USE OF OPEN STREET MAP FOR ACCIDENT INVESTIGATION ON THE ROAD AND MOTORWAYS NETWORKS

Marco Broccoli, Silvia Bruzzone

1. Road Safety performance Indicators and Big data use

Road Safety Performance Indicators (RSPI) give a multidimensional approach for accidents investigation concerning roads, vehicles and persons involved. Combining the use of statistical surveys, administrative geographical information systems (GIS) and Big Data (BD) sources, the result gives new elements on planning infrastructure solutions, applying policies to reduce deaths and serious injuries, reducing social costs on collectivity and estimating efficiency and effectiveness of safety initiatives. Preventing road trauma on public roads is a core responsibility for government, its agencies and stakeholder. It requires a common and shared responsibility.

Nowadays there is a clear information bias as regards the appropriate reference denominators to be placed as basis in construction of statistical indicators linked to road accidents. Resident population is used as a common proxy for exposed at risk in a specific geographical area, but not always an appropriate solution, especially in the light of the seasonal nature of road accidents and concentration, in some periods of the year, in specific locations. Vehicle fleet (Automobile Club of Italy - ACI) can be another administrative source that gives a more accurate information, but the characteristic of the phenomenon implies a deductible distortion on measures due to the mobility of road users. The length of the road network (from Open Street Map) gives for sure a consistent first set of information concerning the different territories.

This information is not available from official statistics at national level, although there are archives and detailed road graphs for each municipality, province and region, a harmonized and systematic national road registry has not been established. The first output of the project is the focus on the exploitation of existing administrative sources, the scouting of new sources and the analysis of integrated and auxiliary data.

The aim of the project is to calculate road accidents rates, mortality and harmfulness indexes, comparing these measures to the correspondent length in meters of carriageway by road direction from the Open Street Map. This approach has been used for the first time, by the authors, to build road accidents indicators and adopted by the “Experimental statistics” at Istat (Broccoli and Bruzzone, 2021).

Although Open Street Map, in fact, is cited as a source of data in many scientific articles, the main purpose was for GIS-Based Traffic Simulation or other aims (Zilske and Neumann, 2011). While the product represents a first result, the final purpose of the project is to expand statistical information with the supply of traffic flows (vehicles / km) on the national road network. This would make it possible to calculate the probability of being involved in the accident, taking into account the different exposure to risk of accidents.

2. The use of GIS systems for a graphical representation

GIS is a geographic system designed to capture, store, manipulate, analyses, manage and present spatial or geographic data. GIS applications are tools that allow users to analyse spatial information, edit data in maps and present the results of all these operations. In order to relate information from different sources, GIS uses a spatial location as the key index variable (key reference by position). Just as a relational database containing text or numbers can relate many different tables using common key index variables, GIS can relate otherwise unrelated information by using location as the key index variable. This key characteristic of GIS has begun an alternative frontier on producing statistical information. Any variable that can be located spatially using an x, y, and z coordinates, representing longitude, latitude, and elevation, respectively. These GIS coordinates may represent other quantified systems of territories (polygons), road networks (lines) and point of traffic (points).

"Join attributes by location" is the algorithm that takes an input vector layer and creates a new vector layer that is an extended version of the input one, with additional characteristics in its attribute table. The additional attributes and their values are taken from a second vector layer. A spatial criterion is applied to select the values, from the second layer, added to each feature from the first layer in the resulting one.

2.1. Istat Census Map localities

The Istat Census Map localities used with the GIS system are provided for the following administrative units: Regions; Provinces; Municipalities (yearly updated); Localities (only 2011 Census). An upgrade of Census Map localities shapes referred to 2011 to municipalities 2016 has been done. It has been built a link table with the aggregation of all 8090 local administrative units' territory, at 2011, to the 7998 municipalities at 2017, included in the Italian territory.

The choice of the localities shapes is due to the harmonization need of the roads graph to the "road type classification" used by the road accidents survey. The localities classification includes four voices:

1. Urban areas
2. Small inhabited areas
3. Productive areas
4. Wide spread houses

As regards the definitions, the Istat[1] descriptions are reported below.

