant entre deux voies. Cependant, sur chaque voie, la proportion de véhicules se trouvant aux extrémités des plages de X est très faible, comme le montre par exemple la distribution des positions latérales pour la voie 1 (Fig V.20). Pour chaque voie, les valeurs moyennes et médianes sont proches, ce qui montre une certaine symétrie dans la distribution des positions latérales.

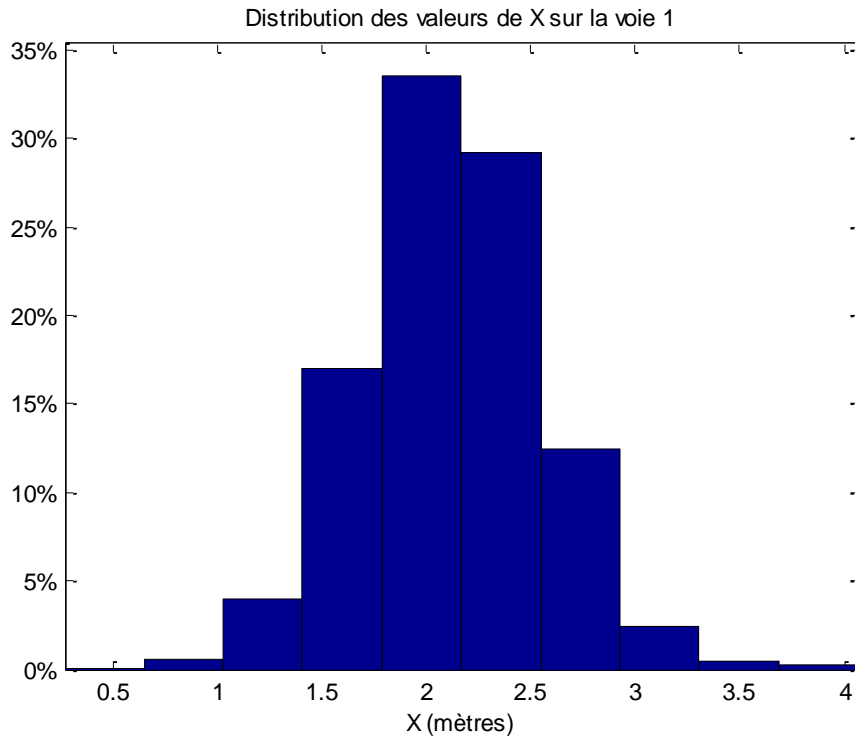


Fig. V.20: Distribution des positions latérales sur la voie 1

Outre la cohérence entre les voies et les positions latérales X, il est également possible de vérifier la cohérence entre les coordonnées des véhicules et les distances intervéhiculaires. Les distances intervéhiculaires sont les distances entre les centres des avants des véhicules, et les coordonnées fournies sont celles des centres des avants des véhicules ; il est donc possible de recalculer les distances intervéhiculaires uniquement à partir des coordonnées.

Comme nous avons vu qu'un véhicule particulier n'est pas caractérisé par une seule ligne mais par un ensemble de lignes consécutives, il convient d'abord d'associer à chaque ligne i la ligne correspondant au véhicule se trouvant devant le véhicule considéré à l'instant considéré. Nous notons cette nouvelle ligne j. Elle est telle que :

$$\left\{ \begin{array}{l} \text{Véhicule Devant}(i) = \text{Identifiant Véhicule}(j), \\ \text{Temps}(i) = \text{Temps}(j). \end{array} \right.$$

Comme les axes permettant de connaître les positions latérales et longitudinales des véhicules sont perpendiculaires, il est possible de recalculer la distance intervéhiculaire grâce à la formule de la distance euclidienne :

$$\text{Distance Intervéhiculaire Recalculée}(i) = \sqrt{(X(i) - X(j))^2 + (Y(i) - Y(j))^2}$$

Nous comparons ensuite les deux distances intervéhiculaires. Une première méthode est de calculer le coefficient de corrélation, qui vaut ici 0.9996. Ce coefficient est calculé sur 280 000 valeurs. Nous utilisons ensuite la loi de Student pour valider la relation entre les deux séries de distances intervéhiculaires. Cette loi donne le résultat suivant : la relation d'égalité entre les deux distances intervéhiculaires est vraie pour un intervalle de confiance supérieur à 99.9%.

Un autre moyen de comparer les deux séries de valeurs est de calculer le rapport de la distance intervéhiculaire recalculée sur la distance intervéhiculaire initialement fournie.

Plus de 99% de ces rapports sont compris entre 0.99 et 1.01 ; il y a donc une certaine cohérence entre les coordonnées et les distances intervéhiculaires.

En revanche, si nous nous intéressons uniquement aux distances intervéhiculaires entre un deux-roues et un autre véhicule, la cohérence est moins présente. En effet, même si le coefficient de corrélation, qui vaut 0.9888, permet la validation de la relation d'égalité entre les deux grandeurs, seuls 51% des rapports de la distance intervéhiculaire recalculée sur la distance intervéhiculaire fournie sont compris entre 0.99 et 1.01. Tous les autres rapports sont supérieurs à 1.01, et ils peuvent atteindre des valeurs très élevées (l'un d'eux est égal à 694 sur la voie 1). Il y a donc un certain nombre d'erreurs de mesure du logiciel lors du passage de deux-roues motorisés, soit au niveau de la mesure des coordonnées, soit au niveau de la mesure de la distance intervéhiculaire.

Dans la suite de notre étude, nous ne prendrons pas en compte les deux-roues. En effet, le type de risque que nous cherchons à étudier n'est pas très adapté aux deux-roues. D'autre part, nous avons vu qu'il est souvent difficile de savoir sur quelle voie ils se trouvent, et que c'est avec les deux-roues que la majorité des erreurs de mesure du logiciel sont commises. Pour cela, il convient non seulement d'éliminer les lignes concernant un deux-roues motorisé, mais aussi de modifier les identifiants des véhicules de devant ou de derrière pour les véhicules se trouvant respectivement derrière ou devant un deux-roues. Par exemple, si nous considérons un véhicule devant lequel se trouve un deux-roues d'après les données, alors le véhicule le précédant sera le véhicule précédant le deux-roues en question.

V.2 Etude du risque de collision avant-arrière

Le risque de collision avant-arrière s'étudie séparément sur chaque voie. Il peut être appréhendé par des indicateurs de risque en temps réel décrits dans le chapitre précédent en utilisant des données individuelles locales. Le fait d'avoir des trajectoires de véhicules permet d'étudier l'évolution des valeurs de ces indicateurs dans le temps et dans l'espace.

V.2.1 Evolution des valeurs des indicateurs par véhicule

Les trajectoires de véhicules permettent, pour chaque véhicule en particulier, de connaître l'évolution dans le temps des valeurs des indicateurs de risque. L'évolution de l'état de risque associé à une situation donne également une idée de ce risque. En effet, si par exemple la valeur du TTC d'un véhicule est faible et qu'elle va en diminuant, alors la situation sera plus risquée que si la valeur de son TTC est aussi faible mais qu'elle va en augmentant. Le même raisonnement est possible pour les autres indicateurs de risque en temps réel. Ainsi, nous pourrions, après avoir fixé un seuil pour chaque indicateur de risque, dire qu'une situation est risquée si la valeur de l'indicateur dépasse ce seuil et que l'indicateur donne une valeur encore plus risquée pour le même véhicule et à l'instant suivant.

Par exemple, sur la voie 1, calculons, pour chaque ligne i , la valeur de l'IBTR. En notant respectivement $V(i)$ et $DIV(i)$ la vitesse en mètres par seconde et la distance intervéhiculaire en mètres correspondant à la ligne i ,

$$IBTR(i) = \max(0, \log_2(V(i)^2 / (2 \times DIV(i) \times 6.25)))$$

Nous plaçons ensuite sur un diagramme espace-temps les lieux et instants pour lesquels la situation est risquée. Dans un premier temps, nous tenons compte de toutes les situations pour lesquelles la valeur de l'IBTR dépasse un certain seuil, puis nous tenons compte uniquement des situations pour lesquelles la valeur de l'IBTR dépasse ce même seuil, et dont la valeur à l'instant suivant pour le même véhicule est plus élevée. Par exemple, dans le cas des données NGSIM et sur la voie 1, si ce seuil est égal à 1, le nombre de situations risquées sur la section considérée entre 8h05 et 8h20 passerait de 60 à 21.

