Applying the Palm distribution for bicycle crash risk assessment and informed policy making

Marcus Skyum Myhrmann[a,*], Stefan Eriksen Mabit[a]

[a] *Transport Division, DTU Management, Technical University of Denmark, Bygningstorvet Building 116B, Kgs. Lyngby, 2800, Denmark*

Abstract

This study analyses factors associated to bicycle accidents by a notion of Palm theory for traffic conditions. The method allows for the comparison of the distribution of conditions as seen by an arbitrary cyclist (the Palm distribution) with those seen by a cyclist subject to an accident (the accident distribution). This allows for a straight forward assessment of the relative risk change given different conditions as well as their statistical significance. The study is based on accident data from police reports of bicycle accidents over 4 years, in Copenhagen and Frederiksberg (N = 1136). Relative risk change compared to the overall risk (Palm frequency) was evaluated given time, weather and seasonality. Relative risk increase was found to be significant at night (0-4) as well as in morning and afternoon peak hours. Upon further evaluation, the night effect is only significant in the weekend, while morning and afternoon peak are only associated to significant increase during the week. This demonstrates how using over-aggregated explanatory variables might cause misleading conclusions with regard to interventions. Furthermore results indicate that precipitation is associated to a general increase in the riskiness. Overall the Palm distribution for traffic conditions offers a novel way to capture factors associated to bicycle accidents.

1. Introduction

With a growing world-wide concern regarding man-made climate change (Stocker and Midgley, 2013), cycling is being endorsed as a means of transportation that improves both urban livability and physical health of users (Infrastructures, 2015).

A reported barrier to increase the mode share of the bicycle is fear of accidents (Horton, 2016; Transport for London, 2014; Vejdirektoratet, 2018). As such a thorough understanding of the factors associated to bicycle crashes is needed.

Previous research into crash count analysis has primarily made use of different count model regressions using multivariate parametric approaches Mannering and Bhat (2014). This allowed researchers to study the effects of different characteristics present at the accident, by analysing estimates in relation to the response in their models. Such models mainly used monthly or daily crash counts as response variable. This means that

*Corresponding author. Tel: +4542956003

Email addresses: mskyum@dtu.dk (Marcus Skyum Myhrmann), smab@dtu.dk (Stefan Eriksen Mabit)

some level of aggregated variables such as median or accumulated rainfall (Theofilatos and Yannis, 2014) (for weather), Average Annual daily traffic (AADT, for traffic flow) (Mannering et al., 2016) were used. These values have various limitations. For example it is well known that weather conditions are confounding factors, affecting both exposure and likelihood of being in an accident (Stipdonk, 2008). Similarly AADT, while being a somewhat telling measure with regard to the size of the road, yields little information about the traffic state at the time of the accident. Furthermore the models are most often based on accident records solely. This data alone provides little information on the standard behaviour of the "system" such as confrontations and near misses, as it effectively does not see them.

The information on standard behaviour is applied by Norros et al. (2016) in the context of car accident analysis. They address how detailed traffic data provide insights in explaining crash data. With that in mind, this study intends to employ a non parametric model for evaluating the bicycle crash risk when taking account for bicycle flow and weather at an hourly level.

The applied method is based on the notion of Palm probability conceptualised for traffic conditions (Norros et al., 2016). The method stems from the theory of count processes and describes conditions from the reference frame of a uniformly randomly chosen random point. In this case the random point will be an arbitrary bicyclist. Thereby the method allows for a straight forward comparison of the distribution of conditions 'seen' by an arbitrary cyclist to that 'seen' by the cyclists subject to an accident, as well as the statistical significance.

Simply put, it allows us to investigate whether the proportional occurrence of accidents given certain conditions is significantly different from the proportional amount of cyclists given the same conditions.

Taking into account the recent years of changing of the weather towards the more extreme (Allan and Soden, 2008), in-depth understanding on which weather conditions lead to increased risk of accidents is crucial for future policy making and preventive efforts.

Therefore the aim of this study is to study the road and weather conditions associated to accident risk for bicycles. To deliver relevant results, enabling meaningful interventions, the analysis will not only assess first-order effects. The study will address and evaluate the interaction of multiple factors that are often neglected in classic parametric multivariate model studies. This is important as the interactions could easily be believed to have non-linear effects, which are important.

