University of Birmingham
Research at Birmingham

# Traffic incidents in motorways: An empirical proposal for incident detection using data from mobile phone operators

Steenbruggen, John; Tranos, Emmanouil; Rietveld, Piet

*DOI:*
10.1016/j.jtrangeo.2016.05.008

*License:*
Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

*Document Version*
Peer reviewed version

*Citation for published version (Harvard):*
Steenbruggen, J, Tranos, E & Rietveld, P 2016, 'Traffic incidents in motorways: An empirical proposal for incident detection using data from mobile phone operators', *Journal of Transport Geography*, vol. 54. https://doi.org/10.1016/j.jtrangeo.2016.05.008

Link to publication on Research at Birmingham portal

**General rights**
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

**Take down policy**
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 02. Jun. 2024

# Traffic incidents in motorways: An empirical proposal for incident detection using data from mobile phone operators

John Steenbruggen, john.steenbruggen@rws.nl
Emmanouil Tranos, e.tranos@bham.ac.uk
Piet Rietveld †

Please cite as:
Steenbruggen, J., E. Tranos, et al. (2016). "Traffic incidents in motorways: An empirical proposal for incident detection using data from mobile phone operators." *Journal of Transport Geography*: in press.

## Abstract

*This paper proves that mobile phone usage data is an easy to use, cheap and most importantly, reliable predictor of motorway incidents. Using econometric modelling, this paper provides a proof of concept of how mobile phone usage data can be utilised to detect motorway incidents. Greater Amsterdam is used here as a case study and the results suggest that mobile phone usage data can be utilised for the development of an early warning system to support road traffic incident management.*

***Keywords**: Road traffic incident management, Mobile phone data, Data science, Collective sensing.*

## 1. Introduction

Increased urbanisation does not come for free and car traffic related congestion and incidents are some of the most pronounced externalities. This paper aims to contribute to the management of these externalities by providing a proof of concept which can assist traffic Incident Management (IM). In brief, this paper proposes the use of mobile phone usage data within an IM system as a tool to detect motorway incidents. By adding a layer of information inferred by mobile phone usage data, which is easily accessible nowadays and free of charge, the efficiency of IM can be drastically increased.

IM involves the cooperation of many public and private actors. To support these tasks in an effective way, advanced information systems and the use of spatio-temporal data are becoming increasingly important (Steenbruggen *et al.*, 2012; Steenbruggen *et al.*, 2014a). Along with the growing ubiquity of mobile technologies, the extensive data logs produced in the course of their usage have helped researchers to create and define new methods of observing, recording, and analysing environments and their human dynamics (O'Neill *et al.*, 2006). In effect, these personal devices create a vast, geographically-aware sensor web that accumulates tracks to reveal both individual and social behaviour in unprecedented detail (Goodchild, 2007). Steenbruggen *et al.* (2013a; b) have identified this phenomenon as *collective sensing*, or, in other words, the reconstruction of "collective human behaviour from individual anonymous digital traces". These traces left by individuals are accumulating at an unprecedented scale (Zhang *et al.*, 2010) resulting in very large data sets known as 'big data'.

The usability of such data has been demonstrated in the relevant literature (Boyd and Crawford, 2012; Steenbruggen, 2014b; Steenbruggen *et al.*, 2014c). In this paper we use various (big) data sets to explore the relationship between motorway traffic incidents and mobile phone usage. Taking into account both the spatial and the temporal dimension, we model how mobile phone activity is related to road traffic incidents.

Traffic incidents may be sensitive to different weather conditions. Therefore, we also include in our models other variables such as meteorological measurements to control for the weather effect on motorway incidents in the Greater Amsterdam area. Within the environmental monitoring domain, the amount and the availability of digital information, based on near real-time sensor measurements, have been rapidly increasing (e.g. see Akyildiz *et al.*, 2002; Hart and Martinez, 2006). Such sensor nodes include highly mobile and intelligent sensor pods (Resch *et al.*, 2010), as well as fixed sensor stations (Alesheikh *et al.*, 2005). Given the increasing accuracy of meteorological monitoring and forecasting, understanding the relationship between weather patterns and traffic incidents can potentially provide valuable insights into understanding and predicting mobility and traffic accidents (Sabir, 2011).

The main research question of this paper is to examine whether we can use mobile phone data to detect motorway traffic incidents (dashed line in Figure 1). The underlying goal is to explain which factors affect mobile phone usage in area *i* at time *t*, and in particular the role of incidents in area *i* at time *t* (straight line in Figure 1).

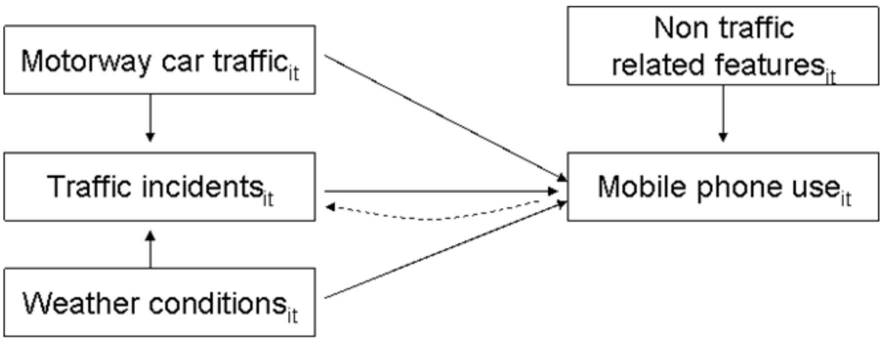



**Figure 1: Graphical representation of our research model**

More specifically, the mobile communication volume depends on the specific land use (e.g. business areas, shopping centres) and other non-traffic-related features, such as weather conditions. We focus on GSM (Global System for Mobile Communications) zones which strongly overlap the motorway infrastructure, which means that motorway traffic intensity and traffic incidents potentially have a substantial influence on mobile phone usage.

The dashed arrow in the figure, pointing in the reverse direction (from mobile phone usage to traffic incidents) is also addressed in the paper. This arrow is not meant to represent a causal relationship, but it is introduced to represent the notion that data on telecom use can be employed to detect the occurrence of incidents, and therefore contribute to a rapid detection of incidents on motorways.

The structure of the paper is the following. The next section describes the data used and then the empirical application is presented. A three step modelling strategy has been designed, which starts by modelling the relation between traffic and mobile phone usage, followed by a model explaining the relationship between car incidents and motorway traffic and finally presenting the marginal effects of this relationship. The paper ends with a conclusion section.

## 2. Data

The study area involves the city of Amsterdam and its surroundings, covering an area of about 1000 km$^2$ (see Figure 2). We used four different types of data sets for our research: mobile phone usage data; traffic incident data; motorway traffic flow data; and meteorological sensor data. These data sets are described below in more detail.