Localities - A more or less wide area, usually known by name, on which one or more houses are grouped or scattered; there are two types of localities: inhabited localities and productive locations. The borders of the inhabited localities (center and inhabited area) are the external limit of the buildings placed at the edges of a grouping of at least fifteen buildings. The confines of the localities include gardens and other accessories areas of the considered buildings, non-built neighboring areas not included (such as fields with or without crops). Buildings located at a distance of more than 70 meters, within built-up areas and 40 meters for small-inhabited areas excluded. If the buildings included in the new locality are adjacent or in proximity (to a max distance of 140 meters for urban areas and 60 meters for small inhabited areas) of transport infrastructures or hydrographic limits, the location border will be extended to the middle of these elements.

Urban areas - Aggregate of contiguous or near houses with roads, squares and similar, or however areas characterized by services or public activities (school, public office, pharmacy, shop or similar), detecting a social life and, generally, also a place of collection for the inhabitants of the neighboring areas. The places of tourists meetings, houses, hotels and similar used for the vacation, inhabited seasonally, are considered as temporary inhabited centers too.

Small-inhabited areas is an area without the place of collection, characteristic of the urban area. It is based on a group of at least fifteen contiguous and near buildings, with at least fifteen families, with roads, paths, squares, farmyards, small gardens and similar, as long as the distance between the buildings does not exceed thirty meters and it is lower than the distance between the center and the nearest of the houses clearly scattered.

Productive areas - Extra-urban area not included in the centers or residential areas with more than 10 local units, or with a total number of employees' upper than 200. The local units are contiguous or close, with roads, squares or similar, or

[1] Istat - Basi territoriali e variabili censuarie https://www.istat.it/it/archivio/104317 ; Istat - Descrizione dei dati geografici e delle variabili censuarie delle Basi territoriali per i censimenti: anni 1991, 2001, 2011 https://www.istat.it/it/files//2013/11/Descrizione-dati-Pubblicazione-2016.03.09.pdf

anyway in a continuous line, not exceeding 200 meters; the minimum area must be 5 hectares.

Wide spread houses - Houses scattered in the municipal territory at a distance not enough to constitute a built-up area.

2.2. Open Street Map and road arch type

Open Street Map (OSM)[2] is a collaborative project aimed on creating free content maps of the world. The project aims at a collection world of geographical data, with the main purpose of creating maps and cartography. The key feature of the geographic data present in OSM is having a free license, the Open Database License. It is therefore possible use them freely for any purpose with the only constraint of mentioning the source. Everyone can contribute by populating or correcting data. The maps are created using the data recorded by portable GPS devices, aerial photographs and other free sources. Most of the Android and iOS GPS navigation software on portable devices are powered by OSM as WisePilot, Maps.me, NavFree, Scout etc.

The Open Street Map vector layers, used in this work, daily updated and free downloadable data, are the following:

- Road graph;
- Point of traffic (POT);
- Added shape: Buildings; Use of the land; Natural; Places; POWF (Point of Worship); POIS (Point of interest); Railways; Transport; Water; Water ways.

Although OSM it is an Open Source tool based on information from a community, the product provides data to be considered reliable and consistent, so much that the major part of GPS Android and iOS navigation software on portable devices are powered by OSM, for example WisePilot, Maps.me, NavFree, Scout, etc. Table 1 contains the list of different type of road arch by Open Street Map.

An innovative method of measuring the length in meters of the road graph is given by the information on the number of carriageways of each road arch of OSM. In the future, in order to provide more detailed information, the use of the number of lanes containing each carriageway is being evaluated.

