Extraction of Naturalistic Driving Patterns with Geographic Information Systems

José Balsa-Barreiro [1] 📧 · Pedro M. Valero-Mora [2] · Mónica Menéndez [3] · Rashid Mehmood [4]

Abstract

A better understanding of *Driving Patterns* and their relationship with geographical driving areas could bring great benefits for smart cities, including the identification of good driving practices for saving fuel and reducing carbon emissions and accidents. The process of extracting driving patterns can be challenging due to issues such as the collection of valid data, clustering of population groups, and definition of similar behaviors. *Naturalistic Driving* methods provide a solution by allowing the collection of exhaustive datasets in quantitative and qualitative terms. However, exploiting and analyzing these datasets is complex and resource-intensive. Moreover, most of the previous studies, have constrained the great potential of naturalistic driving datasets to very specific situations, events, and/or road sections. In this paper, we propose a novel methodology for extracting driving patterns from naturalistic driving data, even from small population samples. We use Geographic Information Systems (GIS), so we can evaluate drivers' behavior and reactions to certain events or road sections, and compare across situations using different spatial scales. To that end, we analyze some kinematic parameters such as speeds, acceleration, braking, and other forces that define a driving attitude. Our method favors an adequate mapping of complete datasets enabling us to achieve a comprehensive perspective of driving performance.

Keywords Big data · Driving patterns · Driving behavior · Geographic information systems · Naturalistic driving · Smart cities

1 Introduction

Smart cities and societies are driven by our ever-growing desires for continual innovations and improvements in every aspect of our life [1, 2]. Transportation, which is the backbone of modern societies, has also been undergoing this continual innovation and improvement process [3]. The environmental, economic, social, and health-related damages caused by transportation are well-known, and demand innovative solutions. Such solutions, in turn, require new methods for modeling and analyzing various aspects of our transportation systems. Understanding driving behavior is one such area that could bring massive environmental, economic, and social improvements.

Driving behavior can be parameterized by defining patterns. Analyzing the patterns allows us to establish different driver groups and vehicle driving models, which can be compared with each other. However, extracting discrete driving patterns from raw data is not trivial due to a large number of elements and factors that may potentially influence our driving performance. Additionally, traditional experimentation methodologies present serious limitations in terms of the volume of data, the number of factors considered, and even in the quality and reliability of the collected data. A detailed literature review of studies related to the extraction of driving patterns is presented in Section 2.

Naturalistic driving (henceforth ND) has great potential for the extraction of driving patterns. This experimental method is based on an exhaustive data collection that aims to characterize the driving behavior of people in real-world situations [4]. For this, it is capable of incorporating most of the factors involved in the driving performance by using a large array of sensors and video cameras inconspicuously installed on-board the vehicle. Data collection in ND trials is based on

📧 José Balsa-Barreiro
jose.balsa@ivt.baug.ethz.ch

[1] Institute for Transport Planning and Systems (IVT), Swiss Federal Institute of Technology, 8093 Zürich, Switzerland

[2] University Research Institute on Traffic and Road Safety (INTRAS), University of Valencia, 46022 Valencia, Spain

[3] Division of Engineering, New York University Abu Dhabi (NYUAD), 129188 Abu Dhabi, United Arab Emirates

[4] High Performance Computing Center, King Abdulaziz University, Jeddah 21589, Saudi Arabia

massive and blind strategies. They are considered massive because they collect (at least theoretically) the largest amount of data related to driving performance. They are considered blind because the data collection is rarely focused on an isolated factor. These two characteristics largely explain why ND data is highly versatile for research, and it allows the evaluation of the different factors that are part of the whole driving experience, i.e., vehicle, road infrastructure, surrounding environment, and driver [5, 6]. Overall, ND based studies offer two main advantages compared to more traditional driving analysis methodologies: (a) the experimental process is unconditioned because the research staff does not interfere in the experiments (at least, theoretically [7]), and (b) it allows registering most of the parameters and indicators that might potentially influence driving performance.

A significant number of ND experiments have been conducted at different scales for different purposes. The most important ND studies are the 100-cars experiment [8] and the SHRP-2 NDS [9] in the United States, and PROLOGUE [10] and UDRIVE [11] in Europe. So far, ND researchers have mainly focused on specific issues related to driving performance, such as the use of mobile devices [12, 13], the frequency of secondary tasks [14], the emergence of any distraction [15], or the effects of anger and other moods on driving performance [16], among others. Usually, most of these studies select some data samples by applying data thinning strategies or, alternatively, limit their focus to some specific situations by restricting the study area.

However, the high potential related to the ND datasets, especially when these data come from large experiments, still poses major research challenges due to the complexity of exploiting these kinds of data. In fact, ND typically involves huge data volumes coming from a high number of instruments, devices, and sensors installed on-board the experimental vehicles. These data are registered at high (or very high) temporal rates for long periods of time. In consequence, ND studies have large requirements in terms of data collection, data handling, and data analysis, combined with the challenges associated with data accuracy in terms of missing data, gaps, and/or anomalous values in the final datasets [17, 18]. Therefore, in order to fully realize the ND potential, there is a critical need for developing a solid strategy for ND research that systematically manages all ND stages, from data preprocessing to analysis.

To address this, we have developed a range of ND methods over the years. In [19, 20], we implemented a methodology for estimating the geoposition of an experimental car in areas where the GPS instrument ceased to function. In [21, 22], we developed a reliable method for checking the quality of this type of data, which allowed us to reduce the number of *false positives* and detect events and incidents more efficiently than with other methods. In [23, 24], we implemented and discussed strategies for mapping kinematic data related to

driving performance by using *Geographic Information Systems* (henceforth GIS).

In this paper, we develop a multi-scale focus where the driving behavior is analyzed as a whole set of factors under different circumstances. Driving behavior is parametrized by means of driving patterns based on very fine-grained data obtained in an ND experiment. At a macro spatial scale, these driving patterns are mainly extracted from kinematic parameters related to driving performance, i.e. acceleration forces and speeds. At a micro spatial scale, we analyze the driving performance in specific road sections by considering some additional indicators such as the engine speed, the braking and acceleration forces, in addition to the position of the gearshift lever. The combination of the two spatial scales for ND analysis allows us to achieve a smarter and more comprehensive perspective of driving patterns compared to other studies focusing on a single spatial scale. Moreover, we exploit multiple features of GIS mapping, a tool rarely considered in previous ND studies. For the macro-scale analysis, we plot histograms and bar charts using software tools such as Microsoft Excel, R Studio, and some statistical software including Statgraphics. For the micro-scale analysis, data are basically represented in the form of maps. These are produced by using the software *ArcGIS* from ESRI.

The contributions of this paper are twofold. First, we present a methodology for extracting driving patterns in different scales with the aid of GIS tools. Second, we demonstrate the great potential of ND data for the extraction of driving patterns. This paper presents a step beyond the traditional analysis of ND data in that, an example is presented to identify, extract, and analyze driving patterns from ND datasets. In order to do that, we have proposed an efficient method for the visualization and extraction of this type of data based on GIS mapping. A better understanding of driving patterns and their relationship with geographical driving areas could bring great benefits for smart cities, including the identification of good driving practices for saving fuel and reducing carbon emissions and accidents.

The rest of this paper is structured as follows. Section 2 presents the background material and reviews the relevant literature. Section 3 describes the data and study area. Section 4 describes the methodology. Section 5 presents an analysis of the results. Finally, Section 6 concludes the paper with an analysis of the most representative results presented in this paper and several directions for future work.

2 Background and literature review

Driving patterns refer to models obtained from the parameterization of performance and behavior while we drive. These patterns have been studied extensively in the past, mainly in relation to fuel consumption and emissions [25]. Most of the

previous studies on this topic based their experimentation set up on one car following another. However, their scope was limited mainly due to ethical and practical reasons, among others. Although some technological advances (i.e. forward-looking lasers) made this type of studies more feasible, its range of applications was rather limited before GPS data was prevalent [26]. In recent years, however, the availability of instrumented cars has made it possible to collect such information in a non-intrusive, comprehensive, and ethically acceptable way. Consequently, nowadays we have more means for studying this topic.

The analysis of common habits allows us to determine general trends related, for example, to a certain group of people. There are multiple approaches for studying driving patterns, mostly considering the dynamic parameters related to driving performance. For example, Egea [27] evaluated driving behavior by checking the cognitive perception of drivers, in addition to some response factors. Delgado et al. [28] proposed a study in the opposite direction, testing how driving influences the physical posture of the driver.

Driving patterns are usually inferred from kinematic parameters. Thus, some scholars have evaluated factors such as the driving efficiency, anomalous behaviors, and the success of some driving license programs. Within the first group, Bratt and Ericsson [29] evaluated the relationship between some driving parameters, fuel consumption, and emission of pollutants for different types of vehicles. Berry [30] evaluated how aggressive driving increases the fuel consumption. Drivers' aggressiveness was categorized according to the speed and other characteristics of the vehicle. This study concluded that to reduce fuel consumption levels, more aggressive drivers should focus on reducing acceleration maneuvers, while less aggressive drivers should drive at lower speeds.

