Investigation of accident modelling data in Greece

George Yannis, Dimitrios Nikolaou, Anastasios Dragomanovits [1]

Department of Transportation Planning and Engineering, National Technical University of Athens, 5, Heroon Polytechniou st., 15773, Athens, Greece, geyannis@central.ntua.gr
Department of Transportation Planning and Engineering, National Technical University of Athens, 5, Heroon Polytechniou st., 15773, Athens, Greece, dnikolaou@mail.ntua.gr
Department of Transportation Planning and Engineering, National Technical University of Athens, 5, Heroon Polytechniou st., 15773, Athens, Greece, dragoman@central.ntua.gr

Abstract

Accident prediction modelling is increasingly used internationally to enhance evidence-based road safety decision making. Yet, the availability of detailed and good quality data on road accidents and related casualties, infrastructure geometric characteristics (e.g. curve radius, lane width, etc.) and traffic attributes (e.g. AADT) consists a basic prerequisite not only for the development of such models but also for their efficient use by road safety practitioners. Within the above context, the aim of this paper is to investigate and discuss the availability and accuracy of accident modelling data in the primary rural road network of Greece, focusing on three types of data that are considered most critical: accident, traffic and road geometry data. With regards to accident data, a case study in the sub-region of Viotia is presented, focusing on the identification of the percentage of accidents that can be accurately geo-located and used for modelling purposes. Concerning traffic data, an exploration of the coverage of the road network by spot traffic measurements also in the sub-region of Viotia is performed and discussed. On geometric design data, the paper focuses on a section of Patras-Pyrgos two-lane highway and examines the accuracy of data that can be obtained through two common Open GIS Data Platforms ("Blender" software and SRTM data through the GPS Visualizer platform), by comparing them with the actual data retrieved from a topographic survey of the highway. From the study it can be concluded that road accident prediction modelling in the rural road network of Greece can potentially be performed only on motorways, provided that an arduous and resource-consuming process is applied to collect and code missing geometric design data. In other rural roads, the absence of traffic volume estimations and even more, of properly geo-located accident data, is an impermeable barrier to such efforts

Keywords: accident data; traffic data; road geometry data;

[1] * Corresponding author. Tel.: +30-210-3301515;
E-mail address: dragoman@central.ntua.gr

1. Introduction

Despite considerable efforts and significant progress in the field of road safety, it remains a major issue that concerns the majority of the countries globally, as road traffic injuries are the eighth leading cause of death for people of all ages. According to the World Health Organisation, it is estimated that about 1,35 million people lose their lives because of traffic accidents annually [1]. Despite the fact that road casualties presented a decreasing trend during last years, the number of road fatalities in several countries remains unacceptable and highlights the need for even greater efforts with respect to road safety [2]. A road accident results from a combination of factors related to the components of the traffic system comprising roads, vehicles and road users, and the way they interact [3].

Budgets for road safety policies and activities are not infinite. Therefore, decision makers and road safety stakeholders have to determine the optimal possible use of available funds. With regards to improvements in the existing road infrastructure, several quantitative methodologies have been developed over the years, to enhance evidence-based decision making. These methodologies include accident analyses, road safety inspections, assessment of the "in-built" safety of roads, use of Accident Prediction Models (APMs), etc. Probably the most detailed approach is offered through the application of APMs, a practice well described in AASHTO Highway Safety Manual (HSM) [4]. Yet, especially this methodology requires high quality data in order to predict accident frequency in specific road elements (segments, intersections, etc.) and produce reliable results. More specifically, the availability of detailed and good quality data on road accidents and related casualties, infrastructure geometric characteristics (e.g. curve radius, lane width, etc.) and traffic attributes (e.g. AADT) consists a basic prerequisite for this kind of modelling. [5]

Within the above context, the aim of the study is to investigate and discuss the availability and accuracy of accident modelling data in the primary rural road network of Greece, focusing on three types of data that are considered most critical: accident, traffic and road geometry data. The paper is structured as follows: the first section concerns accident data availability and presents a case study in the sub-region of Viotia for the period 2011-2015. This analysis focuses on identifying the percentage of accidents that could be accurately geo-located and used for modelling purposes. Then, concerning traffic data, an exploration of the coverage of the road network by spot traffic measurements also in the sub-region of Viotia is performed and discussed. The third section of the paper focuses on geometric design data, which are generally not readily available in official databases in Greece. The investigation focuses on a section of Patras-Pyrgos two-lane highway, and compares the data that can be obtained through two common Open GIS Data Platforms ("Blender" software and SRTM data through the GPS Visualizer platform) with the actual data retrieved from a topographic survey of the highway. Finally, in the last section of the paper the conclusions of the study are summarized.

