# Estimating operating speed for county road segments – Evidence from Italy

Valentina Martinelli, Roberto Ventura, Michela Bonera, Benedetto Barabino *, Giulio Maternini

*Department of Civil, Environment, Land and Architecture Engineering and Mathematics (DICATAM), University of Brescia, 25123 Brescia, Italy*

## ABSTRACT

Vehicle operating speed is a crucial factor for road safety, as it strictly affects occurrence and severity of crashes. Usually, 85th percentile of the operating speed distributions (i.e., $V_{85}$) in free-flow traffic condition is widely accepted as a base value of consistency evaluation for homogenous portion of existing roads. Although the computation of $V_{85}$ is simple, many road authorities cannot collect speed data for each road. Therefore, providing prediction models could be a useful tool to investigate the relationship between $V_{85}$ and road characteristics. The literature proposed several models to account it. However, to the best of our knowledge, the effects of some road geometric characteristics, road markings and signs, traffic data, type of terrain and the simultaneous consideration of different road categories on the $V_{85}$ prediction were not completely analyzed. This paper fills this gap by isolating key variables that mostly affect $V_{85}$. In doing so, 60 000+ car spot speed data were collected along the county road network of the province of Brescia (Italy), and then processed by multiple regression models. The main findings show that $V_{85}$ increases owing to the presence of a wider or paved shoulder, visible road median markings, a higher number of lanes and a higher percentage of cars with respect to the total traffic flow. Conversely, $V_{85}$ decreases as the road axis curvature, the number of accesses and rate of forbidden overtaking increase. In addition, the presence of visible road external markings and the surrounding mountainous terrain contribute to decreasing $V_{85}$. The overall findings may support road authorities to verify roads' operating conditions and, possibly, adjust the speed limits, especially for existing roads.

© 2022 Tongji University and Tongji University Press. Publishing Services by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

## Introduction

Speed is a key factor for road safety, as it can affect the occurrence and severity of crashes. An inappropriate speed, frequently above posted speed limits, is responsible for a high quota of fatalities and disabilities. For instance, in 2019, in Italy 172 183 road crashes were reported and speed was responsible for 12.2% of the total fatal crashes on rural roads (ACI, 2019). Treatments facing unsafe speed have been at the core of the road safety policy for decades, and a significant progress has been made. However, there is still large potential for addressing this longstanding area of road safety at the EU and national

Please cite this article as: V. Martinelli, R. Ventura, M. Bonera et al., Estimating operating speed for county road segments – Evidence from Italy, International Journal of Transportation Science and Technology, https://doi.org/10.1016/j.ijtst.2022.05.007

levels (ETSC, 2019). Therefore, with a strong political support and effective coordination between responsible authorities, speed management strategies, such as the checking of road design consistency, which is strictly related to drivers' behavior, can significantly contribute to achieve road safety targets.

The checking of road design consistency[1] could be regarded as a facet of the potential heterogeneity in drivers' behavior, highlighting safety issues. Moreover, it is particularly useful for roads built before the norms were issued and/or have an 'original' path (or existing roads). Here, inconsistent alignments[2] of geometric road elements (i.e., curve and tangents) may not comply with the drivers' expectations, and, thus, with their behavior (e.g., Eboli et al., 2017). This situation could lead to speed errors, inappropriate driving maneuvers, and/or an undesirable crashes' number. Hence, these alignments could be failing to ensure the consistency with successive road elements and could lead to critical driving errors and crash risk. Conversely, a consistent alignment is preferred to meet drivers' expectations and promote harmonious driving behaviors (Yan et al., 2017).

To point out inconsistencies on successive elements of the road, the estimation of the operating speed in lieu of, or in addition, to design speed is performed (e.g., Russo et al., 2016; Hashim et al., 2016; Eboli et al., 2017; Wang et al., 2018; Lobo et al., 2018). This value is denoted by $V_{85}$ and corresponds to the 85th percentile of the speed distributions in free-flow traffic conditions (at a location in the road alignment). Next, dissimilarities between $V_{85}$ and the design speed on successive elements can be pointed out (e.g., Lamm et al., 1999). Moreover, $V_{85}$ separates acceptable from unsafe speed behavior that disproportionately contributes to the crash risk. For instance, Forbes et al. (2012) have shown that travelling near one standard deviation above the mean operating speed yields the lowest crash risk for drivers. Furthermore, crash risk rapidly increases for drivers travelling two standard deviations or more above or below the mean operating speed (Russo et al., 2016).