2. Method

Palm probability is a notion known from the theory of random point processes. It expresses the distribution of events as "seen" by a randomly chosen point from the point process. In our setting it equivalently describes the distribution of traffic and road conditions as "seen" by an arbitrary cyclist. The idea to apply such a variant of the Palm probability was pioneered by Innamaa et al. (2013) and further conceptualised by Norros et al. (2016).

Assuming we have a time series of continuously observed conditions $S_{t,r} \in S$ (various weather and traffic conditions) and the associated traffic $M_{t,r}$ at a given road sections r and time t. Then the Palm probability of a specific condition $C \in S$ is found by weighing the time-average probability by the traffic $M_{t,r}$.

$$P^0(C) = \frac{1}{M^{tot}} \sum_{i=1}^{R} \int I_{S_{t,r} \in C} M_{t,r} dt \quad \text{where} \quad M^{tot} = \sum_{r=1}^{R} \int M_{t,r} dt \tag{1}$$

As such the palm probability describes the conditions as "seen" by an arbitrary cyclist picked from the combined road sections at time t. However, since the traffic and weather observations in this study are hourly, i.e $t \in \{1, 2, 3, 4, 5, ..., 24\}$, the integral reduces to a sum. And hence the Palm probability of a conditions $C \in S$

$$P^0(C) = \frac{1}{M^{tot}} \sum_{r=1}^{R} \sum_{t=1}^{T} M_{r,t} 1_{\{S_{t,r} \in C\}} \tag{2}$$

Equivalently we describe the empirical accident distribution, i.e. the distribution of conditions seen by a recorded accident, of road conditions as

$$P^{acc}(C) = \frac{1}{N_{acc}} \sum_{r=1}^{R} \sum_{t=1}^{T} k_{r,t} 1_{\{S_{r,t} \in C\}} \tag{3}$$

Taking the ratio of the empirical accident distribution of conditions C and the equivalent Palm distribution we obtain the average proportional change in the intensity of having an accident given condition C (Norros et al., 2016).

$$\frac{P^{acc}(C)}{P^0(C)} \approx E^0 \left[\frac{\alpha(\cdot)}{\alpha_0} \middle| C \right] \tag{4}$$

where α_0 denotes the overall mean intensity of causing an accident for each cyclists and $\alpha(\cdot)$ is the intensity given of causing an accident given the traffic conditions $S_{t,r} \in C$ under the conditions C.
This notion stems from the underlying assumption that these follow a Poisson random process and there being a positive intensity of having an accident. Equivalently equation 4 describes the increase or decrease in the relative risk, when compared to the overall risk level stated by the Palm distribution of the condition. Hence a ratio of 1, means that there is no change in the relative risk associated to a specific condition, while a ratio higher or lower than 1 indicates the fractional effect indicated, i.e 0.5 indicates a 50% reduction in the relative risk given a specific condition. This method therefore presents a non-parametric method for assessing the relative crash risk given any condition C.

Significance testing of the relative risk change given conditions C is performed by obtaining the exact Binomial confidence intervals $[p_l, p_u]$ of the Palm distribution $P^0(C)$ given the number of accidents N_{acc}. The reason for this is due to the low probabilities under some conditions. Under the assumption that any cyclist has a constant stochastic intensity for being involved in an accident, described by the Palm distribution, then if $P^{acc} \notin [p_l, p_u]$ we assess the the effect to be significant at the chosen level. The lower and upper confidence intervals are determined as