[2] OpenStreetMap provides geographic data on thousands of websites, mobile and hardware devices. OpenStreetMap is built by a community of mappers, who contribute, update and monitor data on roads, cafes, railway stations and much more, all over the world - OSM: http://www.openstreetmap.org/about

Table 1 – *Open Street Map road arch classification (a)*

Road type	Road type Description
Secondary link	The link roads (slip roads/ramps) leading to/from or from/to a secondary road or lower class highway.
Tertiary	Roads of local rank. They connect smaller municipalities together. In urban areas, they are side roads to primary and secondary roads with a medium flow of traffic.
Tertiary link	The link roads (slip roads/ramps) leading to/from or from/to a tertiary road or lower class highway.
Unclassified	Classification for some extra-urban road.
Residential	Roads in a residential area, which serve as an access to housing, without function of connecting settlements.
Living Street	Residential road where pedestrians have legal priority over cars, speeds are kept very low.
Pedestrian	Pedestrian areas (roads or squares in urban areas), accessible mainly or exclusively to pedestrians.
Service	Access roads or internal service areas, beaches, camping, industrial areas, shopping centers, parking places etc.
Track	Roads for mostly agricultural or forestry uses
Cycle way	Cycle paths on dedicated carriageway, mainly or exclusively for cycling tourism.
Footway	Paths mainly/exclusively for pedestrians. This includes walking urban tracks, paths in a public park and footpaths
Path	Paths not structured for a public use
Steps	Stairs in steps, exclusively accessible by pedestrians
Unknown	Not classified

(a) Road Archs by Open Street Map – update 16/2/2017.

2.3. Classification of OSM road arches and road accidents localization

The arch road types selected to calculate the indicators analysed are referred to the motorized vehicles flow: motorway, trunk, primary, secondary, tertiary, unclassified, residential, living street, motorway link, trunk link, primary link, secondary link, tertiary link, service, unknown. Pedestrian, track, track_grade, bridleway, cycle way, footway, path, steps are not object of the survey definition.

To build road accidents indicators, with denominator represented by the arches length in Open Street Map, we built a "bridge matrix" between road categories, classified by functional road type, used by OSM, and the categories linked to the

roads holder, used by Istat road accidents survey. This methodology can be defined as a systematic classification technique.

The systematic classification of road arches (Open Street Map), classified in the categories used by Istat, has been modified, in respect of a first release (Broccoli and Bruzzone 2021). A new analytical classification has been adopted, using a more refined technique of attributing single road arches, about three and a half million in total (OSM), to the Istat classification groups (table 2). The operational criterion applied provides the roads classification, through the textual analysis of the Name and Reference attributes, according to the different classes of road arch and spatial attribution of the location type.

Table 2 – *Bridge coding table between roads arches classification by OSM, localities and road type (b)*

Road Arches classification by Open Street Map	Localities at Census 2011			
	Urban areas + Small		Productive areas + Wide	
	Road Localisation by Road accidents survey			
	Motorways	Urban Roads	Motorway	Rural Roads
Motorway	x		x	
Trunk	x		x	
Primary		x		x
Secondary		x		x
Tertiary		x		x
Unclassified		x		x
Residential		x		x
Living Street		x		x
Motorway Link	x		x	
Trunk Link	x		x	
Primary Link		x		x
Secondary Link		x		x
Tertiary Link		x		x
Service		x		x
Unknown		x		x

(a) Istat computing.

In order to use 2017 data, improved by the new technique and to guarantee the comparability to 2016 data, already disseminated, the new 2017 indicators were also completed by a recalculation of the provincial rankings, adopting the 2016 systematic classification (Examples: Main Roads maps Chart 1-4).

Chart 1 – *Motorways. Year 2017 (Istat: analytical classification).*

Chart 2 – *Main Roads. Year 2017 (Istat: analytical classification)*

Chart 3 – *Regional Rural Roads. Year 2017 (Istat: analytical classification)*

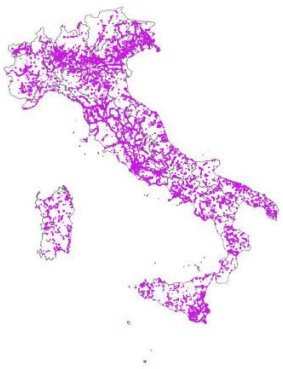

Chart 4 – *Rural Roads. Year 2017 (Istat: analytical classification)*

3. The indicators on road accidents

The proposal to calculate new road accidents indicators was born to provide, new and more suitable measures, for the risk and probability of being involved in a road accident. Even if Resident Population and Vehicles fleet, traditional sizes used, are both denominators considered, at present, *proxy* for the risk exposure to the road accidents, it is clear that they present some critical points. As regards the data updating (numerators of indicators) used to build indicators on the road accidents risk, they refers to Istat "road accidents survey", concerning all road accidents

resulting in deaths or injuries in 2017, involving at least a vehicle circulating on the national road network and documented by a Police authority[3].