The estimation of $V_{85}$ is not a complex task provided that operating speed data are available for each road element. However, this task may be complicated if speed data were not available at all. Indeed, many road authorities cannot collect speed data for each road (or its portion), so that they are not able to measure $V_{85}$ and, therefore, evaluate alignment consistency for them. Nevertheless, because the speed also depends on road characteristics, the development of a prediction speed model may be useful to achieve this target for roads where operating speed measurements are not available (i.e., not monitored roads).

The literature provided relevant models to explain the operating speed as a function of several predictors (variables, factors or determinants) such as roads geometric characteristics, vertical signs, pavement conditions, land use, speed and traffic data (e.g., Medina and Tarko, 2005; Zuriaga et al., 2010; Singh et al., 2012; Himes et al., 2013). Nevertheless, this paper adds a new contribution to the field. More precisely, it evaluates the effects of some road geometric characteristics, road markings and signs, traffic data and type of terrain (i.e., mountain, rolling and flat) on the $V_{85}$ along a segment, that were not completely analyzed in previous studies. In doing so, 60 000+ spot speed data were collected along a portion of the county road network of the province of Brescia (Italy) and then, processed by two multiple regression models.

This paper aims to contribute to both theory and practice. From a theoretical perspective, this paper helps refine previous research by including novel variables that help estimate $V_{85}$. On the practical side, although this study does not focus on the determination of speed limits, the $V_{85}$ estimates could be used as a proxy for the road authorities to enhance road safety conditions. Indeed, the proposed model can be regarded as an aiding support tool for the revision of speed limits especially in the existing roads (Lobo et al., 2018), even in the case of bridges (Ventura et al., 2020). More precisely, for existing roads, the setting of speed limits is a crucial challenge faced by many road authorities, owing to multiple criteria that can be considered. This speed limit may be set according to the speed driven on the road, i.e., $V_{85}$. (e.g., Polus et al., 2000; MIT, 2006; Aarts et al., 2009; Singh et al., 2012; Russo et al., 2016; Lobo et al., 2018). This specification can help drivers in choosing a more appropriate speed with respect to the road conditions, harmonizing the driving speeds along each segment and, possibly, reducing the occurrence and severity of crashes. In addition, the $V_{85}$ estimates enable to include all types of terrains and a wide range model that may help manage different road categories, considering several (and common) geometric characteristics. Thus, public administrations could extend these estimates to the whole territorial context without differentiate per municipalities.

The remaining paper is as follows. Section Literature review provides a background on the most relevant literature of the operating speed prediction models. Section Material and Method introduces the research context and illustrates data and the methods adopted to estimate the operating speed. Section Results and discussion provides and discusses the results in the context of the previous literature. Finally, Section Conclusions and research perspectives presents the conclusions and provides future perspectives.

## Literature review

Operating, percentile and mean speeds (hereafter speed) have been identified as one of the key factors that contributed to road crashes. Therefore, several studies have focused on a better understanding of the main factors affecting the speed by using different data collection tools and analysis methods. All examined studies focused on roads located in rural contexts with scattered buildings and far from intersections or roundabouts. Indeed, in this context drivers may adopt higher speeds

---

[1] The degree to which roads' system is designed and constructed to avoid critical driving maneuvers that can lead to crash risk.
[2] The combinations of roads feature with unusual or extreme characteristics that drivers may drive in an unsafe manner.

ARTICLE IN PRESS

V. Martinelli, R. Ventura, M. Bonera et al.                    International Journal of Transportation Science and Technology xxx (xxxx) xxx

than those of urban roads, because of a lower presence of vulnerable users (i.e., pedestrians and cyclists), the geometric characteristics of the road layout and higher speed limits (Martinelli et al., 2022). Furthermore, in a rural context the uninterrupted-flow condition is evident, and the speed value in the free-flow conditions is appreciable. For instance, uninterrupted-flow facilities have no fixed elements, such as traffic signals, that are external to the traffic stream and may interrupt the flow (HCM, 2016).