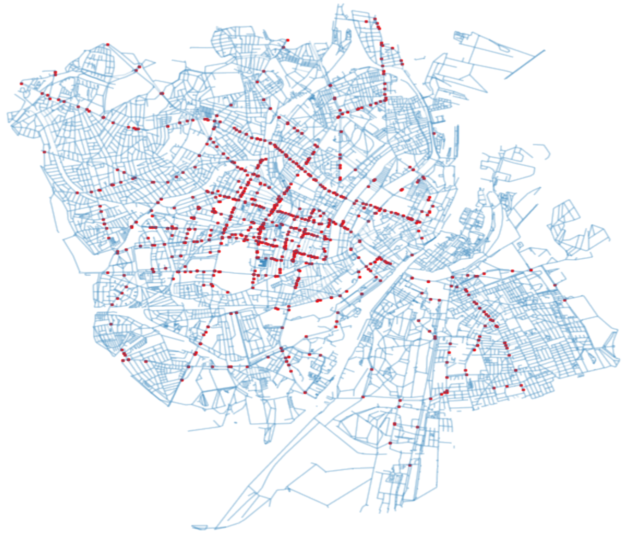

Figure 1: Image of the bicycle network used for model analysis. The red dots show the police registered bicycle accidents, and the blue shows the network structure

$$p_u = 1 - BetaInv(\frac{\alpha}{2}, N - P^0(C) * N_{acc}, P^0(C) * N_{acc} + 1) \tag{5}$$

$$p_l = 1 - BetaInv(1 - \frac{\alpha}{2}, N - P^0(C) * N_{acc} + 1, P^0(C) * N_{acc}) \tag{6}$$

where $P^0(C) * N_{acc}$ describes the expected number of accidents for condition C given Palm distribution $P^0(C)$ and the total number of observed accidents. We set $\alpha = 0.05$, thus assessing significance at the 5% level.

3. Data

3.1. Accident data

The study analysed $N = 1136$ bicycle accidents recorded by the police, having occurred in either Copenhagen or Frederiksberg municipality between 2014 and 2017. Each accident is registered along with its location. Based on the coordinates of the crash location, each accident was assigned to a bicycle lane link in Copenhagen or Frederiksberg, from a "bicycle network" extracted from crowdsourced OpenStreetMap (OSM) data. A plot of the network and accidents used in the study is shown in figure 1.

3.2. Traffic Volume

Far from all bicycle links in the bicycle network are monitored continuously. Previous work however have shown how both model generated estimates of exposure as well as "expert knowledge" from Transport agencies can be used with excellent results (Aldred et al., 2018; Malin et al., 2017)

In this study traffic assignment to the bicycle links in the network is based on the Copenhagen Model for Person Activity Scheduling (COMPAS) that produces daily activity plans for a synthetic population of the area. By using the general traffic calculation method by The Danish Road Directorate we are able to derive the hourly traffic for all days in the period of recorded accidents.

3.3. Weather data

Weather data was obtained from OpenWeatherMap for the entire Copenhagen region, including Frederiksberg. The data contained hourly mean temperatures ($^{\circ}C$), wind-speeds (m/s) as well as weather type categories: "clear", "clouds", "rain", "mist", "fog", "drizzle", "snow", and "thunderstorm". The distribution of the main categories over the 4 years are shown in table 1.

Weather category	Clear	Clouds	Rain	Mist	Fog	Drizzle	Snow	Thunderstorm
Observed Frequency	22.5%	37.0%	18.8%	11.5%	3.7%	3.6%	2.6 %	0.2%

Table 1: Table showing the frequency of the different main weather categories

4. Results and Analysis

4.1. Weather

First we examine the different weather effects that could be thought to be factors. And their related effect on the relative risk and statistical significance.

4.1.1. Main weather classification

The main weather classification are hourly based and explain which overall weather type was the dominating during the hour in question. Furthermore it constitutes an overall classification such that light, moderate and heavy rain combine to form "rain".

On the left plot of figure 2 we see that the weather categories drizzle, mist or snow are associated to average increases of the relative risk of a crash, when compared to the overall risk level. The statistical significance of the effects, shown in the right plot of figure 2, indicates that only the effects of "drizzle" and snow are significant. The findings are consistent in that precipitation is found to be associated to increased relative crash risk, which is in line with previous literature (Theofilatos and Yannis, 2014; Karlaftis and Yannis, 2010). However, the category rain is not correlated with an increase in the relative crash risk, when compared to the overall risk. This might however be related to the splitting of precipitation types.