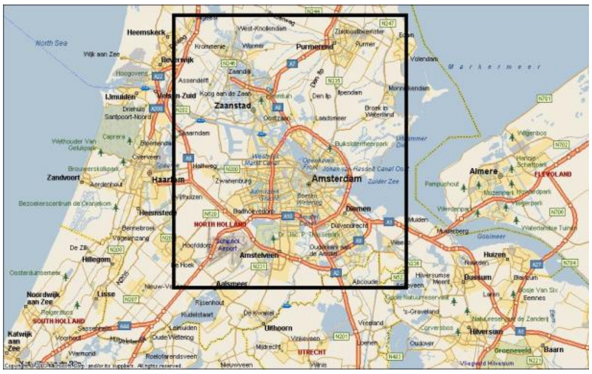



**Figure 2: Overview the Amsterdam test area**
**Note: Owing to the irregular boundaries of the sectors, the network coverage of the study area does not exactly correspond to the above box.**

### 2.1 Mobile phone usage data

The mobile phone usage data that we utilize for this paper was supplied by a major Dutch telecom operator, and provides aggregated information about mobile phone usage at the level of the (GSM) cell zones for the period 2007-2010. The most common format is the '*Call Data Record*' (CDR), according to which subscribers' mobile phone activities are recorded each time a user uses a service (Steenbruggen *et al.*, 2014c). The project uses anonymized data of the mobile network. The raw data contains aggregated CDR information with a temporal dimension of a 1-hour time interval in a certain (GSM) cell zone. In the study area, over 1200 (GSM) cell zones were provided by a Dutch telecom operator (see Figure 2). Based on our criteria, as described at the end of this section, only 109 (GSM) cell zones were used in our modelling exercise (see also Figure 3). The telecommunication operator applied special scripts to extract the necessary data for the project. For the purpose of this case study, we select data from 1 January 2010 (00.00 hr.) through to 20 November 2010 (07.00 hr.). See also Figure 3. The GSM cellular network is built on the basis of radio cells. They define the spatial dimensions of the two best serving cell maps generated by antennas with two different frequencies overlaid on each other: namely, 900 MHz coverage (the basic network with full area coverage), and 1800 MHz (capacity network only in densely populated areas). The size of a wireless cell can vary widely, and depends on many factors, such as land use and urban density. In order to obtain the real mobile phone use-pattern of a certain place in the city, the two *best serving area maps* (900 and 1800 MHz) have been merged.

In the literature, there are a number of different geographical approaches which can be used to handle the raw mobile phone network traffic data (Steenbruggen *et al.*, 2014c). Aggregated CDRs can be represented as: *Voronoi* diagrams (González *et al.*, 2008; Kuusik *et al.*, 2008; Song *et al.*, 2010; Traag *et al.*, 2011), and rasterization (Calabrese *et al.*, 2007; Reades *et al.*, 2009; Girardin *et al.*, 2009). We chose the original best-serving cell maps, because they represent a more realistic representation of the ground truth of the relationship between the original aggregated mobile phone use and the geographical area specified by the telecom operator. The main limitations of CDRs lie in their sparse temporal frequency (data are generated only when a transaction occurs), and on their rather coarse spatial granularity, as locations are based on the granularity of a cell tower (Becker *et al.* 2011). Apart from that, cell towers vary in density (urban vs. rural), which affects estimates, and there are also some concerns regarding the privacy of the use of such data (e.g. Ahas *et al.*, 2007).

The user-generated mobile phone traffic in such large-scale sensor networks reflects the spatio-temporal behavioural patterns of their users. Moreover, depending on a provider's market share and mobile penetration rate, these patterns reflect to some degree the dynamics of the larger population. The anonymized and aggregated volumes of mobile traffic data can be either related with population presence (*Erlang*, *new calls*, *total call lengths*, *SMS*) or with mobility (*handovers*). The variable *total call length* is highly correlated with *Erlang* and therefore is excluded from the analysis. The used variables are defined in Table 1.

**Table 1: Description of the telecom counts used**

| Type of Indicator | Variable | Description | Min. | Max. | Mean | St.dev. |
|---|---|---|---|---|---|---|
| Population presence | Erlang | A standard unit of measurement of traffic volumes, equivalent to 60 | 0 | 73.30 | 3.38 | 4.92 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | minutes of voice | | | | |
| | New calls | The total number of new speech calls initiated in the current (GSM) cell zone | 0 | 2662 | 91.25 | 144.49 |
| | Total call length | The sum of all call lengths | 0 | 263866 | 12173.92 | 17696.82 |
| | SMS | The total number of sent SMS | 0 | 17415 | 106.68 | 216.59 |
| Population movement | Handover | The sum of incoming and outgoing handovers | 0 | 13548 | 434.33 | 583.85 |

**Note: The telecom data used in our case study represents a market share higher than 45 per cent**

In order to derive spatio-temporal information from the high volume of raw mobile network traffic data, a semi-automated (geo-)processing workflow was developed. To ensure that the selected cell zones represent mobile phone usage on highways and not mobile phone usage in other places, we selected only those cells with a maximum percentage of area coverage occupied by motorways. In addition, an important characteristic of the telecom network is that one unique cell zone can consist of multiple geographical polygons. Owing to radio coverage, this can range from 1 to more than 100 polygons per cell zone id. In dense urban areas, the number of polygons is much smaller than in rural areas. In our case study area of Greater Amsterdam (the area within the black rectangle in Figure 3), the range of polygons from one unique cell varies from 1 to 6. For our analysis, we selected only those cells, where the sum of area coverage ($m^2$) of the polygons intersected with the highway, which are related to one GSM zone, is larger than 70 per cent of the sum of all polygons belonging to the same GSM zone. Based on these criteria, we selected 109 from the original data sample including in total 790,865 hourly measurements, corresponding to 322 days and 7 hours, for each of the 109 selected cell zones belonging to the area chosen for the investigation.

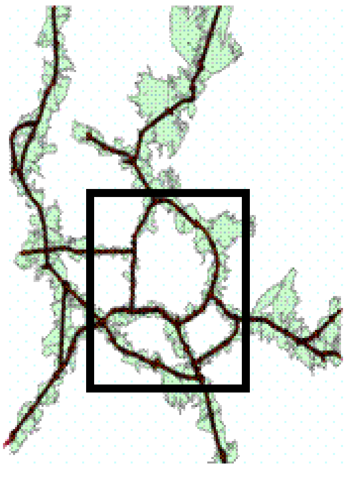


**Figure 3: Selected cells covering the highways of Amsterdam and its surroundings**
Note 1: The Greater Amsterdam area within the black rectangle contains 122 GSM zones.
Note 2: 109 of these GSM zones have an area coverage of > 70% of the highways.

GSM zones are generally characterized by different land-use patterns and population presence. Each type of land use has its own specific spatial signature (see Steenbruggen *et al.*, 2013b). The spread of daily mobile phone patterns, over all weekdays from cells which are related to highways, can be seen in Figure 4. This helps us to understand the spatio-temporal variability of the mobile phone data for the road infrastructure. The diagram shows the average mobile phone intensity per hour for different indicators.

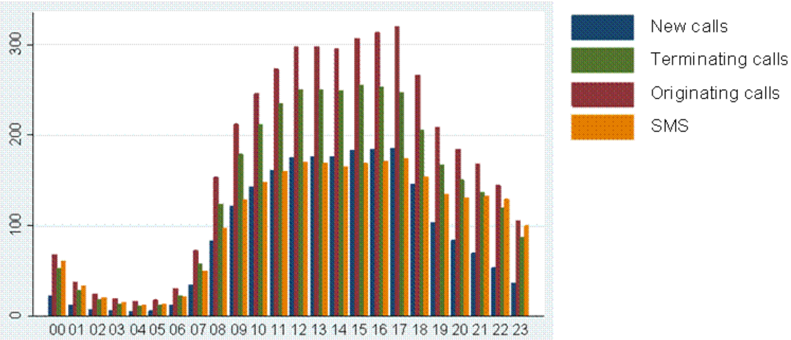



**Figure 4: Spatial distribution of cells covering the highways (the heartbeat of the road infrastructure)**
Note 1: 'Terminating calls' is the number of call attempts to a GSM zone (Mobile Terminated Calls – MTC).
Note 2: 'Originating calls' is the number of call attempts from a GSM zone (Mobile Originating Calls – MOC).