Some recent studies have focused on how electric vehicles influence driving behavior. Pasaoglu et al. [31] and Thiel et al. [32] analyzed the driving patterns associated with several people for improving the driving performance of electric vehicles. Karabasoglu and Michalek [33] compared different driving patterns in electric vehicles by assessing both economic and environmental benefits in comparison with conventional vehicles. They found that, in urban areas such as New York City, electric vehicles reduced the level of emissions by 60% and costs by 20%, whereas significant reductions were barely achieved on high-capacity roads. Fontaras et al. [34] estimated the reduction in fuel consumption for hybrid vehicles was between 40% and 60% compared to conventional vehicles. These levels were even higher in urban environments where vehicles ran at low speeds and had multiple stops (*stop-and-go*) due to traffic congestion. This last study also demonstrated that, at speeds above 95 km per hour, fuel consumption levels were similar in both hybrid and conventional vehicles. Karabasoglu and Michalek [33] also concluded that the road type and the heterogeneity of drivers had a significant influence on the levels of reduction, both in terms of costs and emissions. On the other hand, Sharer et al. [35] showed that hybrid vehicles were much more sensitive to aggressive driving patterns compared to conventional vehicles.

A large number of factors may influence the driving behavior. Some studies have incorporated spatial aspects by assessing different environments and scales. Ericsson et al. [36] studied driving patterns and their emissions in different European environments and cities such as Naples (Italy), Budapest (Hungary), and Malmö (Sweden). Their experiment comprised two parts. The first one defined the theoretical driving patterns for the different cities and roads. The second one evaluated the variability of these patterns based on factors such as the presence of signalized crossings, characteristics, and functionality of the road. Their results determined how driving behavior depends on the environment, showing relevant differences between road types and cities.

Other researchers have analyzed the relation between weather and different driving patterns. Sabback and Mann [37] evaluated the influence of weather on elderly drivers. For that, 40 participants in New York and Florida were involved. The results showed that 60% of the participants in New York altered their driving behavior in winter, while the percentage in Florida was only 20%. James and Goldman [38], and Evans and Rothery [39] just a few years later, analyzed driving patterns in road sections near to signposted crossings, i.e. in the so-called *dilemma zones*. For this study, they considered aspects related to the drivers, such as their gender and whether they were accompanied. With regard to the vehicle, they defined three categories: (a) small, (b) family size, and (c) big size including trucks and buses. Driving performance was evaluated under certain weather conditions at different daytimes. The results of their study showed driving patterns for the different vehicle's categories depending on both weather conditions (basically rainfall) and daytime. This same study also demonstrated how people who drove family size vehicles were the most cautious drivers, especially when they were accompanied. This demonstrates that the vehicle type also influences the driving patterns, not only in terms of speeds but also in the type and duration of the journeys. This was also shown in the CABLED electric driving project carried out in some cities in the UK between 2009 and 2012 [40].

Some other studies have focused on the influence of socio-demographic factors, such as the gender or the age of drivers [41]. In relation to the gender, some studies showed that women tend to drive less at night [42] and to stop more times on the route [43]. In relation to the age, Kington et al. [44] evaluated the influence of some socio-demographic and health-related factors on drivers over 50 years old. Marshall et al. [45] analyzed driving patterns of people over 70 years old by using electronic positioning devices. Hildebrand et al. [46] analyzed the driving patterns and accident rates of elderly people in rural

areas to assess whether they should undergo special programs for renewing their driving licenses. Lotan and Toledo [47] analyzed the driving performance of novice drivers in Israel. Their study showed how these drivers radically changed their driving performance during the apprenticeship program. Williams [48] made an extensive literature review related to driving patterns of young people to determine what factors could explain an increased accident risk. He found that the accident risk was exponentially increased for young drivers at nighttime hours, especially when they were accompanied.

A few other studies have analyzed how health conditions affect driving performance. Fonda et al. [49] evaluated the influence of depressive symptoms in elderly drivers by checking any cessation (or reduction) of the driving activity at any time. Van Landingham et al. [50] conducted a similar research, although focused on patients with glaucoma. Their results showed that 23% of glaucoma patients and 6.9% of suspected cases had experienced a sudden cessation of their driving activity at some moment.

In contrast with the literature described above, in this paper, we focus on developing a methodology for analyzing driving patterns given a data set, rather than describing the driving patterns of a particular group of people. As such, the sample for the experiment that we use, is very small in quantitative terms but quite relevant in qualitative terms (a high number of driving hours with significant data quality – more details are given in Section 3). Notice, however, that the sample size is an important feature of any empirical study in which the goal is to make inferences about a population from that sample. In practice, it is determined considering the trade-offs between cost, time, or convenience of collecting the data, and the need for it to offer sufficient statistical power (representativeness of such data). The latter should depend on the size of the population that will be studied (for example, novice drivers in rural areas, truckers, people with a visual disability, etc.).

3 Data and study area

The data presented in this paper were extracted from PROLOGUE, one of the most ambitious ND experiments carried out in Europe [10]. In Spain, a small pilot study of PROLOGUE was conducted by the INTRAS (*University Research Institute on Traffic and Road Safety*) [51]. The objective of this study was to assess the influence of *In-Vehicle Information Systems* (IVIS) on driving performance. The Spanish experiment was carried out in the surroundings of the city of Valencia, the third most populated city in Spain, during June and July in 2010. The study area was a road section of the V-21 motorway, between the north edge of the city of Valencia and the town of Puzol (Fig. 1). Its length was 15.9 km and it was traveled in both directions being the departure time between 8:00 am and 9:00 am. This same route

was already introduced in previous articles by the same authors [19–24]. The traffic conditions were optimal, with hardly any traffic jams. The trial took place under optimal weather conditions, without any rainfall. The road used in this experiment was mostly straight, although here we subdivide it into four sections based on the same number of curves. The first one is the closest to the city of Valencia on the outbound route, which is precisely the most curved section.

The experiment included five drivers, who participated each over a period of four days. The data for one of the drivers could not be properly registered, so we decided to analyze only the data related to the other four. Those four drivers included two men (A and B) and two women (C and D), all of them middle-aged, between 43 and 45 years old. The group of men had more driving experience. According to previous questionnaires, both men had traveled more than 15,000 km per year, being driver A the most experienced one with more than 30,000 km per year. The group of women were less experienced, with one of them (driver D) driving less than 5000 km per year [51]. Due to the low representativeness of our sample, this paper will be focused on developing a valuable methodology to extract and analyze driving patterns from ND data, rather than producing an exhaustive study of the driving performance of the people involved in this experiment.

The experimental vehicle was a highly instrumented car (HIC) and the drivers were aware of the location of a great number of devices and instruments. They all knew they were being recorded as they signed in advance a scientific and ethical agreement related to the purpose of the experiment and the use of the data. We acknowledge this could have influenced their actual behavior when driving [7]. However, as stated above, in this paper we focus on the methodology to analyze the collected data, not necessarily on the results of this specific experiment. For more information on the whole experiment, the interested reader can refer to Valero-Mora et al. [51].

The people drove for approximately two hours per day. Overall, around 80 parameters related to driving performance were continuously recorded using different instruments, sensors, and devices. Kinematic data related to speed, acceleration, and braking forces were recorded at very high frequencies. Furthermore, a set of cameras installed on-board recorded conditions both inside and outside the vehicle. This allowed to check actions such as how often the driver interacted with any device and what was happening in the road at the same time. The respective indicators and parameters were recorded at different temporal frequencies, depending on the technical specifications of each instrument. Indicators related to kinematics were recorded at very high temporal frequencies, between 10^{-2} s and 10^{-3} s. A more detailed description of the instruments, sensors, and devices is shown in Valero-Mora et al. [51].

In this paper, we will analyze in detail the driving performance in two small road sections, which corresponds to the
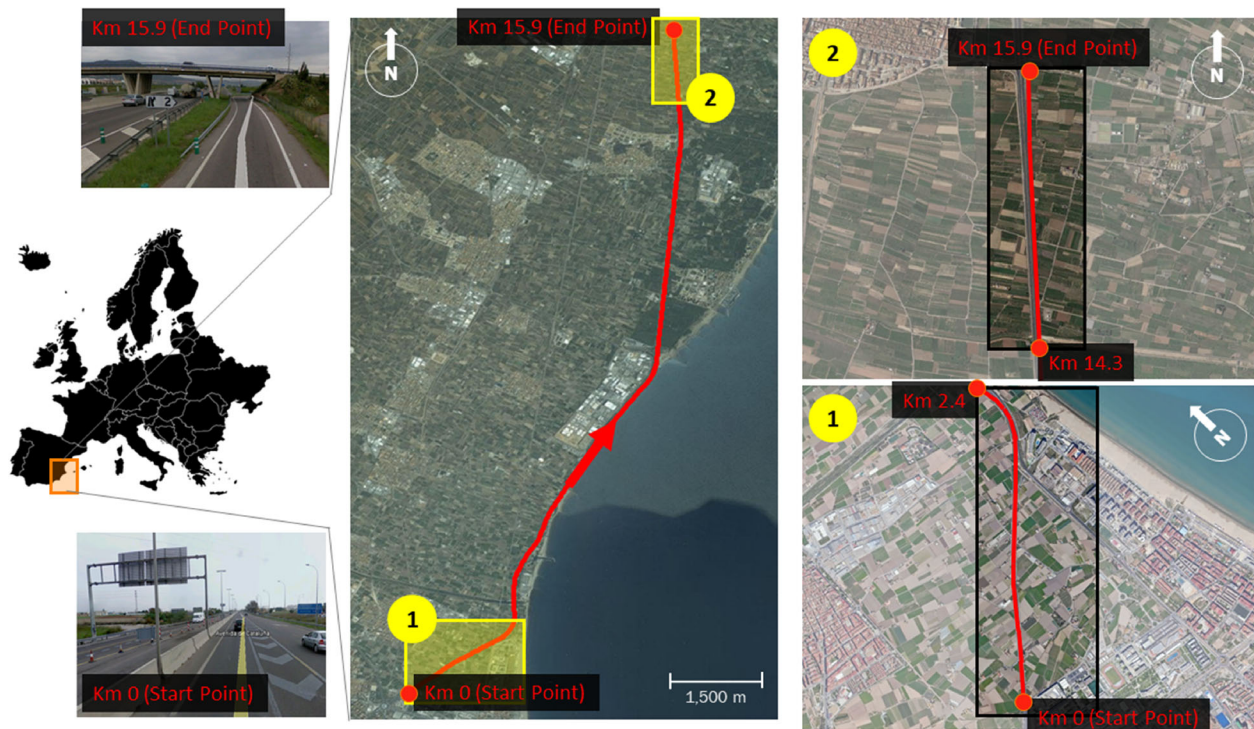
Fig. 1 Illustration of the whole study area. In the subfigure located in the center, the red line corresponds to the complete road between the city of Valencia (Km 0), the southernmost point, and the town of Puzol (km 15.9), the northernmost point. The maps drawn in the section of Results represent the driving performance in the road sections enclosed by the yellow boxes. A more detailed view of these study areas is shown in the subfigures on the right side