As cited before, the Resident Population[4], not always is an appropriate solution to build road accident indicators, because the seasonality of accidents and the concentration in some specific places during the year. The Vehicles fleet by registration province (Automobile Club of Italy – Public Register of Vehicles PRA)[5] provides more suitable information than the population, but it does not consider the mobility of users through the country.

The length in meters of carriageway by road direction of road arch (by OSM)[6] provides, instead, a set of more coherent information referred to different areas, because independent by the seasonality and the mobility users influence.

For this reason, Istat decided to investigate the new frontier of big data and to use the new data source to build road accidents indicators and to start the consolidation process to adopt officially the calculation.

The information on the roads length, provided by OSM, is not available at national level from an institutional source, since, although archives and detailed road graphs by single municipality, province or region exist, a national and harmonised road register has been not created yet.

3.1. Road accident indicators: weighting with Traffic Points (PoT) from Open Street Map

The length in meters of carriageway per direction of the road arch (from Open Street Map), used to calculate the indicators, provides a first coherent set of information relating to the various territories. Although the product constitutes a first usable result, the project's most ambitious goal is to estimate the real traffic flows (vehicles/Km) on the national road network, as widely documented in literature, suitable indicator (Farchi et al 2006; Hakkert et all 2002). This would make it possible to calculate the probability of being involved in a road accident and therefore of real exposure to the risk of accident.

To fill this information gap, other additional information has been obtained from the rich source Open Street Map, in particular, data on the traffic points detected on the road arches (PoT Point of Traffic). A monthly information, downloadable from OSM, is available and the data refers to points over an arch, in which is detected an intensity of traffic. The proposed new road accident indicators, "weighted" with the

[3] Survey on road accidents resulting in death or injuries http://siqual.istat.it/SIQual/lang.do?language=UK
[4] Italian Resident Population - 31/12/2017.
[5] ACI Vehicle fleet (Automobile Club of Italy) all motorized vehicles except the trailers - 31/12/2017.
[6] GIS (Census Map Open Street Map road graph 1/1/2017) by length in meters per carriageway.

information on traffic intensity, was built considering, as a central element, the kilometers of carriageway with the presence of a traffic point on the arch.

The authors calculated the length of roads, considering the extension, in meters, of carriageway with arches on which including the presence of traffic points. The relative frequency (f) of the road length with the presence of traffic points on an arch (by province and type of road) and the complementary frequency (1-f) of arches without point of traffic has been calculated too.

The number of accidents, vehicles involved, deaths and injuries per 100 kilometers of carriageway in the province, excluding the effect on accidents caused by the presence of traffic points on the roads has been processed. The calculation was done by multiplying the value of the indicators for road length and the frequency (1-f) referred to arches without traffic points on the roads of the province.

With reference to the notation used for the calculation, the individual items are defined as follows:

LA Tot $_{i,j,k}$ = Total Length of Arches in meters in the province (i), by functional class of the Open Street Map Arch (j) and Type of locality (k);

LA PoT $_{i,j,k}$ = Length of Arches in meters with presence of Points of Traffic (PoT) in the province (i), by functional class of the Open Street Map Arch (j) and Type of locality (k);

LA Tot $_{i,k,l}$ = Total Length of Arches in meters in the province (i), by Type of locality (k) and Type of road by Istat classification (l);

LA PoT $_{i,k,l}$ = Total Length of Arches in meters in the province with presence Points of Traffic (PoT) in the province (i), by Type of locality (k) and Type of road by Istat classification (l);

