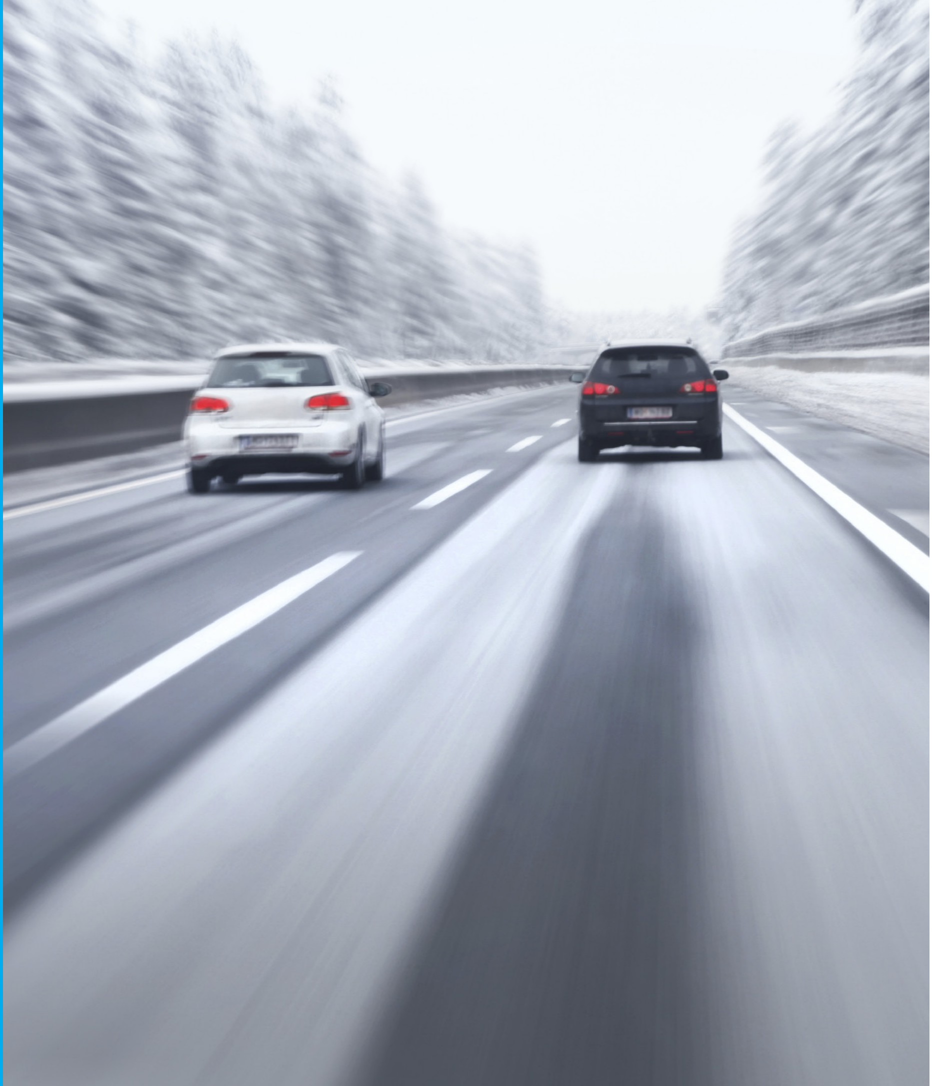

TECHNOLOGY • RESEARCH HIGHLIGHTS • VISIONS • SCIENCE

172

Road traffic incident risk assessment

Accident data pilot on Ring I of the Helsinki Metropolitan Area

Satu Innamaa | Ilkka Norros | Pirkko Kuusela | Riikka Rajamäki | Eetu Pilli-Sihvola

VTT TECHNOLOGY 172

Road traffic incident risk assessment

Accident data pilot on Ring I of the Helsinki Metropolitan Area

Satu Innamaa, Ilkka Norros, Pirkko Kuusela, Riikka Rajamäki & Eetu Pilli-Sihvola

ISBN 978-951-38-8257-0 (URL: http://www.vtt.fi/publications/index.jsp)

VTT Technology 172

ISSN-L 2242-1211
ISSN 2242-122X (Online)

JULKAISIJA – UTGIVARE – PUBLISHER

VTT
PL 1000 (Tekniikantie 4 A, Espoo)
02044 VTT
Puh. 020 722 111, faksi 020 722 7001

VTT
PB 1000 (Teknikvägen 4 A, Esbo)
FI-02044 VTT
Tfn +358 20 722 111, telefax +358 20 722 7001

VTT Technical Research Centre of Finland
P.O. Box 1000 (Tekniikantie 4 A, Espoo)
FI-02044 VTT, Finland
Tel. +358 20 722 111, fax +358 20 722 7001

Road traffic incident risk assessment
Accident data pilot on Ring I of the Helsinki Metropolitan Area

Tieliikenteen häiriöriskin arviointi. Pilotti Kehä I:n onnettomuusaineistolla.
Satu Innamaa, Ilkka Norros, Pirkko Kuusela, Riikka Rajamäki & Eetu Pilli-Sihvola.
Espoo 2014. VTT Technology 172. 49 p. + app. 8 p.

Abstract

The purpose of this project was to apply the Palm distribution to the analysis of riskiness of different traffic and road weather conditions introduced in a previous project (Innamaa et al. 2013), develop the method further, and find factors that statistically significantly affect traffic incident risk.

The method was piloted using data from Ring-road I of the Helsinki Metropolitan Area. The study was based on registered accidents that occurred on Ring-road I in 2008–2012, totalling 1120. In addition to accident data, traffic data from eight automatic traffic measurement stations (inductive loops) and road weather data were also used.

The basic methodological idea was to compare the traffic and weather circumstances just before an accident with the Palm probability of the same circumstances. The notion of Palm probability comes from the theory of random point processes, and means the probability distribution "seen" by a randomly selected point of the point process, i.e. the driver in this case (in contrast to the stationary probability, which is the probability distribution seen at a random time point). If each car driver had a constant stochastic intensity of causing an accident, then the accident circumstances would follow the Palm distribution. The idea of the method applied here is to assess the influence of circumstances on incidents by comparing the incident circumstance distribution with the Palm distribution of circumstances: differences between these distributions hint at effects of circumstances on accidents.

The results showed that there were several specific weather conditions that were more common among drivers who were involved in an accident than among drivers in general. These conditions included air temperature from –6 degrees Celcius down, snowfall or heavy rain, limited visibility, and snowy or wet road surface. The results further showed that the probability of an accident is higher in conditions when a weather alarm is issued by the Transport Agency (the road operator) than in general.

In addition, in weekday afternoon traffic (15–17 o'clock) the risk of accident was found to be 50% higher than generally. In night time traffic (2–5 o'clock) the risk was even higher. The results indicated that the traffic situation correlated poorly with accident risk. However, the results related to the traffic situation can be considered only indicative due to inaccuracies in the accident location information and sparseness of the traffic detector network.

In conclusion, the findings suggest that the proposed method for identifying conditions where accident risk is elevated by comparing the traffic and weather circumstances just before the accident with the “Palm probability” of the same circumstances indeed works. Not all results were statistically significant due to some circumstances being rare. However, with the calculation of risk levels and Kullback-Leibler divergence, it was possible to assess the findings.

Keywords traffic incident risk, Palm probability, driving conditions

Tieliikenteen häiriöriskin arviointi
Pilotti Kehä I:n onnettomuusaineistolla

Road traffic incident risk assessment. Accident data pilot on Ring I of the Helsinki Metropolitan Area. **Satu Innamaa, Ilkka Norros, Pirkko Kuusela, Riikka Rajamäki & Eetu Pilli-Sihvola.** Espoo 2014. VTT Technology 172. 49 s. + liitt. 8 s.

Tiivistelmä

Tämän hankkeen tavoitteena oli soveltaa Palm-jakaumaan perustuvaa aikaisemmassa hankkeessa (Innamaa ym. 2013) kehitettyä menetelmää analysoida erilaisiin liikenne- ja keliolosuhteisiin liittyvää liikenteen häiriöiden riskiä, kehittää menetelmää eteenpäin ja löytää tekijöitä, jotka vaikuttavat liikenteen häiriöiden riskiin tilastollisesti merkitsevästi.

Menetelmää kokeiltiin Kehä I:llä kootulla aineistolla. Tutkimus perustui Kehä I:llä vuosina 2008–2012 tapahtuneisiin, rekisteröityihin onnettomuuksiin, joita oli yhteensä 1120. Tutkimuksessa käytettiin onnettomuusaineiston lisäksi liikennetietoja kahdeksasta liikenteen automaattisesta mittauspisteestä (induktioilmaisimia) ja tiesääaseman tuottamaa tietoa.

Menetelmällinen perusajatus oli verrata liikenne- ja sää-/keliolosuhteita hetkeä ennen onnettomuutta samojen olosuhteiden Palm-todennäköisyyksiin. Palm-todennäköisyyden käsite tulee satunnaisen pisteen prosessoinnin teoriasta ja tarkoittaa satunnaisesti valitun prosessointipisteen, tässä tapauksessa autoilijan, ”näkemää” todennäköisyysjakaumaa (vastakohtana stationaariselle todennäköisyydelle, joka kuvaa todennäköisyysjakaumaa satunnaisella ajanhetkellä). Jos jokaisella autoilijalla on vakiosuuruinen stokastinen intensiteetti aiheuttaa onnettomuus, onnettomuusolosuhteet noudattavat Palm-jakaumaa. Menetelmässä on sovellettu ideaa arvioida olosuhteiden vaikutus liikenteen häiriöihin vertaamalla häiriöiden olosuhdejakaumaa olosuhteiden Palm-jakaumaan: Näiden jakaumien väliset erot viittaavat olosuhteiden onnettomuusvaikutukseen.

Tulokset osoittavat useita sää- ja keliolosuhteita, jotka olivat yleisempiä onnettomuuksiin joutuneiden ajoneuvojen kohdalla kuin yleensä. Tällaisia olosuhteita olivat korkeintaan –6º C ilman lämpötilan lumisade tai rankka vesisade, huono näkyvyys ja luminen tai märkä tienpinta. Lisäksi tulokset osoittivat, että onnettomuuden todennäköisyys on korkeampi olosuhteissa, jolloin Liikennevirasto antaa kelivaroituksen, kuin muuten.

Lisäksi arki-iltapäivien (klo 15–17) liikenteessä onnettomuusriski oli 50 % korkeampi kuin yleensä. Yöaikaan (klo 2–5) riski oli vielä korkeampi. Tulokset osoittivat, että liikennetilanne korreloi huonosti onnettomuusriskin kanssa. Liikennetilanteeseen liittyviä tuloksia voi kuitenkin pitää vain suuntaa-antavina onnettomuuden paikkatiedon epäluotettavuuden ja harvan liikenneilmaisimien verkon takia.

Yhteenvetona voidaan todeta tulosten viittaavan siihen, että ehdotettu menetelmä toimii eli että menetelmä tunnistaa olosuhteet, jolloin onnettomuusriski on kohonnut vertaamalla liikenne- ja sää-/keliolosuhteita hetki ennen onnettomuutta samojen onnettomuuksien Palm-jakaumaan. Kaikki tulokset eivät olleet tilastollisesti merkitseviä olosuhteiden harvinaisuuden takia. Tulokset voitiin kuitenkin arvioida laskemalla riskitasoja ja Kullback-Leibnerin divergenssi.

Avainsanat traffic incident risk, Palm probability, driving condition

Preface

A research project funded by the Finnish Transport Agency launched the development of a method for identifying conditions in which the risk of a traffic incident is elevated. VTT Technical Research Centre of Finland aimed to develop the method further and validate the proposed method by testing it with a wider dataset.

Satu Innamaa led the project, and with Riikka Rajamäki provided the transport engineering perspective. She structured the deliverable and wrote the Introduction and Discussion. Riikka Rajamäki provided raw data for the study and wrote the Data chapter. Mikko Kallio processed the raw data for analysis. Ilkka Norros and Pirkko Kuusela were in charge of developing and applying the mathematical method (Chapters Mathematical Method, Results and Appendix A). Eetu Pilli-Sihvola was responsible for searching the literature for the review.

Contents

List of symbols

D	Traffic density (passenger cars / km)
p	Probability of the value vector
Q	Traffic volume (passenger cars / 5 min)
STD	Standard deviation
V	Speed (km/h)

1. Introduction

1.1 Introduction

Traffic incidents cause secondary accidents and congestion, which at the societal level leads to massive loss of time and increased carbon emissions in addition to fatalities and injuries. Current traffic monitoring systems are reactive and incident management focuses on minimising consequences of incidents rather than trying to proactively prevent incidents from taking place.

Traffic incidents can be caused by a variety of single issues or, more often, by a combination of issues (e.g. human error, adverse weather, large traffic volumes, challenging road geometry). Consequently the analysis of factors that result in increased incident risk requires expert knowledge both of traffic phenomena and of analysis methods. Knowledge of such factors could facilitate the development of an online incident risk assessment model that could be used in real-time operations at traffic management centres. To our knowledge, such models do not exist as yet.

1.2 Literature review

1.2.1 Effect of traffic conditions on incident risk

Pajunen and Kulmala (1995) researched the effect of hourly traffic volumes on traffic safety based on hourly traffic volume and speed data collected in 1991–1993 from automatic traffic measurement stations, combined with accident data from areas where traffic data was available. They found that on two-lane roads and semi-motorways, accident rates generally fell as hourly traffic volume increased. On four-lane roads, accident rates were highest at hourly traffic volumes of 3600–4800 vehicles. However, on motorways, accident rates increased with rising hourly traffic volume, and were highest at very low traffic volumes in both directions. On all road types, accident rates were lower with high hourly traffic volumes when the traffic was unevenly distributed in different directions.

Marchesini and Weijermars (2010) reviewed the literature on the relationship between congestion and safety on motorways. They found that that the likelihood of a crash seems to grow with speed variability, which is a common indicator of unstable traffic conditions. Large differences in speeds between lanes and density variations also appear to make crashes more likely.

When searching for evidence to support the general perception that crash frequency increases with higher congestion levels, Marchesini and Weijermars (2010) came across conflicting results. Some studies found that high volume-to-capacity ratios led to higher crash rates; another study stated that crash rates decreased at high traffic densities. Additionally, one of the studies they examined did not find any relationship between congestion and crash frequency. None of the studies they looked at explicitly provided evidence on the influence of congested traffic conditions on crash rates.

Ishak and Alecsandru (2005) investigated the characteristics of pre-incident, post-incident, and non-incident traffic conditions on freeways. The characteristics were defined by second-order statistical measures derived from spatiotemporal speed contour maps. Four performance measures were used to quantify properties such as smoothness, homogeneity, and randomness in traffic conditions in a manner similar to texture characterization of digital images.

Their study was conducted using data collected from the freeway corridor I-4 in Orlando, Florida. The study corridor was nearly 40 miles long and six lanes wide. The entire corridor was instrumented with 71 inductive dual-loop detectors or stations, spaced approximately half a mile apart. Each detector station collected three traffic parameters – traffic volume, lane occupancy, and speed – from each of the six lanes. The system supported data resolution of 30 seconds. (Ishak and Alecsandru 2005)

With real-world incident and traffic data sets, statistical analysis was conducted to seek distinctive characteristics of three groups of traffic operating conditions: pre-incident, post-incident, and non-incident. Incident data was also collected from various sources, and a total of 116 accidents, reported on different days, were selected for the analysis. Traffic conditions before and after the incident occurrence were separated into two groups: pre-incident conditions and post-incident conditions. Pre-incident conditions were restricted to observations that took up to 10 minutes before the incident happened, whereas post-incident conditions were collected from observations taken up to 10 minutes after the incident. The second-order statistical measures outlined earlier were computed for each group by using speed data collected from loop detectors. An arbitrarily selected time-space window of 5 minutes and three detector stations were used as the basis for calculation of each measure. In addition, non-incident traffic conditions were collected from a total of 5 weekdays in 2001 and used for comparative analysis. (Ishak and Alecsandru 2005)

The statistical analysis showed slight variations among the three groups (pre-, post-, and non-incident conditions) in terms of each of the four measures used. Although the nonparametric tests showed that the distribution of each measure within each group is different, a consistent pattern was not detected within the categories of each measure. Such inconsistency led to the conclusion that the pre-, post-, and non-incident traffic conditions may not be readily discernible from each other and that specific characteristics of precursory conditions to incidents may not be clearly identifiable. Such a conclusion, however, is driven by limited incident and traffic datasets and selected second-order traffic performance measures. Additionally, environmental factors such as inclement weather conditions were not accounted for in this study. (Ishak and Alecsandru 2005)

Ishak and Alecsandru (2005) suggested that further research should be conducted to include a broader sample of data and possibly more sophisticated measures, and to account for factors such as weather conditions and possible inaccuracies in detector data.