2. Accident Data

The Hellenic Statistical Authority maintains the official road accident database in Greece. This database includes road accidents in which at least one involved road user was injured (slightly/seriously) or killed. The case study that will be presented in this section is based on road accident data collected from the Police and codified into the National Road Accident Database by the Hellenic Statistical Authority. The Department of Transportation Planning and Engineering of the National Technical University of Athens (NTUA) has access to this National Road Accident Database. More specifically, in Greece, Traffic Police officers attend the accident site and complete the road accident data in high detail in standardized templates, i.e. the Accident Data Collection Forms, immediately after the occurrence of an accident, providing information on accident conditions, as well as on characteristics related to the road, the involved persons or vehicles. The Accident Data Collection Forms are then forwarded to the Hellenic Statistical Authority, which is responsible for the final checking and codification into the official National Road Accident Database. Copy files of the National Road Accident Database are provided to the Department of Transportation Planning and Engineering of the National Technical University of Athens (NTUA) (with personal identification removed), who developed a system of efficient queries to extract any combination of data. This NTUA database consists of disaggregated data for all road injury accidents in Greece for the period 1985-2019, is updated on an annual basis, and is used for the purposes of this study.

The variables that are included in this database are presented in the following table grouped by accidents', involved road users' and vehicles' characteristics.

Table 1: Variables included in the road accident database

Accident characteristics	Year, Month, Location (geo-code), Area Type, Street number, Kilometer mark (road chainage), Kilometrage direction, Road type, Road code, Road's in junction code, Motorway(Y/N), Week of the year, Day of week, Hour, Date, Number of fatalities, Number of serious injuries, Number of slight injuries, Number of vehicles involved, Pavement type, Weather conditions, Pavement conditions, Pavement state, Night lighting, Traffic directions, Number of lanes for each direction, Direction markings, Lane markings, Left edgeline markings, Right edgeline markings, Median, Central barrier, Left side barrier, Right side barrier, Left side shoulder, Right side shoulder, Pavement width, Straight, Narrowing, Lever crossing, Right turn, Left turn, Turn alteration, Ascent, Descent, Ascent / Descent alternation, Type of accident first impact, Maneuver of vehicle A which likely contributed to the accident, Pedestrian maneuver, Traffic control / signalization, Police / Port Authority, Hit and run accident
Involved road users' characteristics	Road user type, Gender, Age (in years), Nationality, Use of protective equipment, Injury severity, Position in vehicle, Purpose of trip
Involved vehicles' characteristics	Vehicle type and usage, Vehicle plates nationality, With trailer, Vehicle capacity, 1st year of registration, Vehicle Technical inspection, Number of drivers and passengers, Type of alcohol test, Result of alcohol test, Time of alcohol test, Place where alcohol test took place, Driving license, License category, Year of acquisition, Vehicle carried dangerous goods (ADR), Overweight vehicle, Load oversized

Data for all injury road accidents in the sub-region of Viotia were considered for the five-year period 2011-2015. Firstly, a query was executed in the database in which all road accidents in the sub-region of Viotia were searched by year, by type of area, by road code, by station, by infrastructure characteristics (intersection or not, curve or not) and by type of casualties (fatalities, serious injuries, slight injuries).