V.2.2 Calcul du TET et du TIT

Un autre moyen d'appréhender le risque de collision avant-arrière à partir des trajectoires de véhicules est d'utiliser des indicateurs tels que le TET (Time Exposed Time-to-Collision) et le TIT (Time Integrated Time-to-Collision), définis au chapitre 2.1.1.3.

Le calcul de chacun de ces deux indicateurs nécessite au préalable le calcul du TTC pour chaque ligne, c'est-à-dire pour chaque véhicule à chaque instant. Pour chaque ligne i , nous nous servons encore ici des données de la ligne j correspondant au véhicule précédant le véhicule considéré à l'instant considéré. Notons respectivement $V(i)$, $L(i)$ et $DIV(i)$ la vitesse, la longueur du véhicule et la distance intervéhiculaire correspondant à la ligne i . Le TTC se calcule de la façon suivante :

Si $V(i) \leq V(j)$, alors $TTC(i) = +\infty$

Sinon, $TTC(i) = (DIV(i) - L(j)) / (V(i) - V(j))$

Le calcul du TTC donne lieu à quelques valeurs aberrantes : Certaines valeurs de TTC sont négatives, ce qui correspond, d'après la définition du TTC, à une distance intervéhiculaire plus petite que la longueur du véhicule de devant. Ceci ne concerne que 0.2% des véhicules.

Le TET se calcule sur une section donnée et sur une période donnée. Il dépend :

d'une valeur seuil de TTC notée α ,

de la section $n=[d_n \ d_{n+1}]$ sur laquelle il est calculé,

de la période $m=[t_m \ t_{m+1}]$ sur laquelle il est calculé.

Le TET d'un véhicule correspond au temps durant lequel son TTC est inférieur à α . Le TET global est la somme des TET de tous les véhicules se trouvant en au moins un point de la section considérée durant au moins un instant de la période considérée.

Dans le cas des données NGSIM, nous cherchons le nombre de lignes i pour lesquelles :

$$TTC(i) \leq \alpha,$$

$$d_n \leq Y(i) < d_{n+1},$$

$$t_m \leq t(i) < t_{m+1}.$$

Comme chaque ligne correspond au comportement d'un véhicule durant un instant d'un dixième de seconde, nous calculons le nombre de lignes, le $TET(m,n,\alpha)$ est égal au nombre de lignes remplissant simultanément les trois conditions précédentes multiplié par un dixième de seconde.

Afin d'avoir une bonne représentation de la propagation spatiale et temporelle du risque, il convient de prendre des périodes et des sections assez courtes, ce qui permet d'avoir un nombre de valeurs de TET calculées assez important. Nous choisissons un intervalle de temps Δt égal à 1 seconde, ce qui crée un découpage du temps en 900 périodes. Pour la section Δy , nous prenons 12 mètres, car c'est la plus petite valeur qui remplit la condition $CFL \frac{\Delta y}{\Delta t} \geq V_{max}$. Les valeurs seuils de TTC utilisées dans la littérature se situent entre 3 et 5 secondes (Hirst et Graham, 1997). Nous prenons ici un seuil de 5 secondes.

Fig V.21 montre les valeurs de TET sur la voie 1.

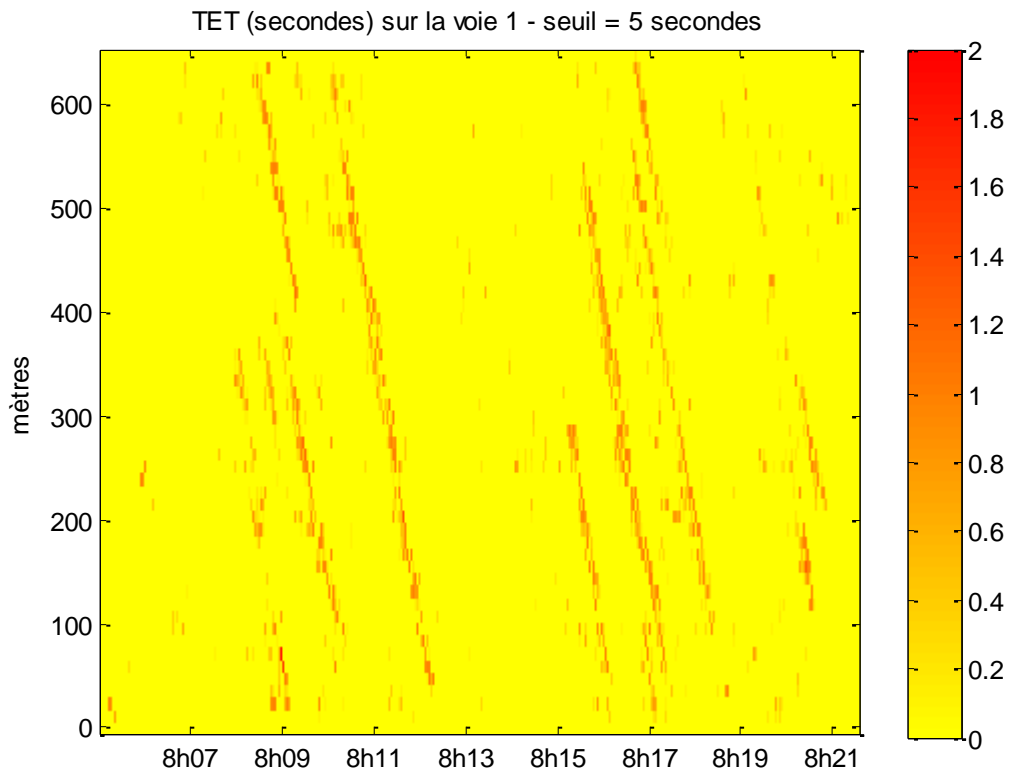


Fig. V.21 TET sur la voie 1

Pour une meilleure interprétation de ce graphique, il peut être intéressant de le superposer aux trajectoires de véhicules (Fig V.22).