A summary of these studies is provided in Table 1, which is concisely commented on what follows for the main points.

Most studies modelled speed for passenger cars, perhaps because they are the dominant quota of vehicles on the road, albeit in a few cases a prediction model for heavy commercial vehicles was developed.

Unlike Bhowmik et al. (2019) and Yan et al. (2017) who considered multilane roads (i.e., four, six and eight lanes), most of studies are generally referred to two-lanes roads with similar characteristics. In addition, speed measurements were performed in different types of terrain such as flat, and/or rolling and/or semi-mountainous (e.g., Medina and Tarko, 2005; Eboli et al., 2017; Russo et al., 2016; Maji and Tyagi, 2018; Lobo et al., 2018; Sil et al., 2020a, 2020b). Besides the type of terrain, speeds were mainly measured on curves or tangents. The formers provided input data that are easier to implement in prediction models than the latter. Indeed, these models depend on widely and clearly analyzed variables such as curvature, curve radius, longitudinal slope, superelevation and side-friction coefficients between road surface and tires (e.g., Polus et al. 2000, Zuriaga et al., 2010, Yan et al. 2017). Conversely, prediction models on tangents might be more complex to implement, because of many predictors needed (Polus et al, 2000; Maji et al., 2018). For instance, Polus et al. (2000) have considered long and short tangents. Others have considered both curves and tangents on their models (e.g., Medina and Tarko, 2005; Zuriaga et al., 2010; Singh et al., 2012; Hashim et al., 2016; Eboli et al., 2017). Russo et al. (2016) have also developed a mixed prediction model that provided better fit than the only tangents model. Finally, a handful of studies focused on speed measurements along a road segment, which is defined as a continuous sequence of design elements composing a stretch of road (Lobo et al., 2018).

Speed measurements were conducted by mean of spot or continuous measures. Spot speed data were collected in fixed points of the network by measuring the individual speed of each vehicle passing a given section and using specific devices (e.g., videorecording). These data may suffer from some disturbances such as changes in driver's behavior owing to the presence of the measurement device, bad data recording during passing vehicles (e.g., Zuriaga et al., 2010). Conversely, continuous speed data were collected by measuring the speed of each vehicle that is tracked by mobile devices (e.g., GPS). Continuous speed data may suffer from several drawbacks as well. First, because vehicles can be tracked individually, the involvement of many test drivers (statistically selected) is required. This is the probable reason of the limited test drivers of each considered study (e.g., 30 drivers for Hashim et al., 2016). Second, the driving tester may be conditioned by the presence of the on-board device, albeit this limitation might be addressed by evaluating the consistency between recorded operating speed by GPS and videorecording (Zuriaga et al., 2010). Third, GPS devices suffer from huge amount of missing and noisy data (e.g., Barabino et al., 2017; Wang et al., 2018). In addition, the use of mobile devices may be expensive when many road kilometers should be surveyed: current literature focused on about 24 km of roads at the maximum (e.g., Lobo et al., 2018).

Different modelling tools were available. Although someone adopted complex modelling such as artificial neural networks (e.g., Singh et al., 2012), and others (e.g., Wang et al., 2018), the main modelling tools were multiple linear regression models. These models are easy to use, simple to interpret, straightforward to assess using basic statistics and can incorporate and analyze many variables (Maji and Tyagi, 2018; Sil et al., 2020a, 2020b). For instance, artificial neural network has limited inferential capabilities as compared to regression analysis (Wang et al., 2018). The other modeling tools such as non-linear analysis, three-stage least-squares or panel data analysis did not meaningfully improve the performance model (e.g., Wang et al.,2018; Sil et al., 2020a).