## 2.2 Traffic incident data

We use six different types of incident categories on the highways provided by the Dutch Ministry of Infrastructure and Environment: object on the highway; accident with injuries; driver is unwell; broken down vehicle; accident with only material damage; and accident with fire. The different types of incident categories are mutually exclusive, for example, if an incident with fire has injuries, it has been classified as an accident with injuries. Table 2 describes the number and relative share of the incident types in the GSM zones observed over the course of about one year. Incidents from the underlying network (local roads) were excluded from our analysis.

**Table 2: Descriptive statistics of hourly traffic incidents of all selected GSM zones in Greater Amsterdam during the observation period 2010**

| Incident type | Description | Number | % Share | Avarage # incidents per hour per GSM zone | Mean incident duration |
|---|---|---|---|---|---|
| 1 | Object on the highway | 261 | 0.110 | 0.00033 | 16 min. |
| 2 | Accident with injuries | 59 | 0.025 | 0.0000746 | 71 min. |
| 3 | Driver is unwell | 32 | 0.013 | 0.0000405 | 24 min. |
| 4 | Broken-down vehicle | 1204 | 0.505 | 0.0015224 | 29 min. |
| 5 | Only material damage | 809 | 0.340 | 0.0010229 | 42 min. |
| 6 | Fire | 17 | 0.007 | 0.0000215 | 57 min. |
| | **Total** | **2382** | **100%** | 0.0030119 | |

**Note: Based on 790,865 hourly observations of 109 GSM zones.**

The temporal variation of the incident occurrence is characterized by some significant temporal regularities. Figure 5 gives an overview of the hourly, daily, and monthly distribution of the incidents. During rush hours there are significantly more incidents, with the highest peak in the evening (Figure 5a). During working days there are substantially more incidents than during the weekend (Figure 5b). Figure 5c gives an overview of the monthly variation. It is important to note that, in our data set, we are missing data from 9 Jan.-25 Jan.

and 21 Nov.-31 Dec. 2010. Since the mobile phone data have a high spatio-temporal (hourly and daily) regularity in terms of data volume, and we have 790, 865 hourly observations, this will not significantly influence the statistical outcomes of our results.

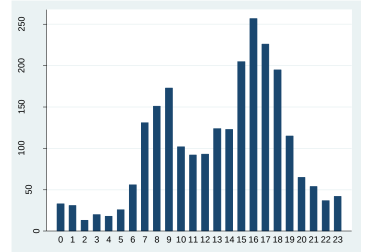


**Fig. 5a: Variation of hourly traffic incidents**

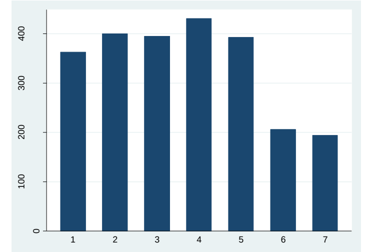


**Fig 5b: Variation of daily traffic incidents over a week**

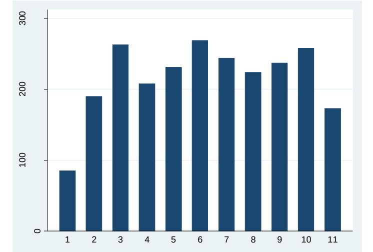


**Fig. 5c: Variation of monthly traffic incidents**

**Figure 5: Temporal distribution of road traffic incidents**

Traffic incidents may be sensitive to the different motorway characteristics. Therefore, we made a distinction between three categories: 1 = intersection point of highways; 2 = highway with exit and entry point; 3 = straight highway. From Figure 6, we can conclude that most motorway traffic incidents take place in categories 1 and 2.

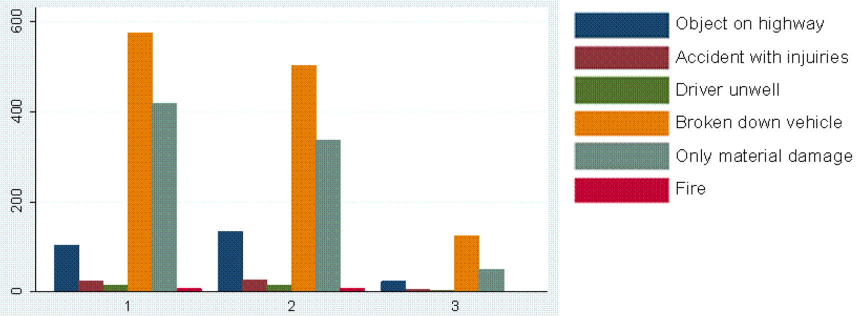



**Figure 6: Number of incidents on different types of infrastructure**
**Note 1: See Table 1 for explanation types of incidents.**
**Note 2: The Figures 1, 2, 3 denote infrastructure categories.**

It is important to realize that the six types of traffic incidents have a different impact on the smoothness of traffic flows. An important aspect is the time needed to handle an incident. Figure 7 gives an overview of the average time taken to handle the different types of incidents.

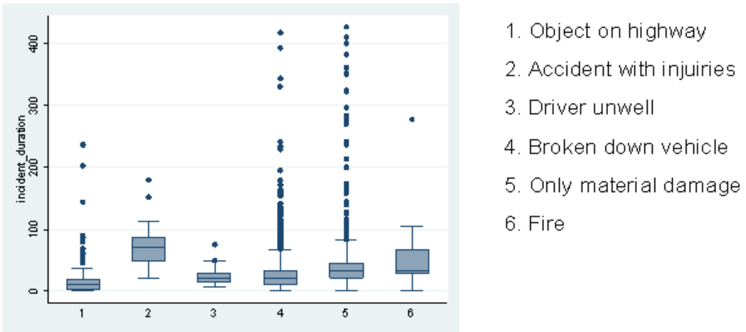



**Figure 7: Average time taken to handle the different types of traffic incidents (in minutes).**

The best serving cells (GSM zones) of the mobile phone data are used the basic measurement for data fusion. Therefore, the location of the road traffic incidents were related these GSM zones.

**2.3 Motorway traffic flow data**

Parts of the Dutch road network, especially in dense urban areas, are equipped with a comprehensive monitoring system based on detection loops with a distribution of between 300 and 500 metres apart. This system allows for the collection, processing, and transmission of dynamic and static traffic data. It contains accurate information about different data types such as traffic flow, average speeds and traffic jam lengths. For the purpose of our case study, we use traffic flow, which is the total number of vehicles which pass per hour on a specific road segment. The accuracy of these measurements lies between 95 and 98 per cent. The data were extracted from the 'MTR+' detection loop application provided by the Dutch Ministry of Infrastructure and Environment (Rijkswaterstaat, 2002). Due to the complexity of the mobile phone best-serving cell maps (GSM zones), we decided to only use traffic measurements of 7 GSM zones which contains an area coverage of only one unique polygon. As the best serving cells of the mobile phone data are used the basic measurement for data fusion, motorway car traffic counts were also related to these GSM zones.