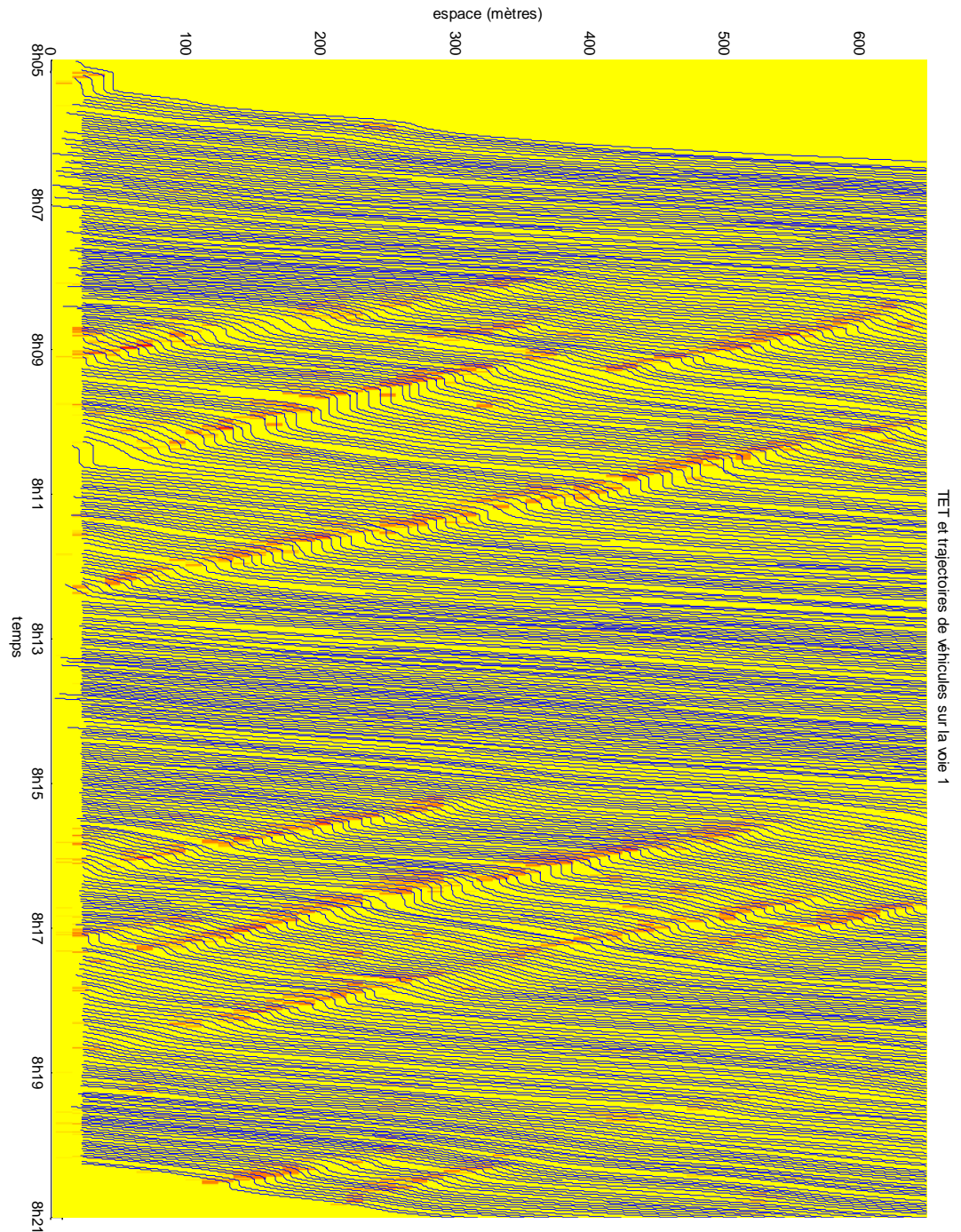


Fig. V.22 TET et trajectoires de véhicules sur la voie 1

Cette représentation graphique montre que, d'après le TET en tant qu'indicateur de risque, les lieux et instants présentant le plus grand risque de collision avant-arrière sont ceux correspondant au commencement d'une congestion. Ce résultat est plutôt logique car au début d'une congestion, les véhicules ralentissent fortement les uns après les autres, et donc le risque de collision avant-arrière est élevé.

Si nous faisons varier le seuil de TTC, les graphiques présentant les valeurs du TET dans un diagramme espace-temps ont le même profil, mais avec des valeurs maximales de TET plus ou moins élevées, comme le montre par exemple Fig. V.23 avec un seuil égal à 3 secondes.

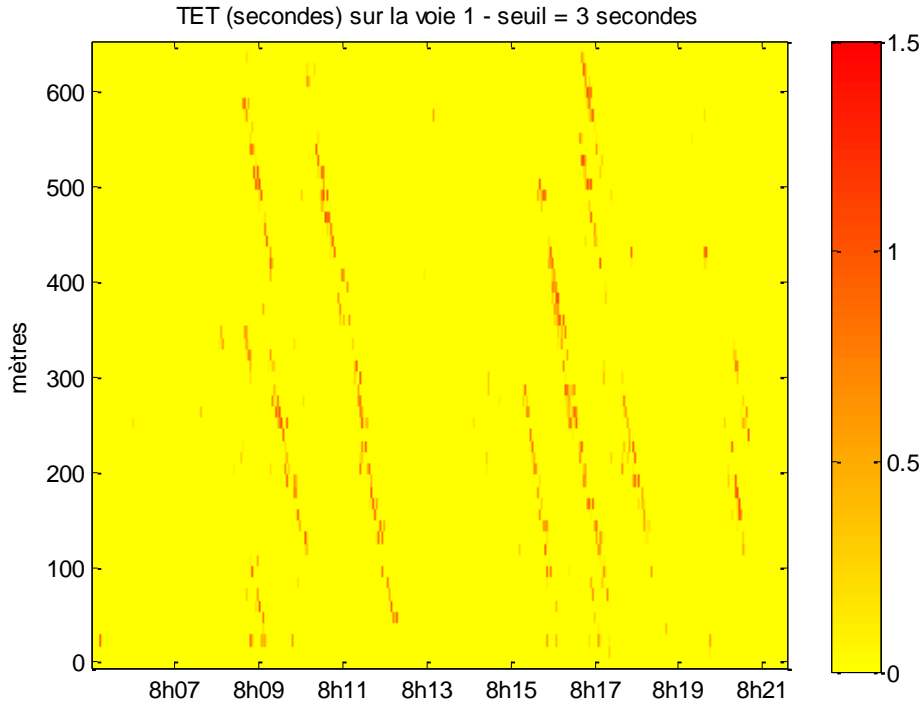


Fig. V.23 TET sur la voie 1 avec un seuil de TTC de 3 secondes

Le même type de graphe peut permettre de représenter les valeurs du TIT, une variante du TET, sur chacune des voies. A l'instar du TET, le TIT se calcule sur une période $[t_m \ t_{m+1}]$ et sur une section $[d_n \ d_{n+1}]$, et il dépend d'un seuil de TTC noté α séparant les situations jugées risquées des situations jugées sûres. Pour un véhicule i , le TIT représente la quantité suivante :

$$TIT(i) = \int_{t_m}^{t_{m+1}} (\alpha - TTC(t)) \delta(t) dt \quad \text{où } \delta(t) = \begin{cases} 0 & \text{si } TTC(t) > \alpha \\ 1 & \text{si } TTC(t) \leq \alpha \end{cases}$$

Le TIT global est ensuite la somme des valeurs des TIT de tous les véhicules se trouvant en au moins un point de la section considérée à au moins un instant de la période considérée.

Dans le cas des données NGSIM, une ligne correspond aux informations relatives à un véhicule particulier durant un dixième de seconde. Le TIT sur la section $[d_n \ d_{n+1}]$ et sur la période $[t_m \ t_{m+1}]$ est égal à $\sum \frac{1}{10} \times (\alpha - TTC(i))$ pour les lignes i telles que :

$$TTC(i) \leq \alpha,$$

$$d_n \leq Y(i) < d_{n+1},$$

$$t_m \leq t(i) < t_{m+1}.$$

De même que le TET, nous calculons le TIT séparément sur chacune des voies, avec une valeur seuil de TTC égale à 5 secondes, sur des sections de 12 mètres et sur des périodes d'une seconde.

La représentation graphique des valeurs du TIT superposées aux trajectoires de véhicules nous permet d'aboutir aux mêmes conclusions que la représentation des TET.

Par ailleurs, le nombre de situations déclarées risquées par ces deux indicateurs est plus important sur la voie la plus à gauche (voie rapide) que sur les autres voies. L'étude précédente a entraîné un découpage du diagramme espace-temps en 54670 cellules, et nous pouvons fixer des seuils de TET et de TIT permettant de distinguer les cellules présentant un risque de collision avant-arrière des cellules sûres. Nous fixons ce seuil de 1 seconde pour le TET, car la durée correspondant à une cellule est d'une seconde, et donc une valeur du TET supérieure à une seconde correspond à au moins un véhicule ayant un comportement risqué durant toute la durée de la cellule. Pour le TIT, nous choisissons un seuil de 3 secondes². Ceci correspond à un véhicule ayant un TTC inférieur à 2 secondes durant toute la durée de la cellule. Le tableau 2 présente, pour chaque voie, le nombre de cellules pour lesquelles le TET et le TIT dépassent ces seuils. Ces nombres sont deux fois plus élevés pour la voie 1 que pour les autres voies.

Fig. V.24 nombres de cellules pour lesquelles le seuil de TET ou de TIT est dépassé

	voie 1	voie 2	voie 3	voie 4	voie 5
nombre de TET > 1 seconde	51	27	27	32	22
nombre de TIT > 3 secondes ²	98	49	52	52	56

V.2.3 Utilisation des indicateurs IBTR et PBTR

Le TET et le TIT représentent des sommes des valeurs d'un indicateur de risque en temps réel se calculant pour chaque véhicule et à chaque instant : le TTC. Il est possible de mettre en œuvre le même genre d'analyse à partir d'autres indicateurs de risque en temps réel, comme par exemple l'IBTR et le PBTR. Ces indicateurs sont définis comme étant nuls lorsque la situation est sûre, et de valeurs croissantes au fur et à mesure qu'elle devient risquée. Nous pouvons donc calculer, sur chaque période et sur chaque section, la somme des valeurs de l'IBTR ou la somme des valeurs du PBTR. Plus cette somme est importante, plus le risque de collision avant-arrière sur la section et la période considérée sera élevé.

