

On the relation between bicycle volumes and individual risks for bicyclists in Berlin.

Leonhard Lücken and Peter Wagner

Institute of Transportation Systems,
German Aerospace Center (DLR),
Rutherfordstraße 2, 12489 Berlin, Germany
e-mail: {leonhard.luecken, peter.wagner}@dlr.de

ABSTRACT

A notoriously difficult problem in traffic safety is the verification of the so-called "safety-in-numbers" (SiN) effect. This term refers to the observation that the individual risk R for a road user is inversely related to the absolute number X of road users of the corresponding type and suggests that a larger number of such road users is the reason for an increase of their individual safety.

We propose a model to identify the causal relation between the correlated phenomena. It includes temporal information to verify which of the correlated phenomena precedes the other. We employ the model for an analysis of bicycle volumes and crashes between bicycles and motorized vehicles in Berlin, where different temporal and spatial aggregation levels are considered. Interestingly, we find a slight indication for a SiN-effect on an annual scale while more fine-grained temporal resolutions rather indicate the opposite effect of increasing risks due to higher volumes.

Keywords: safety in numbers, delay model, causal relation, bicycle crashes, bicycle volume model, ratio variables.

1 INTRODUCTION

Several recent works report on an inverse relation of traffic volumes and the risk that the individual road user faces [9–11, 14–22]. This may refer to the observation that for increasing numbers of pedestrians or bicyclists the risk decreases or, equivalently (from a statistician's perspective), for an increasing risk the number of bicyclists or pedestrians decreases. Given such an observation, one might be tempted to claim that it is trivial that an increasing risk on a specific route induces less bicyclists to use it since most people prefer a lower risk for their physical integrity. On the other hand, researchers have suggested that the actual causal relation may be the opposite one. That is, a decreasing number of bicyclists is conjectured to be the reason for an increasing risk. This latter, hypothesized relation is referred to as the "safety-in-numbers" (SiN) effect. The reasoning behind this approach relies on the idea of an increased experience, or attentiveness, of drivers regarding more frequently encountered types of road users.

The described phenomenon of contra-variant traffic volumes and risks has been discussed lively within the last years. While several works presented results indicating the presence of a SiN-effect [14–18, 20, 21], others criticized the methodology and/or the conclusions drawn from the empirical data [5, 6, 8].

For instance, Bathia and Wier [6] point out that a decreasing individual risk that may accompany increasing modal shares of bicyclists or pedestrians is not necessarily related to a decreasing overall damage caused by road crashes since more severe injuries occur in crashes involving vulnerable

road users. Legitimately, they advise care in the derivation of policies based on the assumption of a SiN-effect. On the other hand, assuming a modal shift from motorised to non-motorized traffic, Elvik [11] did show that the non-linearity of the reported relations may even lead to a decrease of the number of crashes and not only of the individual risk.

Other authors [5, 8, 12] suggested that the apparent inverse relation of volumes and risks might be a statistical artifact owing to a spurious correlation between the quantities X and Y/X for two independent variables X and Y . These concerns are attenuated by the fact that the null-hypothesis in the present case, where X and Y correspond to traffic volumes and crashes, is rather a strong linear relation $X \sim Y$ instead of mutual independence [17].

One major problem remains the establishment of a causative relationship between the appearance of a larger number of, e.g., bicyclists and the decrease of individual risk [13]. Indeed, even if a correlation exists between road user volumes and individual risks, this does in principal not allow to infer an identification of cause and effect. Often, the claim of a SiN-effect largely relies on plausibility arguments, which can equally be brought up in favor of the inverse causal relation, where an improved safety is assumed to be the cause for a larger number of cyclists.

In our work we propose a method for the detection of a causal relation between changes in risk and bicycle volumes. To this end we introduce a model, which includes a "memory" of past bicycle volumes as a predictor. We apply this model to data on crashes and bicycle volumes in Berlin on different temporal and spatial aggregation levels. Interestingly, we find some indication that a SiN-effect on the annual scale seems to coexist with a "hazard-in-numbers" (HiN) effect on a monthly scale. The coexistence of seemingly contradictory effects in the same dataset was also reported in Ref. [12], although there the observation referred to different subsets of the dataset instead of different temporal resolutions.

2 METHODS

2.1 A model with bicycle volume memory for the estimation of risk

In our work, we study the relationship between the exposure of bicyclists $X(t)$ that has occurred within a time frame t and the number of crashes $A(t)$ between bicyclists and cars within the same time frame. The momentary crash rate per unit exposure is the "risk" $R(t) = A(t)/X(t)$. The probability for each bicyclist to be involved in such a crash scales proportional to $R(t)$ and his or her amount of exposure. The "physicist's naive approach" considering crashes as collisions of randomly floating particles suggests that the relation $A(t) \sim \alpha X(t)$ should hold for some coefficient α .