**2.4 In-situ meteorological sensor data**

There is a vast literature on the role of different weather variables in road traffic accidents. For an extensive historical literature review on weather information and road safety, see, for example, Andrey *et al.* (2001a, 2001b and 2003) and SWOV (2009). This literature can be classified in several ways: for instance, by statistical methodology, level of aggregation, time period, geographical location, and explanatory variables; and on the basis of the type of weather measurements (e.g. hourly, daily, or monthly), etc. In the literature, many researchers find that the total number of road accidents increases with different types of weather conditions. An important issue about measurements of weather conditions is the *number of weather factors*. Some studies focus on just one weather factor, while other studies focus on more than one. *Precipitation* is the most significant and most studied weather factor, followed by *snow, temperature, fog, wind*, etc. Examples of such studies are: *precipitation* (Satterthwaite 1976; Brodsky and Hakkert 1988; Andrey and Yagar 1993; Edwards 1996; Andrey *et al.* 2003; Keay and Simmonds 2006; Bijleveld and Churchill 2009); *snow* (Edwards, 1996; Nofal and Saeed, 1997; Brijs *et al.*, 2008); a combination of *snow* and *ice*

(Kallberg, 1996); *temperature* (Stern and Zehavi, 1990; Wyon *et al*. 1996; Nofal and Saeed; 1997); and strong *wind* (Baker and Reynolds, 1992; Young and Liesman, 2007). Many researchers find that the total number of road accidents increases with precipitation. However, there is a range of variation in the empirical findings of the different weather conditions which makes it difficult to generalize the findings of these studies (Sabir, 2011). The *Meteorological* data used for this case study was obtained from the Royal Netherlands Meteorological Institute, KNMI (www.knmi.nl). All measurements are hourly averages, and are measured by accurately calibrated weather stations used for regional weather forecasting. The meteorological measurements are related to the best serving GSM zones. We consider three meteorological variables:

- T: Temperature in units of 0.1$^{o}$ C;
- R: Rainfall (0=no occurrence, 1=occurred during the time of observation);
- S: Snow (0=no occurrence, 1=occurred during the time of observation);

## 3. Empirical application

We adopt a three stage approach. In the first step, mobile phone usage will be regressed against motorway car traffic for those zones where data is available. Here we examine the effect of motorway traffic flow and traffic incidents on mobile phone usage, while controlling for different space-time variables and weather conditions. In the second step, we analyse whether an increase in motorway traffic flows affects the probability of different types of incidents. In the third and last step, the marginal effects different types of mobile phone usage on the probability of having a motorway incident will be estimated. The main objective of the modelling strategy is to test the *usability* of data derived from mobile phone operators as a detector of traffic incidents.

### 3.1 Motorway traffic and mobile phone usage

In this section we estimate the effect of motorway traffic on mobile phone use, given the temporal and spatial dimension and resolution of the data from the mobile phone operator. In order to perform this analysis, different mobile phone usage variables (Erlang, new calls, SMS and handovers) are used here as the dependent variables. The main goal of this exercise is to see which mobile phone variable is significantly related to motorway traffic derived from the detection loop database. The first analysis is limited to 7 GSM zones for which motorway flow data (number of cars per hour) are available. The basic version of this model is:

$$ln(mob_{it}) = b_1 ln(car_{it}) + b_2 incident_{it} + B_1X_i + B_2T_t + B_3W_t + B_4X_i * H_t + \alpha_0 + \varepsilon_{it} . \quad (1)$$

According to Model 1, mobile phone activity ($mob_{it}$) in GSM zone $i$ and time $t$ is affected by: a coefficient $b_1$ of motorway traffic flow ($car_{it}$) in GSM zone $i$ and at time $t$; a coefficient $b_2$ of motorway incidents ($incident_{it}$) in GSM zone $i$ and at time $t$; a vector $B_1$ of fixed effects of GSM zones ($X_i$); a vector $B_2$ of time specific fixed effects ($T$) including hourly, weekday and monthly effects; and a vector $B_3$ of various weather conditions variables ($W_t$). In order to better understand how mobile phone use changes over time, we incorporate into the model the time variability of our observations by introducing hourly interaction terms ($H_t$) for the GSM zones ($X_i$), Note that the variables $mob_{it}$ and $car_{it}$ are introduced in the model via their natural logarithms because they are heavenly skewed.

**Table 3: Prais-Winsten regression of different mobile phone variables for 7 GSM zones using data on motorway traffic**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VARIABLES | ln(erlang) | ln(new_calls) | ln(sms) | ln(handovers) |
| ln(cars) | 0.317 | 0.196 | 0.223 | 0.547 |
| | (11.30)*** | (7.247)*** | (7.626)*** | (29.20)*** |
| incidents | 0.0493 | 0.0283 | 0.0803 | 0.0183 |
| | (1.205) | (0.788) | (1.956)* | (0.692) |
| temperature | 0.000182 | 9.69e-05 | 0.000434 | 0.000188 |
| | (1.193) | (0.548) | (2.550)** | (1.743)* |
| rain | 0.00665 | 0.00827 | 0.0149 | -0.00279 |
| | (0.727) | (0.946) | (1.563) | (-0.459) |
| snow | -0.0485 | 0.00705 | 0.0155 | -0.0102 |
| | (-1.864)* | (0.289) | (0.575) | (-0.593) |
| hour dummies | included | included | included | included |
| weekday dummies | included | included | included | included |
| month dummies | included | included | included | included |
| GSM zone dummies | included | included | included | included |
| hourly interaction terms (GSM zones) | included | included | included | included |
| constant | -1.210 | 2.971 | 2.560 | 1.617 |
| | (-4.900)*** | (12.39)*** | (9.897)*** | (9.788)*** |
| observations | 40,609 | 40,609 | 40,609 | 40,609 |
| R-squared | 0.816 | 0.692 | 0.670 | 0.881 |
| Durbin-Watson (original) | 0.900 | 0.666 | 0.808 | 0.814 |
| Durbin-Watson (transformed) | 2.063 | 2.298 | 2.203 | 2.173 |

t-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Instead of reporting the simple OLS results (which can be obtained upon request), model (1) is estimated using the Prais-Winsten regression in order to address the presence of serial autocorrelation in our data. According to Durbin-Watson test presented in Table 3, our estimation strategy addresses this problem (transformed Durbin-Watson close to 2).

An important finding is that mobile phone usage, measured in Erlang, new calls, sms, and handovers is positively and significantly affected by motorway car traffic. For example, an increase in traffic flow of 1 per cent leads to an increase in Erlang of 0.32 per cent because an increase in traffic flow will result to more cars and therefore more people present in an area. Similarly, an incident is related to an increase in Erlang of 4.9 per cent. An interesting observation is that an incident has a lower impact on handovers (1.8 per cent) than on the other mobile phone uses, which is a signal that an incident makes traffic slower. One major limitation of this model is that we have only for 7 cells motorway car traffic. Because of these data limitations, we cannot test the combined effect of traffic flow and incidents using all the 109 GSM cells. In order to test whether these results are similar when data on car traffic is disregarded, we estimate model (1) without the motorway car traffic variable ($car_{it}$) for the all the 109 GSM zones. Table 4 present this estimation, using again the Prais-Winsten estimator.