Pour chaque ligne i , la valeur de l'IBTR se calcule en fonction de la vitesse $V(i)$ en mètres par seconde et de la distance intervéhiculaire $DIV(i)$ en mètres.

$$IBTR(i) = \max(0, \log_2(V(i)^2 / (2 \times DIV(i) \times 6.25)))$$

Le PBTR relatif à la ligne i dépend de l'IBTR(i) et du PBTR relatif à la ligne j , correspondant au véhicule se trouvant devant le véhicule considéré à l'instant considéré. Le calcul du PBTR(j) doit être fait avant celui du PBTR(i).

$$PBTR(i) = 0 \quad \text{si } IBTR(i) = 0$$

$$PBTR(i) = PBTR(j) + IBTR(i) \quad \text{sinon}$$

Pour chaque voie, les sommes des valeurs des IBTR ou des PBTR dans chaque cellule, caractérisée par une section de 12 mètres et une période de 1 seconde, sont reportées dans un diagramme espace-temps. Fig. V.25 et Fig. V.26 présentent le profil de ces valeurs pour la voie 5.

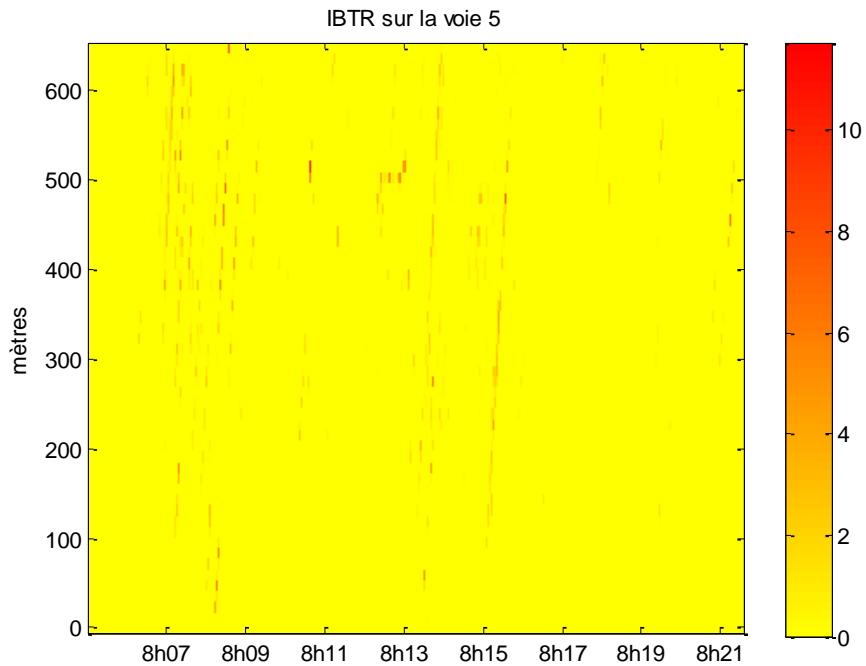


Fig. V.25 Somme des valeurs de l'IBTR sur la voie 5

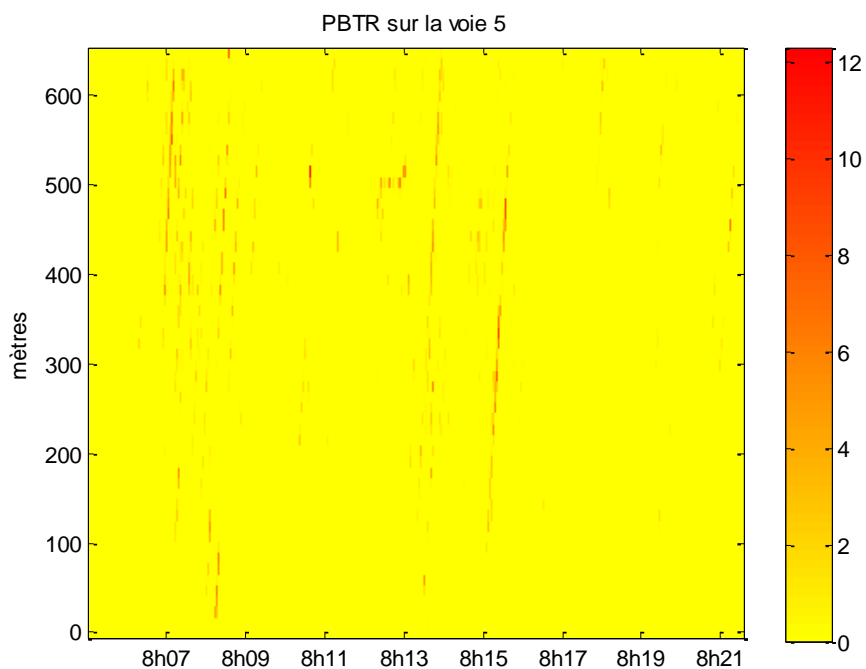


Fig. V.26 Somme des valeurs de PBTR sur la voie 5

Les répartitions des valeurs de ces deux indicateurs dans le diagramme espace-temps ont des profils semblables : Les lignes regroupant les cellules les plus risquées se confondent avec les lignes représentant des trajectoires de véhicules. Ainsi, d'après l'IBTR et le PBTR, si la situation d'un véhicule est risquée, alors elle le sera souvent le long de sa trajectoire.

Les états de risque associés aux différentes voies sont semblables, et ces deux indicateurs ne permettent pas ici de dire que le risque de collision avant-arrière est plus important sur

une voie que sur une autre.

Par contre, même si les lignes composées des cellules présentant les risques les plus importants se confondent avec des trajectoires de véhicules, le risque n'est pas uniformément réparti dans l'espace. En effet, les valeurs des IBTR et des PBTR sont plus élevées à la fin de la section qu'au début. Fig. V.27, qui montre la somme des valeurs de l'IBTR des cellules selon leur localisation dans l'espace, illustre ce phénomène.

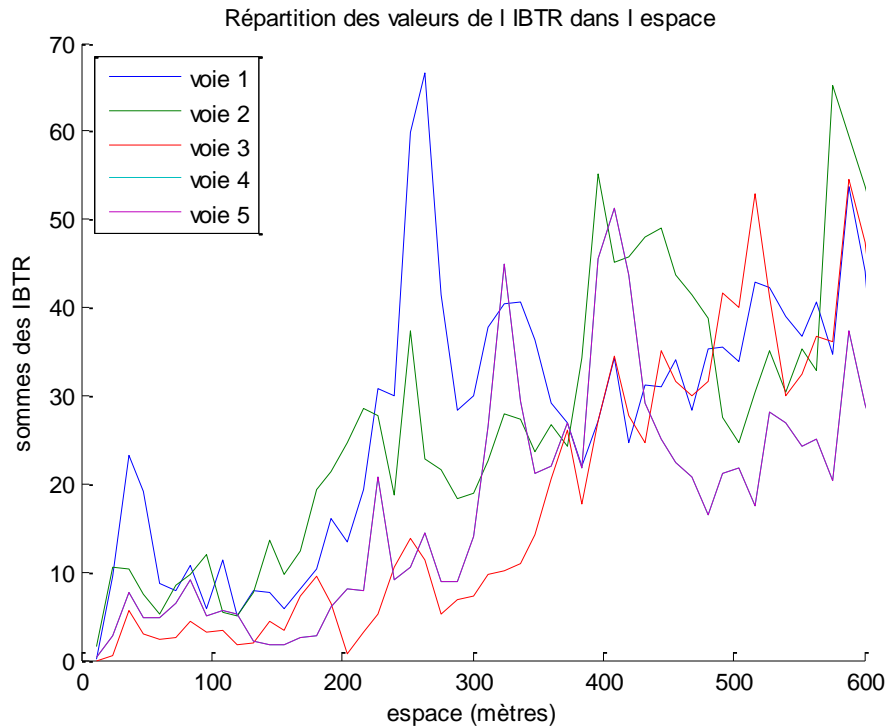


Fig. V.27 Répartition des valeurs de l'IBTR dans l'espace

V.3 Etude du risque de collision latérale

Nous avons vu que le risque de collision latérale sur voie rapide est directement lié à la fréquence des changements de voie. Le fait d'avoir des trajectoires de véhicules permet donc d'étudier ce risque.

Nous pouvons utiliser l'indicateur de risque OAFR. Cet indicateur se calcule sur une section donnée et sur une période donnée.

Sur chaque voie, nous calculons d'abord l'"Average Flow Ratio":

$$AFR_i = \frac{q_{i-1}}{q_i} \times \frac{NL_{i-1,i}}{NL_{i-1,i} + NL_{i-1,i-2}} + \frac{q_{i+1}}{q_i} \times \frac{NL_{i+1,i}}{NL_{i+1,i} + NL_{i+1,i+2}}$$

où :

- les voies (i-1) et (i+1) sont les voies adjacentes à la voie i si elles existent,
- q_i est le débit moyen sur la voie i durant la période considérée,
- $NL_{i,j}$ est le nombre de changements de voie de la voie i vers la voie j durant la période considérée et sur la section considérée.