1.2.2 Effect of weather conditions and other factors

Pajunen and Kulmala (1995) also examined the effect of time of year and daylight on traffic safety. They found accident rates on two-lane roads to be generally higher at times of darkness than during the day. On four-lane roads the personal injury accident rate decreased at night when traffic volumes increased. On semi-motorways accident rates in the dark were found to drop as hourly traffic volumes increased, up to a point, after which the accident rates again started to rise.

Ziemann et al. (2013) examined the occurrence of multi-vehicle rear-end collisions in 2012 in Finland at different times of the year. They examined road accident data from the Finnish Transport Agency's database and found that the time of year had no significant impact on the frequency of multi-vehicle rear-end collisions. When looking at rear-end collisions where only two vehicles were involved, the winter months had an observable increase in the occurrence of these kinds of accidents. The three most common contributing causes to multi-vehicle collisions were found to be driving speed, distance to the preceding vehicle and the focusing on driving task. The use of a mobile phone just before becoming involved in an accident was mentioned as a typical cause related to a lack of focus. The limitations of this examination included a small sample size (one calendar year, 2012) and not taking into account the actual weather and road conditions but only looking at calendar months for the frequency of accidents. The study was performed in co-operation with Liikenneturva, a Finnish association promoting road safety.

Salli et al. (2008) analysed the effects that different wintertime road conditions have on the accident risk in passenger car traffic. They define accident risk as the ratio of the number of accidents to vehicle mileage. Similarly, the accident risk for a specific road condition is defined as the ratio of the number of accidents in those conditions to vehicle mileage in the same conditions. The study consisted of a literature review and statistical accident risk analysis that combined accident information from the Traffic Safety Committee of Insurance Companies (VALT) and road condition data from the Finnish Road Administration. The literature study focused on the effects that road condition, vehicle type, tyres, and driver behaviour have on wintertime accident risk.

The literature study of the Nordic countries (Salli et al. 2008) showed that the rarer specific winter road conditions are, the greater the risk of an accident is. This points to the conclusion that drivers accustomed to driving in winter conditions can better accommodate for them in their own driving behaviour. In addition, the risk of accident was found to be higher in the beginning and end of the winter season. Relating to specific conditions, the relative accident risk was found to be highest on icy and slushy roads.

In the Salli et al. (2008) study, the accident risk in snowy or icy conditions was calculated to be 4.1 times that of bare pavement road conditions. For fatal accidents, the risk in conditions of loose snow or slush was estimated at 4.9 times that of bare pavement conditions. These risk factors were found to correspond quite well with the results of earlier studies on the subject. However, many potential sources of error were identified in evaluating the accident risks for different road conditions. Information about the road condition is often a subjective estimate, the classifications of road conditions are not necessarily comparable, vehicle mileage data in different road conditions is lacking, and the accident data is often small and includes contingency.

Hranac et al. (2006) studied the impacts of weather on different macroscopic traffic flow parameters, but did not look specifically into the effects of weather on the risk of incidents or accidents. The study consisted of a literature review and an analysis that combined archived macroscopic traffic data with historical weather data. Hranac et al. looked at traffic and weather data from three North-American cities: Minneapolis-St. Paul, Baltimore and Seattle. They did not find any impacts on traffic jam density, but both rain and snow did have an impact on traffic free-flow speed, speed at capacity and capacity. In addition, the parameters varied depending on the precipitation intensity. Even though capacity did not vary with snow intensity, capacity reductions between 12% and 20% were found when conditions were snowy.

One interesting finding of the Hranac et al. (2006) study was that Minneapolis-St. Paul experienced more significant reductions in the traffic free-flow speed and speed at capacity under snowy conditions than Baltimore (19% vs. 5%). As Minneapolis-St. Paul receives more snowfall annually, the authors posited that one possible explanation for this was that drivers more accustomed to snow were more aware of its dangers. This corresponds well with the findings of Salli et al. (2008), where the authors examined the impact of weather on traffic in the Nordic countries and came to similar conclusions.

1.2.3 Secondary incident risk estimation

Vlahogianni et al. (2012) introduced a neural network model approach to extract useful information on variables that are associated with the likelihood of secondary accidents. Specifically, traffic and weather conditions at the site of a primary incident were examined. To detect secondary incidents, a dynamic threshold methodology was used that considered real-time traffic information from loop detectors. Two sensitivity measures to evaluate the significance of the variables were used (mutual information and partial derivatives).

As input to the model of Vlahogianni et al. (2012), 3500 incident records between 2007 and 2010 were used. The data was from the Attica Tollway, a 65-km urban motorway connecting two major interurban motorways, Athens International Airport, and Athens city centre in Greece. This incident information was supported by traffic-related information including exact location, number of lanes blocked, total duration of the incident, vehicle type, and number of vehicles involved. Factors such as prevailing traffic conditions (speed and volume) and weather conditions (rainfall intensity) were also considered. As output the model estimated the contribution of different variables to the likelihood of secondary incidents. In addition, the results showed that a multilayer perceptron with a supporting function acting as a general Logit model performed best among the different models.

The likelihood of the proposed model yielding incorrect classification of secondary incidents varied between 6% and 7%. The results suggested that traffic speed, duration of the primary accident, hourly volume, rainfall intensity, and the number of vehicles involved in the primary accident are the top five factors associated with secondary accident likelihood. However, changes in traffic speed and volume, number of vehicles involved, blocked lanes, and percentage of trucks and upstream geometry also significantly influence the probability of having a secondary incident. (Vlahogianni et al. 2012)

Vlahogianni et al. (2012) assessed that the proposed neural network approach is promising as a transport managerial tool for TMCs to support decision-making. It could potentially be extended to other transport applications as well.

1.2.4 Summary of the literature review

From the current literature, it seems that there hasn’t been much research on the impact of traffic conditions on accident risk. Pajunen and Kulmala (1995) found that on two-lane roads and semi-motorways, accident rates generally fell as hourly traffic volume increased. On four-lane roads, accident rates were highest at hourly traffic volumes of 3600–4800 vehicles. However, on motorways, accident rates increased with rising hourly traffic volume, and were highest at very low traffic volumes in both directions. Marchesini and Weijermars (2010) concluded based on literature that that the likelihood of a crash grows with speed variability and that large differences in speeds between lanes and density variations also appear to make crashes more likely. None of the studies they looked at explicitly provided evidence on the influence of congested traffic conditions on crash rates. The study of Ishak and Alecsandru (2005) was based on limited incident and traffic datasets and selected second-order traffic performance measures. Therefore, their conclusion that the pre-, post-, and non-incident traffic conditions may not be readily discernible from each other and that specific characteristics of precursory conditions to incidents may not be clearly identifiable can not be considered final.

More information was found on the riskiness of different road weather conditions. Specifically, Nordic winter conditions have been shown to be risky. Ziemann et al. (2013) found that the winter months had an observable increase in the occurrence of rear-end collisions where only two vehicles were involved. In addition, Salli et al. (2008) showed that the rarer specific winter road conditions are the greater the risk of an accident is. They also found the risk of accident to be higher at the beginning and end of the winter season. Relating to specific conditions, Salli et al. found the relative accident risk to be highest on icy and slushy roads and the accident risk in snowy or icy conditions was calculated to be 4.1 times that of bare pavement road conditions. For fatal accidents, the risk in conditions of loose snow or slush was estimated at 4.9 times that of bare pavement conditions.

In studies of the risk of a secondary accident, similar findings of rainfall intensity affecting the risk have been found as for the overall accident risk. In addition, some indication on the impact of traffic flow were found. Specifically, Vlahogianni et al. (2012) found that traffic speed, duration of the primary accident, hourly volume, rainfall intensity, and the number of vehicles involved in the primary accident were the top five factors associated with secondary accident likelihood. Changes in traffic speed and volume, number of vehicles involved, blocked lanes, and percentage of trucks and upstream geometry also significantly influenced the probability of having a secondary incident. However, the factors that are significant for secondary accident risk are not all necessarily the same for the general incident risk.

1.3 Purpose of the study

The literature review showed that more research should be done on identification of risky driving conditions. Knowledge of such factors could facilitate the development of an online incident risk assessment model that could be used in real-time operations at traffic management centres.

The application of the Palm distribution in analysing different traffic and road weather conditions under which the risk of incident is increased was a new research idea piloted in a previous research project (Innamaa et al. 2013). The purpose of this project was to apply the proposed method, develop it further and find factors that affect traffic incident risk with statistical significance.

Road safety is regarded as a multiplication of three orthogonal factors of exposure, risk of a collision taking place during a trip, and risk of a collision resulting in injuries or death (Nilsson 2004, Figure 1). This study examines the risk of incidents (crash risk). Whether or not a highly risky condition represents a high number of accidents or fatalities is dependent on exposure to the condition and the severity of accidents (risk of fatal injuries in a crash).

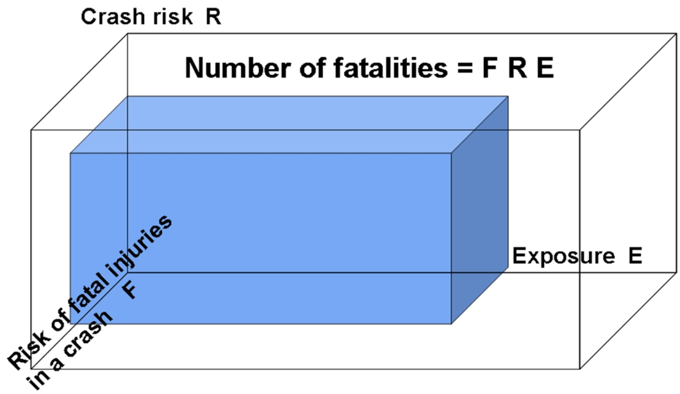

Figure 1. The dimensions of road safety (Nilsson 2004).

2. Method

2.1 Data

The study was conducted on Ring-road I (Ring I) of the Helsinki Metropolitan Area. This beltway has a total length of 24.2 kilometres, two carriageways, and at least two lanes per direction for its entire length. Most of the road has grade-separated interchanges and an 80 km/h speed limit, but near both ends the speed limit is 60–70 km/h and intersections are at-level and light-controlled. Ring I is the busiest road in Finland, carrying 58,000 vehicles per day on average and 90,000 in the northernmost section.

The study was based on registered accidents (from the accident records of the Finnish police) that occurred on Ring I in 2008–2012, totalling 1120 accidents or an average of 0.61 accidents per day. Three of those accidents were fatal and 160 were non-fatal injury accidents. In some cases, accident time, location or conditions were recorded incompletely, and those cases were excluded from this study. The final study data consisted of 1051 accidents.

The most common accident types on this road were rear-end accidents (41%) and accidents related to overtaking and changing lanes (26%). One third of the accidents occurred on the easternmost 4 kilometres of the road. Twenty-six per cent of the accidents occurred between 4 pm and 6 pm (afternoon peak hours). Accidents were fairly evenly divided between the months of the year, only the summer holiday months of June and July having remarkably lower accident numbers. Ten per cent of accidents occurred during snowfall and 11% during rain.

The accident location data lacked reliable information on the direction of travel. Accident locations are based on GPS coordinates recorded by the police, and their accuracy is approximately ±30 metres. In the accident database, accidents are linked to the nearest carriageway; this unreliable carriageway information was used here as no more accurate or reliable source of information was available.

The accident times recorded by the police are mainly exact to within 10 minutes (i.e. 8:00, 8:10, 8:20, etc.). This suggests that accident time accuracy is around ±5 minutes.

In addition to accident data, the following data were used (Figure 2):

i. Traffic data from eight automatic traffic measurement stations (inductive loops) on Ring I in 2008–2012

ii. Road weather data from a station located on highway 3 in Pirkkola, close to Ring I, in 2008–2012.

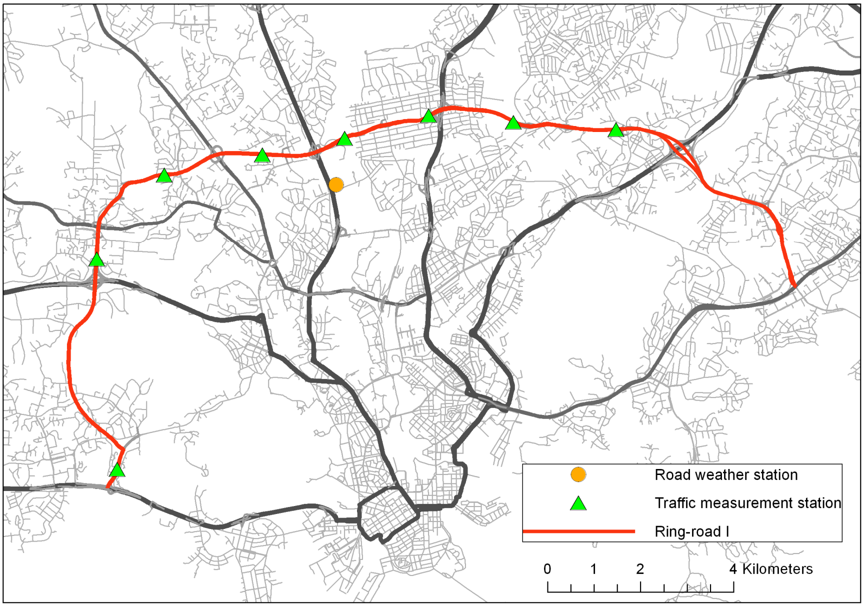

Figure 2. Placement of automatic traffic measurement stations and a road weather station.

Automatic traffic measurement stations are not evenly distributed along this 24 km long road; the easternmost one is located 5.8 km from the eastern end of the road. Therefore the distance between accident location and nearest measurement station is rather long and the traffic situation may be different.

The traffic data consisted of traffic volume and mean speed measurements at 5-minute intervals and standard deviation for speed within that time window. From volume Q (passenger cars/5 min) and speed V (km/h) an estimate of the traffic density was calculated as D = 12Q/V passenger cars/km. Passenger cars and heavy vehicles were counted separately, but in this study they were merged using the passenger car equivalent of 1.5 for heavy vehicles (HCM 1995).

Road weather station data consisted of weather and road surface information at 10–15 minutes intervals. Six variables were used: Air temperature, rain consistency (liquid/crystal) and intensity, visibility, road surface conditions, and warning level. The warnings and alarms given by the system are as follows:

1. Ice warning: No ice on the road surface, but it is likely that there soon will be
2. Ice alarm: Ice or snow on the road surface
3. Frost warning: Frost present on the road surface or likely to appear
4. Rain warning: It is or has been precipitating, and there may be danger of slipperiness (road temperature close to icing temperatures).

Both automatic traffic measurement and road weather station data had some gaps. For missing traffic measurement station records, the data was copied from a neighbour station in the same direction (a later one if such was available, otherwise an earlier), so that if any traffic measurement record existed for a given period, the data showed a set of eight records. The number of missing 5-minute periods was only 65, which is about 0.01% of the whole observation time. For all incidents, there were traffic records for the preceding 5-minute period. However, for about 47% of the registered periods at least some traffic measurement station pair (both directions) was down, and for 9.5% at least two were down. For 111 incidents the traffic data were at least partly inferred from some farther traffic measurement station than the nearest one(s). However, for eight incidents no neighbouring traffic measurement station was alive.