However, various empirical studies have shown that this approach often fails when traffic collisions are considered instead of molecular dynamics. In general, nonlinear models of the form

$$A(t) \sim \alpha X(t)^{1-\gamma} \tag{1}$$

with $0 < \gamma < 1$, provide a better fit (see the references given in the introduction, especially the review [11] and references therein, where values $0.35 \leq \gamma \leq 0.7$ are reported), i.e.,

$$R(t) \sim \alpha X(t)^{-\gamma}.$$

In the present paper we analyze data collected by city authorities on crashes (reporting 73,088 crashes between bicycles and motorized vehicles within the period from 2001 to 2014) and monthly

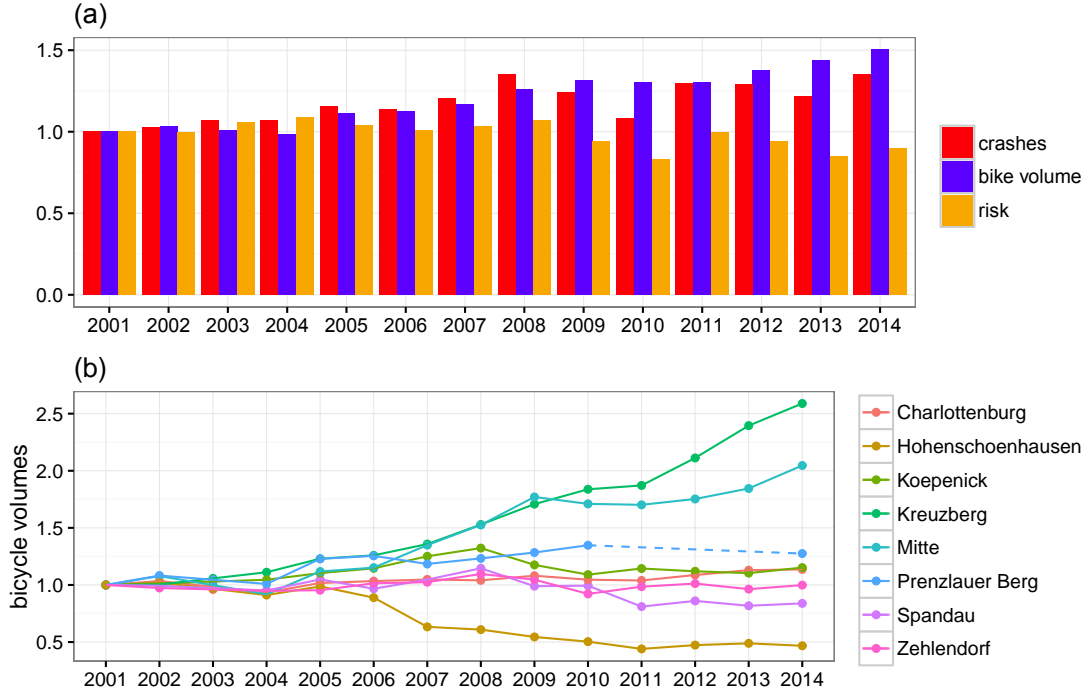


Figure 1. (a) Normalized values (relative to the year 2001) for annual bicycle volumes, bicycle crashes, and cycling risk for the whole city of Berlin during the years 2001-2014. (b) Normalized annual bicycle counts at individual measurement locations.

bicycle counts (i.e., at one day for each month) at eight locations in Berlin¹ [1, 2]. Panel (a) of Fig. 1 shows the historical development of the bicycle counts and crashes involving bicyclists in the years 2001-2014. Panel (b) shows the development of bicyclist volumes for the different measurement locations. Berlin seems to provide a good test field for the SiN effect since the amounts of bicyclists increases steadily in the last years (in 2014 by about 40% since 2001 according to official estimates).

All models employed in this paper will rely on the assumption that the amount of vehicle traffic will stay approximately constant within the considered time span. This seems not very critical for Berlin since the motorization levels did not vary essentially in comparison to the bicycle volumes [4]. In general, one cannot neglect the factor of vehicle amounts. This is usually included as a factor $Y(t)^\gamma$ into models for the crash frequency $A(t)$ [11].

We test for indicators of a SiN-effect and we compare different models to fit the data. In particular, we propose to generalize the model (1) by including a term, which is linearly proportional to $X(t)$, and replacing the term $X(t)^\gamma$ by a quantity $E(t)$ representing a hypothetical increased experience or attentiveness of the road users or an increased visibility of the cyclists. This quantity should therefore exhibit an inverse relationship to the number of crashes and the risk. With this notation, the model (1) can be written as $A(t) \sim \alpha X(t)/E(t)$, with $E = X^\gamma$, and the general non-linear model considered is of the form

$$A(t) \sim \alpha X(t) + \beta X(t)/E(t), \quad (2)$$

¹These counts are collected largely without omissions throughout the years 2007 to 2014 and less continuously since 2001. At the measurement location "Prenzlauer Berg" no counts were conducted for the years 2011-2013 due to construction works

or,

$$R(t) \sim \alpha + \beta/E(t). \quad (3)$$

The value of $E(t)$ can depend on the actual levels of exposure $X(t)$, but also on past levels $X(t-\tau)$ and is considered responsible for any non-linearity in the relation of crashes and bicycle volumes.