L'indicateur OAFR est ensuite défini comme la moyenne géométrique des AFR correspondant à chaque voie. Plus la valeur de cet indicateur est élevée, plus le risque de collision avant-arrière est important.

Le problème rencontré lors de l'étude de cet indicateur sur les données NGSIM est que si la période considérée est trop courte ou que la section considérée est trop petite, il se peut que certains nombres de changements de voies soient nuls, et donc certaines valeurs de l'AFR aussi, ce qui entraînerait que l'OAFR serait nul également. Ces cas ne sont pas intéressants, car, au vu de sa définition, la nullité de l'OAFR ne signifie pas la sûreté de la situation associée, mais signifie que le risque associé à la situation n'est pas mesurable à l'aide de cet indicateur.

Dans un premier temps, nous nous intéressons à la section de 600 mètres dans son ensemble. La plus petite période sur laquelle nous pouvons calculer des valeurs d'OAFR représentatives du risque de collision latérale est de 2 minutes (Fig. V.28).

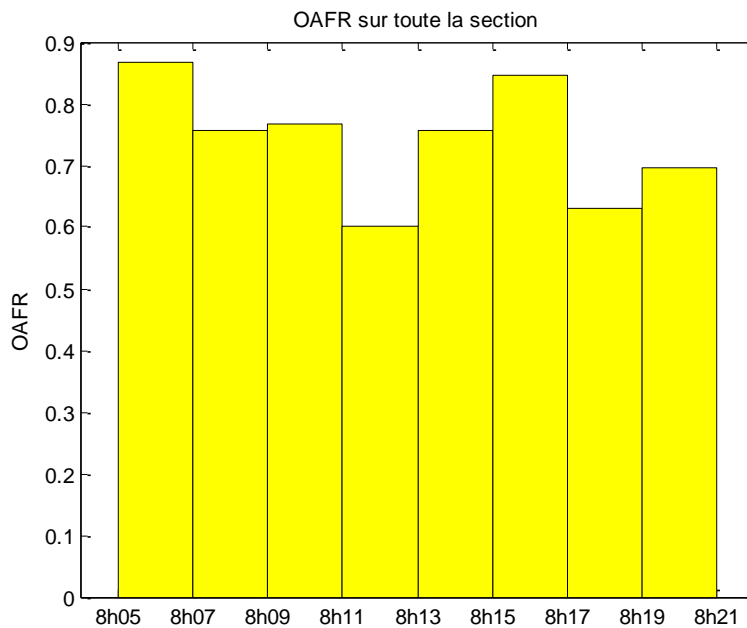


Fig. V.28 OAFR sur toute la section

Le risque de collision latérale est le plus élevé entre 8h05 et 8h07 et entre 8h15 et 8h17, et il est le plus faible entre 8h11 et 8h13.

Nous nous intéressons ensuite à la période de 15 minutes dans son ensemble, et nous essayons d'évaluer le risque de collision avant-arrière sur des sections de même longueur. La longueur minimale qu'il faut prendre pour n'avoir aucun OAFR nul est de 200 mètres (Fig. V.29).

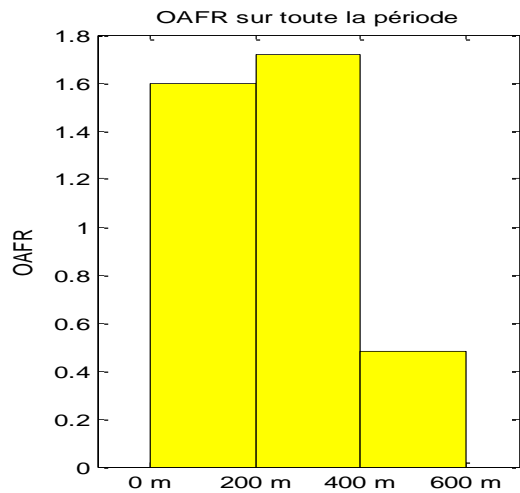


Fig. V.29 OAFR sur toute la période

Le risque de collision latérale est le plus élevé au centre de la section. Il est le plus faible à la fin de la section.

Pour étudier parallèlement les différences spatiales et temporelles au niveau du risque de collision latérale, il convient de trouver un couple de division de l'espace et du temps qui permette de n'avoir que des OAFR non nuls. Si l'espace est divisé en plus de deux sections, aucune division du temps ne permet cela. Il faudrait donc prendre des sections égales à 300 mètres.

Ensuite, la plus petite période permettant d'avoir uniquement des OAFR strictement positifs sur des sections de 300 mètres est de 2.5 minutes. Ainsi, l'OAFR permet de décrire le risque de collision latérale sur des cellules de 300 mètres sur 2.5 minutes, ce qui correspond à une division du diagramme espace-temps en 12 cellules. Cette division est bien trop grossière et ne permet pas une bonne représentation du risque de collision latérale et de sa propagation. Nous ne pouvons donc pas utiliser l'OAFR dans ce but.

Ainsi, l'utilisation de données de type trajectoires de véhicules et d'indicateurs de risque bien construits permet d'avoir une idée, pour chaque situation, d'une part de son état de risque de collision avant-arrière, et d'autre part de son état de risque de collision latérale. Il est alors théoriquement possible de créer un indicateur mixte, caractérisant ces deux risques à la fois. Cet indicateur peut par exemple être défini comme ayant pour valeur la somme des valeurs des deux indicateurs considérés. Il peut aussi se définir à partir de seuils établis pour chacun des deux indicateurs : l'indicateur mixte est alors tel que dès que la valeur d'un des deux indicateurs dépasse son seuil, l'indicateur mixte active une alarme.

Références de l'annexe

ARCHER Jeffery. Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling : A study of urban and suburban intersections. Thèse sci. : Royal Institute of Technology, Stockholm, Suède, 2005, 254 p.

ARON Maurice, BIECHELER Marie-Berthe, PEYTAVIN Jean-François. Temps intervéhiculaires et vitesses : une cause essentielle des accidents de la route. [en ligne]. Les fiches de l'INRETS, n°32, 2004. [ref. du 27 août 2008]. Format PDF : < <http://www.inrets.fr/infos/fiches/axe1/pdf/sec32.pdf>>

BARRAUD Sylvie. Quelques réflexions et méthodes sur l'élaboration de critères ou d'indicateurs. Cours "Systémique et modélisation des systèmes", Master MEGA. Lyon : INSA. 13 mars 2008.

CHANG Gang-Len, KAO Yang-Ming. An Empirical Investigation of Macroscopic Lane-Changing Characteristics on Uncongested Multilane Freeways Transportation Research Part A 25, n° 6, 1991, p. 375-389.

CHUNG Edward. Performance Indicators needs and derivation : Single and multisource solutions. Livrable 4.1 du projet européen INTRO (Intelligent Roads), 2006, 249 p.

CHUNG Edward. Safety indicators for monitoring traffic safety. Livrable 4.3 du projet européen INTRO (Intelligent Roads), 2007, 201 p.

COHEN Simon. Systèmes et méthodes de détection automatique des incidents routiers. Synthèse INRETS n°48, septembre 2005, 112 p.

DE MOUZON Olivier, EL FAOUZI Nour-Eddin, PHAM Minh-Hai, CHUNG Edward. Road safety indicators: Results in Vaud canton. Communication au premier congrès "Road Safety and Simulation", Rome, Italie, 7-9 nov. 2007. 6 p.

DE MOUZON Olivier, PHAM Minh-Hai, EL FAOUZI Nour-Eddin, CHUNG Edward. An effective real-time proactive road traffic risk indicator based on traffic data : Compensated Platoon Braking Time Risk (CPBTR). 7ème Congrès européen et salon des systèmes de Transports Intelligents, Genève, Suisse, 4-6 juin 2008. 11 p.

DE MOUZON Olivier, EL FAOUZI Nour-Eddin, PHAM Minh-Hai, CHUNG Edward. Can safety indicators assess and monitor road traffic risk in real-time? Investigation of two safety indicators on Swiss motorways. Advances in Transportation Studies: An International Journal. Article envoyé le 7 mars 2008, 8 p.