Therefore we combine the variables "rain", "drizzle", "mist" and snow, to form the explanatory variable "precipitation", and examine the relative risk change compared to the overall risk that this variable presents. Furthermore we take into account the duration of precipitation leading up to the accident. This has two reasons: 1) The accident records and weather records are aggregated by the hour and therefore not certain

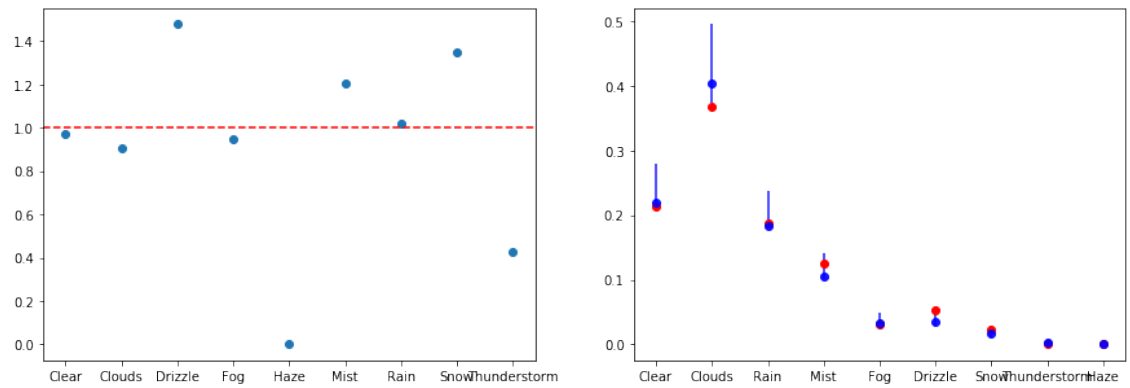

Figure 2: Left: The average proportional increase given the event of different main weather categories $\frac{Q(C)}{P^0(C)}$. Right: The Palm probability $P^0(C)$ and average accident probability observed and calculated given the different weather categories.

that it was raining at the time of the accident. 2) Rain at time $t = 0$ does not constitute that the road is slippery at time $t = 0$. The results are shown in figure 3. The left plot shows the ratio density.

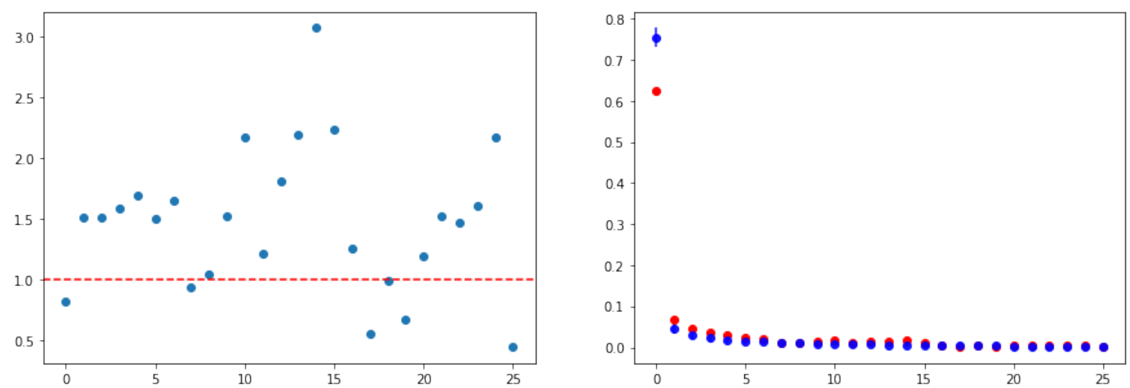

Figure 3: Density plot and assessment of statistical significance for the duration of the precipitation

The ratio densities in figure 3 indicate that risk change given the duration of rain being between 0 and 15 hours is generally in line with the findings in Hermans et al. (2006), which states that especially longer duration of precipitation is significantly associated to the increased number of crashes.

4.2. Accident timing and seasonality

We are interested in evaluating if some days of the week and hours of the day are more indicative of accidents than others. Figure 4 shows the effects of the hours of the day. In the left plot of the figure we see that there is great variability in the relative risk change given the hour of the day. Remarkably we see that the relative risk increase in the hours midnight to 4 am are associated to more than a doubling of the overall risk. These are all statistically significant, as seen in the right plot of figure 4. Accidents in the afternoon peak 16-18 indicate a risk increase of $\approx 50\%$ along with the hour just after, only the effects from 18-19 are significant. In the morning peak hours (06-08) only 8 o'clock is associated to a risk increase. Similar to the afternoon peak, the hour just after the peak period is associated to an increase of the relative risk. Similar findings were made by Norros et al. (2016) for cars on the Helsinki ring road, although they found a decline in the relative risk in the morning peak hours.