For 48 incidents, weather station information was unavailable at the incident time. For all but one of these, the data could be completed using the weather information provided in the accident record instead.

2.2 Mathematical method

2.2.1 Methodological idea

The basic methodological idea was to compare the traffic and weather circumstances just before the accident with the Palm probability of the same circumstances. The notion of Palm probability comes from the theory of random point processes and means the probability distribution "seen" by a randomly selected point of the point process, i.e. the driver in this case (in contrast to the stationary probability, which is the probability distribution seen at a random time point; see *e.g.*, Baccelli and Brémaud (2003)). In our case, we applied the notion of Palm probability as follows:

Palm distribution of road traffic circumstances	
Intuitive idea	Distribution of circumstances seen by a randomly selected driver
Computation from data	1. Estimate traffic information between automatic traffic measurement stations from the pointwise measurement data (time resolution 5 min). The interpolation rules are given in Appendix A. 2. Estimate the traffic density in each road section from the volume Q (passenger cars/5 min) and speed V (km/h) as D = 12Q/V passenger cars/km. Add to each traffic record (per section) a weight computed as D times the length of the road section. 3. Add weather station data to the traffic records. 4. Discretize the numerical quantities with the following granularities: • Traffic volume granularity 10 p.cars/5 min: values 5, 15, 25,... • Traffic mean speed granularity 5 km/h: values 2.5, 7.5, 12.5,... • Traffic speed standard deviation granularity 5 km/h: values 2.5, 7.5, 12.5,... • Air temperature granularity 3°C: values ..., -4.5, -1.5, +1.5, +4.5,.... • Free sight granularity: 200 m • Time granularity: 20 min • Date granularity: Monday, …, Sunday 5. Gather records with identical variable values together and let the weight of a value combination be the sum of the weights of all original records showing that value combination. 6. Normalise the weights into a probability distribution.

Incident-time distribution of road traffic circumstances	
Intuitive idea	Distribution of circumstances seen at time and place of a randomly selected incident
Computation from data	1. Look at the date and time of the incident, take the preceding 5-minute record for the road section where the incident took place, and pick its traffic and weather circumstance values. 2. Give each incident the same weight and gather identical value combinations together. Let the weight of a value combination be the sum of the weights of all (incident) records showing that value combination. 3. Normalise the weights into a probability distribution.

The reason for using the *preceding* 5-minute record in computation of the incident-time distribution of road traffic circumstances was to avoid the influence of the incident itself on the traffic characteristics.

In the implementation with *Mathematica*™, the distributions were presented as lists with elements $\{\{x_1,...,x_k\},p\}$, where the x_i's are values of observables and p is the probability of the value vector. Below is the full list of observables considered in this study:

- Weekday (Monday = 0)

- Time (granularity 20 min)
- Direction (1 = eastward, 2 = westward)
- Road section (between automatic traffic measurement stations), nine values
- Traffic volume
- Mean speed
- Standard deviation (STD) of speed
- Air temperature
- Rain intensity and phase (liquid *vs.* crystal)
- Visibility (in km, 2 = perfect)
- Road surface condition
- Warning state
- Sun up (1 = yes, 0 = no).

The reason for coarse-graining of continuous quantities is that it allows meaningful comparison of the two distributions (see Appendix A). Even with coarse-graining, the total number of all observed combinations was so big that in order to speed-up the computations, the Palm distributions were created separately for three groups of quantities: (i) time and place attributes, (ii) traffic quantities, and (iii) information items provided by the weather station.

If each car driver had a constant stochastic intensity of causing an accident, then the accident circumstances would follow the Palm distribution. The idea of the method applied here is to assess the influence of circumstances on incidents by comparing the incident circumstance distribution with the Palm distribution of circumstances; differences between these distributions hint at effects of circumstances on accidents.

For each quantity, the comparisons of their two distributions were made using three techniques:

1. Plotting the density of the incident distribution with respect to the Palm distribution. A value higher than 1 indicates higher risk, value 2 can be interpreted as 100% higher risk than under Palm distribution, *etc*. This gives a qualitative impression of the relation of the two distributions.

2. Assessment of the statistical significance of the difference of an incident-time frequency of a particular value from its Palm frequency: What is the probability that random sampling from the Palm distribution, with sample size equal to the number of incidents (1051), would yield at least the frequency observed in the incident distribution? More details are given below.

3. Comparison of the whole distributions in terms of the Kullback-Leibler divergence (also known as relative entropy and by several other names; see *e.g*. Cover and Thomas (2006)). If the relative entropy of the incident distribution with respect to the Palm distribution is not significantly higher than that of a random sample (of same size) from the Palm distribution itself, we can conclude that the two distributions are essentially similar. For more details, see Appendix A.

2.2.2 Graphical illustration of the results

In this report, most of the results are presented using two kinds of plots: (i) densities of accident-time distributions of various variables with respect to respective Palm distributions, i.e. comparing the frequency of certain driving conditions at the moments and places where an accident occurred with that for all drivers, and (ii) joint point probability plots of both distributions equipped with 95% confidence intervals around the Palm values.

The first type of plot is illustrated in Figure 3. If the variable in question (in this case the road section) is denoted by *X,* then for each point with coordinate *x* on the horizontal axis, the value on the vertical axis is

$$P_{accident}(X{=}x) \,/\, P_{Palm}(X{=}x).$$

These density ratios are estimates of the relative risk increase/decrease when variable *X* takes the value *x,* compared with the overall risk level. Thus a density ratio 1 stands for the situation where the variable value (e.g. certain driving condition) has no impact on accident risk. Values below 1 represent conditions where the risk of accident is lower than on the average, and values above 1 represent conditions with elevated accident risk.

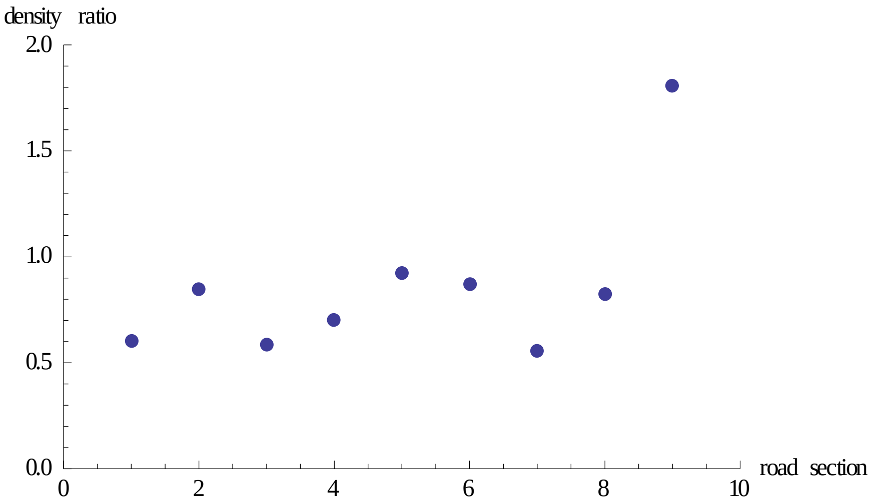

Figure 3. Example of risk levels illustration.

To understand the second type of plot, consider a simplified test outcome of three values A, B and C as illustrated in Figure 4. The Palm distribution presents the frequencies (*i.e.* probabilities, denoted by *p*) of these values obtained from a very large population. This corresponds to the conditions seen by an arbitrary (random) vehicle on the road. In the statistical testing we assume that the Palm frequencies are the "true" frequencies of the various values, illustrated by blue dots in the figure. However, when a small sample with equal value frequencies is drawn and examined, the estimate of frequencies varies. The vertical lines in the figure illustrate how much the small sample frequency estimate can vary on a given confidence level. The vertical line is the (1 - α)% confidence interval [p_{lb}, p_{ub}] that is calculated for a sample the size of the accident vehicle population (1051), where α is the risk level of the statistical testing (0.05 in this case). The calculation of the confidence interval is explained in Appendix A. Note that the blue dots in the figure are always inside the confidence intervals, but the interval need not be symmetric around the blue dot.

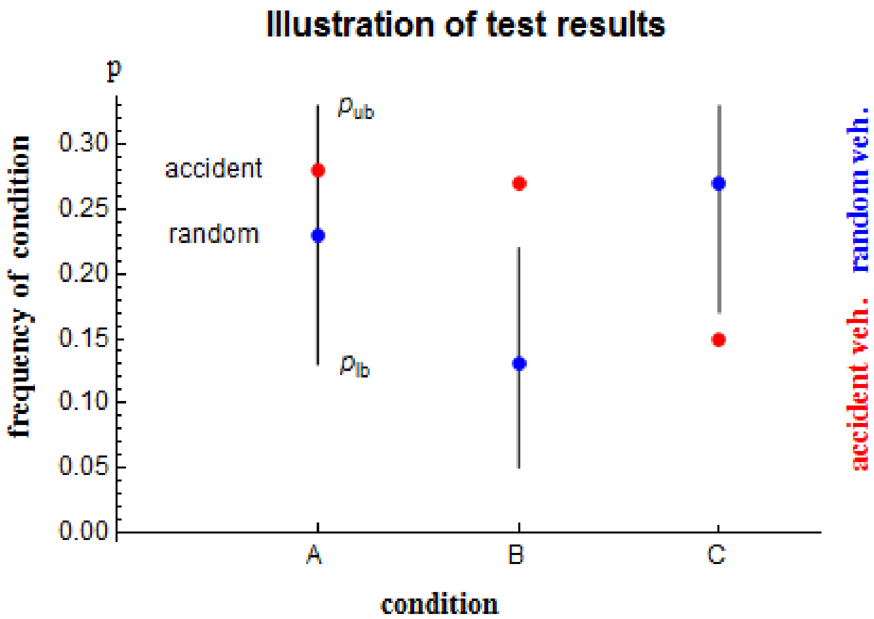

Figure 4. A simple example of how the results of the statistical testing of conditions are illustrated.

Figure 4 summarizes three independent statistical tests, one for each value A, B and C. The vertical axis (p) shows point probabilities of distributions. For value A (e.g. road condition, mean speed value *etc.*), the probability estimate (observed frequency) calculated from the accident vehicle population and illustrated by the red dot is within the confidence interval of the Palm frequency. Thus, it is concluded that value A is not significantly more prone to occur within accident vehicles than within arbitrary vehicles.

For value B, the frequency estimate (red dot) from the accident vehicles is outside the confidence interval of Palm frequencies. It is concluded that at risk (or significance) level α the frequency of condition B differs statistically significantly between arbitrary and accident vehicles. The frequency of value B is larger within the accident vehicle population than in the population of arbitrary vehicles.

For condition C, there is also a statistically significant frequency difference, but now the value C is less frequent among accident vehicles than in general. (Note that because the sum of all probabilities in a distribution is always 1, a heightened accident risk associated with some value must be balanced with lowered risk associated with some other value.)

3. Results

The statistical tests in this section are performed with risk level $\alpha = 0.05$. This means that the probability of rejecting our null hypothesis (that the incident distribution does not differ from the Palm distribution), even if it is true (i.e. by chance), is at most 5%.

3.1 Location, timing and traffic flow conditions

3.1.1 Physical location of incident

First we studied how the incidents are spatially distributed along the whole road, at the resolution of road sections between automatic traffic measurement stations. Figure 5 shows the density of the incident distribution of the road section with respect to the Palm distribution. The left sequence presents the eastward direction, and both start from the west end. Two interesting observations can be made:

i. The eastward direction is more incident-prone than the westward-direction in all nine road sections.

ii. The last (eastmost) section is much more incident-prone than any other road section.

Note that these results tell more than would simple plotting of the incident places along the road, as the varying quantity of traffic along the road is taken into account. Regrettably, the road direction information is known to be unreliable, and thus the direction part of the result is doubtful. If the information on direction were partly uniformly random, this would only decrease the difference between directions, but we cannot exclude the possibility of systematic bias.

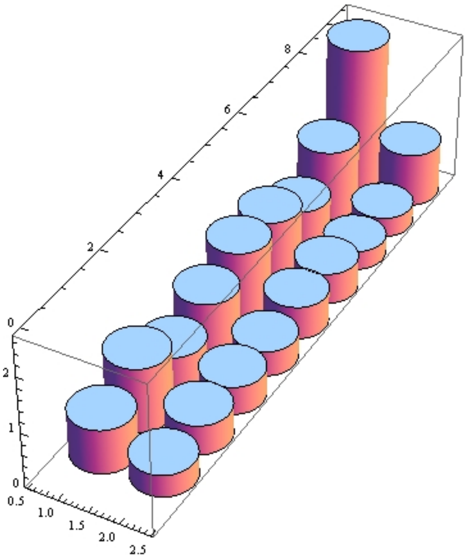

Figure 5. Density of the incident distribution of a road section with respect to the Palm distribution. The left sequence shows the eastward direction, and both start from the west end.

Comparing the parts while neglecting the direction gives the risk levels shown in Figure 6. The result shows that the last road section has 80% higher risk than the average. The Kullback-Leibler test confirms that the incident-time and Palm distributions of the road parts are very different.

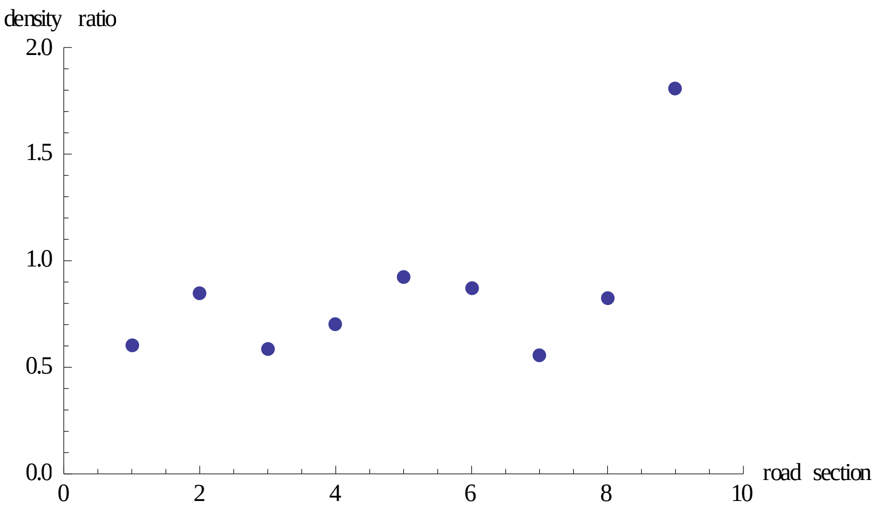

Figure 6. Risk levels of road sections when the direction is not considered.

3.1.2 Incident timing

This subsection studies whether some days of the week or some times of day are more accident-prone than others.

Figure 7 shows that in the statistical testing of the impact of weekday on accident frequency, no weekday indicated a statistically significant difference between a typical vehicle and an accident vehicle. It is interesting to note, however, that on working days the relative risk (seen here as the difference between a red and a blue dot) increases monotonically, although the risks themselves are practically negligible.