**Table 4: Durban-Watson regression of different mobile phone variables for all 109 GSM zones without using car data**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VARIABLES | ln(erlang) | ln(new_calls) | ln(sms) | ln(handovers) |
| incidents | 0.0653 | 0.0558 | 0.0372 | 0.0300 |
| | (6.043)*** | (6.433)*** | (3.984)*** | (3.572)*** |
| temperature | 0.000430 | 0.000402 | 0.000422 | 0.000390 |
| | (8.866)*** | (8.312)*** | (9.362)*** | (9.396)*** |
| rain | -0.0108 | -0.00753 | -0.00121 | -0.00755 |
| | (-4.426)*** | (-3.671)*** | (-0.562) | (-3.897)*** |
| snow | 0.000673 | 0.00845 | 0.0189 | -0.00947 |
| | (0.0953) | (1.431) | (3.056)*** | (-1.693)* |
| hour dummies | included | included | included | Included |
| weekday dummies | included | included | included | Included |
| month dummies | included | included | included | Included |
| GSM zone dummies | included | included | included | Included |
| hourly interaction terms (GSM zones) | included | included | included | Included |
| constant | 0.255 | 0.656 | 0.426 | 4.676 |
| | (5.455)*** | (14.91)*** | (10.02)*** | (120.3)*** |
| observations | 790,865 | 790,865 | 790,865 | 790,865 |
| R-squared | 0.622 | 0.568 | 0.618 | 0.675 |
| Durbin-Watson (original) | 0.688 | 0.517 | 0.626 | 0.607 |
| Durbin-Watson (transformed) | 2.099 | 2.215 | 2.189 | 2.101 |

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The effect of traffic incidents is positive and significant for all the different right hand-side variables. Our interpretation is that if there is an incident on a motorway – especially if it results to traffic disruptions – more drivers will use their phone while they are waiting. The estimated effects of traffic incidents still have approximately the same values, both with and without using the car flow data (coefficients between 0.03 – 0.06), as because the hourly dummies capture the temporal variation of car traffic. Thus, the incident elasticity of about 0.03 to 0.06 (from Table 4) is not just a reflection of exposure. Regarding the weather variables, temperature, and snow are both significant (with a positive effect), rain is significant (with a negative effect).

In a nutshell, the regressions presented in Table 4 indicates that incidents are positively related with mobile phone usage. Because of data limitations we cannot test the combined effect of traffic flow and incidents using all the 109 GSM cells. In regards to Table 3 results, we believe that the variable incidents lack significance because of the small sample of spatial units we have traffic flow data for (7 GSM zones). For instance, if we run a bivariate regression using the natural logarithm of erlang as the LHS variable and incidents as the only RHS variable for all the 109 GSM zones (N=790865), the incidents variable is positive (0.099) and significant (***), while if we run the same regression only for the 7 GSM zones we have traffic flow data for (N=40609) the coefficient (0.063) is not significant. In total, we believe that our strategy of presenting two sets of regressions (Table 3 and Table 4) highlights two important points: traffic flow (ln(cars)) and incidents (incidents) are positively related with mobile phone usage. In order to move a step forward, the next section examines the factors affecting the occurrence of traffic incidents on the highway.

### 3.2 Motorway incidents and motorway traffic

The first step of the analysis provided hard evidence that after controlling for various spatial and temporal characteristics, mobile phone use is statistically related with the amount of traffic in a motorway segment and also with the existence of an incident. The next step is to reverse the direction of causality in order to test whether variables depicting mobile phone usage can be used as detectors of motorway traffic. Equation (2) below describes our logic. According to this model, the probability of having an incident in a motorways segment $i$ at time $t$ is a function of motorway traffic flow ($car_{it}$), mobile phone usage ($mob_{it}$) as well as a number of fixed spatial ($X_i$) and temporal ($T_t$) effects as in equation (1). Since the probability of having an accident is not observable, the left hand-side variable in (2) is a dummy variable ($incident_{it}$) indicating the presence of motorway incident in GSM zone $i$ at time $t$. The usability of this model lies on the fact that if we estimate a significant and positive $b_2$ coefficient after controlling for space (GSM zone) and time (hour of the day, day of the week and month) as well as weather conditions ($W_t$), then it will have been proved that an increase in (the observed) mobile phone use will be related to an increase in the probability of having a motorway incident. In simple English, this model can prove the concept that an IM system which focuses on the rapid identification of motorway incidents will be benefited by the use of a stream of data on mobile phone usage. Equation (2) is estimated for the 7 selected GSM zones for which data on motorway traffic is available.

$$Pr(incident_{it} = 1|x_{it}) = \mathrm{F}[b_1 ln(car_{it}) + b_2 ln(mob_{it}) + B_1 X_i + B_2 T_t + B_3 W_t + \alpha_0], \qquad (2)$$

where the link function F follows from the specification of the probit model.

The results are presented in Table 5. Column (1-4) includes motorway traffic and in Column (5-8) we excluded motorway traffic, because this data was not available. We find that motorway traffic in Column (1-4) has a negative and significantly effect on incidents. We find a positive and significant effect of Erlangs on the probability that an incident occurs during the hour observed. However, the other mobile phone usage indicators (new calls, sms, and handover) are positive but not significant. In regards to the results presented in Table 5 (Column 1-4), we believe that the variable incidents lack significance because of the small sample of spatial units we have traffic flow data for (7 GSM zones). To explore whether this model yields meaningful results when data on car traffic intensity are not available, we estimate equation (2) for all the 109 GSM zones (Column 5-8, Table 5). The most important finding is that the coefficients for mobile phone use, in terms of Erlangs, new calls, sms, and handovers are positive and significant. When the traffic flow data ($car_{it}$) is excluded, the spatial and temporal fixed effects pick up the traffic variation in time and space and the decrease in pseudo R squared is only marginal. Thus, we conclude that when data on traffic flow are absent, a significant positive relation between the probability of a motorway incident and the volume of mobile phone use is found.

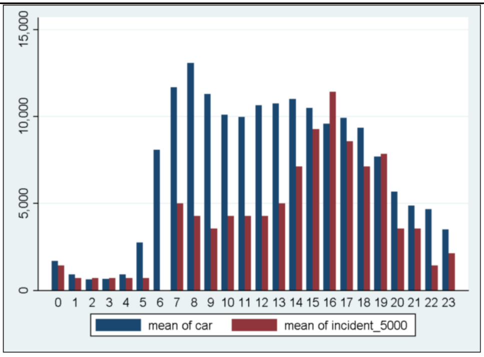
**Table 5: Probit model of the probability of an incident for the 7 (column 1-4) and the 109 (column 5-8) selected GSM zones**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| VARIABLES | incidents | incidents | incidents | incidents | incidents | incidents | incidents | incidents |
| ln(cars) | -0.691 | -0.701 | -0.702 | -0.708 | | | | |
| | (-3.459)*** | (-3.507)*** | (-3.516)*** | (-3.523)*** | | | | |
| ln(trafficerlang) | 0.121 | | | | 0.148 | | | |
| | (1.809)* | | | | (12.34)*** | | | |
| ln(new calls) | | 0.0428 | | | | 0.0998 | | |
| | | (0.813) | | | | (9.445)*** | | |
| ln(sms) | | | 0.0446 | | | | 0.0732 | |
| | | | (0.898) | | | | (6.846)*** | |
| ln(handovers) | | | | 0.00865 | | | | 0.111 |
| | | | | (0.104) | | | | (8.267)*** |
| temperature | 0.000742 | 0.000771 | 0.000766 | 0.000796 | 0.000191 | 0.000203 | 0.000229 | 0.000212 |
| | (0.796) | (0.828) | (0.823) | (0.855) | (0.888) | (0.942) | (1.064) | (0.987) |
| rain | 0.00968 | 0.0119 | 0.0125 | 0.0135 | 0.0410 | 0.0422 | 0.0419 | 0.0427 |
| | (0.122) | (0.149) | (0.157) | (0.170) | (2.286)** | (2.355)** | (2.344)** | (2.386)** |
| snow | 0.271 | 0.274 | 0.273 | 0.277 | -0.0408 | -0.0391 | -0.0363 | -0.0307 |
| | (1.392) | (1.418) | (1.414) | (1.435) | (-0.652) | (-0.625) | (-0.582) | (-0.493) |
| hour dummies | included | included | included | included | included | included | included | included |
| weekday dummies | included | included | included | included | included | included | included | included |
| month dummies | included | included | included | included | included | included | included | included |
| GSM zone dummies | included | included | included | included | included | included | included | included |
| Constant | 0.672 | 0.540 | 0.519 | 0.649 | -2.634 | -2.760 | -2.674 | -3.127 |
| | (0.476) | (0.381) | (0.366) | (0.458) | (-32.90)*** | (-33.36)*** | (-33.11)*** | (-30.01)*** |
| Observations | 38,921 | 38,921 | 38,921 | 38,921 | 790,865 | 790,865 | 790,865 | 790,865 |
| Pseudo R2 | 0.111 | 0.109 | 0.109 | 0.109 | 0.104 | 0.102 | 0.101 | 0.102 |