The total length of the national roads, referred to arches length, with the presence of Points of Traffic (PoT) or in total, is given by the expressions below. Total Length of Arches with Points of Traffic (PoT):

$$\text{LA PoT}_{\text{tot}} = \sum_{i=1}^{111}\sum_{j=1}^{27}\sum_{k=1}^{4} \text{LA PoT}_{\text{i,j,k}} \quad or \quad \sum_{i=1}^{111}\sum_{k=1}^{4}\sum_{l=0}^{9} \text{LA PoT}_{\text{i,k,l}} \tag{1}$$

$$\text{LA Total}_{\text{tot}} = \sum_{i=1}^{111}\sum_{j=1}^{27}\sum_{k=1}^{4} \text{LA Totale}_{\text{i,j,k}} \quad or \quad \sum_{i=1}^{111}\sum_{k=1}^{4}\sum_{l=0}^{9} \text{LA Total}_{\text{i,k,l}} \tag{2}$$

The percentage weight of the Arches Length in meters, with the presence of Points of Traffic (PoT), in the Province (i), by functional class of Open Street Map Arch (j) and Locality type (k) out of the total length is given by the following expression:

$$p\ \text{LA PoT i, j, k} = \frac{\text{LA PoT i,j,k}}{\text{LA Tot i,j,k}} * 100 \quad (3)$$

The percentage weight of the Arches Length in meters, with the presence of Points of Traffic (PoT), at national level, out of of the total length, is given, at last, by the following expression:

$$\text{p LA PoT}_{\text{tot}} = \frac{\text{LA PoT}_{\text{tot}}}{\text{LA Total}_{\text{tot}}} \quad (4)$$

where: **i**=1,…, 103, 108, 109, 110, 111 (Istat Code of province); **j**=1,.....,27 (Functional class of the Open Street Map Arch); **k**=1,.....,4 (Type of locality: 1=Urban areas, 2=Small inhabited areas, 3=Productive areas, 4=Wide spread houses); **l**= 0,……,9 (type of road by Istat classification).

In 2017, there are about 220,000 Traffic Points on the road arches of the national road network and almost 68,000 km of carriageway, considering the extension of all the arches that contained at least one traffic point. This extension represents, in Italy, 6.3% out of the total network. The types of roads, according to the Istat classification, which have the highest percentages of traffic points, are motorways (31.7%), national roads in the built-up area (30.7%) and regional roads in the built-up area (30.2 %). With this criterion, in fact, with the same number of accidents and road kilometers extension, for a similar category of road in two different provinces, for example, the presence of PoT modifies the risk of road accidents, with disadvantage of road sections with less traffic flow, thus they result with an higher danger in terms of vehicles / km.

4. Road accidents indicators computing: main results

The proposed indicators refer to the ratio between road accidents, deaths, injuries and vehicles to the road length, considering the additional information on points of traffic too, the resident population and the vehicle fleet. The following results presented concern indicators out of the road length and the synthesis indicators[7].

The road accidents indicators out of the road length provide a measure of the number of accidents, vehicles involved, dead and injured per 100 kilometers of carriageway in the province. As regards the results, some evidences show that a maximum exposure to the risk of accidents and number of vehicles involved, for motorways and urban roads, is stated mainly in main cities. Milan records a maximum for accidents and vehicles rates per 100 km of carriageway, on Motorways (respectively 144.69 and 313.54 for the 2017 road arches classification) and Genoa on Urban roads, for accidents and vehicles per 100 km (152.91 and 256.66). On

[7] For more details: Istat Experimental Statistics https://www.istat.it/en/archivio/257384

Rural roads, in contrast, the medium-sized provinces result most affected, Monza Brianza has the maximum value.