entry and exit of the experimental vehicles in the motorway. These study areas are represented by yellow boxes in Fig. 1. The first study area (box 1) focuses on the initial road section, which is just located in the north edge of the city of Valencia. This road section is mainly straight until the emergence of a wide-open curve in the northern direction. It has an approximate length of 2.3 km, and the curve appears at 1.9 km from the starting point. Within this road section, the experimental car enters the motorway, where it is theoretically expected to progressively increase its speed. The second study area (box 2) corresponds to the end road section, near to the town of Puzol, where the experimental car exits the motorway. With a length of about 1.6 km, this road path is mostly straight. In this road section, the drivers take the deceleration lane, which corresponds to the last 400 m of the road path. Here, the experimental car is expected to progressively reduce its speed.

4 Methodology

In this paper, we introduce a novel methodology for extracting driving patterns based on the kinematic data related to four people. The main indicator for defining driving behavior is the vehicle speed. Similar studies consider the *average speed*, a value estimated by dividing the distance traveled by the time invested in doing so. In addition to this parameter, we use the *instantaneous speed* registered by the speedometer installed in the car. Moreover, here we analyze additional kinematic parameters such as the *acceleration* or the *engine speed*, as well as the *position of the accelerator* and *brake pedals* at any time.

The extraction of driving patterns is perfomed by checking the driving performance at macro and micro spatial scales. In the first case, we mostly use indicators related to average and instantaneous speeds, which are shown in bar charts. Average speeds show a simple and fixed perspective of the whole dataset. This allows us to define a very global perspective of how someone really drives. The values of instantaneous speeds add more information by observing the driving behavior along the road section. These values are represented by linear plots where the X-axis corresponds to the complete distance traveled. Likewise, GIS mapping of a relevant number of indicators is performed in specific road sections (Fig. 2). GIS mapping allows us to define speed profiles by considering the spatial location of the vehicle at any time. At a macro spatial scale, we evaluate road sections where drivers exceed the maximum and minimum legal speeds. At a micro spatial scale, we analyze the acceleration forces in specific road sections. As a concrete example, we make an exhaustive
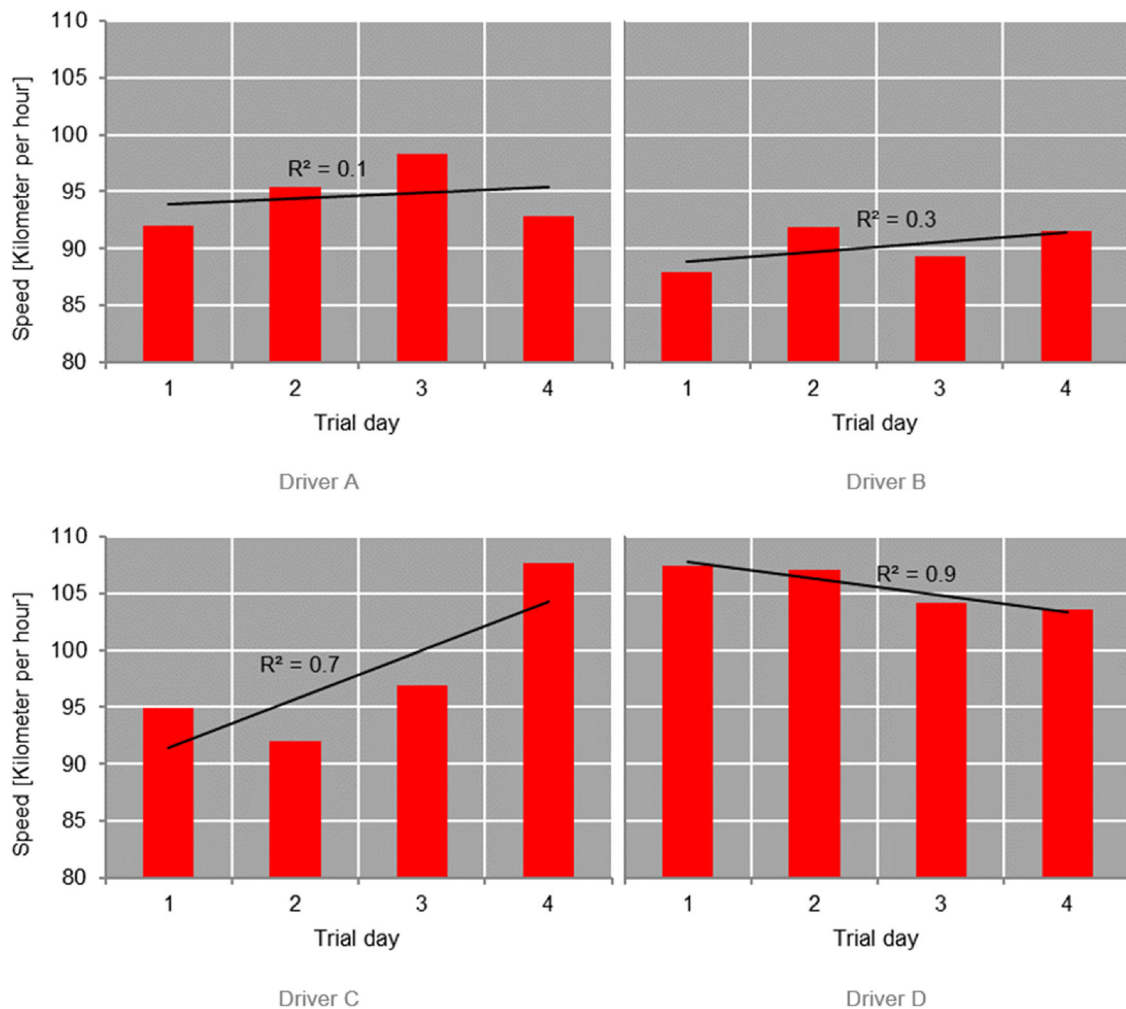
Fig. 2 Average speeds for every driver on each day of the trial. The black line shows the regression (R^2) of average speeds over time

evaluation of the driving performance for one of the drives by checking all the maneuvers and actions he conducted while he tried to enter and exit the route.

Mapping with GIS systems uses three basic and simple entities, i.e. points, lines, and polygons. However, to plot massive data such as that collected in this experiment one must apply some pre-processing steps. Initially, the route line must be digitized such an array of nodes that are spatially equidistant, which simplifies and defines the complete route. The values of each indicator in any area must be spatially interpolated for assigning values to these nodes [19].

The kinematic data registered and the working methodology implemented in this paper allow us to define and extract driving patterns. Understanding these patterns help us simplify the complex driving behavior in real-world conditions, which consider the surrounding traffic environment at any time and anywhere. Although our dataset is very reduced having a small group of drivers, our working methodology can be implemented for extracting comprehensive driving patterns in much larger populations as well.