These data were used to investigate in which way they could be used for microscopic modelling analysis and identify if these data are appropriate for the development of APMs. An important issue for consideration during the accident analysis is the treatment of road accidents with unknown location. In many cases of road accidents included in the database, there is no indication of the road on which the accident occurred and/or the specific location of the accident. The following table (Table 2) presents the number and the respective percentage of road accidents occurred on unknown roads during the period 2011-2015 in sub-region of Viotia. Based on Table 2, it can be observed that 51% (232/451) of total injury accidents in Viotia from 2011 to 2015 were coded as occurring on unknown road.

Table 2: Accidents with unknown road for the years 2011-2015 in the sub-region of Viotia

Year	Total Accidents	Unknown Road	Unknown Road (%)
2011	118	57	48%
2012	92	53	58%
2013	101	55	54%
2014	75	35	47%
2015	65	32	49%
Total	451	232	51%

Even for accidents on known roads, the specific location of some road accidents is unknown and is not included in the database. Table 3 demonstrates the number and the respective percentage of accidents that have occurred on known roads but there is no indication of the accident's specific location. In a further 9% (42/451), although the road code was available, the specific location (road chainage) was unknown.

Table 3: Accidents on known road and unknown station for the years 2011-2015 in sub-region of Viotia

Year	Accidents – Known Road	Known Road – Unknown Station	Known Road – Unknown Station (%)
2011	61	9	15%
2012	39	14	36%
2013	46	8	17%
2014	40	8	20%
2015	33	3	9%
Total	219	42	19%

In a more detailed level of analysis, 14 rural roads were isolated and the geo-located accidents were analyzed in order to identify whether the infrastructure characteristics as recorded in the road accident database are identical to the actual infrastructure characteristics of the site. These roads are namely E.O.03, E.O.29, E.O.44, EP.5, EP.11, EP.17, EP.21, EP.23, EP.24, EP.28, EP.30, EP.31, EP.36 and EP.37. The following table (Table 4) presents the total number of road accidents on these roads and the number of road accidents on these roads with unknown accident's specific location for the five-year period 2011-2015.

Table 4: Accidents on known and codified roads and unknown station for the years 2011-2015 in the sub-region of Viotia

Year	Accidents – Known codified Road	Known codified Road – Unknown Station	Known codified Road – Unknown Station (%)
2011	16	1	6%
2012	14	2	14%
2013	12	1	8%
2014	9	0	0%
2015	4	1	25%
Total	55	5	9%

An additional table (Table 5) was created to identify whether the infrastructure characteristics as recorded in the road accident database match to the actual basic infrastructure characteristics retrieved from Google Earth aerial imagery. It was found that the basic geometric characteristics (e.g. intersection, curve or straight segment, presence of lighting) matched in only 54% of the cases.

Table 5: Accidents on known-codified roads and accidents with identical infrastructure characteristics

Year	Accidents – Known codified Road and known Station	Matching of infrastructure characteristics (accident database and road coding)	(%)
2011-2015	50	27	54%

As a conclusion, and taking into account the results of the four previous tables (Tables 2-5), out of a total of 451 recorded accidents in the road network of Viotia, only for 177 (39%) is both the road code and the road station recorded. Furthermore, based on the detailed analysis of a sample of roads, it can be assumed that for approximately half of these accidents (46%) there are obvious discrepancies between basic geometric characteristics of the accident location, as recorded in the database, compared to Google Earth data, leading to the deduction that no more than 21% of available injury accidents data is usable for purposes of accident analysis and modelling that requires precise accident location.

On the contrary, motorway concessionaires in Greece usually maintain their own databases in which road accident data with exact location of accidents are recorded, commonly also including accidents with material damage only.

3. Traffic Data

In Greece, there is no official national database for traffic data, either traffic volumes (AADT) or traffic synthesis. Regularly updated datasets exist only for urban areas (e.g. in Athens greater area) and on toll-operated motorways. However, even these datasets are usually not openly and readily available to researchers and practitioners. Traffic data on lower class rural roads (national and/ or regional) are usually collected on a per-case basis by regional road authorities, using spot traffic counts.