Previous modellings showed that variables related to road axis geometry (e.g., curve radius, curvature change rate – $CCR_S$, the sum of the absolute values of angular changes – $CC$, length of the investigated element, and longitudinal slope), geometrical cross section (e.g., carriageway width, lane width, median width, shoulder width, guardrail presence), roadside configuration (e.g., presence of a sidewalk or cycle path), road signs (e.g., posted limits, chevron signs, advisory speeds) and other variables (e.g., presence of later accesses, land use, type of terrain, safety distance, traffic and accident data) affected all forms of speed. More precisely, as for road horizontal and vertical alignment, a greater curve radius increased vehicles' operating speed (e.g., Polus et al., 2000; Zuriaga et al., 2010; Eboli et al., 2017; Hashim et al., 2016; Wang et al., 2018), as opposed to $CCR_S$ (e.g., Zuriaga et al., 2010; Russo et al., 2016; Yan et al., 2017). In addition, a greater curve or tangent length increased operating speed (e.g., Polus et al., 2000; Zuriaga et al., 2010; Eboli et al., 2017; Hashim et al., 2016; Russo et al., 2016), while a greater longitudinal slope decreases operating speed (e.g., Medina and Tarko, 2005; Himes et al., 2013; Yan et al., 2017). As for cross section geometry, the carriage width usually increased the speed (e.g., Russo et al., 2016; Yan et al., 2017) but not always (e.g., Lobo et al., 2018), whereas the lane width did not affect the speed. In addition, a larger shoulder usually increased vehicles' speed, whereas the effect of the pavement type was controversial. Indeed, while Singh et al. (2012) did not show effect, Medina and Tarko (2005) reported the strongest effect on the increasing speed in the case of gravel shoulder as opposed to paved one. Someone has considered the number of lanes for each traffic direction, but this variable was not significant (e.g., Semeida, 2013; Yan et al., 2017; Bhowmik et al., 2019). Finally, the effect of the median width was not significant (Bhowmik et al., 2019). As for roadside configuration, the presence of sidewalk or cycle path had a negative effect on vehicle speed (Bhowmik et al., 2019). The presence of lateral access or intersections reduced speed, being turnings difficult maneuvers that require a slowdown (e.g., Medina and Tarko, 2005; Himes et al., 2013; Russo et al., 2016; Lobo et al.,

**Table 1**
Summary of considered literature.

| Author, year | Country | Vehicle type | Cross section type | Road length [km] | Survey locations | Locations number [#] | Obs. variable | Survey instrument | Obs. number [#] | Model structure | Estimated variable | Variables (Significant) | $R^2$ (or $R^2_{adj}$)* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sil et al., 2020a | India | Cars or LCV or HCV | 4L | N/A | C | 29 | SS | Video recording | >1 015 | MLR | OS | RAG | *0.72 (cars) *0.62 (LCV) *0.58 (HCV) |
| Sil et al., 2020b | India | Cars | 4L | 100 | C | 24 | SS | Video recording | N/A | MLR | MS and DS | RAG | *0.86 (MS) *0.90 (DS) |
| Bhowmik et al., 2019 | United States | N/A | MR | N/A | S | 368 | SS | Video recording | 27 600 | OP | Proportions of VSC | RAG and CR, traffic data, adjacent land use, WC | N/A |
| Lobo et al., 2018 | Portugal | Cars | 2L | 23.5 | S | 9 | CS | GPS device | 675 | NLR | PS | RAG and CR, traffic data | N/A |
| Maji et al., 2018 | United States | N/A | 2L | N/A | T | 251 | SS | N/A | 50 200 | MLR | OS | PC, PSL | *0.85 (all highways) *0.27 (lower speed limit highways) *0.71 (higher speed limit highways) |
| Maji et al., 2018 | India | Cars or LCV or HCV | 4L | N/A | C | 13 | SS | Video recording | >650÷1 339 | MLR | OS and 98th PS | RAG, OS, OS of preceding point | *0.85÷0.95 (cars) *0.40÷0.82 (LCV) *0.46÷0.93 (HCV) |
| Maji and Tyagi, 2018 | India | Cars | 4L | N/A | C | 11 | SS | Video recording | N/A | MLR | OS | RAG, OS of preceding point | *0.47÷0.99 |
| Wang et al., 2018 | N/A | Cars and trucks | 2L | N/A | C | 219 | CS | Mixed instruments | 9 906 | NLR | MS | RAG and CR, PSL, traffic data, vertical signs, driver's sex, WC | N/A |
| Yan et al., 2017 | China | Cars or trucks | MR | N/A | C | 304 | CS | GPS device | >30 400 | R | OS | RAG | 0.65 (cars) |
| Eboli et al., 2017 | Italy | Cars | 2L | 6 | C and T | 29 | CS | GPS devices | N/A | MLR | OS | RAG and OS of preceding element | 0.94 (C) 0.91(T) |
| Hashim et al., 2016 | Egypt | Cars | 2L | 23 | C and T | 32 | CS | GPS devices | N/A | MLR | OS | RAG | 0.70÷0.73 (C) 0.85 (T) |
| Russo et al., 2016 | Italy | Cars | 2L | 184 | C and T | 509 | SS | Laser detectors | >50 900 | NLR | OS | RAG and CR | 0.70 (T) 0.99 (C) 0.89 (T + C) |
| Himes et al., 2013 | United States | Cars | 2L | N/A | S | 9 | SS | On-pavement sensors | N/A | MLR | MS and DS | RAG and CR, PS, adjacent land use, traffic data, speed characteristics | 0.88 (MS) 0.57 (DS) |
| Singh et al., 2012 | United States | Cars and trucks | 2L | N/A | T | 241 | SS | N/A | N/A | ANN | OS | RAG, traffic data, PSL, PC, accident data | N/A |
| Zuriaga et al., 2010 | Spain | Cars | 2L | 26 | C and T | 134 | CS | GPS devices | 582 | NLR | OS | RAG | 0.76÷0.84 (C), 0.52 (T) |
| Medina and Tarko, 2005 | United States | N/A | 2L | N/A | C and T | 99 | SS | Mixed instruments | > 9 900 | MLR | PS | RAG and CR, HVC, PSL, standardized PS | *0.93 (C) *0.84 (T) |
| Polus et al., 2000 | United States | Cars | 2L | N/A | T | 162 | SS | Mixed instruments | > 16 200 | NLR | OS | RAG | 0.68 (short T) 0.84 (long T) |