5 Results

First, we present the results obtained at a macro spatial scale. The analysis of the average speed values shows how the driver D had a more regular driving performance, although he was also the fastest, always exceeding 100 km per hour. Checking his driving behavior over time, we can see that he reduced his speed over the days of the experiment. In contrast, driver B was the slowest one, while drivers A and C showed similar average speeds. Drivers A, B and C showed a more irregular driving performance in comparison with driver D (Fig. 2). A more detailed view of their driving performance is shown in Fig. 3. This chart represents the percentage of time that each person spent driving within a specific range of speeds. These speed values are grouped into a series of intervals. Two of these intervals show values outside the speed limits set by the national law, i.e. slower than 60 km per hour and faster than 120 km per hour. The mapping of the speed values for the respective drivers ratify that driver D was the fastest during the
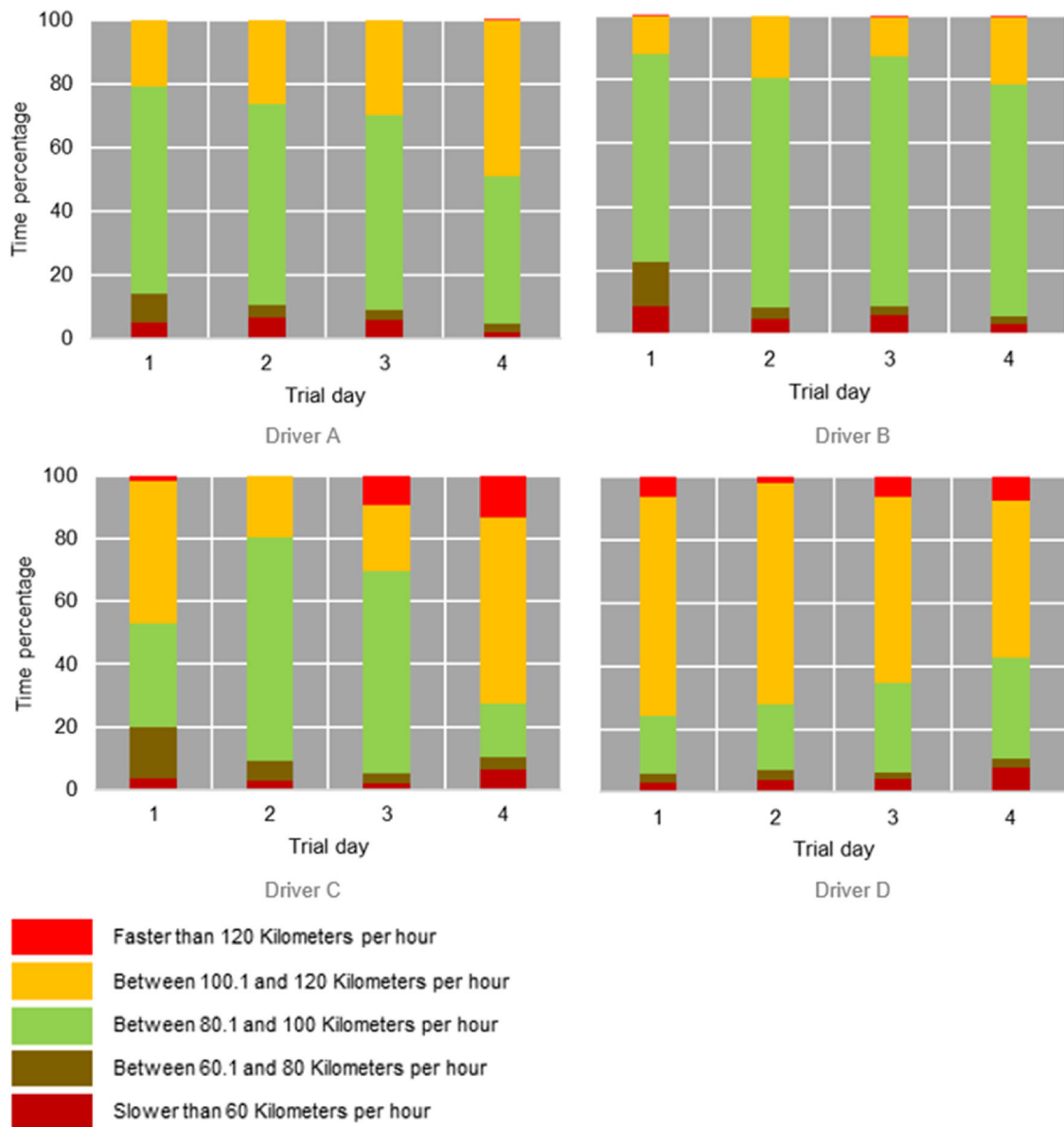
Fig. 3 Distribution of total time that drivers spent traveling at different speeds. The first and the last intervals show speed values below or above the allowed limits established by the Spanish law

first three days. On the fourth day, driver C becomes the fastest, exceeding the speed limit in multiple occasions.

The profiles drawn by speed values recorded by speedometers (i.e., *instantaneous speeds*) show a more detailed representation of the actual driving performance. These values over time are depicted in Figs. 4 and 5 for each driver, being the X-axis the traveled distance. Figure 4 shows these values for each day of the experiment, while Fig. 5 shows the average across the four days (with a solid black line). Figure 5 also includes the deviation and variability of this driving performance. It is estimated in each point of the route by considering the difference (a) between the maximum and minimum speed values (area in yellow), and (b)

adding/subtracting the standard deviation values (dashed red lines). The results show how drivers C and D had a more irregular driving performance in many road sections.

Although this perspective is valuable, it does not consider the route itself. This is precisely one of the main advantages of GIS mapping, which allows us to relate these driving performance observations to the different characteristics of the road. Figure 6 shows the specific road locations where the drivers used extreme speeds, in particular those speeds outside the limits set by the national law.

Only two of the drivers, C and D, exceeded the maximum speeds set by Spanish law at any time. They did so in three of the four days of the trial.
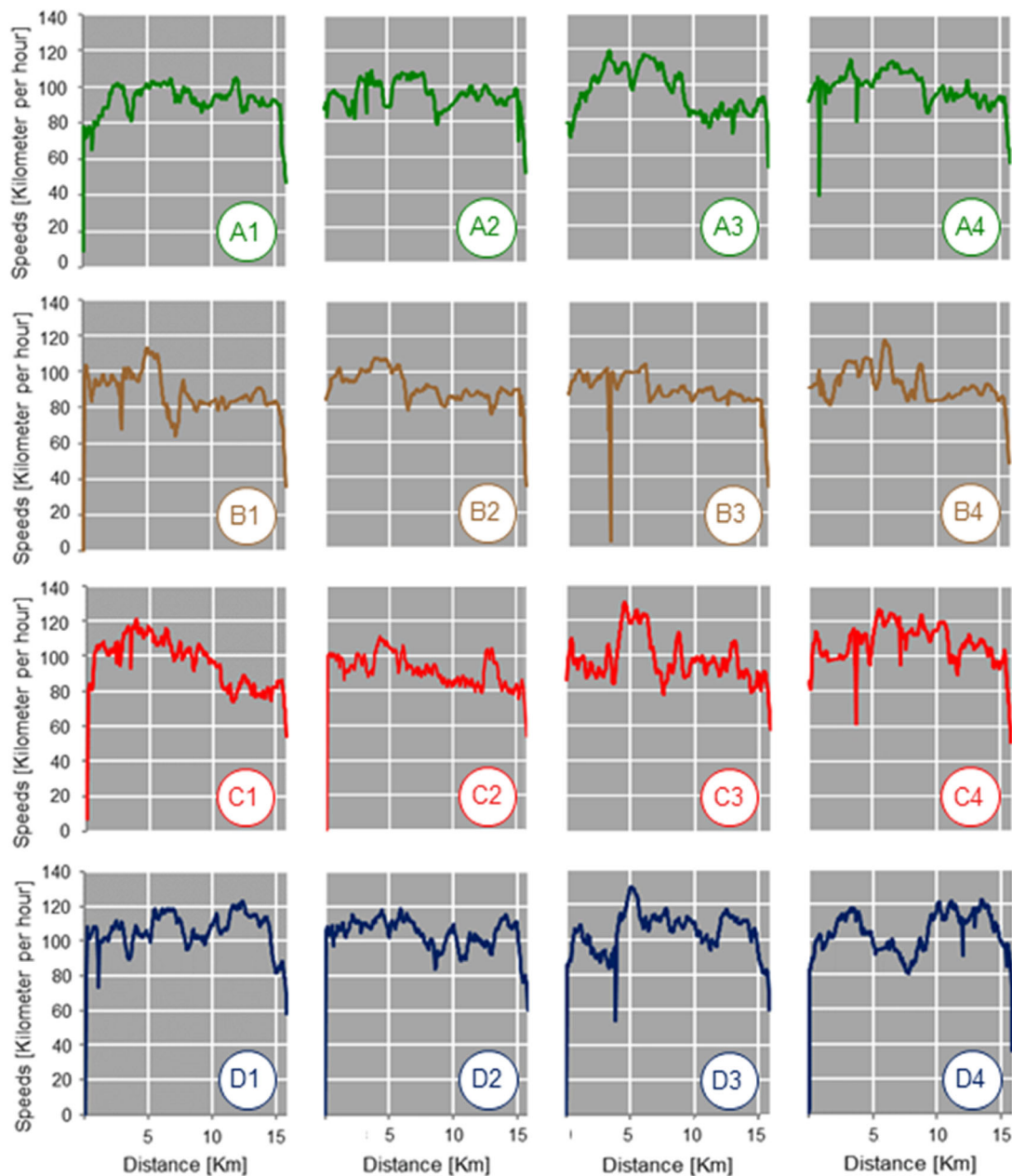
Fig. 4 Instantaneous speeds for each driver during the experiment. The codes in the circles identify each driver (A/B/C/D) and experiment day (1/2/3/4). The X-axis represents the distances traveled from the starting point of the route, whereas the Y-axis represents the speed values

Driver C exceeded 120 km per hour in the same road section, while driver D did it in different road sections, although mostly in the second part of the route.

In the other extreme, abnormally low speeds were mostly registered in the first and last road sections, where the experimental car was entering and leaving the motorway. In certain cases, some of the drivers enter the motorway with lower speeds than expected. This was the case for driver A during the fourth day and for driver D during the third day. By checking the cameras on-board of the experimental car, we realized that the

drivers were forced to reduce their speeds because of the surrounding traffic conditions.