Note that according to the proposed interpretation the term $\alpha X(t)$ in (2) accounts for crashes, which are not influenced by an enhanced attentiveness. Its importance indicates a split of crashes into an avoidable part, that can be diminished if the driver's experience and attention towards cyclists increase, and a part that is not significantly affected by an increased exposure. In the representation as individual risk this manifests as a saturation of $R(t)$ towards some constant if the number of bicyclists becomes large and $\gamma > 0$.

For the simple nonlinear model (1) we have an instantaneous relationship $E(t) = X(t)^\gamma$. Substituting this by a delayed relation allows us to test for a causal relationship between changes in bicycle counts and crash rates based on temporal relation of their changes. Indeed, supposing a SiN-effect caused by a elevated attentiveness of the drivers towards an increased amount of bicyclists, the preceding adaptation process of the drivers can be supposed to take a finite amount of time. If such a causal relation is present it should be reflected by the data via a higher positive correlation of crash rates with an average over past bicycle volumes (in comparison to the immediate numbers). In particular, we consider for the attentiveness $E(t)$ the form

$$E(t) = \left(\sum_{s=0}^{\tau} w(s)X(t-s) \right)^\gamma, \quad (4)$$

with weights $w(s)$ representing the magnitude of the influence exerted by the amount of bicyclists present at time $t-s$ on the actual crash rate at time t . In the following we will assume a uniform distribution of weights $w(t) \equiv \text{cst.}$ since it lead to the most significant results (compared to exponentially, resp. linearly, decaying weights). Where the nonlinear models (2) and (3) do not lead to an improvement compared to a simple linear regression, we state the results for the linear model, e.g. $R \sim \alpha + \beta E(t)$, where a negative coefficient β serves as a possible indication of a SiN-effect.

Note that the exponent γ is invariant under affine transformations, e.g. $X \rightarrow aX + b$, of the variables in (2) and (3), respectively. This allows to derive qualitative conclusions also from rescaled data. In the following we mainly use relative quantities (ratios in comparison to a particular year), this is mainly because estimates on bicycle volumes are difficult to obtain as absolute numbers.

As often stated as a critique to the SiN-hypothesis, an opposite causality might hold if both, traffic volumes as well as the individual safety improve where the safety increases firstly. For instance, the construction of a cycle path can provoke that more bicyclists choose a certain route in consequence. This direction effects may be indicated elevated correlations of the crash rates with a predictor as, e.g., $P(t) = (\sum_{s=0}^{\tau} X(t+s))^\gamma$. However, it may also be convenient to use s or risk as the predictor variable to probe this case.

2.2 A weather dependent model for the daily bicyclist variation

To estimate bicyclist volumes on a daily basis for the years 2001-2014 we employ a non-linear model based on weather data collected at the station of Berlin-Tempelhof [3].

More precisely, we have constructed a model which allows the prediction of the changes in the variation $W(t)$ of bicyclist numbers on a daily basis. If the number of average bicyclists $P(t)$ in an area is known, the variation can be used to predict the actual numbers $X(t)$ as