As a result, the lack of traffic data is also an important obstacle in microscopic road infrastructure safety research in Greece and in many cases it actually defines the type and magnitude of research that can realistically be conducted. In order to gain an understanding of the extent of available data, a case study investigation of traffic data availability was performed in the national and regional road network of the sub-region of Viotia. Contact with the road management authority of Viotia resulted in identifying a set of spot traffic count results, covering a 12h per day period (8am to 8pm) for a period of three days: Wednesday 10/9/2014, Friday 12/9/2014 and Saturday 13/9/2014, for only four locations, combined for both directions of travel: on Thiva-Livadeia Road, Livadeia-Lamia Road, Thiva Ring Road and Elefsina-Thiva Road (Figure 1).

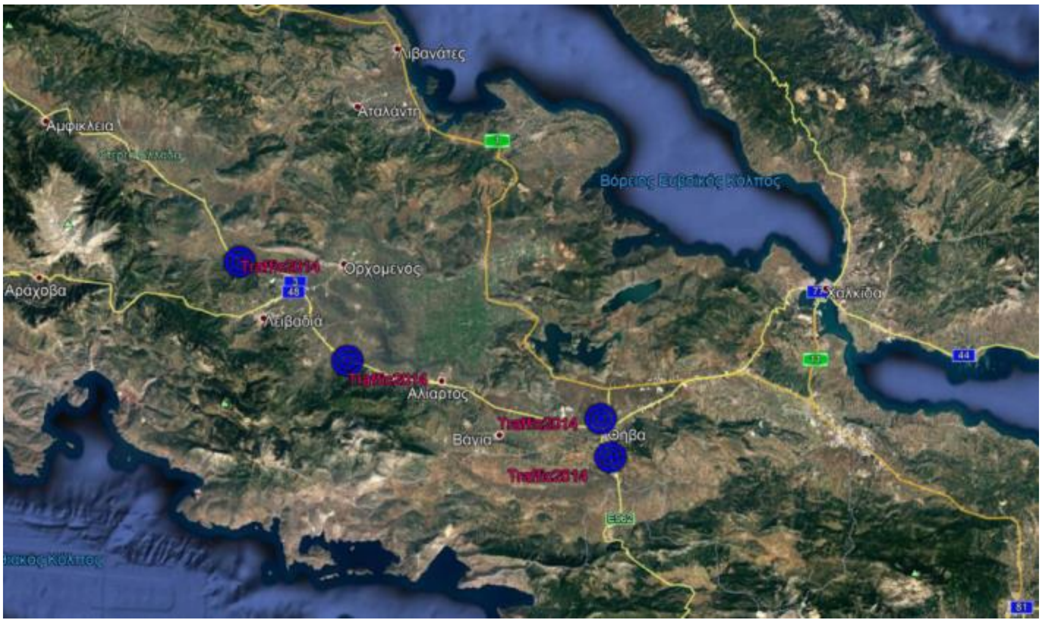

Figure 1: Locations of available traffic data in the sub region of Viotia. Source: Road management authority of Viotia sub-region - field surveys in September 2014.

It can be expected that traffic data with a similar level of detail and extent can be obtained also for other sub regions of Greece. The above traffic data could be potentially useful for road safety analyses, after suitable elaboration to estimate the Average Annual Daily Traffic (AADT). The available detailed information on traffic synthesis (passenger cars, buses, light trucks, 2-axes heavy trucks, 3-axes heavy trucks, and heavy trucks with trailers) may also provide qualitative information for the causes of road accidents during the road safety inspections. However, the data cover a very small fraction of the road network in Viotia sub-region, thus, severely limiting the scope of the analyses.

On the contrary, on toll operated motorways, toll stations data can provide a very comprehensive and detailed dataset for traffic volumes and synthesis of traffic, that are fully appropriate for road safety analysis and modelling.

4. Geometric Design Data

The development and the application of road infrastructure APMs is inherently related to the availability of data on the examined road infrastructure, including geometry (e.g. horizontal curvature, vertical curvature and slope), cross section elements (e.g. presence of central median, number of lanes, lane width, shoulder type and width, etc.), roadside conditions (e.g. distance of hazards, road safety barriers, etc.) and other road features and equipment (e.g. rumble strips, condition of markings and signs, road lighting, etc.). Not all types of road infrastructure data are necessary at all times; the selection of the parameters that need to be considered as independent variables in the models is probably the most critical decision that affects the robustness of the approach.