A more detailed perspective in some specific road sections allows us to observe more carefully how drivers perform certain maneuvers and actions. Figure 7 shows the acceleration/deceleration of the vehicle at the entrance and exit of the route, which basically corresponds to the road sections depicted in the yellow boxes in Fig. 1. Vehicle acceleration/deceleration is estimated from the changes in speed along the road (speed at cell [i] in comparison to speed at the upstream cell [i-1], with cells being 50 m long). Speed values at each cell are based on
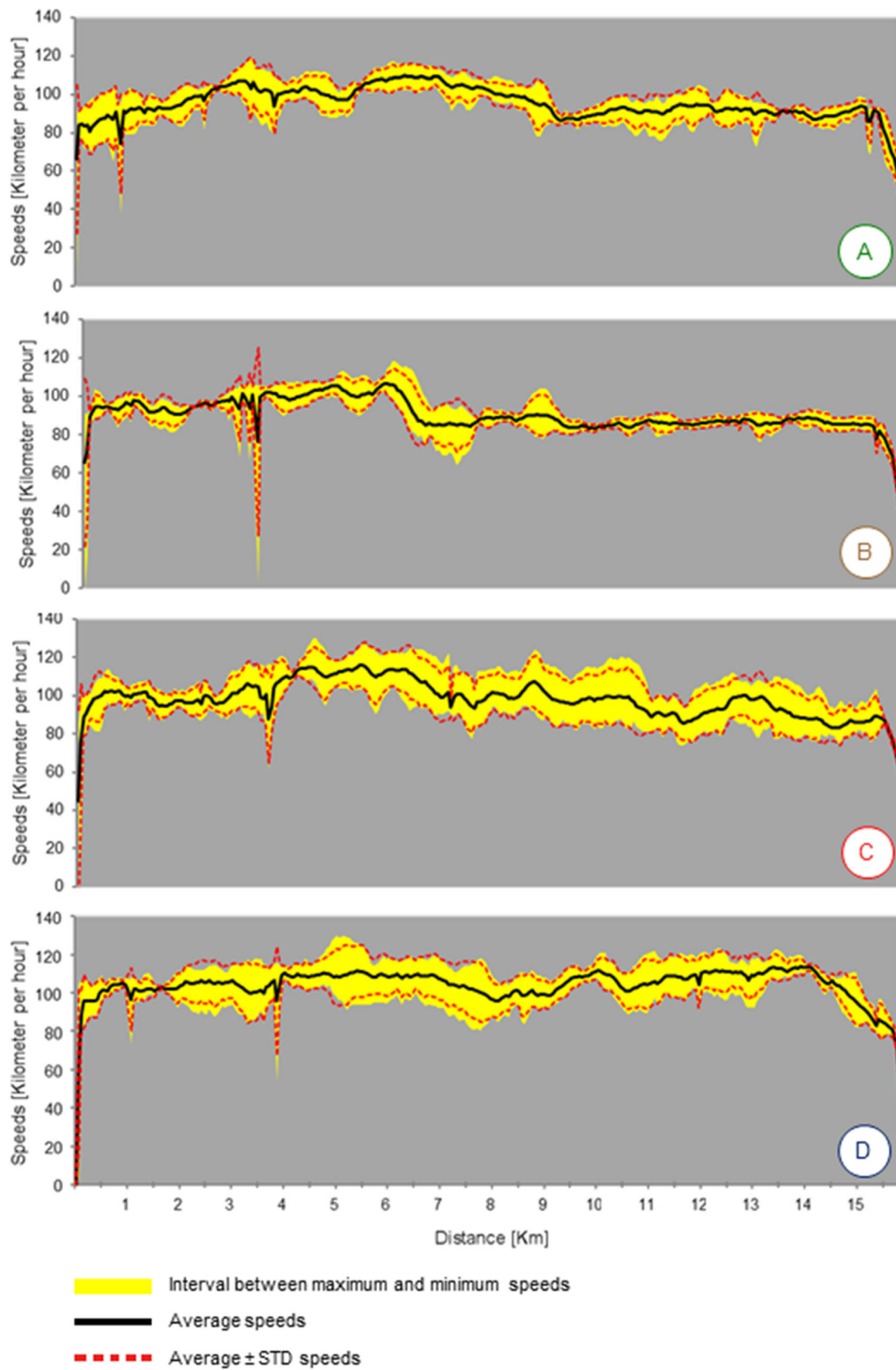
Fig. 5 Average instantaneous speeds for each driver (code A/B/C/D in the circle) across the four days. The X-axis represents the distances traveled from the starting point of the route, whereas the Y-axis represents the speed values. The continuous black line represents the average speeds in each point along the route. The dashed red lines result from adding/subtracting the standard deviation (STD) to the average speeds. The yellow area indicates the difference between the maximum and minimum speeds in each point of the route

Fig. 6 Road sections where extreme speeds were registered. In red, road sections where the drivers exceed 120 Kilometers per hour. In yellow, road sections where the drivers drove slower than 60 km per hour. The codes in the circles identify each driver (A/B/C/D) and experiment day (1/2/3/4)
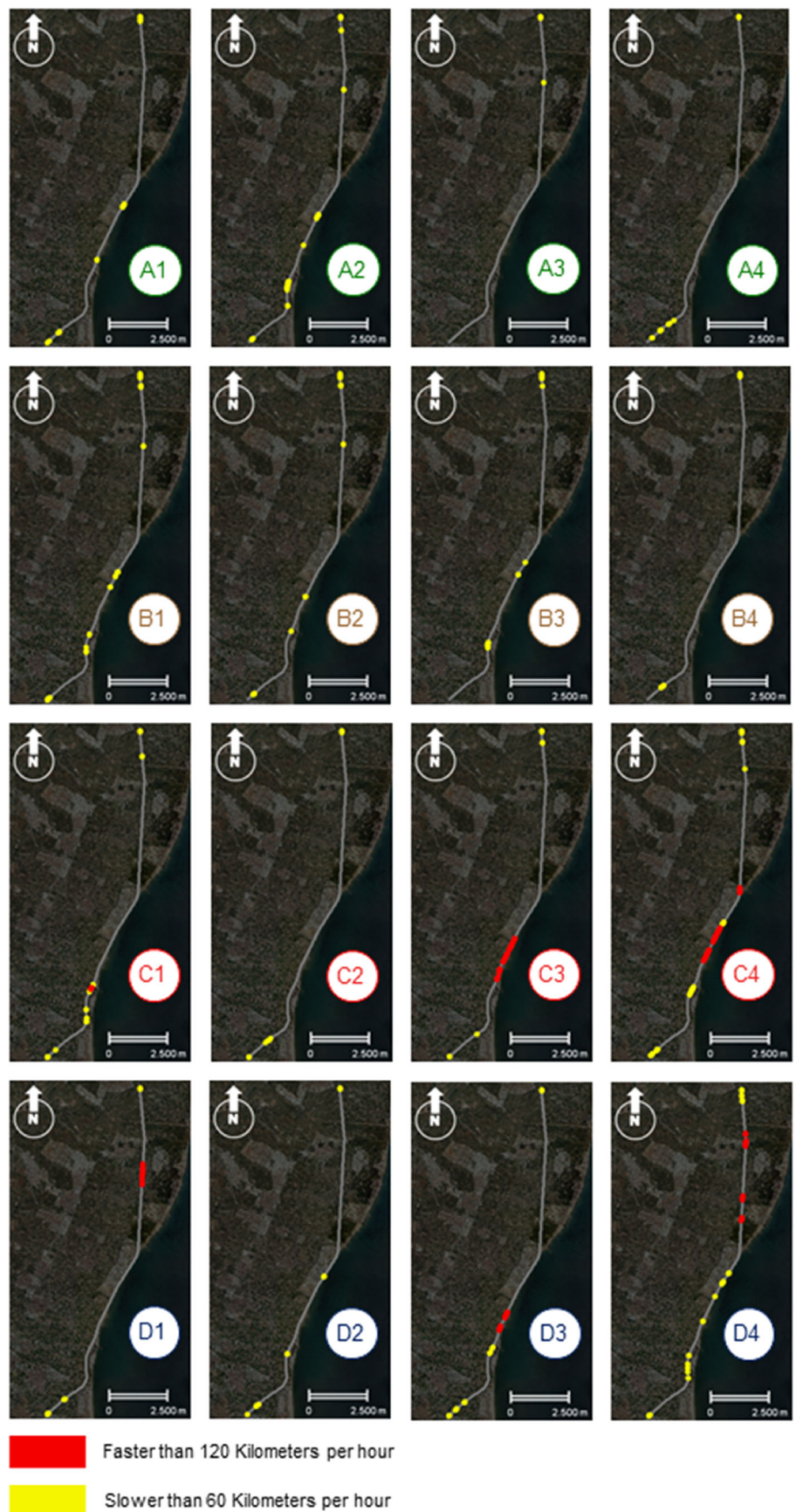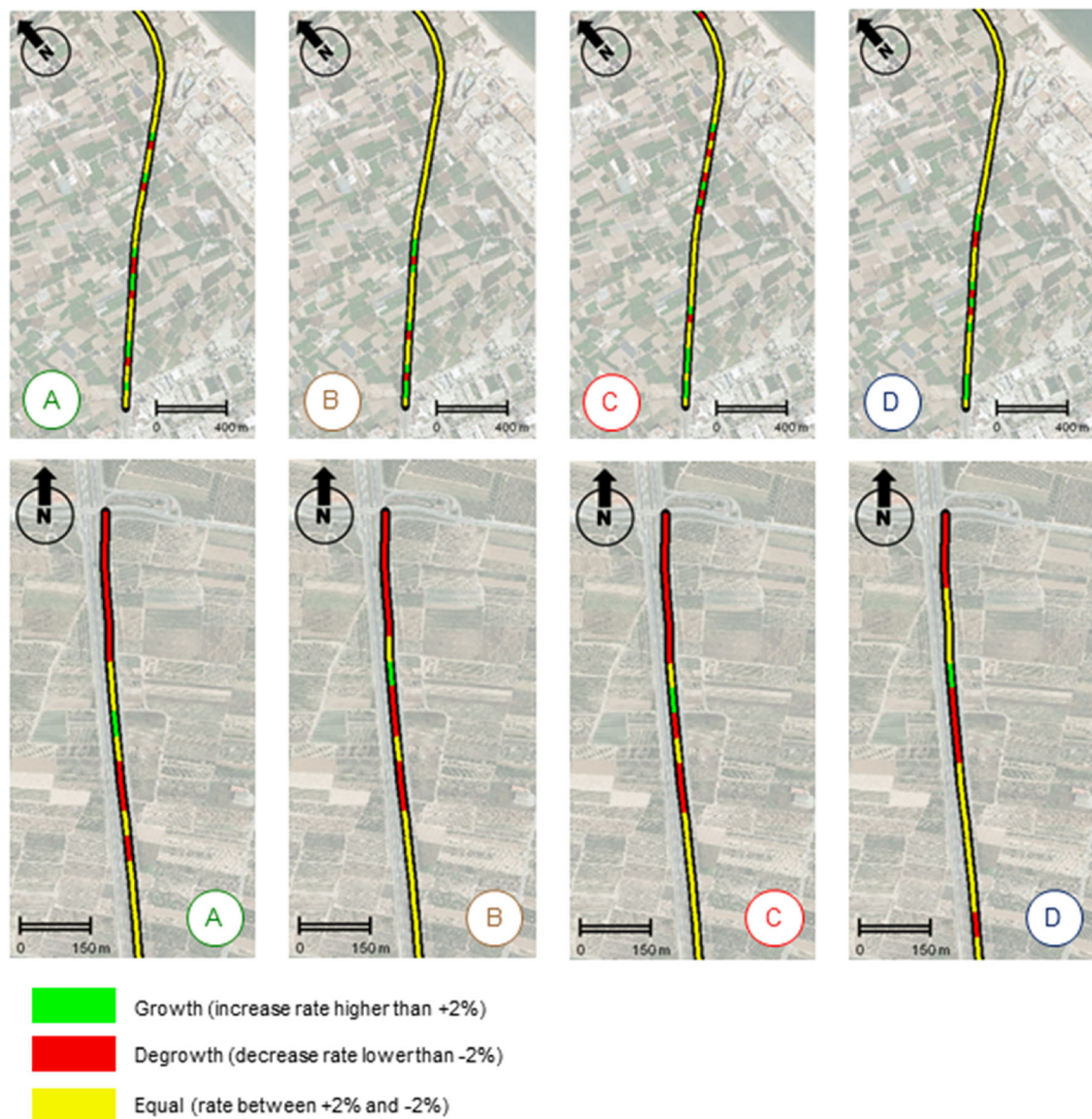
Fig. 7 Acceleration and deceleration of the vehicle during the entry to (upper row) and exit from (bottom row) the motorway. Driving performance is evaluated by considering the average speeds of each driver (code A/B/C/D within the circle) along the road