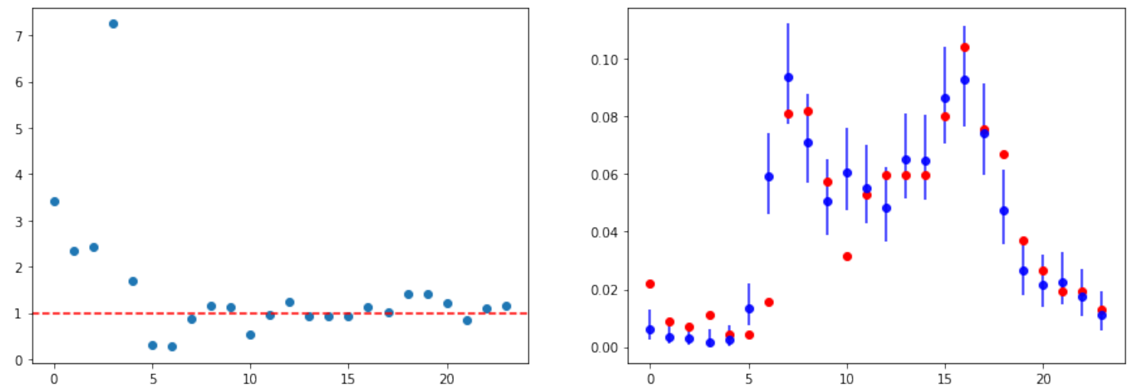

Figure 4: Left: The average relative risk change given the hour of the day $\frac{P^{acc}(C)}{P^0(C)}$. Right: The Palm probability $P^0(C)$ (blue) and $P^{acc}(C)$ (red) given the hour of the day.

The relative risks and significance of the days of the week are shown in table 2. This shows an increase in relative risk as a function of day during the working week, with friday being significant, and a significant decrease in the weekend days

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Palm frequency	0.15	0.15	0.15	0.15	0.15	0.11	0.12
Density ratio	1.04	1.05	1.04	1.13	1.17	0.78	0.65
Significant	N	N	N	N	Y	Y	Y

Table 2: Relative risk change and statistical significance given day of the week.

Previous research has shown significant differences in the exposure and risk profile of working week and weekend travel (Dozza, 2017). Therefore we evaluate the difference in the riskiness related to the hours of the day in weekends and working week. The results of this are shown in figure 5

The results shown in figure 5 are day and night in comparison with the results for the aggregated hour of the day in figure 4. Most importantly the effects observed in figure 5 paint a much more telling picture.

For the weekend, we observe a significant increase in relative risk for the period from 0-4 in the morning, $> 300\%$. This dominated the effects for the aggregated week seen in figure 4, as no significant risk change is observed during the working week for these times. Interpreting this to be related to going out and intoxicated behaviour is now much more plausible. Furthermore the weekend has no morning and afternoon peak and ”those times” are not associated to increased risk. This is incremental to later intervention efforts. During the working week, we observe significant increases in the relative risk, compared to baseline at morning peak 8-9 and afternoon peak 16-19, with the exception of 17:00. In the morning peak, the relative risk is increased