$$X(t) = P(t)(1 + W(t)). \quad (5)$$

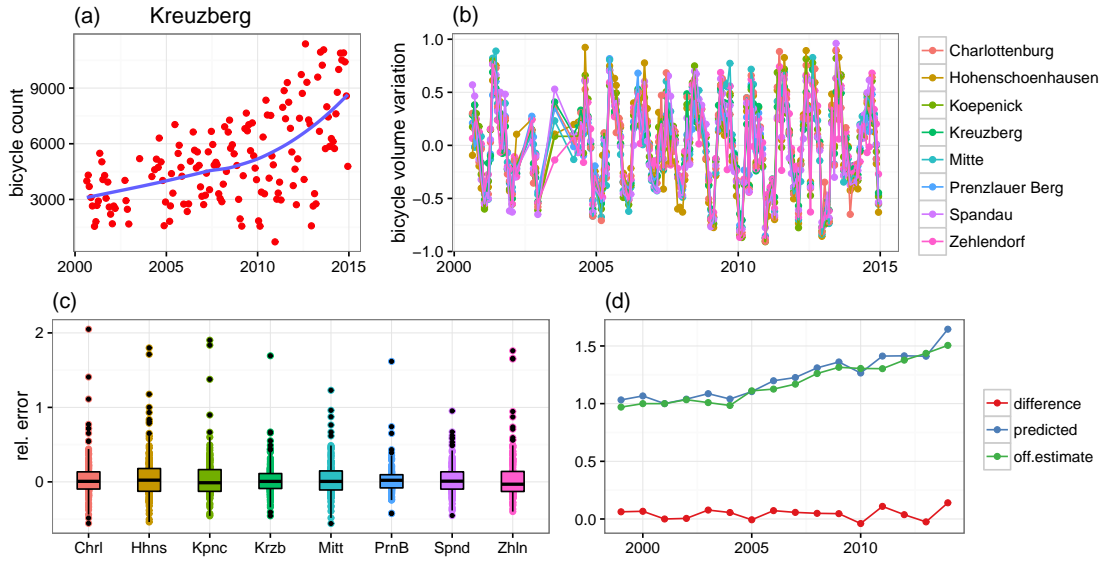


Figure 2. A weather based bicycle volume prediction model. (a) Monthly bicycle counts at the measurement location in Kreuzberg and the LOESS fit for $P(t)$ (blue line), see text. (b) Derived weather-dependent variations $W(t)$ of bicyclist numbers at the different measurement sites. (c) Relative prediction errors of the model (7) at different measurement locations. The lower/upper hinges denote the 25%/75% quantiles, the whiskers extend to the lowest/highest value within 1.5 times the distances of the hinges. (d) Comparison of the official estimate of annual bicycle volumes and the predictions by (7).

Since we do not have the exact knowledge of $P(t_j)$ at the times t_j of measurements, we estimate it by local regression (LOESS) $P \sim X$ giving a slowly varying fit on the observed bicyclists x_j at time t_j (about one data point per month is available), see Fig. 2(a) for an example. The variation is then estimated as

$$W(t_j) = x_j/P(t_j) - 1 \quad (6)$$

Figure 2(b) shows the obtained estimates $w_j = W(t_j)$ for the different measurement locations. One may notice these derived values are largely independent of the particular measurement locations exposing a more universal characteristic of the data. We use the estimates w_j to predict the daily variation of bicycle volumes $W(t)$ for each day t .

As predictor variables for $W(t)$ the daily mean temperature $T(t)$ (in $^{\circ}\text{C}$), the daily hours of sunshine $S(t)$ and the daily amount of precipitation height $N(t)$ (in mm) measured at the weather station Berlin-Tempelhof are considered². They are combined in a nonlinear model of the form

$$W(t) \sim \alpha + \beta f(c_N N(t)) f(c_T (T(t) - \theta_T)) (1 + a_S f(c_S S(t))) / (1 + a_S), \quad (7)$$

where $f(x) = (1 + \exp(x))^{-1}$ is the standard logistic sigmoid. A least squares fit on w_j given by (6) yields the solution $\alpha = -0.843$, $\beta = 2.7662$, $c_N = 0.0627$, $a_S = 2.5758$, $c_S = -0.31$, $c_T = -0.2283$, and $\theta_T = 3.1903$. The residual standard error is $\sigma^2 = 0.1838$ and the maximal p -value is $3.42 \cdot 10^{-4}$ for a_S . The relative errors $\varepsilon_j^{\text{rel}} = ((1 + w_j)P(t_j) - x_j)/x_j$ of the model's predictions are shown in Fig. 2(c).

Given the daily prediction of the bicycle volume variation, we obtain the estimated daily changes of volume in all Berlin by assuming the average $P(t)$ being proportional to a linear interpolation

²The model was not improved by including more data as, e.g., wind strength, or the predictors' values on preceding days.

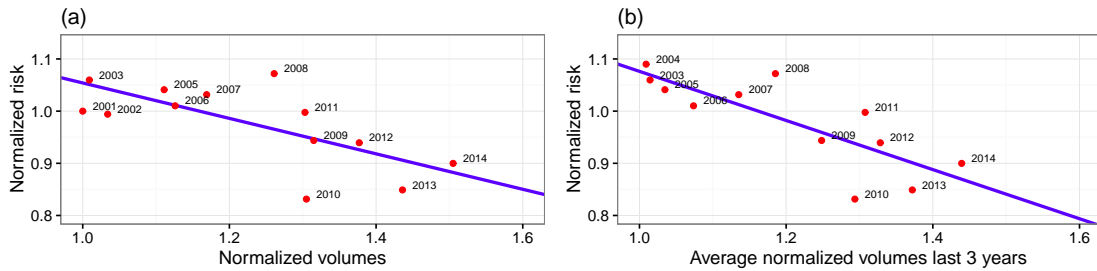


Figure 3. (a) Annual, normalized bicycle volumes versus annual, normalized risk. (b) Average normalized bicycle volumes during the last 3 years versus annual, normalized risk. The blue lines show the best least squares fit of the form (8) with (a) $\tau = 0$ and (b) $\tau = 2$, respectively.

of the official annual estimates. See Fig. 2(d) for a comparison of the differences between the predictions of (7) and the official estimates on the annual bicycle volumes relative to the volumes in 2001 (deviations below 10%).