those average from the four days they drive. Acceleration refers to a relative increase in speed values, i.e. higher than +2% between successive cells, while deceleration refers to a relative decrease, i.e. lower than −2%. Speeds are considered constant when the change is between −2% and +2% in successive cells. Although one would expect a process of continuous acceleration at the entrance of the motorway and a continuous deceleration at the exit of it, the dynamics represented in Fig. 7 seem much more complex.

There are relevant differences between drivers due to particular traffic conditions in successive days as well as personal preferences. These differences are more evident in the initial road stretch. In the ending section, all the drivers follow a quite similar pattern of deceleration by performing the same sequence of maneuvers: (a) initial deceleration, (b) speed

adaptation, and (c) final deceleration. However, we can note some relevant differences regarding the time and duration of these actions. The initial deceleration usually consists of a double reduction of speeds, which sometimes tends to be a single action. Concerning the final deceleration maneuver, it happens in the last 150–300 m of the route. This behavior is similar for drivers A, B, and C compared to driver D, who slows down more abruptly.

A micro spatial scale approach favors a much more exhaustive analysis. The next figures (Figs. 8 and 9) represent the driving performance of driver A in two concrete road sections. Figure 8 corresponds to the entrance of the motorway (box 1 in Fig. 1), a road stretch 2.4 km in length. The spatio-temporal mapping of ND data allows us to define common driving patterns but also to observe how deviations emerge over time. The

Fig. 8 Driving profile of driver A at the entrance of the motorway. In columns, kinematic parameters related to engine speed, gearshift, vehicle speed, and position of clutch/accelerator/brake pedals are shown. In rows, his driving performance over time (experiment day 1/2/3/4 in the circle)
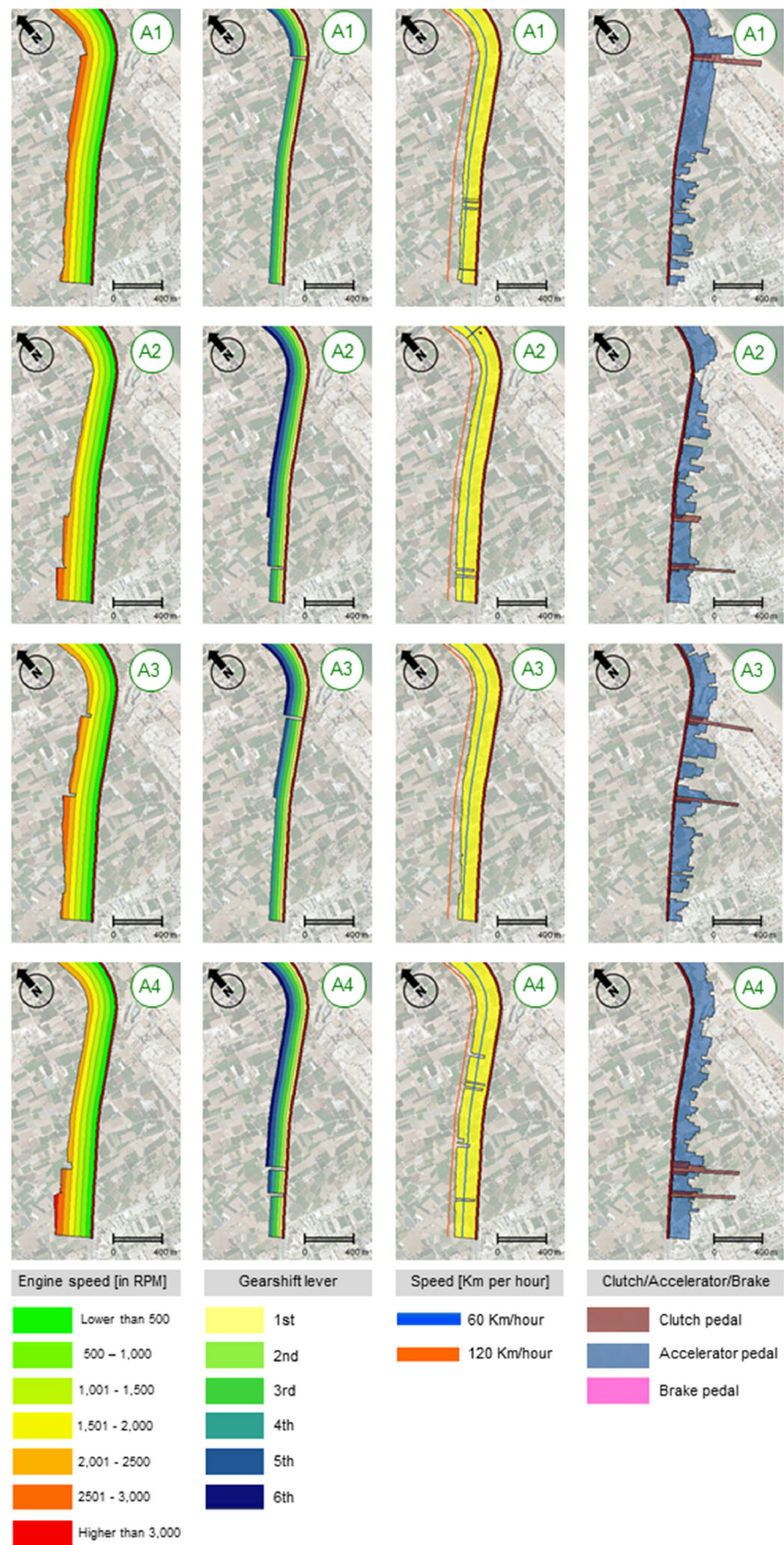
Fig. 9 Driving profile of driver A at the exit of the motorway. In columns, kinematic parameters related to engine speed, gearshift, vehicle speed and position of clutch/accelerator/brake pedals are shown. In rows, his driving performance over time (experiment day 1/2/3/4 in the circle)

order of the successive sub-figures presents a matrix structure, where each row represents one single day. In columns, we represent different kinematic parameters such as (a) the engine speed, (b) the gearshift lever, (c) the vehicle speed, and (d) any forces applied to clutch, accelerator, and brake pedals.

Notice that in the road section covering the entrance of the motorway (Fig. 8) Driver A does not use the brake pedal, as he is mostly accelerating. In contrast, in the road section covering the exit of the motorway (Fig. 9), Driver A uses the accelerator pedal first, and the brake pedal later as he exits the motorway. Despite these similarities across days, we can also observe some differences over time.

At the entrance to the motorway ((Fig. 8), during the first day, the speed increases progressively from 60 to 110 km per hour. The accelerator pedal is mostly pressed, albeit with certain interruptions. Driver A drives most of the time using the fourth gear. Before the curve, he stabilizes his speed and, once inside the curve, he accelerates again. A similar trend is observed for engine speeds, which progressively increase before the curve, where he decides to shift the gear up leading to a significant reduction of revolutions per minute. In contrast, on the second day, the speed values follow a more regular profile, with some reductions before the curve. In the presence of the curve, the driver gradually decelerates the vehicle and, once inside, he accelerates again in a staggered way. The driver performs two gear shifts, both at the beginning of the route. While the first action is carried out in a slow manner, the second one is a much faster maneuver, although the driver only steps partially on the clutch. Thus, he substantially reduces the engine speed during three intervals. The use of longer gear positions allows the driver to reduce engine speeds and this derives, consequently, in a more stable and efficient driving model, less fuel-demanding, and more environmentally friendly. The speed profile of the third day is quite similar to the first day. The driver tends to press the accelerator pedal, although with some interruptions. He shifts two gear upwards. First, he changes from the fourth to the fifth gear in the middle of the initial road section. After that, he changes from the fifth to the sixth gear just before the curve. He drives with a high engine speed most of the time due to the fast speeds and a delayed gear shifting. On the fourth day, the driver shows again a very regular speed profile. He shifts up the gear two times, one after the other in the first part of the road section. Several sudden drops in speed values are registered, which may be explained by eventual malfunctions in the sensor. We can also see that the accelerator pedal is always pressed. Engine speeds are high during the first part, but these are reduced after the driver shifts to the sixth gear, showing a smooth profile.