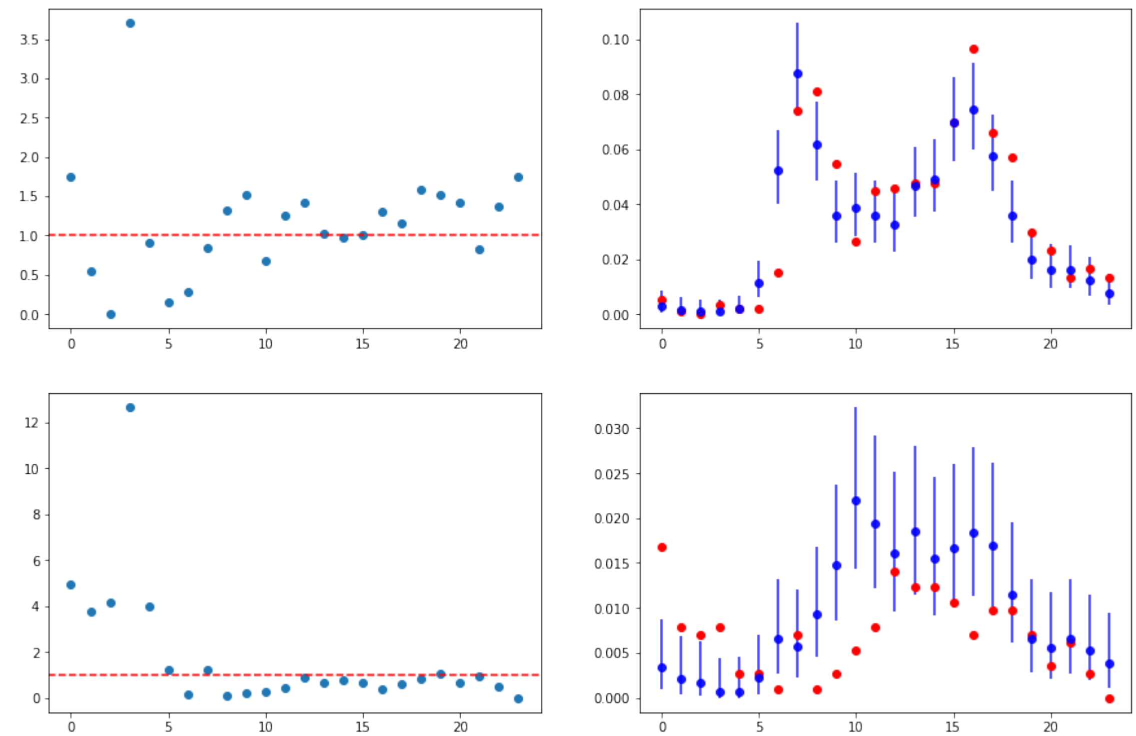

Figure 5: Top: Density ratio and significance of the hour of the day effect, conditioned on being during the working week. Bottom: Density ratio and significance of the hour of the day effect, conditioned on being during the weekend

by $\geq 30\%$ and in the afternoon $\approx 50\%$. Furthermore evening hours are also associated to increased crash risk, although only the effect centred at 23:00 is significant. This could be somewhat attributed to darkness, but light hours in Copenhagen vary a lot between winter and summer and therefore this cannot be said with certainty. The relative risk change in the peak hours could be attributed to people being tired respectively in the morning and after work. Similar results have been found by Dozza (2017); Pack et al. (1995). An interesting finding is the increased risk after the maximum Palm frequency peak in the morning. This could indicate that people are late for work and hence being distracted and driving faster.

Regarding darkness separately, we find that darkness is associated to a significant relative risk increase of 17%, the result is shown in table 3. This is in line with previous research (Johansson et al., 2009), which however compares risk in darkness to risk in light.

Conditions	Palm frequency	Density Ratio	Significant
dark	0.44	1.17	Y
light	0.57	0.87	Y
dry	0.70	0.87	Y
wet	0.30	1.30	Y

Table 3: Relative risk change and significance given dark/ light and wet or dry

Season	Winter	Spring	Summer	Autumn
Palm frequency	0.16	0.28	0.33	0.24
Density	1.3	0.9	0.9	1.06
Significant	Y	Y	Y	N

Table 4: Seasons based on sun positions, their associated relative risk change compared to the overall risk and assessment of significance

When taking a broader look and evaluating the seasonality we find that summer and spring are associated to significant decreases and winter to a significant increase in the relative risk, when compared to the overall risk as seen in table 4.