z-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The negative relation between car traffic and incident can be interpreted by the hourly variation of these phenomena. As indicated in Figure 8a, the temporal variation of motorway incidents does not follow the temporal variation of motorway traffic. For instance, the probability of having a car incident in the morning, when the car traffic peaks, is rather small. More specifically, the morning rush hours (between 6:00 hr and 10:00 hr) show more motorway traffic flow than the evening rush hour (between 16:00 hr and 20:00 hr). However, there are clearly more incidents during the evening rush hours, with fewer cars, than in the morning rush hours, with more cars on the road. Although this analysis contains only 7 GSM zones, the spatial signature of hourly distribution of traffic incidents shows approximately the same pattern for all the 109 GSM zones (see Figure 8b).



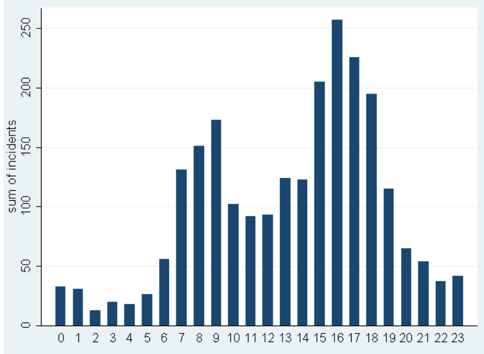




**Figure 8a: Distribution of the hourly car traffic and sum of traffic incidents (7 selected GSM zones)**

**Figure 8b: Distribution of the hourly sum of the traffic incidents (109 selected GSM zones)**

### 3.3 Marginal effects for various telecom activity and types of incidents

In this section we go one step further to analyse more specifically whether mobile phone usage can be used as a detector of different types of traffic incidents for all 109 GSM zones. Furthermore, to present the estimation results in a way that is easy to understand, we make use of marginal effects (ME) based on probit models. It is important to note that these ME coefficients are not directly comparable to output generated by OLS regressions. In OLS regressions the marginal effect can be directly obtained from the estimated coefficients. Since probit models are inherently non-linear, the marginal effects depend on the level of the independent variable, and also on the levels of other independent variables. Therefore, marginal effects have a 'ceteris paribus' interpretation. They tell what happens if a given variable varies, while all the other variables remain unchanged. Here, we confine ourselves to a presentation of the estimates by means of the marginal effects based on the mean value of all independent variables. Equation (3) presents the probit model presented above (2) without the car traffic flow variable, which is estimated for all the 109 selected GSM zones:

$$Pr(incident_{it} = 1|x_{it}) = \mathrm{F}[b_1 ln(mob_{it}) + B_1X_i + B_2T_t + B_3W_t + \alpha_0] \qquad (3)$$

The results are presented in Table 6 for the total number of accidents and various types of mobile phone measures (the result for Erlang corresponds with Column (2) in Table 5).

**Table 6: Marginal effects of mobile phone use on incident probability; based on data for 109 GSM zones.**

| VARIABLES | (1) incidents | (2) incidents | (3) incidents | (4) incidents |
|---|---|---|---|---|
| ln(erlang) | 0.000615 | | | |
| | (12.34)*** | | | |
| ln(new_calls) | | 0.000426 | | |
| | | (9.445)*** | | |
| ln(sms) | | | 0.000316 | |
| | | | (6.846)*** | |
| ln(handovers) | | | | 0.000474 |
| | | | | (8.267)*** |
| temperature | 7.94e-07 | 8.66e-07 | 9.86e-07 | 9.09e-07 |
| | (0.888) | (0.942) | (1.064) | (0.987) |
| rain | 0.000177 | 0.000187 | 0.000188 | 0.000190 |
| | (2.286)** | (2.355)** | (2.344)** | (2.386)** |
| snow | -0.000160 | -0.000158 | -0.000149 | -0.000126 |
| | (-0.652) | (-0.625) | (-0.582) | (-0.493) |
| hour dummies | Included | Included | Included | Included |
| weekday dummies | Included | Included | Included | Included |
| month dummies | Included | Included | Included | Included |
| GSM zone dummies | included | Included | Included | Included |
| observations | 790,865 | 790,865 | 790,865 | 790,865 |
| Pseudo R2 | 0.104 | 0.102 | 0.101 | 0.102 |

z-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1

The main findings of this model can be summarized as follows. Even after controlling for hourly effects, coefficients for Erlang, new calls, and sms are still positive and significant for traffic incidents. The marginal effects can be interpreted as follows. A 1 per cent increase in new calls, increases the probability of an incident in a specific GSM zone by 0.000426 $*(0.01) = 4.26*10^{-6}$. This is equal to $4.26*10^{-4}$ per cent. This is clearly a very low figure, but note that the average probability of an incident is 0.00301 (see Table 1). So the relative increase in the probability of an incident related to 1 per cent increase in new calls is about 0.142 per cent. Similarly, a 0.1$^{o}$ C. increase in temperature increases the probability by $8.66*10^{-7}$, which is indeed a very small effect.

The last modelling step looks in more detail at the different types of incidents, as described earlier in Section 2.1, written as:

$$Pr(incident\_type_{it} = 1|x_{it}) = \mathrm{F}[b_1 ln(mob_{it}) + B_1 X_i + B_2 T_t + B_3 W_t + \alpha_0] \quad (4)$$