4.1 Potential data sources

Potential road geometric design data sources commonly include:

- National Road Authorities Databases: Road infrastructure and road design data are commonly collected and maintained in the asset management databases of National Road Authorities (NRAs). In Greece however, the road registry maintained by the Ministry of Infrastructure and Transport includes mostly administrative data and there is no road geometry database exists with sufficient detail to be able to provide useful and meaningful data for road infrastructure analyses.

- Data from vehicle mounted cameras and road survey vehicles: Vehicle mounted cameras can be used for surveys of road infrastructure: a road is recorded in high resolution while driving at a constant speed appropriate for recording. Weather condition for this type of survey should be ideal, and it is typically performed during the day. The primary purpose of this type of survey is to collect geo-referenced images of road segments which can be used for road attribute coding. Furthermore, equipping the vehicle with various sensors enhances data collection and analysis of multiple data types such as road element data, operating data, and traffic volume data.

  An extensive use of such a road infrastructure data collection methodology took place during the period 2012-2015 by Egnatia Odos SA, in the framework of the Greek Road Rehabilitation and Safety Project. A large part of the national and regional rural road network of Greece (excluding motorways) was surveyed, including 4.200 km of national roads and 10.800 km of regional roads, covering the 13 regions of the country, in order to identify potential sections for road rehabilitation and safety works. In the data collection phase of the above project, vehicle mounted video cameras were used in conjunction with GPS and georeferenced AutoCAD drawings were developed with the horizontal and vertical alignment of the examined roads and the respective road station. Satellite images were also used as a background of the horizontal alignment drawings. Using these drawings and the video footage, the following data were collected and coded in databases, on the basis of the start/ end station: road gutter, drainage ditch, pavement width, unsealed shoulders, high embankments, high cuts, additional traffic lanes, medians, sidewalks, technical works (culverts, bridges, etc.), traffic signs, road safety barriers, delineators, lighting posts, other posts, at-grade intersections, interchanges, access facilities, pavement deficits, bus stops, etc.

  These data are adequately detailed and appropriate for road infrastructure analysis; yet they are somewhat outdated as road improvements have already taken place in some locations.

- Data from High Definition (HD) maps: A high-definition map (HD map) is a highly accurate 3D map containing details not normally present on traditional maps. Such maps can be precise at a centimeter level. HD maps are captured using an array of sensors, such as LiDARs, radars, digital cameras and GPS. HD maps can also be constructed using aerial imagery. High-definition maps usually include map elements such as road shape, road marking, traffic signs and barriers. An example of HD mapping suppliers includes TomTom, Here, Navtech, MobilEyE etc.

- Open GIS road geometry data: A series of online utilities provides coordinates along the road network of many countries, including Greece. In order to investigate the potential and accuracy of Open GIS Data in effectively describing road geometry (horizontal elements and elevations) a pilot assessment study was performed as presented in the following paragraphs.

4.2 Pilot evaluation of Open GIS road geometry data

Data extraction and assessment was based on comparing road geometry data retrieved from OPEN GIS sources to the actual data for the road axis of Patra-Pyrgos National Road in the area "Vrachneika", as derived from a detailed topographic survey at scale 1: 500 (Figure 2).

The investigation included the use of:

- Blender software (free software available at: https://www.blender.org/) with GIS tracking of road data, and
- the GPS Visualizer platform that retrieves data from the Shuttle Radar Topography Mission (SRTM) database - same to Open Street Maps (OSM) data that is also accessible via API.