4.3. Combined effects of multiple factors

Our findings with regard to the hourly effects of respective working week and weekend domains illustrate the importance of considering effects that might only be apparent due to aggregation.
The literature on weather effects proposes that it has an effect in terms of both slippery road surface as well as impairing vision (Theofilatos and Yannis, 2014). Other studies equally report impaired vision and darkness to be key drivers in crash rates (Twisk and Reurings, 2013). In the study by Konstantopoulos et al. (2010) the effect of rain and darkness is even studied in combination with regard to accident risk. Such studies however have primarily regarded car drivers.
Rain on a bicycle does not only impair visibility and surface grip, but also impairs the mobility of the cyclists. Similar to cold weather, one will often observe that cyclists in rainy weather try to take a more compact position on the bike, which in turn limits mobility on the bicycle.

As a final investigation we analyse the combined effects of precipitation with darkness and seasonality. The results of these combinations are summarised in table 5, where the relative risk is seen to be non-linear depending on the combinations of explanatory variables. This becomes clear as the density ratios in table 5 do not necessarily match the linear combinations of equivalent density rations in table 4 and 3.

5. Discussion and Conclusion

This study uses the Palm probability on traffic and weather conditions in a study bicycle of accidents in Copenhagen, to perform an in-depth analysis of weather and time conditions as crash factors. The results highlight that the method of contrasting the distribution of conditions from a crashed cyclist's point of view to those of an arbitrary cyclist, yields more relevant outcomes.

While the methodological frontier of statistical and econometric modelling with regard to accident analysis is more and more adept at identifying factors and unobserved heterogeneity associated to crash risk and injury severity outcomes (Mannering and Bhat, 2014; Mannering et al., 2016), these do not always prove the easiest conclusions to base interventions on. The non-parametric approach conceptualised by Norros et al. (2016),

Combined effects	Palm frequency	Density Ratio	Significant
dark (dry)	0.29	0.99	N
dark (wet)	0.14	1.46	Y
light (dry)	0.41	0.79	Y
light (wet)	0.16	1.16	Y
autumn (dry)	0.15	0.90	N
autumn (wet)	0.10	1.76	Y
spring (dry)	0.21	0.89	N
spring (wet)	0.07	0.82	Y
summer (dry)	0.24	0.80	N
summer (wet)	0.07	0.90	N
winter (dry)	0.10	0.97	N
winter (wet)	0.06	1.58	Y

Table 5: Relative risk change and significance when accounting for variable interaction and their effect of the risk of being subject to a bicycle accident

applied in this study, delivers straightforward results that are easily interpreted and therefore actioned upon. A simple model comes at the cost of not accounting for a lot of the unobserved heterogeneity discussed in Mannering et al. (2016), however it is believed that the finer ”granularity” of the variables in this study make up for it as it allows for a much finer picture with regard to factors associated to bicycle crashes in Urban settings. Alternatively future research may use our approach combined with methods allowing for heterogeneity.

5.1. Conclusion

The specific results on precipitation given different seasons, and visibility as well as accounting for the differences in traffic of the working week and weekends reveal the importance of accounting for 2nd order interactions, and the fallacy of some aggregated variables. This is especially important to account for if meaningful interventions and policy changes are to be made, ensuring further safety of the bicyclists. Specifically results indicate that continued precipitation leading up to the accident shows a general increase in the relative risk of having an accident, compared to the overall risk. The risk change associated to the time of day revealed significant increases in the relative risk in early morning hours (0-4) as well as post-afternoon peak (18-19), compared to the overall risk. However, upon splitting time of day into working week and weekends, the results indicate significantly increased crash risk in the late morning peak (8-9), afternoon peak and early evening (16, 18-20), while the weekend shows that early morning hours (0-4) are associated to significant increases in the relative risk compared to the overall crash risk.

References

Aldred, R., Goodman, A., Gulliver, J., Woodcock, J., 2018. Cycling injury risk in London: A case-control study exploring the impact of cycle volumes, motor vehicle volumes, and road characteristics including speed limits. Accident Analysis & Prevention 117, 75–84. URL: https://www.sciencedirect.com/science/article/pii/S0001457518301076?via%3Dihub, doi:10.1016/J.AAP.2018.03.003.

Allan, R.P., Soden, B.J., 2008. Atmospheric warming and the amplification of precipitation extremes. Science 321, 1481–1484. doi:10.1126/science.1160787.