**Table 7: Marginal effects of mobile phone use on incident probability for various types of incidents; based on data for 109 GSM zones.**

| VARIABLES | (1) Object on the highway | (2) Accidents with Injuries | (3) Driver unwell | (4) Broken-down vehicle | (5) Only material damage | (6) Fire |
|---|---|---|---|---|---|---|
| ln(erlang) | -1.08e-05 | 4.13e-05 | 6.28e-05 | 0.000328 | 0.000301 | -4.79e-06 |
| | (-0.482) | (2.351)** | (3.613)*** | (9.164)*** | (9.286)*** | (-0.266) |
| temperature | 4.18e-07 | -2.87e-07 | -2.58e-08 | 1.08e-06 | -6.62e-07 | 1.25e-07 |
| | (0.918) | (-0.873) | (-0.0932) | (1.745)* | (-1.149) | (0.351) |
| rain | -4.32e-05 | -1.90e-05 | -4.66e-05 | 1.94e-06 | 0.000256 | -2.57e-05 |
| | (-1.126) | (-0.717) | (-1.900)* | (0.0358) | (5.079)*** | (-0.867) |
| snow | -0.000174 | 0.00218 | | 2.58e-05 | -7.82e-05 | |
| | (-1.402) | (1.594) | | (0.145) | (-0.530) | |
| hour dummies | Included | Included | Included | Included | Included | Included |
| weekday dummies | Included | Included | Included | Included | Included | Included |
| month dummies | Included | Included | Included | Included | Included | Included |
| GSM zone dummies | Included | Included | Included | Included | Included | Included |
| observations | 579,431 | 204,089 | 139,232 | 739,833 | 717,952 | 42,270 |
| Pseudo R2 | 0.0711 | 0.0700 | 0.0800 | 0.104 | 0.0939 | 0.0632 |

z-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1

An ‘object on the highway’ is not a major event with large disruption for traffic and this is why it does not result to a significant coefficient. The categories accidents with injuries, driver being unwell, broken-down vehicles and incidents with only material damage are all positive and significant as shown in Table 7. The marginal effects can be interpreted as follows. A 1 per cent increase in Erlang, increases the probability of an incident with only material damage in a specific GSM zone by $0.000301*0.01=3.01*10^{-6}$. This is equal to $3.01*10^{-4}$ per cent. We find rather higher effects for broken-down vehicles than for the other types of incident.

There are a number of reasons which may explain this outcome. The A10 Amsterdam ring road has the characteristic that the number of cars during the day is close to its maximum capacity. For safety reasons, before 2011, the Traffic Management Centre completely closed 1 driving lane so it could be used for emergency aid. This directly caused traffic jams, even just for broken-down vehicles compared with other types of accidents. Since 2011 (so after our study period), because of major congestion problems, this policy has been changed. Furthermore, our database has only 59 incidents with injuries. Only a part (approximately 40 per cent) of these incidents took place during rush hours (see Figure 9).

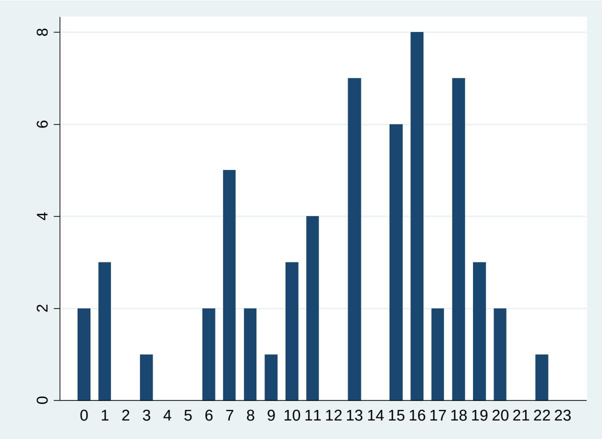



**Figure 9: Temporal distribution of road traffic incidents with injuries**

Another interesting finding is that temperature is only significant (and positive) for broken-down vehicles ($p<0.1$). One might expect that low temperatures would affect broken-down vehicles. The plausible explanation is that, in cold weather conditions, people are likely to have more problems with their batteries. This kind of problem occurs when drivers start their cars from their homes. Cars will need to be repaired before they can enter the highway.

Rain is only positive and significant for material damage (collisions between cars) ($p<0.01$). This means that wet weather conditions significantly influence the safety on the roads. They affect incidents with only material damage but not accidents with injuries. The explanation is that in the rain drivers reduce speed (see Sabir, 2011), so serious accidents are less probable, but still the probability of less-serious accidents does increase when it rains.

## 4. Conclusions

The above analysis provided a proof of concept, according to which data on mobile phone usage can be utilised within an IM system as a detector of traffic incidents and provide the basis of an early warning system. The new layer of information provided by mobile phone usage data as a means to detect motorway incidents can improve IM systems without adding any addition cost that other live traffic measurement systems may impose. Data on mobile phone use can be obtained free of charge as it is collected by mobile phone operators for billing purposes.

At a more detailed level, an important finding is that mobile phone usage in telecom GSM zones is strongly affected by motorway traffic flow, and also by the occurrence of traffic incidents. This makes mobile phone usage in GSM zones that are crossed by motorways a promising proxy for traffic flow and the occurrence of incidents on these roads. The estimated effects of traffic incidents on mobile phone use still have approximately the same values both with and without using the car traffic flow variable. The high resolution of the mobile phone data enables us to extract vital information at a very fine-grained spatial scale.

The main limitation of our study is that the temporal resolution as the one hour interval is quite crude. A reduction to 5 or 10 minutes would enable us to provide information which answers the required timelines of the end-users in the traffic management centres. Nevertheless, even with such sparse temporal resolution the results of the analysis support our proposal. Moreover, the data quality of the mobile telecom network also deserves further attention. A telecom network consists of a complicated technical structure, where the mobile phone use data are extracted from. A valuable exercise would be to compare the different approaches which commonly found in the relevant literature, such as *rasterization and Voronoi* diagrams. Another limitation of our approach is that we use only the data of one telecom operator. To enrich our data, it would be highly interesting to use the data of other telecom operators in the Netherlands. Next to that, the increasing use of smart phones also generates new events in terms of data usage. A next step in our approach is to enrich our modelling exercise with data related network events. Other interesting themes would be to apply concepts such as 'Dynamic Data Driven Application Systems' (DDDAS), to handle real-time data flows from the telecom network, and to develop a simulation system for evacuation and the effect of different emergency scenarios and types of agent behaviour (e.g. Darema, 2004; Madey *et al.*, 2007). Moreover, with more sophisticated data, it would be also interesting to combine other application such as traffic speed, travel times, traffic flows, and traffic congestion related to traffic incidents. Another research approach is to use individual probes instead of aggregated data as have been explored in this paper. Such research direction is highly depending on the willingness of telecom operators to share such data and the constraints in terms of privacy regulations. Moreover, this also requires special hardware in the telecom network. However, such data will enable the research field to address new research question, such as Origin Destination transportation modes, using advanced spatial data analytics.

Finally, the process of data fusion, which combines information originating from multiple sources, could be further explored. Overlapping information and comparative data sets, such as detection loop data, estimation of traffic flow, weather conditions, social media data, and floating car data could be used to detect, identify, and track relevant objects in a region to support Situational Awareness for traffic incident management.

## REFERENCES


Ahas, R., Aasa, A., Roosea, A., Mark, Ü., Silm, S. (2007). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management* 29(3), pp. 469-486.

Akyildiz, I.F., Su, W., Sankarasubramaniam, Y. and Cayirci, E. (2002). A survey on sensor networks, *IEEE Communications Magazine* 40, pp. 102-114.

Alesheikh, A.A., Oskouei, A.K., Atabi, F. and Helali, H. (2005). Providing interoperability for air quality in-situ sensors observations using GML technology, *International Journal of Environment Science and Technology* 2, pp. 133-140.

Andrey, J. and Yagar, S. (1993). A temporal analysis of rain-related crash risk, *Accident Analysis and Prevention* 25(4), pp. 465-472.

Andrey, J., Mills, B, and Vandermolen, J. (2001a). Weather information and road safety. Department of Geography, University of Waterloo. Adaptation and Impacts Research Group, MSC Environment Canada. Paper Series – No. 15. August 2001.

Andrey, J., Suggett, J., Mills, B. and Leahy, M. (2001b). Weather-related road accident risks in mid-sized Canadian cities, in: *Proceedings of Canadian Multidisciplinary Road Safety Conference XII*, June 11-13, London.