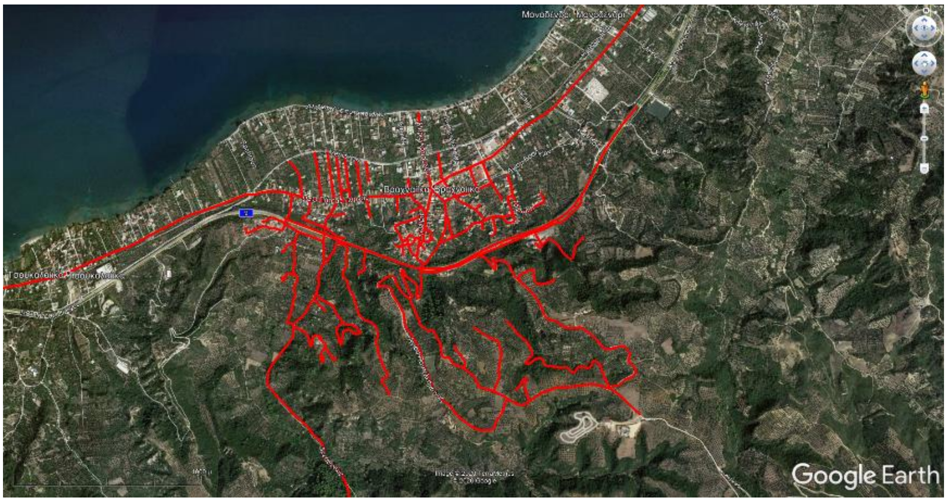

Figure 2: Blender and GPS Visualizer data assessment area.

4.2.1 Blender Software

Blender (https://www.blender.org/) is a free software released under the GNU General Public License. It supports the entirety of the 3D pipeline—modeling, rigging, animation, simulation, rendering, compositing and motion tracking, video editing and 2D animation pipeline. Using the add-on ‘’Blender GIS’’ (https://github.com/domlysz/BlenderGIS), Blender software can retrieve and process geographic information in standard GIS file formats e.g. shapefile vector, raster image, geotiff DEM, OpenStreetMap xml.

The steps the research team followed using the Blender software were:

1. Using the add-on, the area where the topographic survey was made, was visually identified and the background map was retrieved (GIS tab → webgeodata → Basemap, source google – satellite level).
2. Digital model was retrieved (GIS tab → webgeodata → Get SRTM)
3. Open street Map (OSM) data was saved for the existing roads (GIS tab → webgeodata → Get OSM, highway level).
4. The highway level information was extracted (*.shp file) in the form of lines with elevation data (GIS tab → Export ,feature: line).
5. Import of the *.shp file in Autocad software and comparison of the elevations of the imported lines to the topographic survey elevations.

4.2.2 GPS Visualizer platform and Shuttle Radar Topography Mission database

GPS Visualizer is an online utility that creates maps and profiles from geographic data (https://www.gpsvisualizer.com/). It is free and easy to use, yet powerful and extremely customizable. Input can be in the form of GPS data (tracks and waypoints), driving routes, street addresses, or simple coordinates.

The elevations of the same eight sampling points considered for the Blender software were also estimated by using the DEM database (https://www.gpsvisualizer.com/elevation). The procedure of converting the Autocad points (*.dwg file) to kml/kmz files was the following:

1. Export of the Autocad points to a *.shp file.
2. Import of the *.shp file in Google Earth and then export as *.kml file.
3. Import of the *.kml file in DEM database of GPS Visualizer site (www.gpsvisualizer.com).
4. Export in *.txt file.

4.2.3 Comparison of Open GIS Data to topographic survey data

The data extracted from Blender software and from GPS Visualizer platform were compared against the respective points on the topographic survey, with regards to their elevation as shown in Tables 6 and 7 that follow.

Table 6: Accuracy assessment of road centerline points - Blender software

Point no.	X (Easting)	Y (Northing)	Elevation (Blender)	Elevation (Survey)	Difference in elevations (m)
1	294999,85	4225789,11	29,18	25,98	3,20
2	295066,33	4225763,10	42,33	43,17	0,84
3	295230,16	4225760,95	48,94	49,60	2,36
4	295506,68	4225736,57	49,35	47,40	2,34
5	295867,39	4225772,21	68,33	71,20	2,94
6	295901,81	4225838,99	66,06	64,40	4,56
7	295917,74	4225759,28	82,82	87,10	13,55
8	296081,10	4225921,02	58,82	56,80	0,82