Dozza, M., 2017. Crash risk: How cycling flow can help explain crash data. Accident Analysis & Prevention 105, 21–29. URL: https://www.sciencedirect.com/science/article/pii/S0001457516301464, doi:10.1016/J.AAP.2016.04.033.

Hermans, E., Brijs, T., Stiers, T., Offermans, C., 2006. Journal of intelligent systems. Transportation Research Board. URL: https://doclib.uhasselt.be/dspace/handle/1942/1365.

Horton, D., 2016. Fear of Cycling , 149–168doi:10.4324/9781315575735-13.

Infrastructures, M.d.D.d.e.d., 2015. Declaration on Cycling as a climate friendly Transport Mode Preamble. Technical Report. URL: www.gouvernement.lu.

Johansson, Ö., Wanvik, P.O., Elvik, R., 2009. A new method for assessing the risk of accident associated with darkness. Accident Analysis and Prevention 41, 809–815. doi:10.1016/j.aap.2009.04.003.

Karlaftis, M.G., Yannis, G., 2010. Weather Effects on Daily Traffic Accidents and Fatalities: A Time Series Count Data Approach. Transportation Research Board 89th Annual Meeting URL: http://147.102.154.193/gyweb/wp-content/uploads/geyannis-pc102.pdf.

Konstantopoulos, P., Chapman, P., Crundall, D., 2010. Driver's visual attention as a function of driving experience and visibility. Using a driving simulator to explore drivers' eye movements in day, night and rain driving. Accident Analysis and Prevention 42, 827–834. doi:10.1016/j.aap.2009.09.022.

Malin, F., Norros, I., Innamaa, S., 2017. Accident risk of road and weather conditions on different road types. Road Safety & Simulation International Conference 2017 122, 181–188. URL: https://doi.org/10.1016/j.aap.2018.10.014, doi:10.1016/j.aap.2018.10.014.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic Methods in Accident Research 1, 1–22. URL: https://www.sciencedirect.com/science/article/pii/S2213665713000031, doi:10.1016/J.AMAR.2013.09.001.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research 11, 1–16. URL: https://www.sciencedirect.com/science/article/pii/S2213665716300100, doi:10.1016/J.AMAR.2016.04.001.

Norros, I., Kuusela, P., Innamaa, S., Pilli-Sihvola, E., Rajamäki, R., 2016. The Palm distribution of traffic conditions and its application to accident risk assessment. Analytic Methods in Accident Research 12, 48–65. URL: http://dx.doi.org/10.1016/j.amar.2016.10.002, doi:10.1016/j.amar.2016.10.002.

Pack, A.I., Pack, A.M., Rodgman, E., Cucchiara, A., Dinges, D.F., Schwab, C.W., 1995. Characteristics of crashes attributed to the driver having fallen asleep. Accident Analysis and Prevention 27, 769–775. doi:10.1016/0001-4575(95)00034-8.

Stipdonk, H., 2008. Deliverable D7.10: Time series applications on road safety developments in Europe Title: Building the European Road Safety Observatory Integrated Project, Thematic Priority 6.2 "Sustainable Surface Transport". Technical Report.

Stocker, T.F., D.Q.G.K.P.M.T.S.A.J.B.A.N.Y.X.V.B., Midgley, P., 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Technical Report. URL: https://www.ipcc.ch/report/ar5/wg1/.

Theofilatos, A., Yannis, G., 2014. A review of the effect of traffic and weather characteristics on road safety. Accident Analysis & Prevention 72, 244–256. URL: https://www.sciencedirect.com/science/article/pii/S0001457514001900#bib0260, doi:10.1016/J.AAP.2014.06.017.

Transport for London, 2014. Attitudes towards cycling Annual report 2014 TfL , 1 – 198URL: https://tfl.gov.uk/cdn/static/cms/documents/attitudes-to-cycling-2014-report.pdf.

Twisk, D.A.M., Reurings, M., 2013. An epidemiological study of the risk of cycling in the dark: The role of visual perception, conspicuity and alcohol use. Accident Analysis and Prevention 60, 134–140. doi:10.1016/j.aap.2013.08.015.

Vejdirektoratet, 2018. Hvorfor stiller vi cyklen? , 16.