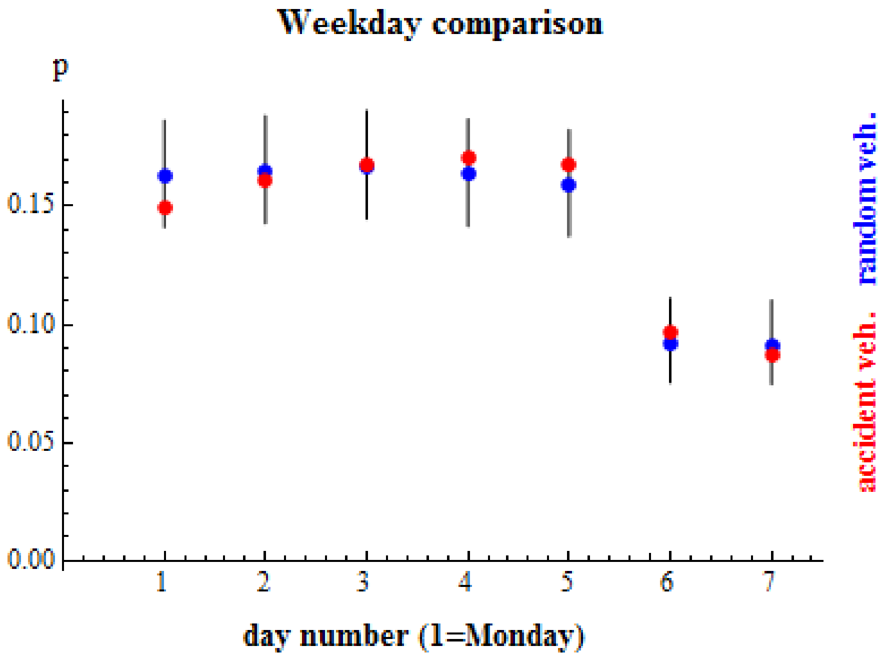

Figure 7. Impact of weekday on road accidents.

Next we considered the impact of hour of the day. Hours range from 0 to 23, and hour *h* refers to time interval (*h*,*h*+1). Figure 8 shows the risk level of each hour in terms of the density ratio between incident-time distribution and Palm distribution. Now the variation is remarkable. It is interesting to see that normal afternoon traffic involves about 50% higher risk than the base level and at night time the risk is even higher. The distributions are significantly different also by the Kullback-Leibler criterion.

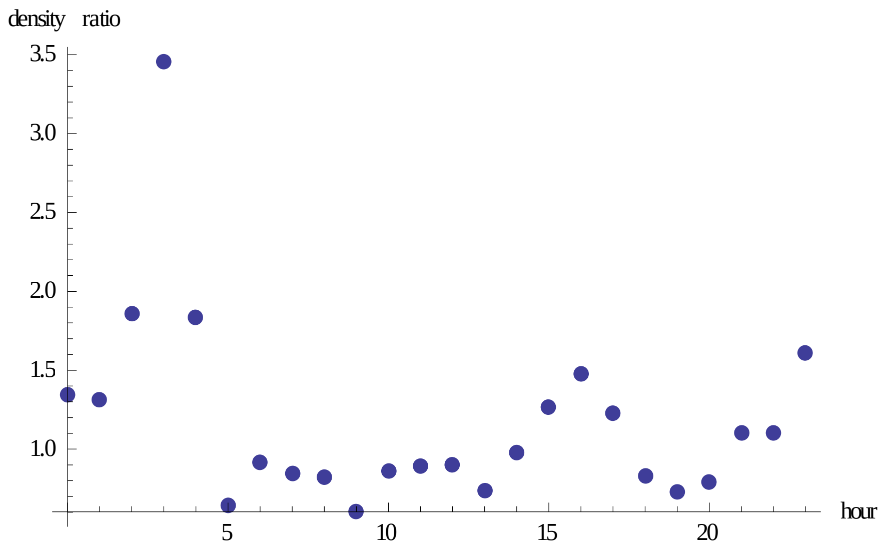

Figure 8. Risk level estimates for hours of the day in terms of the density ratio between incident-time distribution and Palm distribution.

Statistical testing of accident condition distributions indicated similarly significantly differing accident frequencies for hours 3, 9, 13, 15, 16 and 19 (Figure 9). During the evening peak period (15–17 o'clock) the probability of accident is higher than otherwise. In addition, the hour after the morning peak (9–10 o'clock) would be safe. The difference during night time is less interesting from e.g. the traffic management point of view although statistically significant because of very low traffic volumes.

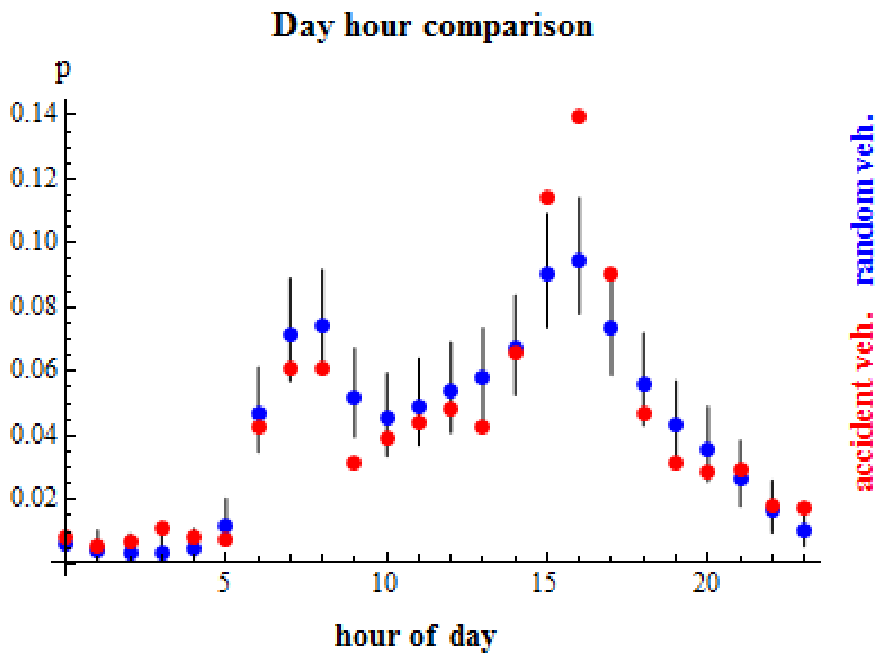

Figure 9. Comparison of the impact of hour of the day on road accidents.

The length of daylight varies in the Helsinki region from less than 6 hours to almost 19 hours per day and, consequently, the hour of the day does not by itself tell whether an accident occurred in daylight or in darkness (lighting by street lights). Therefore we studied how the availability of daylight affects the accident frequencies (Figure 10). The risk of accident turned out to be only slightly higher in darkness than in daylight, but the difference is statistically significant.

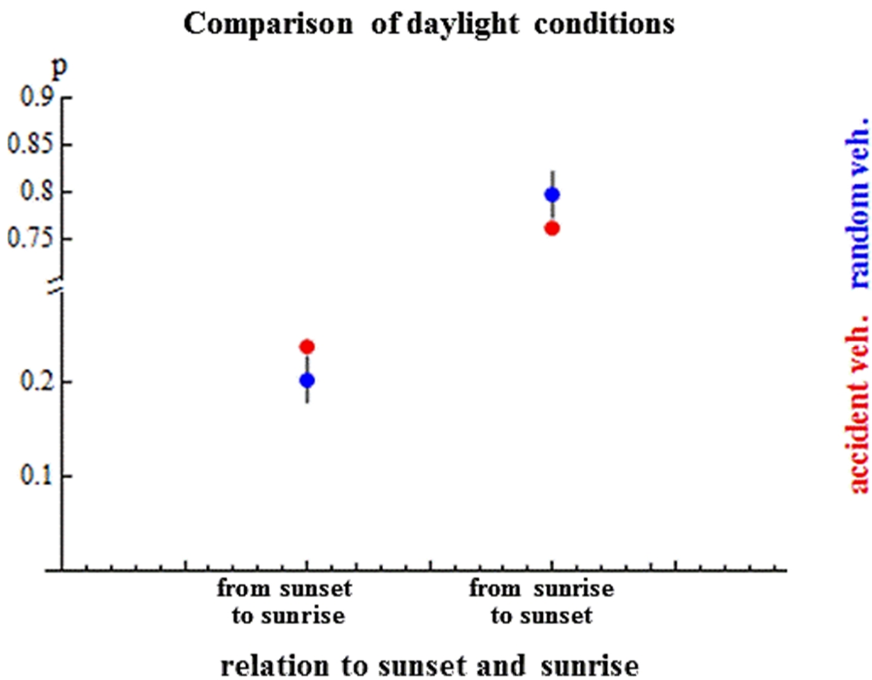

Figure 10. Test of the impact of daylight on road accidents.

3.1.3 Traffic volume and density

In this study, with the data available originating from the sparse loop detector network and including uncertainties in the direction information of the accident database, the results on the traffic situation can be considered indicative. Nevertheless, the results (Figure 11) indicate that in general traffic volumes have no impact on the frequency of road accidents. Only very low volumes (5 and 15 p.cars/5 min) and volume 135 p.cars./5 min were overrepresented in accident vehicle data compared to those seen by an arbitrary vehicle on the ring road. The smallest traffic volumes correlated with the night time result (hourly timing of incident and driving in darkness). The isolated volume value with significantly higher probability of accident (135 p.cars/5 min) may be a coincidence without further explanation. The Kullback-Leibler test confirms that the difference between the two distributions is insignificant.

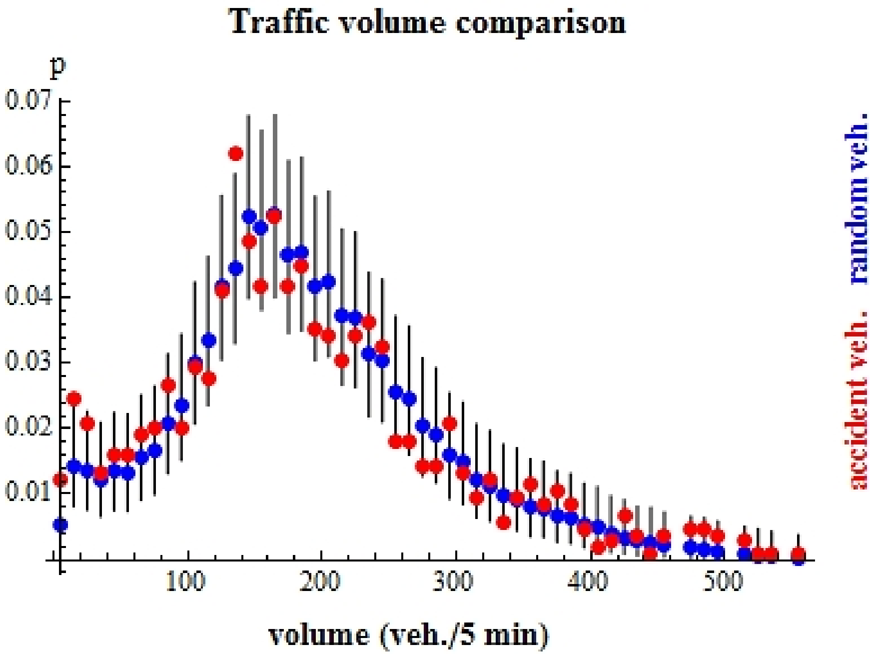

Figure 11. Test of the impact of traffic volumes on road accidents.

As regards traffic density, Figure 12 seems to suggest that high densities of 90–100 passenger cars/km would increase the incident risk by 50–100%, whereas still higher densities would reduce it below average. However, Figure 13 shows that only densities of 5 and 25 p.cars/km show a statistically significant difference in accident frequencies. Density 5 p.cars/km is seen to be over-represented for accident vehicles in comparison with random vehicles. This is most likely related to night time traffic. On the other hand, the density of 25 p.cars/km was less common among accident vehicles than generally. According to the Kullback-Leibler test, the overall difference between the two distributions is borderline significant.

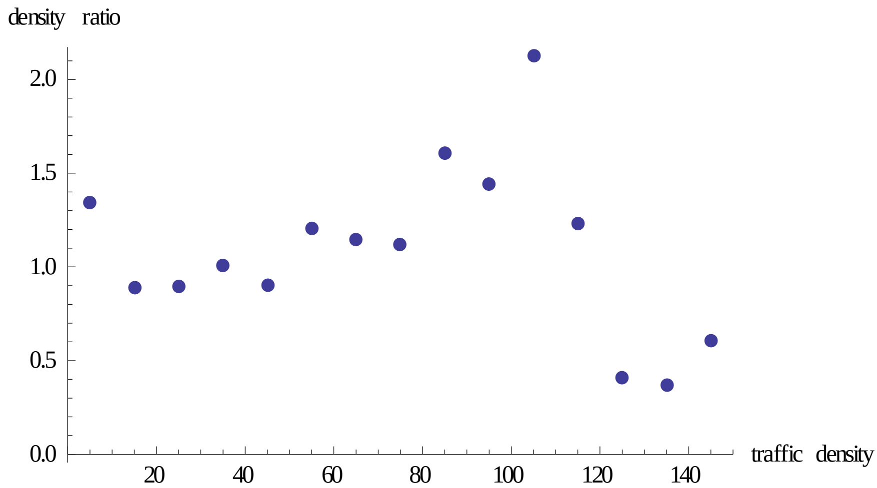

Figure 12. Estimated risk levels in relation to traffic density.

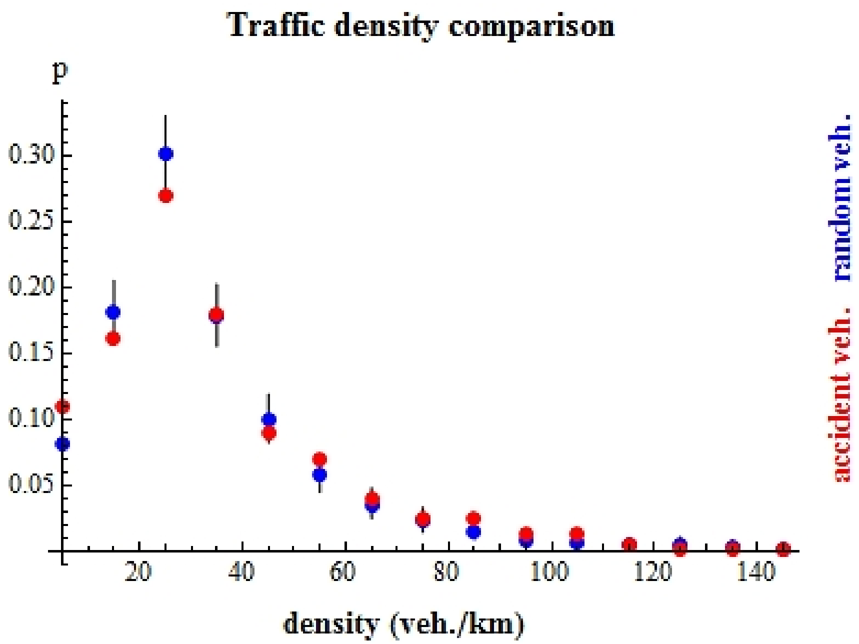

Figure 13. Test of the impact of traffic density on road accidents.

3.1.4 Speed and speed variation

Measured vehicle speeds were grouped into a granularity of 5 km/h. The statistical testing outlined in Figure 14 divides speed groups into two almost equally-sized groups. Speed groups centred at 22.5, 37.5, 42.5, 47.5, 57.5, 67.5, 77.5, 82.5 and 87.5 km/h show a statistically significant difference in accident frequency. Normal, fluent traffic with speed slightly above the limit (80 km/h) was seen to have statistically very significantly lower accident probability than otherwise. Higher speeds are related to situations with low traffic volume and good driving conditions. Lower speeds are thus related to circumstances where the driving is in some respect inconvenient, and for them the accident risk is seen to be heightened, for many speed intervals significantly. This pattern is quite strong as red is above blue 10 consecutive times.