Table 7: Accuracy assessment of road centerline points - GPS Visualizer platform

Point no.	Latitude	Longitude	Elevation (GPS Visualizer)	Elevation (Survey)	Difference in elevations (m)
1	38,15933492	21,661877543	28,90	25,98	2,92
2	38,15911579	21,662643230	43,50	43,17	0,33
3	38,15913362	21,664512330	49,60	49,60	1,70
4	38,15897678	21,667673047	47,40	47,40	0,39
5	38,15937938	21,671776748	71,20	71,20	0,07
6	38,15998853	21,672150183	64,40	64,40	2,90
7	38,15927434	21,672354699	87,10	87,10	9,27
8	38,16076776	21,674171548	56,80	56,80	1,20

From the above analysis it is evident that no accurate information for vertical alignment and road elevations can be collected from both Open GIS data sources (Blender and GPS Visualizer). Specifically, street surface elevations obtained from Open GIS applications have very large deviations, both between applications (e.g. Blender data compared to GPS Visualizer data) and (more importantly) when compared to actual surveyed elevations. In more than half of the randomly selected examined points (6 out of 8 for Blender data and 5 out of 8 for GPS Visualizer data), elevation differences from the survey exceed 1m. The problem seems to be is intensified in cases where the road is at a cut or fill section of considerable height (e.g. point 7), where differences up to 13,5m were observed.

On the other hand, with regards to the horizontal alignment, qualitative evaluation of data for the road centerline location from both Blender and GPS Visualizer reveals small differences compared to the surveyed road centerline and these data can potentially be used to build a road geometry database for the purpose of road safety analyses.

5. Conclusion

Based on the results of the above pilot studies, for non-motorway rural roads in Greece the absence of traffic volume information and even more, of properly geo-located accident data, seems to be an impermeable obstacle for accident prediction modelling efforts. Specifically, for non-motorway rural roads it was found that:

- approximately 80% of the injury accidents recorded in the official National Road Accident Database has either missing or obviously inaccurate accident location information,
- existing traffic volume data on the rural road network are largely unavailable and derived from scarce spot counts performed several years ago,
- geometric (road design) data are available only in the deliverables of the Greek Road Rehabilitation and Safety Project performed on behalf of Egnatia Odos SA in 2012-2015. Limited data can be retrieved from Open GIS sources, keeping however in mind that elevation information is largely inaccurate.

On the other hand road accident prediction modelling can potentially be performed on motorways, using accident and traffic data maintained by road operators, provided that an arduous and resource-consuming process is applied to collect and code missing geometric design data.

The present study however exhibits certain limitations that need to be considered along with the findings and conclusions. Firstly, the study focuses only on the rural road network; availability and accuracy of road safety data for urban roads may significantly differ and is outside the scope of the present study. Secondly, both the investigation of crash data reliability and location information with regards to the official Greek National Road Accident Database and of traffic data availability, is focused on a single prefecture of Greece. Although this prefecture (Viotia) is considered quite representative of average conditions, it may be true that different conditions may prevail in other prefectures (particularly in island prefectures). Finally, with regards to traffic and geometry data on motorways, a single road axis (Olympia Odos) has been examined. Therefore, further research is required to examine on a wider scope road safety data availability and accuracy in Greece, in order to fully realize current status in higher detail.

Acknowledgment

This research was performed within the research project i-safemodels, been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the Call "Bilateral and Multilateral R&T Cooperation between Greece and China" (project code: T7ΔΚΙ-00253).

References

1. WHO (2018). Global Status Report on Road Safety 2018. World Health Organization, Geneva (WHO/NMH/NVI/18.20)
2. Organisation for Economic Co-operation and Development (2008). European Conference of Ministers of Transport. Speed Management. Paris, OECD.
3. Haddon Jr, W. (1980). Advances in the epidemiology of injuries as a basis for public policy. Public health reports, 95(5), 411.
4. AASHTO (American Association of State and Highway Transportation Officials) (2010) Highway Safety Manual, 1st edn. AASHTO Washington, DC, USA.
5. Ambros, J., Jurewicz, C., Turner, S., & Kieć, M. (2018). An international review of challenges and opportunities in development and use of crash prediction models. European transport research review, 10(2), 1-10.