3 RESULTS

In this section we review the Berlin data focussing on the presence of SiN- or HiN-effects while taking into account a bicycle memory in the models. Where it performs better, we apply the proposed non-linear model (3) for the bicyclist's risk. This is done on various temporal and spatial aggregation levels with differing outcomes as shown in the following table:

	city	district
annual	SiN	rather SiN
monthly	HiN	rather HiN
daily	rather HiN	NA

3.1 Annual aggregation

Starting with the most coarse perspective, we consider the statistics for the whole city of Berlin since 2001. Figure 3(a) shows the normalized annual risk $R_n(t) = R(t)/R(t = 2001)$ to be involved in a collision in dependence of the normalized annual amount of bicyclists $X_n(t) = X(t)/X(t = 2001)$. A value for the risk above one indicates a higher risk than in the reference year 2001, a value below one indicates a lower relative risk. A SiN-effect would correspond to a negative slope of the data points, which seems to be present to some degree. In Fig. 3(b) the annual risk is plotted against the averaged values for the bike counts within the last three years, i.e.,

$$R_n(t) \sim \alpha + \frac{\beta}{(\tau + 1)} \sum_{j=0}^{\tau} X_n(t - j), \quad (8)$$

for $\tau = 2$.

The minimal residual standard error $\sigma^2 = 0.0538$ and the minimal p -value $p = 0.0015$ are both attained for $\tau = 2$. Although the values are only marginally different from the ones at $\tau = 0$ ($\sigma^2 = 0.0574$ and $p = 0.0033$), one might take this as a weak indication of a causative SiN-effect

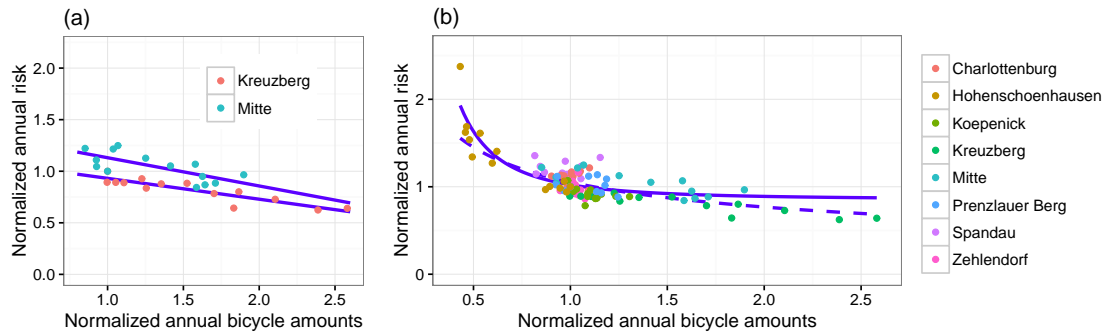


Figure 4. (a) Annual, normalized bicycle volumes at the measurement locations in Kreuzberg and Mitte versus annual, normalized risk referring to the enclosing district. The blue solid lines indicate linear fits on the data with $\tau = 0$. (b) Annual, normalized bicycle volumes at different measurement locations versus the annual, normalized risk within the enclosing district.

with an "inertia" at the timescale of 1-3 years. Further, we observe that the model yields better fits for $\tau > 0$ than for $\tau < 0$ perhaps also rather supporting the SiN hypothesis (data not shown).

The above numbers are strongly aggregated. If one intends to investigate the presence of a SiN-effect on smaller temporal or spatial scales, e.g. seasonal or local variations, a closer view at the data on less aggregated levels is necessary. The statistical assessment bears several difficulties. For instance, the crash data at the immediate neighborhood of the measurement locations for the bicycle is sparse and seems insufficient for a meaningful analysis. If aggregating the crashes on a monthly basis one finds for 95% of the cases two or, more frequently, less crashes were recorded in the measurements location neighborhood within one month.

Even, if the bicycle crashes in the immediate neighborhood of the measurement location are temporally aggregated for each year, significant deductions seem impossible since the number of data points at each location is too low to allow significant deductions.

One (questionable) remedy is to increase the area, which is taken into account around the measurement location and use the variation of bicycle amounts at the measurement location as an approximation of the variation in that enlarged area. This potentially disregards some effects that are local specifics of the measurement locations, e.g., infrastructural changes³.

Keeping these remarks in mind, we present some of the resulting statistics in Fig. 4. Again for the majority of the measurement locations no significant fits are attained. The only locations, where good fits are available (p -values ≤ 0.01) on the individual data are the sites in the districts of Kreuzberg and Mitte where the largest differences in bicycle volumes as well as the highest absolute numbers at the measurement locations were observed within the considered time span. The corresponding, linear fits are shown in Fig. 4(a). The optimal linear fits (lowest p -values) are obtained for $\tau = 0$ in Kreuzberg and for $\tau = 2$ in Mitte [cf. Eq.(8)]. Both cases show an inverse relation of risk and bicyclist numbers. A nonlinear approach does not improve the linear models in these cases.

We emphasize that the combined data consisting of normalized values from different locations is strongly depended on the chosen reference year for the normalization and the locally differing

³Besides the construction work at the site "Prenzlauer Berg" we could identify one possible case, where the construction of an alternative cycling path in the year 2006 seems to have influenced the bicycle volumes at the measurement location in the district of Hohenschönhausen, cf. Fig. 1(b).