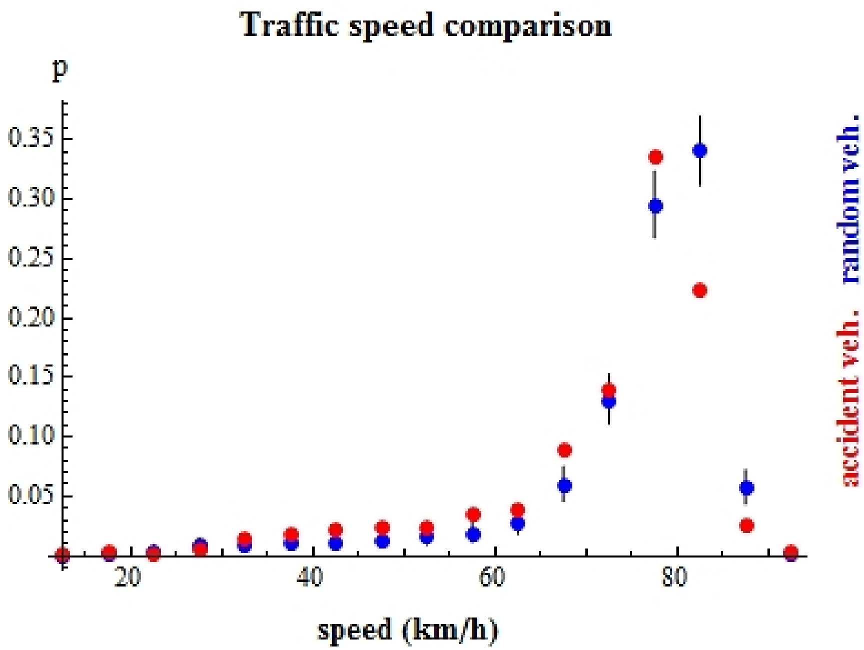

Figure 14. Test of the impact of vehicle speeds on road accidents. Speeds are grouped into a granularity of 5 km/h.

Figure 15 plots actual risk levels (density ratios between incident-time and Palm distributions), which are considerable when mean speeds below 70 km/h raise the risk by 50–100%. Figure 16 shows that the incident time distribution of mean speed is indeed stochastically smaller (except for some extreme values) than the Palm distribution (that is, one cumulative distribution function is dominated by another).

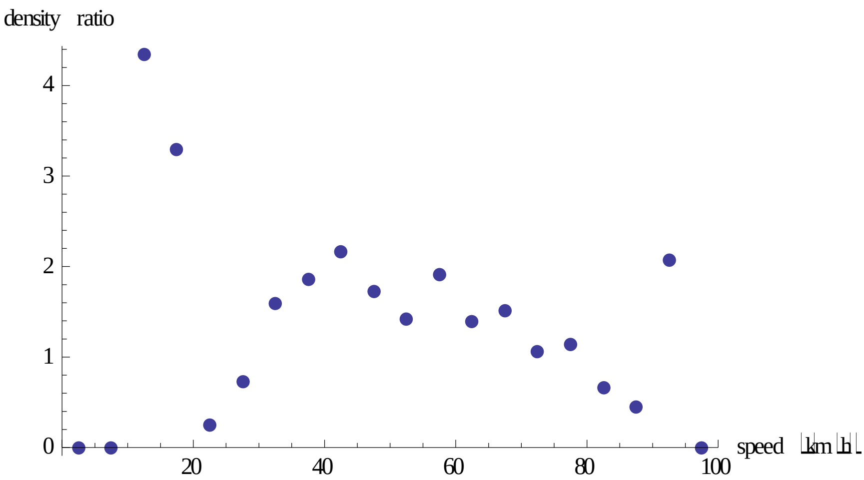

Figure 15. Estimated risk levels at different speeds. Here density values equal to 0 represent speeds not present during accident times.

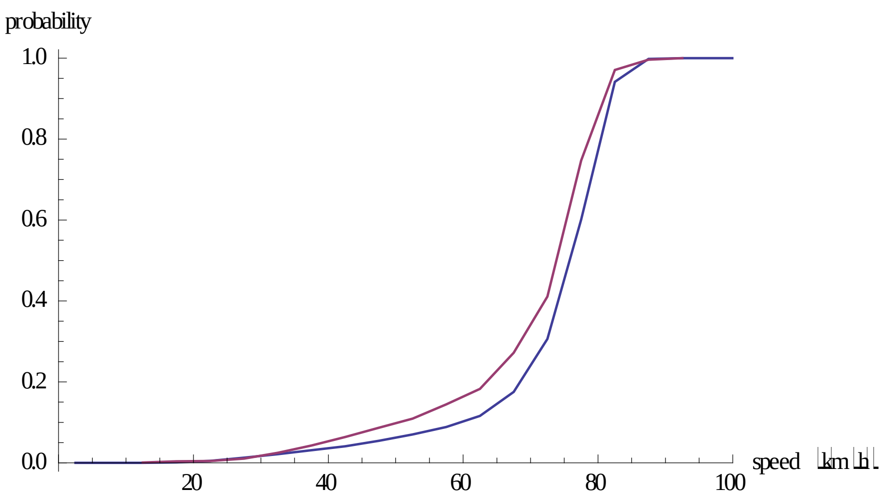

Figure 16. Cumulative distribution functions of the Palm (lower, blue) and incident (higher, red) distributions of mean speed.

It is natural to expect that a steady traffic flow would be safer than when the vehicle speeds have considerable variation. Therefore also the standard deviations of vehicle speeds were computed for each 5 minute interval. Figure 17 shows that traffic speed variation groups centred at 12.5, 17.5, and 27.5 km/h indicate significant difference in accident frequency. Variation group 12.5 km/h with the highest Palm probability is significantly safer, whereas higher speed variations indicate higher accident frequency. According to the Kullback-Leibler criterion the two distributions differ highly significantly.

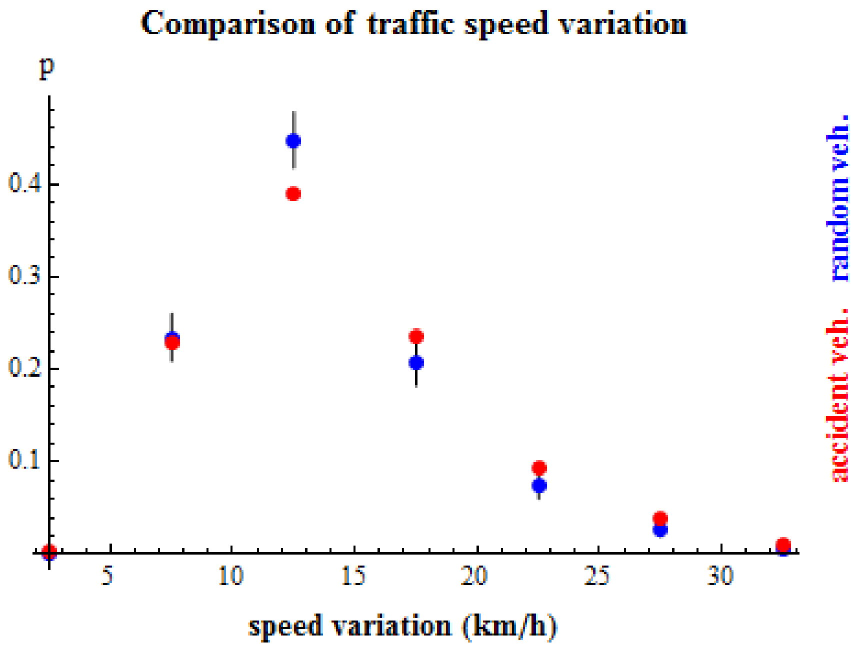

Figure 17. Test of the impact of vehicle speed variation on road accidents. Speed variations are grouped into a granularity of 5 km/h.

Figure 18 shows that the risk becomes monotonously higher with speed variation above the typical level (represented here by 12.5 km/h). The actual risk levels are not very high but not negligible either. For example, the standard deviation value of 27.5 km/h, whose Palm frequency is about 10% (i.e. 10% of drivers drive in such conditions), increases the incident risk by about 40%. Finally, Figure 19 shows that in fact the standard deviation of vehicle speeds at incident times dominates stochastically those seen in the Palm distribution. The difference between the whole distributions according to the Kullback-Leibler criterion is significant, but not as pronounced as it was in the case of speed.

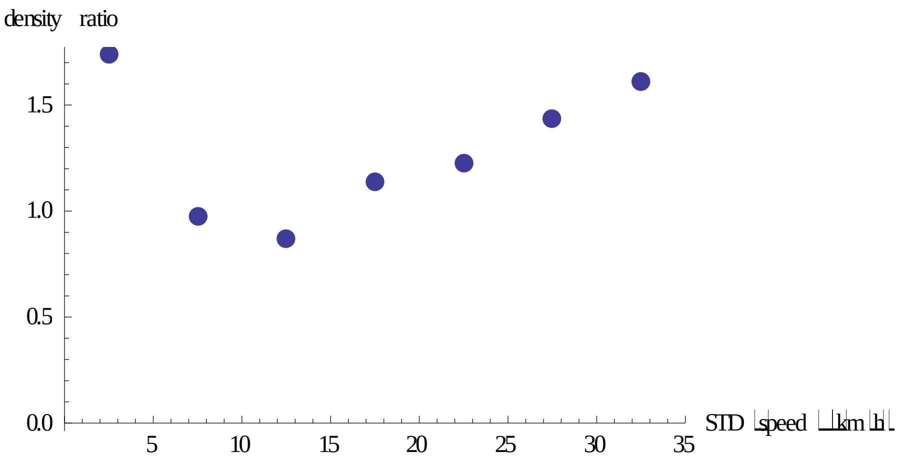

Figure 18. Incident-time *vs.* Palm density ratios of the standard deviation of speed.

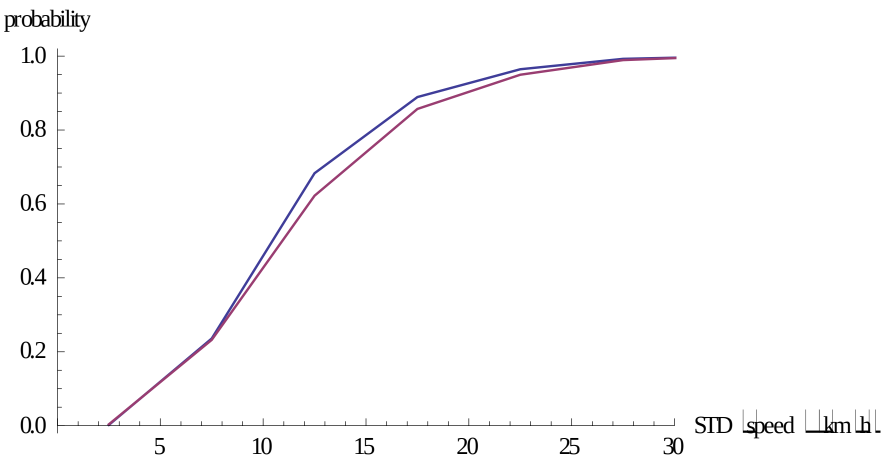

Figure 19. Cumulative distribution functions of the Palm (higher, blue) and incident (lower, red) distributions of standard deviation of speed.

3.2 Road and weather conditions

3.2.1 Air temperature

Air temperatures are grouped with a granularity of 3 degrees Celsius (Figure 20). Only the temperature groups centred at -7.5 and 4.5 degrees show a statistically significant difference in accident frequencies. The profile of risk levels (density ratios) shown in Figure 21 shows that the risk increases at -7.5 by about 50% and suggests (observations are too few to be statistically significant) that temperatures lower than -15 degrees Celsius may bring considerably heightened incident risk (by more than 100%). On the other hand, it is somewhat surprising that the risk does not increase at all when the temperature sinks below freezing. According to the Kullback-Leibler criterion, the incidence-time distribution of temperature differs significantly from the Palm distribution, but not as heavily as with some other observables.

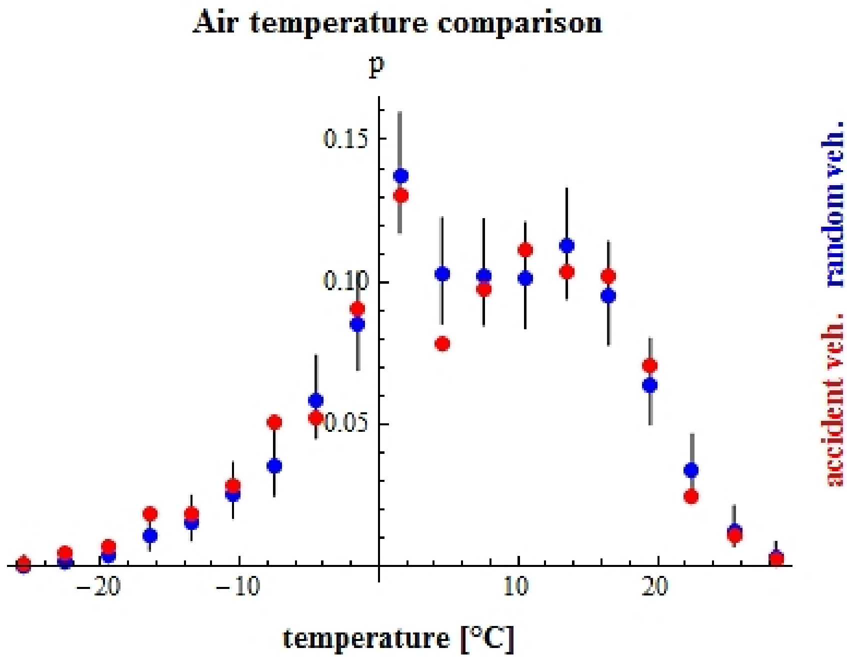

Figure 20. Test of the impact of air temperature on road accidents.

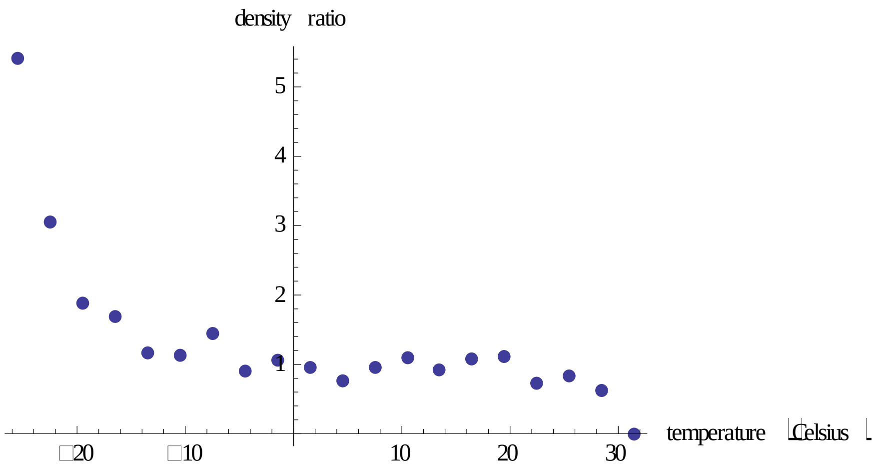

Figure 21. Incident-time *vs.* Palm density ratios of the standard deviation of air temperature.

3.2.2 Rain types

The findings on the effect of six different types of rain on incidents are summarised in Table 1 and the Palm frequency results are shown also in Figure 22. It is notable that weak and moderate liquid rain has no effect on incident risk and even weak snowfall raises it only by 30%. However, heavy liquid rain as well as moderate and heavy snowfall raise it, respectively, by 190%, 470% and 740%. The Kullback-Leibler test shows a highly significant difference in whole distributions, but the heaviest rain conditions are rare enough to make the significance test fail for these.

Table 1. Effect of different types of rain on incident risk.