The Road accidents indicators by road length "weighted" by the PoT (Point of Traffic) information measures the number of accidents, vehicles involved, deaths and injuries per 100 kilometers of carriageway in the province, "adjusted " by the effect on accidents of the presence of traffic points on the roads. The ratios are based on the product between the value of the indicators by road length and the complementary percentage (1-p) to the presence of arches with traffic points on the roads of the province. It is interesting to observe as often the same province presents different positions in the ranking of the indicator by road length, compared to the ranking of indicators adjusted by traffic effect. This is the case of the road accidents on the highways in Genoa and Milan, e.g, the position of Genoa, in fact, reveals a disadvantaged position, in respect of Milan, considering the real danger of the roads of the province with traffic. Finally, a group of *composed and synthetic road accidents indicators* completes the set of 2017 indicators. The method used is the arithmetic mean (z-scores) MZ method (Software Ranker[8] by Istat), (Mazziotta and Pareto, 2013). The values of indices have been calculated according to the MZ (arithmetic average Z scores)[9] method for road arches; vehicle fleet and population. This method allows calculating indicators excluding the traffic volumes on the road and to consider the cause of the accident in a different light, linked to the driving behavior and the characteristics of the infrastructures and vehicles, with or without the traffic flow influence. The results shows that the ranking of the Italian provinces, in respect of the new synthetic indicator, is modified if compared with the ranking of the single indices and underlines a new profile of accidents risk at territorial level.

5. Conclusions

Big Data are not generated directly for statistical purposes, for a statistical use, a suitable methodology must be able to link events in order to refer mainly to units of the population of interest for official statistics. Other elements required are: to process data with the aim to collect them coherently with the statistical framework, to give weights to data so as to guarantee representativeness with respect to the target population and to estimate aggregates of interest including quality measures by spatial and time dimensions. The project of the use of Open Street Map for the

[8] Istat Ranker Tool desktop software http://www.istat.it/en/tools/methods-and-it-tools/analysis-tools/ranker; I.Ranker web application https://i.ranker.istat.it

[9] The three synthesis methods evaluated are: MZ - arithmetic mean (z-scores); MR - relative indices method (IR); MPI - Mazziotta-Pareto Index method (De Muro, 2010). The criterion applied after the robustness analysis is MZ.

investigation of road accidents pattern is adherent to the features cited and the activities in progress are going to provide new developments in the future.

References:

BROCCOLI M., BRUZZONE S. 2021. *Use of the open street map to calculate indicators for road accidents on the Italian roads. Updating with 2017 data*, Istat Roma, Experimental Statistics https://www.istat.it/en/archivio/257384

FARCHI S., MOLINO N., GIORGI ROSSI P. et al. 2006. *Defining a common set of indicators to monitor road accidents in the European Union*. BMC Public Health 6, 183.

HAKKERT A.S., BRAIMAISTER L.2002. *The uses of exposure and risk in road safety studies,* SWOV Institute for Road Safety Research, Leidschendam (NL).

HARBISON F.H., MARUHNIC J., RESNICK J. R. 1970. *Quantitative Analyses of. Modernization and Development*, Princeton University Press, New Jersey.

MAZZIOTTA M., PARETO A. 2013. Methods for constructing composite indices: one for all or all for one?, *Rivista Italiana di Economia Demografia e Statistica*, Vol. 62, No. 2.

RETALLACK A.E., OSTENDORF B. 2020. Relationship between Traffic Volume and Accident Frequency at Intersections, *International Journal of Environmental Research and Public Health*, Vol. 17, No. 4.

SUMMARY

Use of open street map for accident investigation on the road and motorways networks

The aim of the paper is to provide new methods and indicators for road accidents analysis. Traditionally, the road accidents, mortality and harmfulness rates use, as standard denominators, the resident population or vehicles fleet. These measures are not always adequate to rank correctly the exposure to risk of accidents; due to neither the resident population nor the registered vehicles represent correctly the traffic units circulating on the roads. The present research proposes a new approach to the traditional measures, using the length in meters of carriageway by road direction from the Open Street Map, adding the information on the Points of Traffic too. Although the product represents an intermediate result and provides a new ranking of the road accidents risks, the final purpose is to expand the statistical information with the supply of traffic flows (vehicles/km) on the national road network, as widely used in literature, but still missing in Italy.

Marco BROCCOLI, Istat - Istituto Nazionale di Statistica broccoli@istat.it
Silvia BRUZZONE, Istat - Istituto Nazionale di Statistica bruzzone@istat.it