A similar approach can be observed in Fig. 9. The distribution and the parameters presented in this figure are the same as the previous one. This road section is 1.6 km in length and corresponds to box 2 in Fig. 1. On this road section, the driver slows down drastically to exit the motorway through the deceleration lane. However, this manoeuver is performed differently over time. The first day, he presses the brake two times. Eventually, he shifts the gear lever down from the sixth to the fourth, avoiding the fifth position. The engine speed draws a similar profile compared to the vehicle speed. The reduction during the second day is more progressive. The profile drawn by vehicle speeds shows a convex shape, slightly different from the first day. The driver decelerates for a while before stepping on the brake pedal all at once. Then, he shifts down the gear lever, from the sixth to the third position. Again, the profile drawn by the engine speed is very similar to the vehicle speed. The reduction in speeds during the third day is similar to the previous day, although this action is performed on a smaller road section. The driver steps on the brake pedal twice, almost without any time in between. The driver shifts the gear lever down from the sixth to the third position in a continuous and staggered manner, although slower than the day before. During the fourth day, a very progressive reduction in speed values is observed again. There, the driver suddenly ceases to accelerate the vehicle and, immediately after that, presses the brake pedal. He repeats this action twice, although he accelerates slightly in between. He then shifts the gear lever down abruptly. The profile drawn by engine speed also shows a very progressive reduction.

6 Discussion and conclusions

In this paper, we present a novel methodology for the extraction of driving patterns. Our working proposal is innovative, as no similar research exists. It is based on a double axis: (a) raw driving data and (b) mapping data. Our dataset comes from a naturalistic driving (ND) experiment. This type of data has an enormous potential for research. Among other advantages, these data provide more information than ever related to the whole set of conditions that affect and determine driving behavior. Thus, we count with valuable information related to the driver's performance, but also to the traffic conditions, the type and characteristics of the road, and the surrounding environment, among other factors. This type of data is adequate for the extraction of driving patterns, in a more reliable and accurate way compared to traditional methodologies.

In relation to mapping, we have implemented a double approach based on a multi-scale spatial analysis. At a macro spatial scale, we display mostly bar charts and histograms, which help to describe the driving behavior in general terms. This approach allows us to check some driving tendencies. For example, some people tend to drive slower or faster than the average, some people are very consistent in their speeds, while others show very irregular driving performance over

time. That being said, this macro spatial scale approach does not give any insigths into a number of factors related to the specific actions behind those general trends. The micro spatial scale approach then supplements and reinforces the previous observations. With this in mind, we conduct a more comprehensive analysis of the driving patterns associated with the first driver. This micro scale analysis is conducted by mapping with GIS some relevant kinematic parameters in certain road sections. This type of representation allows us to estimate metrics such as those related to how safe or environmentally friendly someone drives.

Our working methodology is essentially based on mapping with GIS tools. So far, GIS tools have a limited set of applications in road traffic research [52]. Most of the studies in traffic and transportation issues related to GIS tools have been limited to map the spatial distribution of some hotspots, usually crashes, or to highlight some vulnerable road sections. For that, they used simple features and entities such as points for hotspots and lines for road sections. In the case of the current paper, our methodology is much more ambitious by seeking how to exploit the huge potential of ND data. To that end, GIS tools allow us to correlate the road layout to different driving performance metrics. In addition, these tools make it possible to plot simultaneously a great number of kinematic driving parameters – a functionality that can be enabled/disabled depending on the operator requirements –. This facilitates the subsequent phases of analysis and interpretation of results by allowing the analyst to incorporate different perspectives. Thus, it is possible to involve professionals with different levels of expertise and technical background in multidisciplinary working teams. This, in turn, encourages the development of more creative solutions for dealing with multi-faceted problems such as those related to road safety and the management and control of traffic flows [53].

This study, however, presents certain limitations. Regarding the experiment, our data comes from a small-scale experiment with a shoestring budget. In fact, only five drivers participated in the Spanish experiment of PROLOGUE. Our sample dataset was unrepresentative for defining any group of drivers, and some differences in the driving performance might have been motivated by different traffic conditions at any particular time. Nonetheless, the dataset was also extensive in data records, which made it adequate for implementing our proposed methodology. Also, as stated before, drivers were aware of the experiment, which could potentially affect their driving behavior and, therefore, reduce the naturalistic validity of the final data [7]. This, in our case, was not a significant issue, as in this paper we focus on the methodology and not the insights from the data itself. However, it does limit our ability to derive any major conclusions from the observed driving patterns. Last, the original data had some gaps. These were easily detected in the case of kinematic indicators

because these correspond to anomalous fluctuations represented along the route [21, 22]. Other gaps might not be so easy to recognize.

In general, a better understanding of driving patterns and their relationship with geographical driving areas could bring great benefits for smart cities, including the identification of good driving practices for saving fuel and reducing carbon emissions and accidents; all leading to more sustainable development. Advances in electronics and information and communication technologies (ICT) are rapidly enhancing our ability to sense and monitor the various phenomena and processes at both the micro- and macro-levels. Smart cities and organizations are characterized by exploiting advanced sensing and decision-making technologies, as well as by deploying dynamic actuation technologies. All these technologies also improve our capabilities for using and interpreting naturalistic driving data. The type of analysis presented in this paper, can be used to educate citizens about various driving patterns and their positive and negative impacts on fuel consumption, carbon emissions, vehicle wear and tear, and accident risks. Similarly, the generated insights can be used to inform insurance companies regarding liability so they can implement strict measures for social intervention and/or to modify public behavior [54].

As we move into the future, with more advanced sensing and instrumentation technologies, it will be possible to identify driving patterns dynamically, in real-time, and issue warnings to the driver when appropriate [55]. Some of the insights generated with naturalistic driving data could also be used to design automated cars that drive in a more sustainable manner. Notice that many countries around the world have already passed regulations that allow some form of autonomous vehicles on the road. Thus, most probably, we will have vehicles with various levels of automation sharing the roads in the near future [56, 57]. In such complex driving environments where human-driven vehicles have to interact with computer-driven vehicles, a good understanding of driving behavior will be crucial to maintain both safety and efficiency. In the long run, it is expected that we will have roads or lanes with fully autonomous vehicles, potentially reaching a time when human driving is completely banned on public roads. In such an era, vehicle manufacturers and/or vehicle operators may like to add facilities for multiple humanistic driving patterns in the vehicles, or vehicle trips, to amuse customers or as a means to provide service differentiation. Certainly, these humanistic driving pattern services would be curtailed under some government policies for the public good. Our future work will explore these directions for smart city and autonomous vehicle designs.

In short, understanding driving patterns allows us to reduce and simplify the complex behavior of drivers in a way that many insights can be generated. In this paper, we presented an adequate methodology for extracting driving patterns from ND data. We developed different visualization strategies that are adapted to a multi-scale approach. The main contributions

include (a) better exploiting the high potential of fine-grained ND data, and (b) demonstrating the adequacy of mapping these data by using GIS tools. The low representativeness of our sample prevents us from drawing relevant conclusions about the profiles of any group of people. However, the methodology developed in the paper at hand can be applied to similar studies with a larger scope. Finally, relevant issues, including open challenges on ND data, have also been discussed in this paper.

Looking to the future, our method is a significant step towards understanding the actual behavior of people driving. Future studies should implement strategies focused on exploiting the huge potential of ND data. Certainly, one of the critical aspects will be to develop methodologies for data mapping that further enhance our understanding of this data.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest.

References

1. Mehmood R, See S, Katib I, Chlamtac I (2020) Smart infrastructure and applications. Foundations for smarter cities and societies. Springer International Publishing, Cham

2. Mehmood R, Bhaduri B, Katib I, Chlamtac I (2018) Smart societies infrastructure technologies and applications. Springer International Publishing, Cham

3. Alomari E, Mehmood R, Katib I (2019) Road traffic event detection using twitter data, machine learning, and apache spark. The 3rd IEEE international conference on Smart City innovations (SCI 2019), Leicester, pp 1888–1895

4. Regan M, Williamson A, Grzebieta R, Tao L (2012) Naturalistic driving studies: literature review and planning for the Australian Naturalistic Driving Study. Australasian College of Road Safety Conference 2012, Sydney

5. Backer-Grøndahl A, Lotan T, van Schagen I (2011) Summary and integration of a series of naturalistic driving field trials. Promoting real life observations for gaining understanding of road behaviour in Europe, PROLOGUE (deliverable D3.7). Institute of Transport Economics (TØI), Oslo

6. McLaughlin S, Hankey J, Dingus T (2008) A method for evaluating avoidance systems using naturalistic driving data. Accid Anal Prev 40:8–16

7. Valero-Mora P, Tontsch A, Welsh R et al (2013) Is naturalistic driving research possible with highly instrumented cars? Lessons learnt in three research centres. Accid Anal Prev 58:187–194

8. Dingus T, Klauer S, Neale V et al (2006) The 100-car naturalistic driving study, phase II: results of the 100-car field experiment. National Highway Traffic Safety Administration, Washington DC

9. Campbell K (2012) The SHRP 2 naturalistic driving study. Addressing driver performance and behavior in traffic safety. TR News Magazine 282:30–35