	Rain type						
	No rain	Weak, water	Moderate, water	Heavy, water	Weak, snow	Moderate, snow	Heavy, snow
Palm frequency	84.29%	6.82%	0.59%	0.1%	7.72%	0.25%	0.05%
Significance of difference in Palm frequency	T	F	F	F	T	T	F
Risk level (density ratio)	0.94	1.1	0.96	2.9	1.3	5.7	8.4

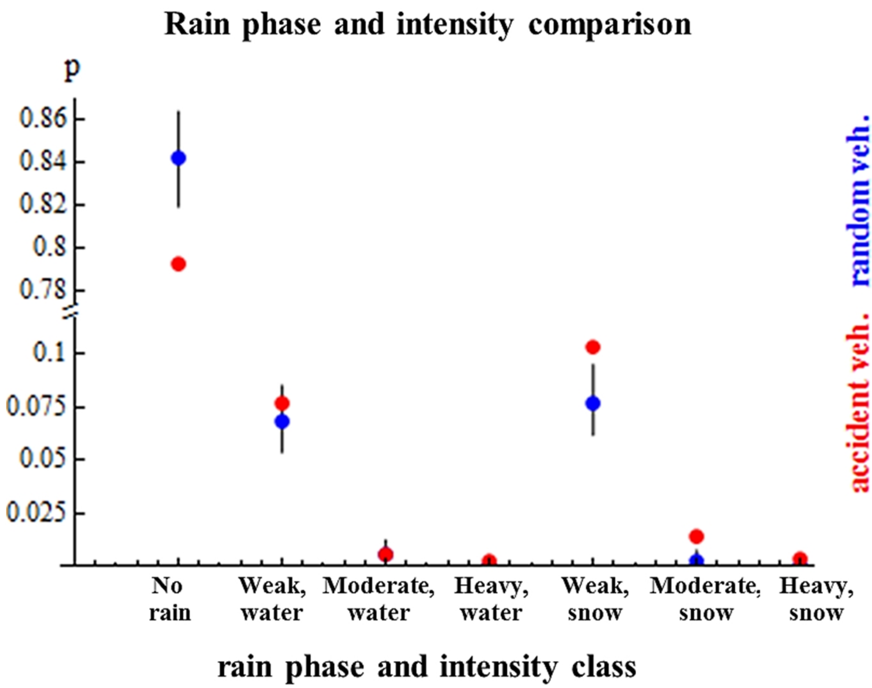

Figure 22. Test of the impact of rain phase on road accidents.

3.2.3 Visibility conditions

The weather station announced good visibility (at least 2 km) with Palm frequency 97% (i.e. 97% of vehicles driving when visibility is good). Figure 23 plots the risk levels (density ratios) of impaired visibility values at resolution 200 m. Reduced visibility (less than 2 km) seems to increase the risk, and the most extreme situations (visibility 600 m or less) increased it considerably. However, these conditions are rare, and the incident numbers show a statistically significant difference from the Palm distribution only in the visibility groups 1.2 and 1.6 km (Figure 24). The distributions are, however, significantly different by the Kullback-Leibler criterion.

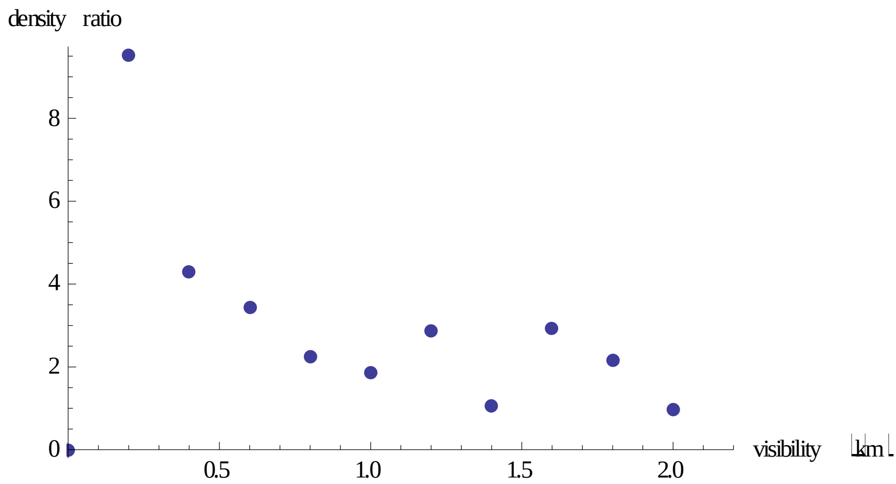

Figure 23. Incident-time *vs.* Palm density ratios of visibility.

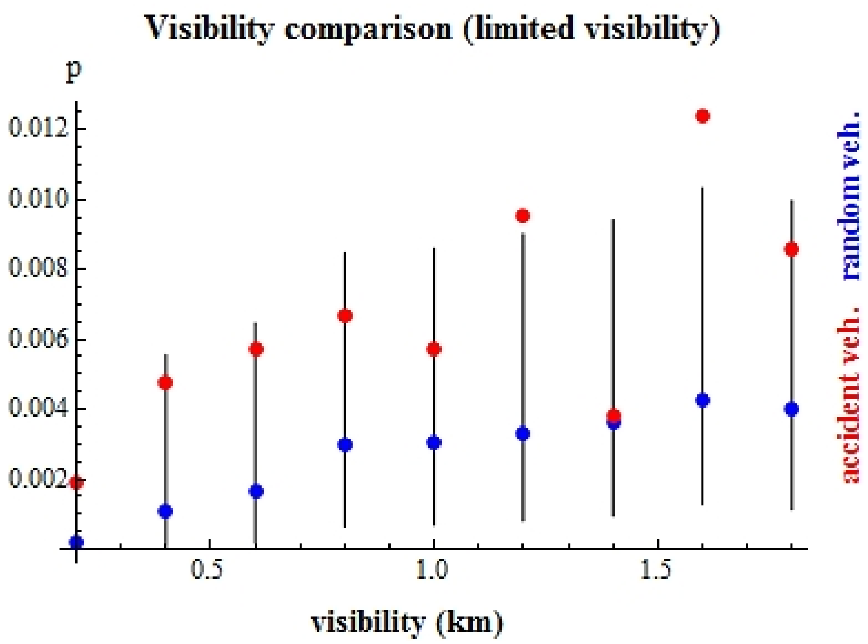

Figure 24. Test of the impact of reduced visibility on road accidents. The last class, 2 km visibility, was dropped from the figure as it is the most typical case with probability around 0.95. The cases in the graph form the rest, i.e. 5% of the data.

3.2.4 Road conditions and traffic warnings

Table 2 and Figure 25 summarise the findings on the effect of road surface conditions. The high accident probability on snowy road (risk increase 130%, statistically significant difference from Palm) is related to the similar impact of snowfall. The effect of ice is smaller (computed risk increase 70%, difference from Palm not statistically significant, however). On the other hand, the risk increase of wet road surface is rather low (30%), but this condition is quite common and the difference from Palm is statistically significant. As a whole, the distributions differ very significantly in terms of the Kullback-Leibler divergence.

Table 2. Effect of road surface conditions on incident risk.

Road surface condition	Palm frequency	Significance of difference in accident frequency	Density ratio
Icy	0.9%	F	1.7
Moist	13.8%	F	0.9
Dry	51%	F	0.99
Snowy	3.1%	T	2.3
Wet	9.4%	T	1.3
Wet and salted	3.7%	F	1.3
Probably wet and salted	12.5%	F	0.9

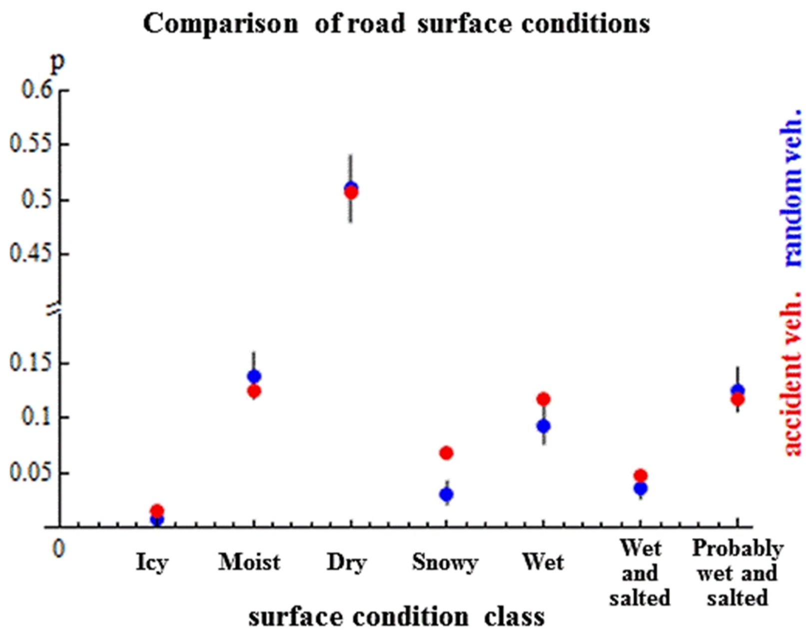

Figure 25. Test of the impact of road surface conditions on road accidents.

The final study object was to examine the impact of traffic warning status. Here, 'warning' means a situation where the road weather station gives the road maintenance authority a warning that there may be frost or rain, or the surface is starting to freeze and salting may be needed (see Chapter 2.1 for more detail).

The results are summarised in Table 3 and Figure 26. When the warning level was 'OK', there were indeed (statistically significantly) slightly less accidents than otherwise. The 'Alarm' and 'Rain' warning states were linked to 100% and 40% rise of incident risk, respectively, both observations being statistically significant. The warning for frost left the incident risk completely intact. The distributions differed significantly also in terms of the Kullback-Leibler divergence.

Table 3. Relation of traffic warnings to incident risk.

Warning level	Palm frequency	Significance of difference in accident frequency	Density ratio
Empty	5.6%	F	–
Alarm	0.7%	T	2.0
Frost	3.3%	F	1.0
OK	79.4%	T	0.94
Rain	8.5%	T	1.4
Warning	2.6%	F	1.4

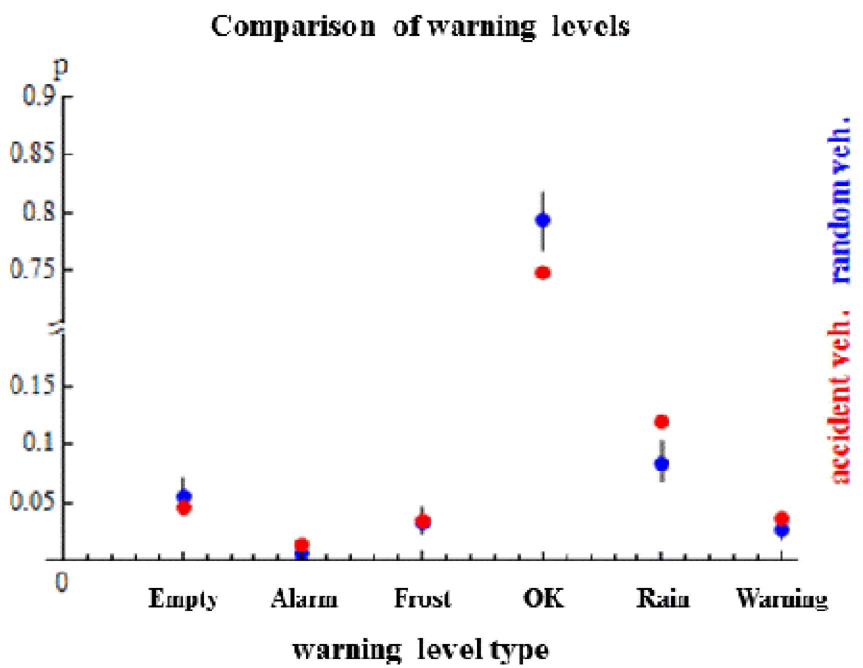

Figure 26. Test of the impact of warning levels on road accidents.

4. Discussion

This study was designed to apply the method proposed by Innamaa et al. (2013) to identify conditions when the risk of a traffic incident is elevated and develop the method further. A further aim was to find factors that do indeed affect traffic incident risk with statistical significance by piloting the method on Ring road I in the Helsinki Metropolitan Area.

The main results showed that with the proposed method the riskiness of weather and traffic conditions could be assessed and risky conditions identified. Specifically, there were several specific weather conditions that were more common among drivers who were involved in an accident than among drivers in general. These conditions included air temperature from –6 degrees Celsius down, snowfall or heavy rain, limited visibility, and snowy or wet road surface. Furthermore, the results showed that the probability of an accident is higher in conditions when a weather alarm is given by the Transport Agency (the road keeper) than in general.

Examples of risky weather conditions include the following specific findings: With air temperature –6 – –9 degrees Celsius, the risk of an accident is about 50% higher than generally, and even higher with air temperature below –15 degrees Celsius. Even weak snowfall increases the risk by 30%. However, heavy rain and moderate to heavy snowfall increase the accident risk substantially. These weather conditions are related to the time at which the road is wet or snowy despite the highest winter maintenance class. A snowy road increases the risk by 130%, icy road by 70% and wet road by 30%. Poor visibility (less than 2 km) increases the accident risk, but with visibility less than 300 metres the risk is very high. Such risky weather conditions are somewhat rare in the Helsinki region. Nevertheless, the high risk level is seen in a number of accidents, as e.g. 10% of accidents occurred during snowfall and 11% during rain.

The results are in line with those of Salli et al. (2008), who showed that the rarer specific winter road conditions are, the greater the risk of an accident. Relating to specific conditions, Salli et al. found the relative accident risk to be highest on icy and slushy roads, or 4.1 times that of bare pavement road conditions (a higher estimate than in our results). For fatal accidents, Salli et al. (2008) estimated the risk in conditions of loose snow or slush to be 4.9 times that of bare pavement conditions.

In this study, factors other than weather were shown to be related to a higher probability of an accident than in general. Although the day of the week was not found to be relevant from an accident risk point of view, the time of day was. In weekday afternoon traffic (15–17 o'clock) the risk of accident was 50% higher than generally. In night time traffic (2–5 o'clock) the risk was even higher, but traffic volumes are so low that night time accidents are still rather rare. The correlation of night time, darkness and very low traffic explains the finding of elevated accident risk for the two latter. The high accident rate at night when the traffic volume was lowest was in agreement with the earlier results of Pajunen and Kulmala (1995).

Ishak and Alecsandru (2005) concluded that pre-, post-, and non-incident traffic conditions may not be readily discernible from each other and that specific characteristics of precursory conditions to incidents may not be clearly identifiable. Such a conclusion, however, was (also) driven by limited incident and traffic datasets and selected second-order traffic performance measures. Additionally, environmental factors such as inclement weather conditions were not accounted for in this study.

The results of this study support the conclusion of Ishak and Alecsandru (2005) that traffic situation correlated poorly with accident risk. However, also our results related to the traffic situation can be considered only indicative. The accident data available for this pilot did not include reliable information on the direction of traffic where the accident took place, thus an unreliable estimate had to be used in defining the traffic situation prior to the accident. Nevertheless, traffic conditions identified as risky included traffic density 90–100 vehicles/km and speeds below the limit (under 70 km/h). However, normal fluent traffic with speed slightly over the limit (80 km/h) was seen to have a statistically very significantly lower accident probability than otherwise. Higher speeds are related to situations with low traffic volume and good driving conditions. Lower speeds are thus related to circumstances where driving is in some respect inconvenient, and here the accident risk is seen to be heightened, for many speed intervals significantly. The result is not in line with the finding of Pajunen and Kulmala (1995) that on four-lane roads, accident rates were highest at hourly traffic volumes of 3600–4800 vehicles; in our results high traffic volume alone did not increase the relative accident risk when traffic density is taken into account.

One would naturally expect a steady traffic flow to be safer than when vehicle speeds have considerable variation, as shown by Marchesini and Weijermars (2010). This was found to be the case also here; a standard deviation of speed (among vehicles passing a certain point within a 5 min period) above 15 km/h was related to a somewhat higher accident risk.

The most eastern section of Ring I was shown to have an 80% higher accident risk than other parts of the road. That was seen also in the number of accidents, as one third of them occurred on the easternmost 4 km of road. This part of the road has at-level intersections with traffic lights and a significant proportion of heavy traffic due to e.g. a harbour nearby.

5. Conclusions

In conclusion, the findings suggest that the proposed method for identifying conditions where accident risk is elevated, by comparing the traffic and weather circumstances just before an accident with the Palm probability of the same circumstances, indeed works. Not all the results were statistically significant due to some circumstances being rare. However, with the calculation of risk levels and Kullback-Leibler divergence, it was possible to assess the findings.