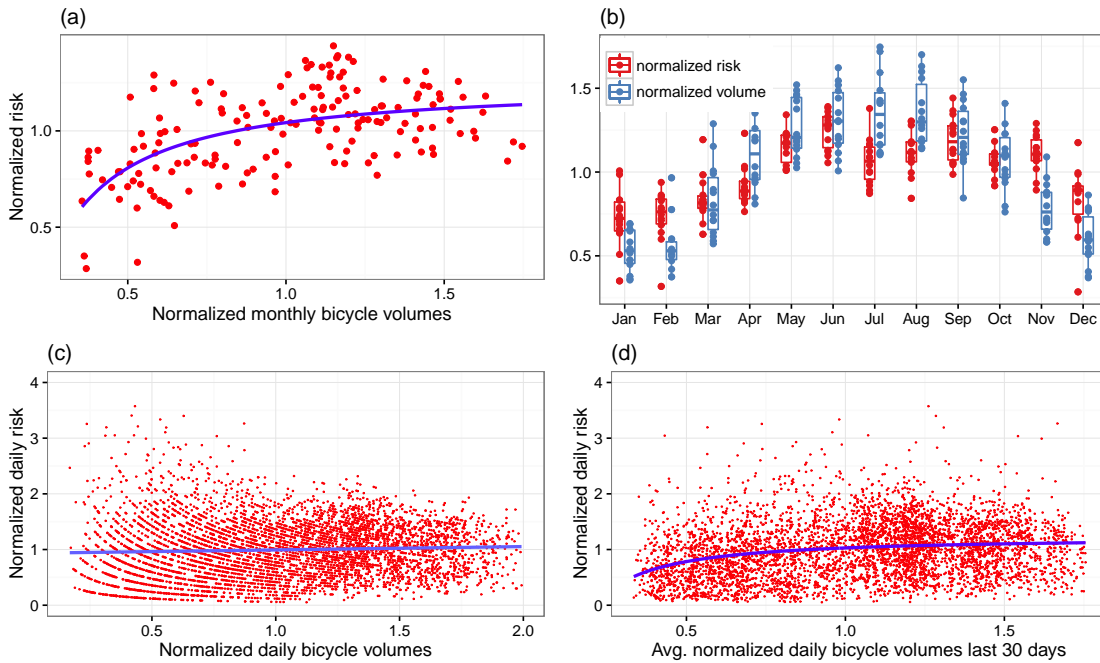


Figure 5. (a) Monthly risks vs. predicted relative bicycle volumes in Berlin. (b) Seasonal fluctuations of both values, see Fig. 2 for an explanation of the boxplots. (c) Daily risks vs. normalized bicycle volumes in Berlin. (d) Daily risks vs. average normalized bicycle volumes in the last 30 days.

relative development of the bicycle volumes. However, it might still be possible to identify a general trend from the combined data. The related statistical justification is an interesting problem but beyond the scope of this paper. If the normalized data from all districts is combined, the picture given in Fig. 4(b) is obtained. The blue solid line indicates the best fit by a model of the form (3) with $E(t) = X(t)$ (exhibiting $\sigma^2 \approx 0.1161$ and a maximal p -value of ≈ 0.0008 for β). The dashed line indicates the corresponding fit when the data from Hohenschönhausen³ is disregarded illustrating the sensitive dependency of the fit on the data ($\beta = 1$ fixed with $\sigma^2 \approx 0.1004$ but a maximal p -value of ≈ 0.367 for γ).

3.2 Monthly and daily resolution

In this section we seek to verify the presence of effects that act on a faster timescale than the annual changes, which were considered so far. To this end it is required to increase the granularity of the bicycle volume data allowing to account for daily or monthly varying crash and bicycle counts instead of the annually aggregated estimates or single occasion counts provided by official sources. For this purpose we utilize the weather based model introduced in Sec. 2.2 to predict daily bicycle volumes.

We first test for an SiN-effect within monthly average bicycle volumes on a city wide level. Surprisingly, we rather find the inverse (i.e., HiN) effect, that an increasing monthly number of bicyclists correlates to a *higher* risk for the individual bicyclists, see Fig. 5(a). Without presenting the details, we note that the observation persists when taking memory into account for the prediction according to (4) with $\tau > 0$ being the number of months. The best fit for a model of the form (3) with $E(t) = X(t)^\gamma$ is obtained for $\gamma > 0$ but $\beta < 0$. In this case the above interpretation of $E(t)$ (as being largely determined by the driver's attentiveness) is obviously not valid. Also an

inspection of the seasonal variations of bicycle volumes and risks shows that higher volumes in the warmer season is seemingly accompanied by higher risk, see Fig. 5(b). These findings contrast with the slight indication of the presence of an SiN-effect at an annual aggregation level.

Concerning the monthly crash data and bicycle volume prediction restricted to the different districts associated with each measurement location we observe a similar relation. For some cases the risk increases with the bicycle volumes and it never decreases significantly.