10. van Schagen I, Welsh R, Backer-Grøndahl A et al (2011) Towards a large-scale European naturalistic driving study: main findings of PROLOGUE. Promoting real life observations for gaining understanding of road behaviour in Europe, PROLOGUE (deliverable D4.2). Institute for Road Safety Research (SWOV), Leidschendam

11. Barnard Y, Utesch F, van Nes N, Eenink R, Baumann M (2016) The study design of UDRIVE: the naturalistic driving study across Europe for cars, trucks and scooters. Eur Transp Res Rev 8(14)

12. Tivesten E, Dozza M (2015) Driving context influences drivers' decision to engage in visual–manual phone tasks: evidence from a naturalistic driving study. J Saf Res 53:87–96

13. Metz B, Landau A, Hargutt V (2015) Frequency and impact of hands-free telephoning while driving. Results from naturalistic driving data. Transport Res F: Traffic Psychol Behav 29:1–13

14. Metz B, Landau A, Just M (2014) Frequency of secondary tasks while driving. Results from naturalistic driving data. Saf Sci 68:195–203

15. Foss R, Goodwin A (2014) Distracted driver behaviors and distracting conditions among adolescent drivers: findings from a naturalistic driving study. J Adolesc Health 54(5):S50–S60

16. Precht L, Keinath A, Krems J (2017) Effects of driving anger on driver behavior. Results from naturalistic driving data. Transport Res F: Traffic Psychol Behav 45:75–92

17. Backer-Grøndahl A, Phillips R, Sagberg F, Touliou K, Gatscha M (2009) Topics and applications of previous and current naturalistic driving studies. Promoting real life observations for gaining understanding of road behaviour in Europe, PROLOGUE (deliverable D1.1). Institute of Transport Economics (TØI), Oslo

18. Pilgerstorfer M, Runda K, Brandstätter C et al (2011) Small-scale naturalistic driving pilot. DaCoTa (deliverable D6.3). Kuratorium für Verkehrssicherheit (KfV), Vienna

19. Balsa-Barreiro J, Valero-Mora P, Pareja I, Sánchez M (2013) Geo-referencing naturalistic driving data using a novel method based on vehicle speed. IET Intell Transp Syst 7(2):190–197

20. Balsa-Barreiro J, Pareja I, Tontsch A, Sánchez M (2012) Preprocessing of data for recovery of positioning data in naturalistic driving trial, vol 14-15. Proc. 3rd European conference on human Centred Design for Intelligent Transport Systems, Valencia, pp 235–245

21. Balsa-Barreiro J, Valero-Mora P, Pareja I, Sánchez M (2015) Proposal of GIS methodology for quality control procedures (QC) of data obtained in naturalistic driving studies. IET Intell Transp Syst 9(7):673–682

22. Balsa-Barreiro J, Valero-Mora P, Pareja I, Sánchez M (2014) Quality control procedure for naturalistic driving data using Geographic Information Systems. Proc. 4th European Conference

on Human Centred Design for Intelligent Transport Systems, Vienna (Austria)

23. Balsa-Barreiro J, Valero-Mora P, Berne J, Varela-García A (2019) GIS mapping of driving behavior based on naturalistic driving data. ISPRS Int J Geo Inf 8(5):226

24. Balsa-Barreiro J, Valero-Mora P, Berne J, Varela-García A (2016) Representation of kinematic driving behaviour using Geographic Information Systems. Proc. 5th European Conference on Human Centred Design for Intelligent Transport Systems, Loughborough (UK)

25. Ericsson E (2000) Variability in urban driving patterns. Transp Res Part D: Transp Environ 5:337–354

26. Wolf J, Hallmark S, Oliveira M, Guensler R, Sarasua W (1999) Accuracy issues with route choice data collection by using global positioning system. Trans Res Record: J Trans Res Board 1660(1): 66–74

27. Egea A (unkown) El comportamiento humano en conducción: Factores perceptivos, cognitivos y de respuesta. Private communication Universidad de Murcia. https://www.um.es/docencia/agustinr/pca/textos/cogniconduc.pdf Accessed 23 July 2020

28. Delgado A, Maradei M, Castellanos J (2013) Influencia de los patrones posturales en la conducción y la antropometría en la carga biomecánica del raquis. Iconofacto 9(12):38–55

29. Bratt H, Ericsson E (1999) Measuring vehicle driving patterns. Estimating the influence of different measuring intervals. Proc. 2nd KFB research conference, Lund

30. Berry I (2010) The effects of driving style and vehicle performance on the real-world fuel consumption of U.S. light-duty vehicles," Master Thesis Dissertation, Massachusetts Institute of Technology

31. Pasaoglu G, Fiorello D, Martino A, Scarcella G, Zubaryeva A, Thiel C (2012) Driving and parking patterns of European car drivers. A mobility survey. JRC Scientific and Policy Reports. European Commission, Brussels

32. Thiel C, Alemanno A, Scarcella G, Zubaryeva A, Pasaoglu G (2012) Attitude of European car drivers towards electric vehicles: a survey. JRC scientific and policy reports. European Commission, Brussels

33. Karabasoglu O, Michalek J (2013) Influence of driving patterns on life cycle cost and emissions of hybrid and plug-in electric vehicle powertrains. Energy Policy 60:445–461

34. Fontaras G, Pistikopoulos P, Samaras Z (2008) Experimental evaluation of hybrid vehicle fuel economy and pollutant emissions over real-world simulation driving cycles. Atmos Environ 42:4023–4035

35. Sharer P, Leydier R, Rousseau A (2007) Impact of drive cycle aggressiveness and speed on HEVs fuel consumption sensitivity. Argonne National Laboratory, Lemont

36. Ericsson E, Mocsári T, Rapone M, Brundell-Freij K (2004) Variability in driving patterns over street environments in three European cities. Technical report. Lund Institute of Technology

37. Sabback F, Mann W (2005) The influence of climate and road conditions on driving patterns in the elderly population. Phys Occupational Therapy Geriatrics 23(2/3):63–74

38. James J, Goldman M (1971) Driving patterns for different-sized vehicles. Ergonomics 14(3):363–369

39. Evans L, Rothery R (1976) Comments on effects of vehicle type and age on driver behaviour at signalized intersections. Ergonomics 19:559–570

40. The Ecologist (2010) Electric cars can 'fit with current driving patterns'. The Ecologist. Available online at https://theecologist.org/2010/jun/18/electric-cars-can-fit-current-driving-patterns. Accessed 23 July 2020

41. Sullman M, Paixon J, Stephens A (2017) Gender roles, sex and the expression of driving anger. Accid Anal Prev 106:23–30

42. Healey J (2001) Small SUVs catch on, especially with women USA Today http://usatoday30usatodaycom/money/autos/2001-07-31-suvhtm Accessed 23 July 2020

43. McGuckin N, Murakami E (1999) Examining trip-chaining behavior: comparison of travel by men and women. J Transp Res Board 1693:79–85

44. Kington R, Reuben D, Rogowski J, Lillard L (1994) Sociodemographic and health factors in driving patterns after 50 years of age. Am J Public Health 84(8):1327–1329

45. Marshall S, Wilson K, Molnar F, Man-Son-Hing M, Stiell I, Porter M (2007) Measurement of driving patterns of older adults using data logging devices with and without global positioning system capability. Traffic Injury Prevention 8(3):260–266

46. Hildebrand E, Myrick B, Creed T (2000) The rural elderly: driving patterns and accident involvement. Proc. 35th Annual Conference of the Canadian Transportation Research Forum, Charlottetown, pp 175–191

47. Lotan T, Toledo T (2007) Driving patterns of young drivers within a graduated driver licensing system. Proc. 86th annual meeting of the Transportation Research Board, Washington, DC

48. Williams A (2003) Teenage drivers: patterns of risk. J Saf Res 34: 5–15

49. Fonda S, Wallace R, Herzog A (2001) Changes in driving patterns and worsening depressive symptoms among older adults. J Gerontol: Social Sci 56B(6):343–351

50. van Landingham S, Hochberg C, Massof R, Chan E, Friedman D, Ramulu P (2013) Driving patterns in older adults with glaucoma. BMC Ophthalmol 13(4)

51. Valero-Mora P, Tontsch A, Pareja I, Sánchez M (2010) Using a highly instrumented car for naturalistic driving research: a small-scale study in Spain. Promoting real life observations for gaining understanding of road behaviour in Europe, PROLOGUE (deliverable D3.5). Instituto de Tráfico y Seguridad vial (INTRAS), Valencia

52. Aguero-Valverde J, Jovanis P (2008) Analysis of road crash frequency with spatial models. Transport Res Record 2061:55–63

53. NurFazzillah M, Mahmud A, Ahmad N (2011) Value-adding safety audit with GIS. Geospatial World https://wwwgeospatialworldnet/article/value-adding-road-safety-audit-with-gis Accessed 23 July 2020

54. Sasidharan L, Menendez M (2019) Application of partial proportional odds model for analyzing pedestrian crash injury severities in Switzerland. J Trans Safety Secur 11(1):58–78

55. Tilg G, Yang K, Menendez M (2018) Evaluating the effects of automated vehicle technology on the capacity of freeway weaving sections. Trans Res Part C: Emerg Technol 96:3–21

56. Yang K, Guler I, Menendez M (2016) Isolated intersection control for various levels of vehicle technology: conventional, connected, and automated vehicles, Trans Res Part C: Emerg Technol, 72 (2016): 109–129

57. Yang K, Menendez M, Guler I (2019) Implementing transit signal priority in a connected vehicle environment with and without bus stops. Transportmetrica B: Transport Dynamics 7(1):423–445