DUPRE Guy. Accidents en courbe et caractéristiques de l'infrastructure. [en ligne]. Séminaire "Caractéristiques routières et Sécurité", 17-19 mai 2005. [ref. du 5 mai 2008]. Format PDF : < http://www.lcpc.fr/ext/pdf/sem/2005_secu_17-02.pdf>

FEDERAL HIGHWAY ADMINISTRATION. US Highway 101 Dataset. [en ligne], bulletin d'information de janvier 2007. [ref. du 18 août 2008]. Format PDF : <<http://www.tfhr.gov/about/07030.pdf>>

GARBER Nicholas J., GADIRAJU Raju. Factors affecting speed variance and its influence on accidents. Transportation Research Record, n°1213, 1989, p. 64-71.

GARBER Nicholas J., GADIRAJU Raju. Impact of differential speed limits on the speed of traffic and the rate of accidents. Transportation Research Record, n°1375, 1992, p. 44-52.

GRAS R., BENOIT M., DEFFONTAINES J.P., DURU M., LAFAGE M., LANGLET A., OSTY

P.L. Le fait technique en agronomie - Activité agricole, concepts et méthodes d'étude. Paris : L'Harmattan, 1989. 183 p.

HIRST Stephen, GRAHAM Robert. The format and presentation of collision warnings. Ergonomic and safety of intelligent drivers interfaces, 1997, p. 203-219.

HUPFER Christoph. Deceleration to Safety Time (DST) – a useful figure to evaluate traffic safety ? Conférence ICTCT, 3ème session, Lund, Suède, 1997. 6 p.

JANSSEN Wiel. From driver behaviour/driver state to accident risk. [en ligne]. Stakeholders Forum 2008. 2 avril 2008, Lyon, France. [ref. du 27 août 2008]. Format PPT : <http://www.noehumanist.org/documents/presentations_stackholders_lyon2008/07_HUMANIST-SF2008_WielJ.pdf>

KEAY Kevin, SIMMONDS Ian. Road accidents and rainfall in a large Australian city. Accident Analysis and Prevention, vol. 38, 2006, p. 445-454.

LABOUZE Eric, LABOUZE Robert. La comptabilité de l'Environnement. Revue Française de Comptabilité, n°272, novembre 1995. 92 p.

LEE Chris, ABDEL-ATY Mohamed, HSIA Liang. Potential Real-Time Indicators of Side-swipe Crashes on Freeways. Communication au Transportation Research Board, 85ème meeting annuel, Washington D.C. 22-26 jan. 2006. 16 p.

LORD Dominique, MANAR Abdelaziz, VIZIOLI Anna. Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. Accident Analysis and Prevention, vol. 37, 2005, p. 185-199.

LUCAS Jean-Paul. Indicateurs et tableaux de bord qualité en ingénierie informatique. [en ligne]. 18 avril 2002. [ref. du 17 juin 2008]. Format PPT : <www.spin.org/presentations/2002/April_SPIN_presn_french.ppt>

MARTIN Peter T., PERRIN Joseph, HANSEN Blake, KUMP Ryan, MOORE Dan. Incident Detection Algorithm Evaluation. [en ligne]. Université de l'Utah, mars 2001. [ref. du 28 août 2008]. Format PDF : <www.mountain-plains.org/pubs/pdf/MPC01-122.pdf>

MAURIZI Brigitte, VERREL Jean-Louis. Des indicateurs pour les actions de maîtrise des pollutions d'origine agricole. Ingénieries, n°30, juin 2002, p. 3-14.

MINDERHOUD Michiel M., BOVY Piet H.L. Extended time-to-collision measures for road traffic safety assessment. Accident Analysis and Prevention, vol. 33, jan. 2001, p. 89-97.

MITCHELL G., MAY A., MACDONALD A. PICABEU : a methodological framework for the development of indicators of sustainable development. The International Journal of Sustainable Development and World Ecology, vol. 2, 1995, p. 104-123.

NILSSON Gorän. Traffic safety measures and observance. [en ligne]. Swedish National Road and Transport Research Institute (VTI), 2004. [ref. du 27 août 2008]. Format PDF : <<http://www.vti.se/EPIBrowser/Publikationer%20-%20English/M951A.pdf>>

OCDE/CEMT. La gestion de la vitesse : Document de synthèse. [en ligne]. Centre de recherche sur les transports, 2007. [ref. du 27 août 2008]. Format PDF : <www.international-transportforum.org/jtrc/safety/SpeedSummaryF.pdf>

ORGANISATION MONDIALE DE LA SANTE. Rapport mondial sur la prévention des traumatismes dus aux accidents de la circulation, 2004, 217 p.

THEILLOUT Jean Noël, GUICHON Dominique, LERAY Valérie, MONMOUSSEAU Isabelle. CSA vitesse-N118 : Analyse du comportement des usagers avant et après la pose d'un contrôle section automatisé "vitesse". 5 janvier 2006, 65 p.

UNO Nobuhiro, IIDA Yasunori, ITSUBO Shinji, YASUHARA Shinji. A microscopic analysis of traffic conflict caused by lane-changing vehicle at weaving section. Communication à la 13ème conférence mini-Euro Handling Uncertainty in the Analysis of Traffic and Transportation Systems, Université Polytechnique de Bari, Italie, 10-13 juin 2002, p. 143-148.

VOGEL Katja. A comparison of headway and time to collision as safety indicators. Accident Analysis and Prevention, vol. 35, 2003, p. 427-433.

Abréviations

Concept	Signification
AFR	Average Flow Ratio
ANN	Artificial Neural Network
ASTRA	Bundesamt für Strasse
ATC	Automatic Traffic Count
BFU	Beratungsstelle für Unfallverhütung
CPBTR	Compensated Platoon Braking Time Risk
DETEC	Federal Department of the Environment, Transport, Energy and Communications
DMD	Délai Moyen de Détection
DUI	Driving Under the Influence of alcohol or other drugs
EPFL	Ecole Polytechnique Fédérale de Lausanne
FEDRO	Federal Road Office
FSO	Federal Statistical Office
FUSAIN	Fusion of safety indicator
HGV	Heavy Goods Vehicle
IBTR	Individual Braking Time Risk
IFSTTAR	Institut français des sciences et technologies des transports, de l'aménagement et des réseaux
INTRO	Intelligent Road
JDL	Joint Directors of Laboratories
LAVOC	Laboratoire de voies de circulation
LICIT	Laboratoire Ingénierie Circulation Transport
MARS	Multi-Agent Real-time Simulator
NGSIM	Next Generation SIMulation
NHTSA	National Highway Traffic Safety Administration
NTS	Non-crash Traffic Situation
OAFR	Overall Average Flow Ratio
OCDE	Organisation de coopération et de développement économiques
PBTR	Platoon Braking Time Risk
PET	Post-Encroachment-Time
PICUD	Potential Index for Collision with Urgent Deceleration
PTS	Pre-crash Traffic Situation
RC	Risk Criteria
RI	Risk Indicator
RIM	Risk Identification Model
SI	Safety Indicator
SOSL	Speed Over Speed Limit
TD	taux de détection
TET	Time Exposed Time-to-Collision
TFA	taux de fausses alarmes
TIT	Time Integrated Time-to-Collision
TR	Traffic Regime
TS	Traffic Situation
TTC	Time-To-Collision
UK	United Kingdom