Also for daily data, we do not find any significant trace of a SiN-effect. A comparison of the daily risk with the daily predicted variations on the basis of the model (3) with $E(t) = X(t)^\gamma$ does not improve a simple linear fit essentially. The non-linear model fit is shown in Fig. 5(c) although it can hardly be distinguished from a linear one as its parameters are $\alpha = 0.9394$, $\beta = 0.05638$, and $\gamma = -0.6724$, with a residual standard error of $\sigma^2 = 0.5302$. Once more, we find $\gamma\beta < 0$ contradicting an alleged SiN-effect. Interestingly, the nonlinear fit can be improved for a choice of $E(t)$ according to (4) with an optimum at $\tau = 30$ days, see Fig. 5 for the corresponding model fit. It yields parameters $\alpha = 1.21881$, $\beta = -0.1899$, and $\gamma = 1.19743$, with a residual standard error of $\sigma^2 = 0.5159$ and much smaller p -values for β and γ than for the case $\tau = 0$. A choice of $\tau < 0$, i.e., an association of the current risk with future bicycle volumes, does not improve the model's fit.

4 CONCLUSIONS

We have presented an approach to study the relation of bicyclist volumes and the individual risk for a bicyclist to be involved in a crash. In particular, we presented a memory model (3) for non-linearities that has the potential to differentiate between a "safety-in-numbers" effect and the opposite "numbers-due-to-safety" effect.

This approach was applied to data on bicycle counts and crashes in Berlin using different levels of spatial and temporal aggregation. The most noticeable results from the data analysis is that a decreasing risk for bicyclists apparently exists on an annual time scale for the whole city and, presumably, for some districts. These figures are indeed accompanied by a steadily growing number of bicyclists, but although the model fits tend to a slight improvement if bicycle counts of the previous year are taken into account in a model of the form (3), we do not declare this a clear evidence for a SiN-effect.

There exists an analogy between different timescales for the correlation of bicycle volumes and risks and the different timescales of the weather model introduced in Sec. 2.2. The slower timescale, where an SiN effect seems conceivable, corresponds to the quantity $P(t)$ describing the slowly changing average number of bicyclists.

The faster timescale of daily changing bicycle volumes corresponds to seasonal or rather daily fluctuations of the weather. In contrast to the annual figures, on such faster time scales we do not observe any indication for a decreasing risk with increasing bicycle volumes. Rather the opposite relation, i.e., a HiN-effect, is found in some cases. This indicates that changes in the drivers' attentiveness induced by daily changes in bicycle volumes are not strong enough to dominate other, more important factors acting on a faster timescale. For instance, such a factor could be the disproportional high increase in the volumes of inexperienced bicyclists if good weather conditions are present. We remark that this hypothesis could not be approved by taking a closer look at the crash causes in the present data. For instance, the proportion of crashes *caused* by the involved bicyclist does not vary substantially with the daily volumes. This is no definite disproof, though, since inexperienced bicyclists may be involved in crashes more frequently without causing them more frequently.

Furthermore, varying weather conditions could induce changes in risks directly, independent of the changes induced in bicycle volumes. A closer investigation of the relation between the weather conditions and the risk is a possible direction for further research, and the analysis of the relation between bicycle volumes and risks given a fixed weather condition could remove the interference to some degree. In any case we have to admit that, so far, we could not validate any explanation for the HiN effect statistically. The same holds for the observation that some data restricted to district level indicates a maximal risk for intermediate bicycle volumes, which may suggest that a non-monotonic model as proposed by Borsos et al. [7] might be appropriate.

Although the sought-after phenomenon of SiN was not present on monthly or daily scales, we have tested the memory model on the data and found that a finite memory of approximately 30 days improves the prediction of daily risks. Currently the explanation of this phenomenon is an open problem. It might merely be a consequence of the smoothing that is imposed on the predictors as a side effect of the temporal aggregation. However, the fit does not improve if the aggregation is carried out using future value of the predictors. Therefore a causal factor may still await its identification. Sticking to the interpretation of the quantity $E(t)$ as an attentiveness of drivers in this situation would be equivalent to claim that an increased exposure diminishes the attentiveness. This may hold for some monotone and frequently repeated tasks, but is disputable in the context of encounters between bicyclists and drivers in urban traffic.

It may be justified to object the presented data analysis on grounds of its lack of its accuracy. We agree on the point that several components of the employed models can be improved, although this might affect the clarity of the presentation. Perhaps most importantly for the phenomenon under consideration, the models (2) and (3) for the prediction of crashes and risk does not take into account volumes of motorized traffic. Although the level of motorization in Berlin did only change marginally this does not necessarily imply that the same holds for the amount kilometers traveled, as commuters and variations in the individual vehicle usage are disregarded. Especially concerning the daily variations in the modal shares it may be of importance to include a model for the variation of motorized traffic volumes as well. Moreover, the model (7) for the variation of bicycle volumes is still rather imprecise in its predictions [cf. Fig. 2] and it may be elaborated in more detail to improve them. For instance, one might include information on special events (e.g., strikes of public transportation services), extreme weather, school vacations, tourism (dependent on the location), etc.