Andrey, J., Mills, B., Leahy, M. and Suggett, J. (2003). Weather as a chronic hazard for road transportation in Canadian cities. *Natural Hazards,* 28, pp. 319-343.

Baker, C.J. and Reynolds, S. (1992). Wind-Induced accidents of road vehicles. *Accident Analysis and Prevention* 24(6), pp. 559-575.

Becker, R.A., Caceres, R., Hanson, K., Loh, J.M., Urbanek, S., Varshavsky, A., and Volinsky, C. (2011). A tale of one city: Using cellular network data for urban planning. *Pervasive Computing* 10(4), pp. 18-26.

Bijleveld, F. and Churchill, T. (2009). The influence of weather conditions on road safety SWOV Institute for Road Safety Research, Report No. R-2009-9, The Netherlands.

Boyd, D. and Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society* 15(5), pp. 662-679.

Brijs, T., Dimitris, K. and Greet, W. (2008). Studying the effect of weather conditions on daily crash counts using a discrete time-series model. *Accident Analysis and Prevention* 40(3), pp. 1180-1190.

Brodsky, H. and Hakkert, A.S. (1988). Risk of a road accident in rainy weather, *Accident Analysis and Prevention* 3, pp. 161-176.

Calabrese, F., Kloeckl, K. and Ratti, C. (2007). WIKICITY: Real-time urban environments. *Pervasive Computing, Mobile and Ubiquitous Systems* 6(3), pp 52-53. July - September 2007.

Darema, F. (2004). Dynamic data driven applications systems: A new paradigm for application simulations and measurements. In Computational Science-ICCS 2004, pp. 662-669. Springer Berlin Heidelberg.

Edwards, J.B. (1996). Weather-related road accidents in England and Wales: A spatial analysis, *Transport Geography* 4, pp. 201-212.

Girardin, F., Vaccari, A., Gerber, A, Biderman, A. and Ratti, C. (2009). Quantifying urban attractiveness from the distribution and density of digital footprints, *International Journal of Spatial Data Infrastructures* 4, pp. 175-200.

Goodchild, M.F. (2007). Citizens as voluntary sensors: Spatial data infrastructure in the world of web 2.0. *International Journal of Spatial Data Infrastructures Research* 2, pp. 24-32.

González, M.C., Hidalgo, C.A. and Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature* 453(5), pp.779-782, June 2008.

Hart J.K. and K. Martinez (2006). Environmental Sensor Networks: A revolution in the earth system science? *Earth-Science Reviews* 78, pp. 177-191.

Kallberg, V.-P. (1996). Experiment with reduced salting of rural main roads in Finland, *Transportation Research Record* 1533, pp. 32-43.

Keay, K. and Simmonds, I. (2006). Road accidents and rainfall in a large Australian city, *Accident Analysis and Prevention* 38, pp. 445-454.

Kuusik, A., Ahas, R. and Tiru, M. (2008). Analysing repeat visitation on country level with passive mobile positioning method: An Estonia case study, University of Tartu, Positium LBS.

Madey, G., Barabási, A.-L., Chawla, N.V., Gonzalez, M., Hachen, D., Lantz, B., Pawling, A., Schoenharl, T., Szabó, G., Wang, P. and Yan, P. (2007). Enhanced situational awareness: Application of DDDAS concepts to emergency and disaster management, in: *International conference on computational science, serial lecture notes in computer science (LNCS 4487)*, Y. Shi, G. D. van Albada, J. Dongarra, and Sloot, P. M. A. (eds), May 2007, pp. 1090-1097.

Nofal, F.H. and Saeed, A. (1997). Seasonal variation and weather effects on road traffic accidents in Riyadh City, *Public Health* 111, pp. 51-55.

O'Neill, E., Kostakos, V., Kindberg, T., Schieck, A.F., Penn, A., Fraser, D.S., and Jones, T. (2006). Instrumenting the city: Developing methods for observing and understanding the digital cityscape. *Ubicomp*, pp. 315-332.
Reades, J., Calabrese, F. and Ratti, C. (2009). Eigenplaces: analysing cities using the space-time structure of the mobile phone network. *Environment and Planning B: Planning and Design* 36, pp. 824-836.
Resch, B. Mittlboeck, M. and Lippautz, M. (2010), Pervasive monitoring - An intelligent sensor pod approach for standardised measurement infrastructures. *Sensors* 10, pp. 11440-11467.
Rijkswaterstaat (2002). Productcatalogus basisinformatie. Adviesdienst Verkeer en Vervoer, hoofdafdeling basisgegevens.
Sabir, M. (2011). Weather and Travel behaviour. Dissertation, research series Vrije Universiteit Amsterdam, March 2011.
Satterthwaite, S.P. (1976). An Assessment of Seasonal and Weather Effects on the Frequency of Road Accidents in California Accident *Analysis and Prevention* 8, pp. 87-96.
Song, C., Koren, T., Wang,, P. and Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. *Nature Physics* 6, pp. 818-823.
Steenbruggen, J., Nijkamp, P., Smits, J.M. and Grothe, M. (2012). Traffic incident management: A common operational picture to support situational awareness of sustainable mobility. *International Journal Transport Economics* 1 (2012).
Steenbruggen, J., Borzacchiello, M.T., Nijkamp, P. and Scholten, H. (2013a). Data from telecommunication networks for Incident Management: An exploratory review on transport safety and security, *Transport Policy* 28, pp. 86-102.
Steenbruggen, J., Borzacchiello, M.T., Nijkamp, P. and Scholten, H. (2013b). Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: a review of applications and opportunities. *GeoJournal* 78(2), pp. 223-243.
Steenbruggen, J., Nijkamp, P. and van der Vlist, M. (2014a). Urban traffic incident management in a digital society: An actor–network approach in information technology use in urban Europe. *Technological Forecasting and Social Change* 89, 245-261.
Steenbruggen, J. (2014b). Tourism geography: Emerging trends and initiatives to support tourism in Morocco. Vrije Universiteit Amsterdam, Faculty of Economics and Business Administration, Research Memorandum 2014-2, pp. 1-29
Steenbruggen, J., Tranos, E. and Nijkamp, P. (2014c). Data from mobile phone operators: A tool for smarter cities. *Telecommunication Policies* 39(3-4), pp. 335-346.
Stern, E. and Zehavi, Y. (1990). Road safety and hot weather: A Study in applied transport geography, *Transactions of the Institute of British Geographers* 15(1), pp. 102-111.
SWOV (2009). Fact sheet: Influence of weather on road safety, Institute for Road Safety Research, The Netherlands.
Traag, V.A., Browet, A., Calabrese, F. and Morlot, F. (2011). Social event detection in massive mobile phone data using probabilistic location inference, submitted to IEEE SocialCom'2011.
Wyon, D. P., Wyon, I. and Norin, F. (1996). Effects of moderate heat stress on driver vigilance in a moving vehicle, *Ergonomics* 39(1), pp. 61-75.
Young, R.K. and Liesman, J. (2007). Estimating the relationship between measured wind speed and overturning truck crashes using a binary logit model, *Accident Analysis and Prevention* 39, pp. 574-580.
Zhang, D., Guo, B., Li, B. and Yu, Z. (2010). Extracting social and community intelligence from digital footprints: an emerging research area, in: *Proceedings of the 7th international conference on Ubiquitous Intelligence and Computing (UIC'10).* Springer-Verlag Berlin, Heidelberg 2010 ISBN:3-642-16354-8 978-3-642-16354-8.