Concept	Signification
vph	vehicles per hour

Bibliographie

-
- accident, a. C. i. N. (2011), A Crash Is No Accident.
-
- Aguero-Valverde, J. & Jovanis, P. (2008), Analysis of Road Crash Frequency with Spatial Models, *Transportation Research Record: Journal of the Transportation Research Board*, 2061(-1), 55-63.
-
- Aguero-Valverde, J. & Jovanis, P. P. (2006), *Spatial Analysis of Fatal and Injury Crashes in Pennsylvania*, Elsevier, Oxford, ROYAUME-UNI.
-
- Arnedt, J. T., Wilde, G. J. S., Munt, P. W. & MacLean, A. W. (2001), How Do Prolonged Wakefulness and Alcohol Compare in the Decrements They Produce on a Simulated Driving Task?, *Accident Analysis & Prevention*, 33(3), 337-344.
-
- ARON, M., BIECHELER, M.-B. & PEYTAVIN, J.-F. (2007), Temps Intervéhiculaires Et Vitesses : Une Cause Essentielle Des Accidents De La Route.
-
- Bloche, I. (2002), Techniques De Fusion De Données Et Applications a L'exploitation Du Trafic, *SURPRISE 96 Journal*, Recueil multiforme de données en circulation routière 59-90.
-
- Blum, J. J. & Eskandarian, A. (2006), Managing Effectiveness and Acceptability in Intelligent Speed Adaptation Systems, *Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE*, pp. 319-324.
-
- Boschung. (2010), *Measuring Stations*.
-
- BPU. (2011), *Statistics on Non-Occupational Accidents and the Level of Safety in Switzerland. Status 2010: Road Traffic, Sports, Home and Leisure*.
-
- Breiman, L. (2001), *Random Forests, Machine Learning*, 45(1), 5-32.
-
- DfT. (2009), *Reported Road Casualties Great Britain: 2008. Annual Report*, Department for Transport, London.
-
- Donnell, E. T. & Mason, J. J. M. (2006), Methodology to Develop Median Barrier Warrant Criteria, *Journal of Transportation Engineering*, 132(4), 269-281.
-
- EuroRAP. (2009), *European Campaign for Safe Road Design*.
-
- FEDRO. (2009), *Instruction: Traffic Counting Stations (Title Translated from French)*, Swiss Federal Roads Office.
-
- FEDRO. (2009), *Roads & Traffic - Facts & Figures 2009 : Annual Publication of the Swiss Federal Roads Office.*, Swiss Federal Roads Office
-
- FSO. (2005), *Road Traffic Accidents - Instructions (Title Translated from French)*.
-
- FSO. (2005), *Road Traffic Accidents - Survey Form (Title Translated from French)*, Federal Statistics Office.
-
- FSO. (2010), *Definitions (Title Translated from French)*, Federal Statistics Office.
-
- Golob, T. F., Recker, W. W. & Alvarez, V. M. (2004), Freeway Safety as a Function of Traffic Flow, *Accident Analysis & Prevention*, 36(6), 933-946.
-
- Hayward, J. C. (1971), *Near Misses as Ameasure of Safety at Urban Intersections*, The Pensilvania State University, Department of Civil Engineering.
-
- Horst, R. v. d. & Hogema, J. (1993), *Time-to-Collision and Collision Avoidance Systems*.
-
- Klein, L. (2004), *Sensor and Data Fusion: A Tool for Information Assessment and Decision Making*, SPIE Press Book.
-
- Klingender, M., Ramakers, R. & Henning, K. (2009), In-Depth Safety Impact Study on Longer and/or Heavier Commercial Vehicles in Europe, *Power Electronics and Intelligent Transportation System (PEITS), 2009 2nd International Conference on*, pp. 368-373.
-

Lord, D. & Mannering, F. (2010), The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives, *Transportation Research Part A: Policy and Practice*, 44(5), 291-305.

Minderhoud, M. M. & Bovy, P. H. L. (2001), Extended Time-to-Collision Measures for Road Traffic Safety Assessment, *Accident Analysis & Prevention*, 33(1), 89-97.

Mulder, M., Abbink, D. A. & Boer, E. R. (2008), The Effect of Haptic Guidance on Curve Negotiation Behavior of Young, Experienced Drivers, *Systems, Man and Cybernetics*, 2008. SMC 2008. IEEE International Conference on, pp. 804-809.

NHTSA. (1997), *Crashes Aren't Accidents*.

Pande, A. (2005), Estimation of Hybrid Models for Real-Time Crash Risk Assessment on Freeways, Department of Civil and Environmental Engineering, University of Central Florida, Orlando, FL 32816-2450, United States.

Pham, M.-H., Mouzon, O. d., Chung, E. & Faouzi, N.-E. E. (2008), Sensitivity of Road Safety Indicators in Normal and Crash Cases, 10th International Conference on Application of Advanced Technologies in Transportation (AATT), Athens, Greece.

RoadPeace. (2011), *Road Crash - Not Road 'Accident'*.

Rumar, K. (1985), *The Role of Perceptual and Cognitive Filters in Observed Behavior, Human Behavior and Traffic Safety*, Plenum Press, New York, pp. 151-165.

Rzepecki-Smith, C. I., Meda, S. A., Calhoun, V. D., Stevens, M. C., Jafri, M. J., Astur, R. S. & Pearlson, G. D. (2010), Disruptions in Functional Network Connectivity During Alcohol Intoxicated Driving, *Alcoholism: Clinical and Experimental Research*, 34(3), 479-487.

SAN. (2011), *Trial Driving License Et 2-Phases Training (Translated from French)*, Service des automobiles et de la navigation, Lausanne.

Steinberg, A. N., Bowman, C. L. & White, F. E. (1999), Revisions to the Jdl Data Fusion Model.

Stergiou, C. & Siganos, D. (1996), *Neural Networks*, SURPRISE 96 Journal.

Stine, J. S., Hamblin, B. C., Brennan, S. N. & Donnell, E. T. (2010), Analyzing the Influence of Median Cross-Section Design on Highway Safety Using Vehicle Dynamics Simulations, *Accident Analysis & Prevention*, 42(6), 1769-1777.

Umedu, T., Isu, K., Higashino, T. & Toh, C. K. (2010), An Intervehicular-Communication Protocol for Distributed Detection of Dangerous Vehicles, *Vehicular Technology, IEEE Transactions on*, 59(2), 627-637.

Wang, C., Quddus, M. A. & Ison, S. G. (2009), Impact of Traffic Congestion on Road Accidents: A Spatial Analysis of the M25 Motorway in England, *Accident Analysis & Prevention*, 41(4), 798-808.

White, F. E. (1991), *Data Fusion Lexicon*.

Yu, S., de Moor, B. & Moreau, Y. (2008), Clustering by Heterogeneous Data Fusion: Framework and Applications, *Learning from Multiple Sources Workshop (NIPS 2008)*.

Clôture du projet



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für
Umwelt, Verkehr, Energie und Kommunikation UVEK
Bundesamt für Strassen ASTRA

FORSCHUNG IM STRASSENWESEN DES UVEK Formular Nr. 3: Projektabschluss

Strassen, Brücken, Tunnel

erstellt / geändert am: 20/05/2012

Grunddaten

Projekt-Nr.: ASTRA 2006/014

Projekttitel: Fusion des indicateurs de sécurité routière

Enddatum: 30/10/2009

Texte

Zusammenfassung der Projektergebnisse:

This study aims to contribute to improve motorway traffic safety by developing models that can identify real-time traffic risk. The approach is to classify potential risk indicators and test them with historical data including traffic, meteorological and crash data. First, risk indicators are short-listed based on the test of their sensitivity with real data. After that, two types of traffic conditions are distinguished: the conditions leading to crashes, called pre-crash cases and the conditions when no crash occurs, called non-crash cases. Pre-crash and non-crash cases are matched into groups such that under each group, pre-crash and non-crash cases are comparable. Each of such groups is considered as a traffic regime. Single-variable models are then developed under each traffic regime and provide inputs to the next step when a final model is developed under each traffic regime by fusing results of single-variable models.

A further step to predict short-term traffic crash risk is undertaken by combining the risk status of several most recent traffic cases in real-time. The test with historical data show that the developed prediction model can perform with high accuracy.

Although more test needs to be done, the system of short-term traffic crash risk prediction prove the potential of assisting traffic operators or drivers in a risk-oriented traffic management scheme.



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für
Umwelt, Verkehr, Energie und Kommunikation UVEK
Bundesamt für Strassen ASTRA

Zielerreichung:

This study proposes a methodology to make use of risk indicators to provide traffic crash risk assessment and short-term prediction in a real-time framework. The sensitivity test of risk indicators was done using real data indicating how good each of risk indicators can do in assessing traffic crash risk. Thereafter, the performance of risk indicators is combined to produce fused models with higher performance.

The achievement in this study includes:

1. A methodology to combine the performance of risk indicators by fusing them together.
2. A methodology to divide traffic conditions into traffic regimes such that typical traffic conditions leading to rear-end and sideswipe crashes can be identified.
3. With a selected study site, models for identifying real-time traffic crash risk are developed. These models can provide additional information to traffic operators in order to better monitor traffic risk at the location of the study site.
4. A list of critical factors who can influence the occurrences of crashes. With this list, appropriate preventive measure can be identified and tested.

Folgerungen und Empfehlungen:

Although there is still a lot of work needed to be done, the current study shows the very encouraging results. The developed models for the selected study present high accuracy in identifying traffic risk via the test with historical data.