While the circumstances in Berlin do not allow to make decisive claims about a SiN-effect, the proposed method or its generalization to more detailed models can be useful for further research. In situations, where a stronger correlation between bicycle volumes and risk is present it may allow to derive a clearer temporal relation and distinguish between cause and effect. Indeed, if the method was applied to sufficiently fine-grained and extensive data it seem intelligible that it could even help to detect the *timescale* of an alleged SiN-effect. One might hypothesize elevated correlations at three distinct scales. Firstly an immediate effect, caused by an increased visibility of bicyclists in a group, secondly, an intermediate effect, as an inert societal attentiveness to a seasonal fluctuation of bicycle volume, and finally a long-term familiarization that could, for instance, be observable in a society where a substantial modal shift towards cycling takes place.

REFERENCES

- [1] Fahrradverkehr-Pegelzählungen Berlin - Jahresbericht 2014. Tech. rep., Verkehrslenkung Berlin (VLB), 2014. (in german).
- [2] Sonderuntersuchung "Radverkehrsunfälle" in Berlin 2014. Tech. rep., Der Polizeipräsident in Berlin, Stab des Polizeipräsidenten, 2014. (in german).

- [3] Historische tägliche Stationsbeobachtungen, Station Berlin-Tempelhof. Tech. rep., Climate Data Center of the Deutscher Wetterdienst (DWD), 2015. (in german).
- [4] Statistisches Jahrbuch 2015. Tech. rep., Amt für Statistik Berlin-Brandenburg, 2015. (in german).
- [5] ANDREASSEN, D. Population and registered vehicle data vs. road deaths. *Accident Analysis & Prevention* 23, 5 (1991), 343–351.
- [6] BHATIA, R., AND WIER, M. Safety in Numbers re-examined: Can we make valid or practical inferences from available evidence? *Accident Analysis & Prevention* 43, 1 (2011), 235–240.
- [7] BORSOS, A., KOREN, C., IVAN, J., AND RAVISHANKER, N. Long-term safety trends as a function of vehicle ownership in 26 countries. *Transportation Research Record: Journal of the Transportation Research Board*, 2280 (2012), 154–161.
- [8] BRINDLE, R. Lies, damned lies and 'automobile dependence'-some hyperbolic reflections. In *Australasian Transport Research Forum* (1994), vol. 19, pp. 117–131.
- [9] BRÜDE, U., AND LARSSON, J. Models for predicting accidents at junctions where pedestrians and cyclists are involved. how well do they fit? *Accident Analysis & Prevention* 25, 5 (1993), 499–509.
- [10] EKMAN, L. *On the Treatment of Flow in Traffic Safety Analysis,-a non-parametric approach applied on vulnerable road users*. PhD thesis, Lund University, 1996.
- [11] ELVIK, R. The non-linearity of risk and the promotion of environmentally sustainable transport. *Accident Analysis & Prevention* 41, 4 (2009), 849–855.
- [12] ELVIK, R. Can a safety-in-numbers effect and a hazard-in-numbers effect co-exist in the same data? *Accident Analysis & Prevention* 60 (2013), 57–63.
- [13] ELVIK, R., AND BJØRNSKAU, T. Safety-in-numbers: a systematic review and meta-analysis of evidence. *Safety Science* (2015).
- [14] FYHRI, A., AND BJØRNSKAU, T. Safety in numbers - uncovering the mechanisms of interplay in urban transport with survey data. In *Proceedings, International Cycling Safety Conference* (2013), pp. 20–21.
- [15] FYHRI, A., SUNDFØR, H., AND BJØRNSKAU, T. Safety-in-numbers - combining a panel design and cross cultural survey to examine the suggested mechanisms. In *International Cycling Safety Conference. Gothenburg* (2014).
- [16] GÅRDER, P., LEDEN, L., AND PULKKINEN, U. Measuring the safety effect of raised bicycle crossings using a new research methodology. *Transportation Research Record: Journal of the Transportation Research Board*, 1636 (1998), 64–70.
- [17] GEYER, J., RAFORD, N., RAGLAND, D., AND PHAM, T. The continuing debate about safety in numbers-data from Oakland, 2006.
- [18] JACOBSEN, P. L. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury prevention* 9, 3 (2003), 205–209.
- [19] JONSSON, T. *Predictive models for accidents on urban links-A focus on vulnerable road users*. PhD thesis, Lund University, 2005.

- [20] NORDBACK, K., MARSHALL, W. E., AND JANSON, B. N. Bicyclist safety performance functions for a US city. *Accident Analysis & Prevention* 65 (2014), 114–122.
- [21] ROBINSON, D. L. Safety in numbers in Australia: more walkers and bicyclists, safer walking and bicycling. *Health promotion journal of Australia* 16, 1 (2005), 47–51.
- [22] VANDENBULCKE, G., THOMAS, I., DE GEUS, B., DEGRAEUWE, B., TORFS, R., MEEUSEN, R., AND PANIS, L. I. Mapping bicycle use and the risk of accidents for commuters who cycle to work in Belgium. *Transport Policy* 16, 2 (2009), 77–87.