Poor weather conditions including e.g. heavy rain, snowfall, unmaintained road surface, poor visibility and lowered speed levels were shown to be related to elevated accident risk. Also night time and afternoon peak hour were risky. The identification of specific conditions in which the traffic management centre should be alerted could be applied in traffic status applications, especially in centres that are monitoring traffic and conditions over a large area.

With the data available in this study, which originated from a sparse loop detector network and included uncertainties in direction information in the accident database, the results regarding traffic situation can be considered indicative. The question remains whether having more precise location information for accidents and a denser traffic detector network would have made it possible to identify (more) traffic conditions that are related to elevated accident risk. Now one third of accidents occurred on the easternmost part of the road where the traffic monitoring network is the sparsest. Consequently, a study on risky traffic flow characteristics is left for future research.

A similar method based on Palm probabilities could be applied to identify road sections where more accidents take place than statistically should, taking into consideration e.g. traffic volumes and road maintenance class. By analysing the resulting sections, physical road environment characteristics resulting in elevated accident risk could be identified.

Acknowledgements

The authors would like to thank the Finnish Transport Agency for funding the early stages of this research (Innamaa et al. 2013) and VTT for funding the further development and testing of the method.

References

Baccelli, M. & Brémaud, P. (2003). Elements of Queueing Theory. Springer-Verlag.

Cover, T.M. & Thomas, J.A. (2006). Elements of Information Theory. 2nd edition. John Wiley & Sons.

Hranac, R., Sterzin, E., Krechmer, D., Rakha, H. & Farzaneh, M. (2006). Empirical studies on traffic flow in inclement weather. Federal Highway Administration, Washington, D.C.

Innamaa, S., Pilli-Sihvola, E. & Norros, I. (2013). Travel time and incident risk assessment. Research reports of the Finnish Transport Agency, 31/2013. Finnish Transport Agency, Finland. 74 p. + appendices 3 p.

Ishak, S. & Alecsandru, C. (2005). Analysis of freeway traffic incident conditions by using second-order spatiotemporal traffic performance measures. Transportation Research Board National Research Council, pp. 20–28.

Marchesini, P. & Weijermars, W. (2010). The relationship between road safety and congestion on motorways: a literature review of potential effects. R-2010-12. SWOV Institute for Road Safety Research, Leidschendam. 28 p.

Nilsson, G. (2004). Traffic Safety Dimensions and the Power Model to describe the effect of speed and safety. Bulletin 221. Department of Technology and Society. Lund University. Sweden.

Pajunen, K. & Kulmala, R. (1995). Tuntiliikenteen vaikutus liikenneturvallisuuteen [*The effect of hourly traffic volumes on traffic safety*]. Research reports of the Finnish National Road Administration, 37/1995. Finnish National Road Administration, Helsinki. 42 p. + app. 23 p.

Salli, R., Lintusaari, M., Tiikkaja, H. & Pöllänen, M (2008). Keliolosuhteet ja henkilöautoliikenteen riskit [*Wintertime road conditions and accident risks in passenger car traffic*]. 28 April 2008, Tampere University of Technology, Department of Business Information Management and Logistics.

Vlahogianni, E., Karlaftis, M. & Orfanou, F. (2012). Modeling the Effects of Weather and Traffic on the Risk of Secondary Incidents. Journal of Intelligent Transportation Systems, 16, pp. 109–117.

Ziemann, M., Salminen, J. & Tebest, T. (2013). Liian kovaa, liian lähellä ja keskittymättä – täällä tapahtuivat Suomen ketjukolarit vuonna 2012 [*Too fast, too close and without concentration – multi-vehicle rear-end pile-ups in Finland in 2012*]. 25 October 2013, Yle – Finnish Broadcasting Company. Available at: http://yle.fi/uutiset/liian_kovaa_liian_lahella_ja_keskittymatta__taalla_tapahtuivat_suomen_ketjukolarit_vuonna_2012/6899588 [accessed 11 February 2014].

Appendix A: Mathematical methods

APPENDIX
Mathematical methods for Road traffic incident risk assessment

Ilkka Norros and Pirkko Kuusela
VTT, Technical Research Centre of Finland
ilkka.norros@vtt.fi, pirkko.kuusela@vtt.fi

1 Introduction

This Appendix to VTT's report **Road traffic incident risk assessment** presents the mathematical notions and methods applied therein. The central notion is the "Palm distribution" of road traffic conditions, whose definition and practical computation are discussed in Section 2. Statistical testing of the significance of differences between the Palm distribution of conditions and the distribution of conditions seen at the time and place of an incident is considered in Section 3. Finally, we study the comparison of entire distributions in terms of the Kullback-Leibler divergence in Section 4.

2 "Palm distribution" of road traffic conditions

2.1 Palm probability

We first recall some basic notions and relations of the Palm theory of stationary point processes, following the presentation and notation of Baccelli and Brémaud [1].

A *counting measure* on $\mathbb{R}$ is a measure taking finite integer values on compact sets, and a random counting measure N is called a *point process*. For any Borel set $C \subseteq \mathbb{R}$, $N(C)$ is a random variable presenting the number of points in C. We assume all our point processes to be simple, i.e. $N(\{x\}) \in \{0,1\}$ for all $x \in \mathbb{R}$. Denote the epochs of N by random variables T_n, numbered as

$$\cdots < T_{-1} < T_0 \leq 0 < T_1 < \cdots ,$$

so that $N(C) = \sum_{n\in\mathbb{Z}} 1_{\{T_n \in C\}}$ for any Borel set $C \subseteq \mathbb{R}$.

Let N be a point process defined on the probability space $(\Omega, \mathcal{F}, P)$, and let $(\theta_t)_{t\in\mathbb{R}}$ be a *flow*, i.e. a group of measurable applications from $(\Omega, \mathcal{F})$ to itself such that $\theta_s \circ \theta_t = \theta_{s+t}$. The process N is said to be compatible with the flow (θ_t), if $(N \circ \theta_t)(C) = N(C + t)$ for any Borel set C. If $P \circ \theta_t = P$ for all t, then N is said to be *stationary*, and the probability P is called a *stationary probability* of N. If N is stationary, the quantity

$$\lambda_N = E\{N((0,1])\}$$

is called its *intensity*.

The *Palm probability measure* associated with N is the probability measure P_N^0 on $(\Omega, \mathcal{F})$ defined by

$$P_N^0(A) = \frac{1}{\lambda_N t} E\left\{\int_0^t (1_A \circ \theta_s) N(\,\mathrm{d}s)\right\}, \quad A \in \mathcal{F}, \tag{1}$$

where $t > 0$ is arbitrary. Intuitively, the Palm measure shows Ω from the viewpoint of an arbitrary point of the point process N, placing that point to the origin. In particular, since $T_0 \circ \theta_{T_n} = 0$ a.s. for any n, we always have $P_N^0(T_0 = 0) = 1$. Conversely, the stationary probability P can be expressed in terms of the Palm probability through Ryll-Nardzewski's inversion formula:

$$E\{f\} = \lambda_N E_N^0 \left\{\int_0^{T_1} (f \circ \theta_t)\,\mathrm{d}t\right\}. \tag{2}$$

Choosing $f = 1$ yields $\lambda_N = 1/E_N^0\{T_1\}$.

It is important to note that the stationary stochastic model $(\Omega, \mathcal{F}, P, \theta)$ may contain many more characteristics than just the points of N. The Palm probability associated with N gives the distribution of all these characteristics from the point of view of an arbitrary (typical, "randomly chosen" (although there is no uniform distribution on the infinite set of integers)) point of N.

2.2 Definition and computation of road traffic Palm distribution

For the purposes of the present study on the effect of road traffic conditions on traffic incidents, we adopt the idea of Palm probability to obtain a usable notion of *road and traffic conditions seen by an arbitrary vehicle*. To understand the need of such a notion, consider for example the fact that most incidents happen in daytime — can we infer from this that daytime is more dangerous for driving than nighttime?

Our data on daily traffic conditions is composed as follows:

- The traffic data originate from 8 LAM stations along the road Ring 1, denoted as LAM(i, d), where $i = 1, \ldots, 8$ is the number of the station and $d = 1, 2$ the direction.

- Each 5-minute period of the raw LAM measurement data is summarized in a record with the following fields:

$N_t(i,d,L)$	amount of light vehicles at LAM(i,d) in time slot t
$N_t(i,d,H)$	amount of heavy vehicles at LAM(i,d) in time slot t
$S_t(i,d,L)$	mean speed of light vehicles at LAM(i,d) in time slot t
$S_t(i,d,H)$	mean speed of heavy vehicles at LAM(i,d) in time slot t
$V_t(i,d,L)$	Standard deviation (STD) of speed of light vehicles at LAM(i,d) in ti
$V_t(i,d,H)$	STD of speed of heavy vehicles at LAM(i,d) in time slot t

- The weather and road condition data originate from a measurement station near Ring 1 and provide certain characteristics

$$W_t(1), \ldots, W_t(k)$$

  at the same 5 minute resolution as the traffic data.

- The data cover 5 years (2008-2012).

Although the classical Palm theory focuses on the relations between the stationary view and the view from an arbitrary point, stationarity plays no role in the approach developed here. Road traffic is not a stationary process, nor do we need to model it by a stationary process. In a finite setting, the counterpart of Palm probability is just the *mean over all points of their "viewpoints"*. The spatio-temporal "points" in our case are individual vehicles on the road. Since we observe them spatially just at 8×2 LAM stations and temporally as aggregates over time slots of length 5 minutes, we can only *estimate* our Palm probability from the data described above. This estimation proceeds in five stages:

- for each time slot, the *traffic density* (unit: vehicles/km) is estimated on each of the 9×2 road sections into which the LAM stations split the road;

- each of these densities is multiplied by the length of the respective road section (unit: (vehicles/km)×km = vehicles);

- localised traffic quantitites on each road section are estimated using simple interpolation formulae (Subsection 2.3);

- numerical values are quantized (discretized to take relatively few values) to make their resolution comparable to that "seen" by incidents (Subsection 2.4);

- finally, the "Palm distribution" (we drop the quotes from now on) is built by weighing the quantized value vector of each pair (time slot, road section) by its estimated amount of vehicles.

2.3 Synthesis and interpolation formulae

In this study, we do not distinguish between vehicles of different weight but synthetize them using the equivalent "1 heavy vehicle = 1.5 × light vehicle", i.e.

$$N_t(i,d) := N_t(i,d,L) + 1.5 N_t(i,d,H). \tag{3}$$

The mean speed and the standard deviation of speeds are synthetized using the following heuristic formulae, derived by thinking vehicle speeds as independent and identically distributed random variables:

$$S_t(i,d) := \frac{N_t(i,d,L)}{N_t(i,d)} S_t(i,d,L) + \frac{1.5N_t(i,d,H)}{N_t(i,d)} S_t(i,d,H), \tag{4}$$

$$V_t(i,d) := \sqrt{\frac{N_t(i,d,L)}{N_t(i,d)} E_t^2(i,d,L) + \frac{1.5N_t(i,d,H)}{N_t(i,d)} E_t^2(i,d,H) - S_t(i,d)^2}, \tag{5}$$

where

$$E_t^2(i,d,L) = V_t(i,d,L)^2 + S_t(i,d,L)^2, \quad E_t^2(i,d,H) = V_t(i,d,H)^2 + S_t(i,d,H)^2.$$

Using the same principle of independent speeds, the traffic volume, mean speed and standard deviation of speeds on road section $i+1$, limited by LAM stations i and $i+1$, are estimated by the following interpolation formulae:

$$\bar{N}_t(i+1,d) := \frac{1}{2}(N_t(i,d) + N_t(i+1,d)), \tag{6}$$

$$\bar{S}_t(i+1,d) := \frac{N_t(i,d)S_t(i,d) + N_t(i+1,d)S_t(i+1,d)}{N_t(i,d) + N_t(i+1,d)}, \tag{7}$$

$$\bar{V}_t(i+1,d) := \sqrt{\frac{N_t(i,d)E_t^2(i,d) + N_t(i+1,d)E_t^2(i+1,d)}{N_t(i,d) + N_t(i+1,d)} - \bar{S}_t(i+1,d)^2}, \tag{8}$$

where

$$E_t^2(i,d) = V_t(i,d)^2 + S_t(i,d)^2.$$

The first and last road section are limited by a single LAM station each, so for them we choose simply

$$\begin{array}{lll} \bar{N}_t(1,d) = N_t(1,d), & \bar{S}_t(1,d) = S_t(1,d), & \bar{V}_t(1,d) = V_t(1,d), \\ \bar{N}_t(9,d) = N_t(8,d), & \bar{S}_t(9,d) = S_t(8,d), & \bar{V}_t(9,d) = V_t(8,d). \end{array}$$

Finally, our traffic density estimate $\bar{D}_t(i,d)$ on road section i and direction d in time slot t is

$$\bar{D}(t,i,d) := \frac{12\bar{N}_t(i,d)}{\bar{S}_t(i,d)}, \quad i = 1,\ldots,9, \quad d = 1,2. \tag{9}$$

(Here the factor 12 is the number of 5 minute slots in an hour.)

2.4 Palm probability with quantization

For each index triple (t,i,d) (where i now refers to road section, not to LAM station) we now have a vector of traffic and weather characteristic values

$$X(t,i,d) = (X_1(t,i,d),\ldots,X_m(t,i,d)),$$

with $m = 13$ and the meanings as listed on page 22 of the main report, and its weight, the traffic density estimate (9). Let the domain of variable X_i be $\mathcal{X}_i$, $i = 1, \ldots, m$. Our Palm distribution is initially defined on $\mathcal{X} = \prod_{i=1}^m \mathcal{X}_i$ through

$$P^0_{\mathcal{X}}(X_1 \in A_1, \ldots, X_m \in A_m) = \frac{\sum_{t,i,d} \bar{D}(t,i,d) 1_{\{X_1(t,i,d) \in A_1, \ldots, X_m(t,i,d) \in A_m\}}}{\sum_{t,i,d} \bar{D}(t,i,d)}.$$

Although finite, this object is much too big for practical computation, and it is also too fine-grained to be reasonably compared with our distribution of conditions at incident times. Therefore we make the quantization described in the table on page 21 of the main report. Formally, this means that we select appropriately sparse subsets $\Omega_i \subset \mathcal{X}_i$ and quantization functions $\phi_i : \mathcal{X}_i \to \Omega_i$, $i = 1, \ldots, m$, and consider the coarser-grained variables $Y_i = \phi_i(X_i)$ instead of the original data. Our final Palm probability on $\Omega = \prod_{i=1}^m \Omega_i$ is defined, similarly as above, by

$$P^0_{\Omega}(Y_1 \in B_1, \ldots, Y_m \in B_m) = \frac{\sum_{t,i,d} \bar{D}(t,i,d) 1_{\{Y_1(t,i,d) \in B_1, \ldots, Y_m(t,i,d) \in B_m\}}}{\sum_{t,i,d} \bar{D}(t,i,d)},$$

but the atoms of $P^0_{\mathcal{X}}(\cdot)$ are much fewer than the those of $P^0_{\Omega}(\cdot)$.

3 Statistical testing of a single traffic condition

The Palm distribution expresses various traffic environmental conditions or states from the perspective of an arbitrary vehicle on the road. Although there is extensive traffic data available only a few vehicles get into an accident. In accident cases also the traffic environment conditions have been recorded as accurately as possible. Our research question is: Are some traffic environmental conditions more prone to cause accidents? This calls for statistical comparison of observed environmental conditions from the arbitrary vehicle and accident vehicle point of view.

The derivation of the Palm and accident distributions has been described earlier in this appendix. The challenge in comparing Palm and accident vehicle conditions lies in the fact that the Palm distribution is estimated from a huge population whereas there are few accidents. Several traffic conditions have no recorded accidents and even if accidents exist, there are very few of them.

Traffic conditions can be seen as binomially distributed; we observe that a certain number of accident vehicles falls in a given traffic condition class. This gives the probability (frequency) of the class. In addition to the point estimate for the probability, namely the number of accident vehicles in the class divided by the population size, we wish to build a confidence interval around our estimate. Typically this is done by making a normal distribution approximation of the binomial distribution.