The performance of developed models can be further improved if more other risk indicators are considered, which depend on the availability of data at the study site. There exist also potentials to improve the performance of developed models by optimizing design choices at each methodological steps. With current developed models and the achievement stated above, the following conclusions and recommendations can be withdrawn:

1. Although crashes are rare events, there are always typical patterns of traffic conditions, called risky patterns, that can lead to crashes. The difficulty to identify risky patterns in real-time is that there are many traffic patterns similar to risky patterns yet do not end up with crashes.
2. Speed - related factors are more frequently identified to be involved in crash occurrences. An example is the speed difference between two lanes (in case of 2x2 motorway sections) which stimulates more risky traffic patterns. As such a preventive measure to harmonize the speed on both of the lanes might reduce the crash risk.
3. The trade-off between false alarm rate and detection rate is interesting as there is no way to maximize both rates simultaneously. However, false alarm rate should be minimized as much as possible in order to maintain the confidence of road users to the system. As such, detection rate is accepted to be lower.
4. The developed models can provide useful information to traffic operators in traffic operation center or to on-road traffic signal to directly guide traveller.

Publikationen:

1. Pham, M.-H., Faouzi, N.-E.E., Dumont, A.-G., 2011. Real-time identification of risk-prone traffic patterns taking into account weather conditions 90th Transportation Research Board annual meeting, Washington DC.
2. Pham, M.-H., Bhaskar, A., Chung, E., Dumont, A.-G., 2011. Methodology for Developing Real-time Motorway Traffic Risk Identification Models Using Individual Vehicle Data, 90th Transportation Research Board annual meeting, Washington DC.
3. Pham, M.-H., Bhaskar, A., Dumont, A.-G., 2010. Vers un modèle proactif pour identifier des risques de trafic, Conférence Sécurité routière: prévention des risques et aide à la conduite, PRAC2010 Paris, France.
4. Pham, M.-H., Bhaskar, A., Chung, E., Dumont, A.-G., 2010. Random Forest Models for Identifying Motorway Rear-End Crash Risks Using Disaggregate Data, 13th International IEEE Annual Conference on Intelligent Transportation Systems, Madeira Island, Portugal.
5. Pham, M.H., Bhaskar, A., Chung, E., Dumont, A.G., 2010. Towards a pro-active model for identifying motorway traffic risks using individual vehicle data from double loop detectors, Road Transport Information and Control Conference and the ITS United Kingdom Members' Conference (RTIC 2010) - Better transport through technology, IET, pp. 1-9.
6. Pham, M.-H., Chung, E., Dumont, A.-G., 2009. Methodology for traffic risks identification, Swiss Transportation Research Conference Ascona, Ticino, Switzerland.
7. Pham, M.-H., Mouzon, O.d., Chung, E., Faouzi, N.-E.E., 2008. Sensitivity of road safety indicators in normal and crash cases, 10th International Conference on Application of Advanced Technologies in Transportation (AATT), Athens, Greece.
8. Pham, M.-H., Mouzon, O.d., Chung, E., Dumont, A.-G., 2008. Sensitivity of risk indicators under motorway traffic regimes clustered by self-organizing map, 7th European Congress on ITS, Geneva, Switzerland.

Der Projektleiter/die Projektleiterin:

Name: André-Gilles

Vorname: Dumont

Amt, Firma, Institut: LAVOC, EPFL

Unterschrift des Projektleiters/der Projektleiterin:



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für
Umwelt, Verkehr, Energie und Kommunikation UVEK
Bundesamt für Strassen ASTRA

FORSCHUNG IM STRASSENWESEN DES UVEK

Strassen, Brücken, Tunnel

Formular Nr. 3: Projektabschluss

Beurteilung der Begleitkommission:

Beurteilung:

L'étude menée aborde la problématique de la prévision du risque dans un flux de véhicules. En traitant un grand nombre de données réelles, les chercheurs arrivent à une évaluation du risque potentiel sur le court terme.
Le rapport décrit très en détail les approches possibles et un modèle a été développé pour quantifier ce risque. L'approche scientifique est rigoureuse et le travail de traitement des données peut servir de modèle pour d'autres études.
On peut regretter l'utilisation de l'anglais pour le centre du rapport et le français pour les annexes. Cette situation découle de la coopération de deux centres de recherche : le LAVOC et l'IFSTTAR.
La difficulté de prendre en compte les conditions météorologiques dans l'évaluation du risque est ici mise en évidence.
La recherche montre aussi les limites de la prévision du risque. D'une part on peut développer des modèles fiables pour évaluer le risque de collision entre plusieurs véhicules, mais difficilement pour les accidents de véhicules isolés. D'autre part il s'applique surtout aux zones à haut risque.

Umsetzung:

Le modèle développé doit permettre d'appliquer les résultats dans deux cas pour informer préventivement :
-Le conducteur qu'il est à un moment précis en situation de risque
-L'opérateur du centre de gestion du trafic que la situation sur un tronçon donné est en cours de dégradation et que le risque augmente. Il peut alors agir avec des mesures proactives de gestion du trafic.
Le modèle peut aussi servir à évaluer les effets de la mise en service d'un système de gestion du trafic ou d'un nouveau aménagement.
L'application ne fait pas partie de l'étude mais le modèle développé est à disposition pour des tests en situation réelle.

weitergehender Forschungsbedarf:

Les thèmes de la sécurité routière et de la gestion du trafic restent une priorité. Les applications télématiques ouvrent de nouvelles perspectives de recherche.
Aucun besoin supplémentaire direct n'est identifié à ce stade.

Einfluss auf Normenwerk:

Cette recherche n'a pas d'influence directe sur la normalisation.

Der Präsident/die Präsidentin der Begleitkommission:

Name: Patrick

Vorname: Maillard

Amt, Firma, Institut: Robert-Grandpierre et Rapp SA

Unterschrift des Präsidenten/der Präsidentin der Begleitkommission:

Index des rapports de recherche en matière de route

Bericht-Nr.	Projekt Nr	Datum	Tite
1323	VSS 2008/205	2011	Ereignisdetektion im Strassentunnel
1323	VSS 2008/205	2011	Ereignisdetektion im Strassentunnel Détection d'incidents dans les tunnels routiers (Seulement disponible en allemand)
1326	VSS 2006/207	2011	Erfolgskontrolle Fahrzeugrückhaltesysteme
1326	VSS 2006/207	2011	Erfolgskontrolle Fahrzeugrückhaltesysteme Contrôle de l'efficacité des dispositifs de retenue de véhicules (Seulement disponible en allemand)
1328	VSS 2005/302	2011	Grundlagen zur Quantifizierung der Auswirkungen von Sicherheitsdefiziten
1328	VSS 2005/302	2011	Grundlagen zur Quantifizierung der Auswirkungen von Sicherheitsdefiziten Principes pour la quantification des effets des déficits de la sécurité (Seulement disponible en allemand)
1331	VSS 2005/501	2011	Rückrechnung im Strassenbau Analyse inverse pour la construction routière (Seulement disponible en allemand)
1331	VSS 2005/501	2011	Rückrechnung im Strassenbau
1332	VSS 2006/905	2011	Standardisierte Verkehrsdaten für das verkehrsträgerübergreifende Verkehrsmanagement
1332	VSS 2006/905	2011	Standardisierte Verkehrsdaten für das verkehrsträgerübergreifende Verkehrsmanagement Standardisation des données de trafic pour gestion intermodale du trafic (Seulement disponible en version allemand)

1335	VSS 2007/502	2011	Stripping bei lärmmindernden Deckschichten unter Überrollbean-spruchung im Labormaßstab. Désenrobage des enrobés peu bruyants des couches de roulement sous sollicitation de roulement en laboratoire (Seulement disponible en version allemand)
1335	VSS 2007/502	2011	Stripping bei lärmmindernden Deckschichten unter Überrollbean-spruchung im Labormaßstab.
1343	VSS 2009/903	2011	Les technologies de base pour l'enregistrement automatique des usagers de moyens de transports (Seulement disponible en version allemand)
1343	VSS 2009/903	2011	Basistechnologien für die intermodale Nutzungserfassung im Personenverkehr
1344	VSS 2009/709	2011	Initialprojekt für das Forschungspaket "Nutzensteigerung für die Anwender des SIS"
1344	VSS 2009/709	2011	Projet initial pour le paquet de recherche "Augmentation de l'utilité pour les usagers du système d'information de la route" (Seulement disponible en version allemand)
1334	ASTRA 2009/009	2011	Was treibt uns an? Antriebe und Treibstoffe für die Mobilität von Morgen
1334	ASTRA 2009/009	2011	Was treibt uns an? Antriebe und Treibstoffe für die Mobilität von Morgen Transports de l'avenir ? Moteurs et carburants pour la mobilité de demain (seulement disponible en version allemand)