However, in our case such an approximation would be poor, as the probability of falling into the same traffic condition, p, can be very small. A typical rule of thumb suggests that $np > 5$ or $n(1-p) > 5$, where n is the population size and p is the occurrence probability of the inspected outcome. In addition, estimation is poor around $p = 0$ and $p = 1$. Unfortunately, this is the case in this data and we need to build a so-called "exact" confidence interval [2].

In our statistical testing of accident condition frequencies we take our H_0 hypothesis to be that the occurance probabilities of traffic environment conditions are the same with an arbitrary or an accident vehicle point of view. This means that the "true" occurrence probabilities in traffc accident cases would be given by the Palm distribution, which is calculated from a very large sample. However, the accident vehicle sample is very small. We ask what is the acceptable range of observed probabilities if we draw a small random sample from the Palm distribution. For this we derive the exact binomial confidence interval described next.

Assume that we have an observation of size n and k vehicles fall into a given traffic condition (i.e. "success"). The exact binomial $(1-\alpha)$ confidence interval for p is an interval $[p_{ub}, p_{lb}]$ such that

$$\sum_{k=0}^{x} \binom{n}{k} p_{ub}^{k} (1-p_{ub})^{n-k} = \frac{\alpha}{2}$$

and

$$\sum_{k=x}^{n} \binom{n}{k} p_{lb}^{k} (1-p_{lb})^{n-k} = \frac{\alpha}{2}.$$

These two expressions define the lower and the upper tails of the binomial probability density function (PDF), and with risk level α we have required those tails to have probabilities $\alpha/2$, as illustrated in Figure 1 of this appendix. The lower tail sets a condition to p_{ub} whereas the upper tail sets a condition for p_{lb}.

The upper and lower tails can be expressed by the cumulative distribution function (CDF). With binomial distribution CDFs can be calculated using so-called beta functions. When x is an integer then

$$\sum_{k=0}^{x} \binom{n}{k} p^{k} (1-p)^{n-k} = \mathrm{BetaReg}(1-p, n-x, 1+x),$$

where

$$\mathrm{BetaReg}(z, a, b) = \frac{B_z(a,b)}{B_1(a,b)}$$

with

$$B_z(a,b) = \int_0^z t^{a-1}(1-t)^{b-1} dt.$$

InvBetaReg is the inverse of the regularized betafunction BetaReg, i.e., InvBetaReg$[s, a, b]$ is a solution for z in $s = \mathrm{BetaReg}[z, a, b]$. Thus

$$\mathrm{BetaReg}(1-p_{ub}, n-k, 1+k) = \frac{\alpha}{2}$$

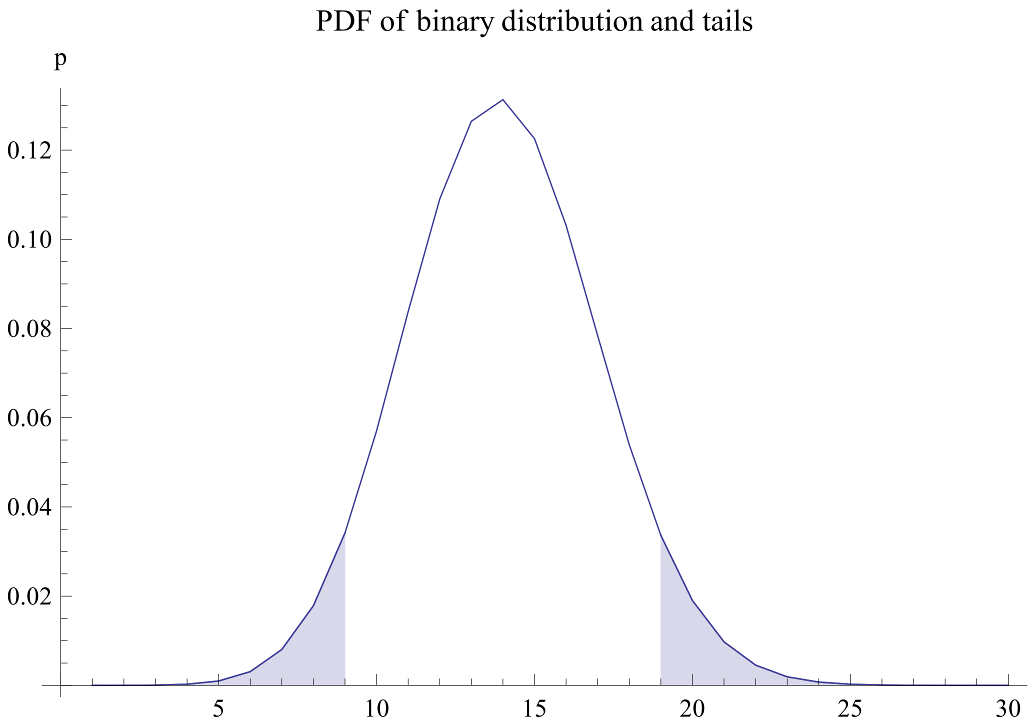

Figure 1: Illustration of the PDF of binary distribution. Upper and lower tails are illustrated by shaded areas. A shaded area corresponds to the risk level of the statistical test.

and expressing the upper tail using the cumulative distribution function (i.e., changing the summation to run from 0 to $k-1$ instead of from k to n) gives rise to

$$\mathrm{BetaReg}(1-p_{lb}, n-k+1, k) = 1 - \frac{\alpha}{2}.$$

Using InvBetaReg to solve for p_{lb} and p_{ub} yields to expressions

$$\begin{aligned} p_{ub} &= 1 - \mathrm{InvBetaReg}(\frac{\alpha}{2}, n-k, k+1) \\ p_{lb} &= 1 - \mathrm{InvBetaReg}(1-\frac{\alpha}{2}, n-k+1, k). \end{aligned} \tag{10}$$

Thus given a traffic condition and its Palm probability, we calculate according to the null hypothesis our estimate for vehicles in this class (i.e., number of expected successes, k) when we draw a random sample of size n. Here n is the size of the accident vehicle population. From this we estimate the 95% confidence interval (i.e. $\alpha = 0.05$) $[p_{lb}, p_{ub}]$, where p_{lb} and p_{ub} are given by (10). If the probability obtained from the accident data falls outside this confidence interval, we reject our null hypothesis and conclude that accident conditions differ significantly from conditions faced by an arbitrary vehicle.

4 Kullback-Leibler divergence (relative entropy)

The Kullback-Leibler divergence, alias relative entropy, of a probability measure Q on a measurable space $(\Omega, \mathcal{F})$ with respect to another probability measure P on $(\Omega, \mathcal{F})$ is defined as

$$D(Q||P) = \begin{cases} \int_\Omega \log \frac{\mathrm{d}Q}{\mathrm{d}P}(\omega) Q(\mathrm{d}\omega), & \text{if } Q \text{ is absolutely continuous w.r.t. } P \\ \infty, & \text{otherwise,} \end{cases}$$

where $\mathrm{d}Q/\mathrm{d}P$ denotes the density of Q with respect to P. Intuitively, the number $D(Q||P)$ tells how much information (in bits) one obtains when learning that the true distribution is Q, having so far believed that it is P. The quantity $D(Q||P)$ is not symmetric w.r.t. P and Q, and in particular it is not a metric.

In our study, we compare the accident probability Q with the Palm probability P^0 for different characteristics separately. The relative entropy is simple to compute, but what is less obvious is what the relative entropy value tells us. Here we used a pragmatic approach: we generate 20 samples from P^0 that each has the size of our set of incidents, and compute for each sample the relative entropy $D(\widehat{P}||P^0)$, where $\widehat{P}$ denotes the empirical distribution given by the sample. If these numbers are all smaller than $D(Q||P^0)$, we infer that Q differs significantly (confidence level 95%) from P^0, since the randomness of incident times does not explain why Q differs this much from P^0.

References

[1] F. Baccelli and P. Brémaud, *Elements of Queueing Theory.* Berlin: Springer Verlag, 2003.

[2] S. Wallis, "Binomial confidence intervals and contingency tests: mathematical fundamentals and the evaluation of alternative methods." *Journal of Quantitative Linguistics*, vol. 20, no. 3, pp. 178–208, 2013.

Series title and number
VTT Technology 172

Title	**Road traffic incident risk assessment** **Accident data pilot on Ring I of the Helsinki Metropolitan Area**
Author(s)	Satu Innamaa, Ilkka Norros, Pirkko Kuusela, Riikka Rajamäki & Eetu Pilli-Sihvola
Abstract	The purpose of this project was to apply the Palm distribution to the analysis of riskiness of different traffic and road weather conditions introduced in a previous project (Innamaa et al. 2013), develop the method further, and find factors that statistically significantly affect traffic incident risk. The method was piloted using data from Ring-road I of the Helsinki Metropolitan Area. The study was based on registered accidents that occurred on Ring-road I in 2008–2012, totalling 1120. In addition to accident data, traffic data from eight automatic traffic measurement stations (inductive loops) and road weather data were also used. The basic methodological idea was to compare the traffic and weather circumstances just before an accident with the Palm probability of the same circumstances. The notion of Palm probability comes from the theory of random point processes, and means the probability distribution "seen" by a randomly selected point of the point process, i.e. the driver in this case (in contrast to the stationary probability, which is the probability distribution seen at a random time point). If each car driver had a constant stochastic intensity of causing an accident, then the accident circumstances would follow the Palm distribution. The idea of the method applied here is to assess the influence of circumstances on incidents by comparing the incident circumstance distribution with the Palm distribution of circumstances: differences between these distributions hint at effects of circumstances on accidents. The results showed that there were several specific weather conditions that were more common among drivers who were involved in an accident than among drivers in general. These conditions included air temperature from –6 degrees Celcius down, snowfall or heavy rain, limited visibility, and snowy or wet road surface. The results further showed that the probability of an accident is higher in conditions when a weather alarm is issued by the Transport Agency (the road operator) than in general. In addition, in weekday afternoon traffic (15–17 o'clock) the risk of accident was found to be 50% higher than generally. In night time traffic (2–5 o'clock) the risk was even higher. The results indicated that the traffic situation correlated poorly with accident risk. However, the results related to the traffic situation can be considered only indicative due to inaccuracies in the accident location information and sparseness of the traffic detector network. In conclusion, the findings suggest that the proposed method for identifying conditions where accident risk is elevated by comparing the traffic and weather circumstances just before the accident with the "Palm probability" of the same circumstances indeed works. Not all results were statistically significant due to some circumstances being rare. However, with the calculation of risk levels and Kullback-Leibler divergence, it was possible to assess the findings.
ISBN, ISSN	ISBN 978-951-38-8257-0 (URL: http://www.vtt.fi/publications/index.jsp) ISSN-L 2242-1211 ISSN 2242-122X (Online)
Date	May 2014
Language	English, Finnish abstract
Pages	49 p. + app. 8 p.
Name of the project	
Commissioned by	
Keywords	Traffic incident risk, Palm probability, driving condition
Publisher	VTT Technical Research Centre of Finland P.O. Box 1000, FI-02044 VTT, Finland, Tel. +358 20 722 111

Julkaisun sarja ja numero
VTT Technology 172

Nimeke	**Tieliikenteen häiriöriskin arviointi** **Pilotti Kehä I:n onnettomuusaineistolla**
Tekijä(t)	Satu Innamaa, Ilkka Norros, Pirkko Kuusela, Riikka Rajamäki & Eetu Pilli-Sihvola
Tiivistelmä	Tämän hankkeen tavoitteena oli soveltaa Palm-jakaumaan perustuvaa aikaisemmassa hankkeessa (Innamaa ym. 2013) kehitettyä menetelmää analysoida erilaisiin liikenne- ja keliolosuhteisiin liittyvää liikenteen häiriöiden riskiä, kehittää menetelmää eteenpäin ja löytää tekijöitä, jotka vaikuttavat liikenteen häiriöiden riskiin tilastollisesti merkitsevästi. Menetelmää kokeiltiin Kehä I:llä kootulla aineistolla. Tutkimus perustui Kehä I:llä vuosina 2008–2012 tapahtuneisiin, rekisteröityihin onnettomuuksiin, joita oli yhteensä 1120. Tutkimuksessa käytettiin onnettomuusaineiston lisäksi liikennetietoja kahdeksasta liikenteen automaattisesta mittauspisteestä (induktioilmaisimia) ja tiesääaseman tuottamaa tietoa. Menetelmällinen perusajatus oli verrata liikenne- ja sää-/keliolosuhteita hetkeä ennen onnettomuutta samojen olosuhteiden Palm-todennäköisyyksiin. Palm-todennäköisyyden käsite tulee satunnaisen pisteen prosessoinnin teoriasta ja tarkoittaa satunnaisesti valitun prosessointipisteen, tässä tapauksessa autoilijan, ”näkemää” todennäköisyysjakaumaa (vastakohtana stationaariselle todennäköisyydelle, joka kuvaa todennäköisyysjakaumaa satunnaisella ajanhetkellä). Jos jokaisella autoilijalla on vakiosuuruinen stokastinen intensiteetti aiheuttaa onnettomuus, onnettomuusolosuhteet noudattavat Palm-jakaumaa. Menetelmässä on sovellettu ideaa arvioida olosuhteiden vaikutus liikenteen häiriöihin vertaamalla häiriöiden olosuhdejakaumaa olosuhteiden Palm-jakaumaan: Näiden jakaumien väliset erot viittaavat olosuhteiden onnettomuusvaikutukseen. Tulokset osoittavat useita sää- ja keliolosuhteita, jotka olivat yleisempiä onnettomuuksiin joutuneiden ajoneuvojen kohdalla kuin yleensä. Tällaisia olosuhteita olivat korkeintaan –6º C ilman lämpötilan lumisade tai rankka vesisade, huono näkyvyys ja luminen tai märkä tienpinta. Lisäksi tulokset osoittivat, että onnettomuuden todennäköisyys on korkeampi olosuhteissa, jolloin Liikennevirasto antaa kelivaroituksen, kuin muuten. Lisäksi arki-iltapäivien (klo 15–17) liikenteessä onnettomuusriski oli 50 % korkeampi kuin yleensä. Yöaikaan (klo 2–5) riski oli vielä korkeampi. Tulokset osoittivat, että liikennetilanne korreloi huonosti onnettomuusriskin kanssa. Liikennetilanteeseen liittyviä tuloksia voi kuitenkin pitää vain suuntaa-antavina onnettomuuden paikkatiedon epäluotettavuuden ja harvan liikenneilmaisimien verkon takia. Yhteenvetona voidaan todeta tulosten viittaavan siihen, että ehdotettu menetelmä toimii eli että menetelmä tunnistaa olosuhteet, jolloin onnettomuusriski on kohonnut vertaamalla liikenne- ja sää-/keliolosuhteita hetki ennen onnettomuutta samojen onnettomuuksien Palm-jakaumaan. Kaikki tulokset eivät olleet tilastollisesti merkitseviä olosuhteiden harvinaisuuden takia. Tulokset voitiin kuitenkin arvioida laskemalla riskitasoja ja Kullback-Leibnerin divergenssi.
ISBN, ISSN	ISBN 978-951-38-8257-0 (URL: http://www.vtt.fi/publications/index.jsp) ISSN-L 2242-1211 ISSN 2242-122X (verkkojulkaisu)
Julkaisuaika	Toukokuu 2013
Kieli	Englanti, suomenkielinen tiivistlmä
Sivumäärä	49 s. + liitt. 8 s.
Projektin nimi	
Toimeksiantajat	
Avainsanat	Traffic incident risk, Palm probability, driving condition
Julkaisija	VTT PL 1000, 02044 VTT, puh. 020 722 111

Road traffic incident risk assessment

Accident data pilot on Ring I of the Helsinki Metropolitan Area

ISBN 978-951-38-8257-0 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN-L 2242-1211
ISSN 2242-122X (Online)

VTT TECHNOLOGY 172 Road traffic incident risk assessment