LCV = light commercial vehicle; HCV = heavy commercial vehicle; 2L = Two-lane roads; 4L = Four-lane divided roads; MR = Multilane roads; N/A = not available; S = segments; C = curves; T = tangents; SS = spot speed; CS = continuous speed; R = regression analysis; MLR = multiple linear regression; NLR = non-linear regression; ANN = artificial neural networks; OP = ordered probit; OS = operating speed; MS = mean speed; DS = speed standard deviation; PS = percentile speed; VSC = vehicle speed categories; RAG = Road axis geometry; CR. = cross section; PSL = posted speed limits; PC = pavement conditions; WC = weather conditions.

ARTICLE IN PRESS

2018; Bhowmik et al., 2019). Moreover, Wang et al. (2018) showed that a guardrail could reduce speed. As for other variables, increasing the stopping sight distance, the speed increases as well (e.g., Medina and Tarko, 2005). The effect of the inclusion or exclusion of posted speed limits in the operating speed models was also investigated. For example, Himes et al., (2013) showed that the posted speed limits slightly favored the increase of mean speed and speed deviation. Moreover, they argued that the posted speed limits should be included in speed prediction models unless posted speed limits do not change across sites. Conversely, other authors observed that posted speed limits were not significant for vehicular speed prediction (e.g., Bhowmik et al., 2019; Maji et al., 2018a). Furthermore, Wang et al. (2018) showed that vertical signs to signalize a curve contributed to reduce vehicular speeds. Furthermore, some authors showed that vehicles speed decreased with an increase in average daily traffic (e.g., Singh et al., 2012; Lobo et al., 2018; Bhowmik et al., 2019). Finally, the effect of the land use on speed produced contrasting results. Himes et al. (2013) showed that the presence of wooded or residential areas increased vehicular speed. Conversely, higher proportion of industrial area or shopping centers were negatively associated with the vehicular speed (Bhowmik et al., 2019). Singh et al. (2012) have demonstrated that an increase in collision rate causes a reduction in 85th percentile speed.

Undoubtedly, all these studies have contributed to the estimation of speed and provided valuable results for research and practice. However, some gaps persist.

First, almost all studies estimated the operating speed along curves and tangents that may result in a subjective choice of the representative sites. Conversely, the use of a segment may avoid this choice by considering the aggregated characteristics of the entire zone (Lobo et al., 2018). Although Bhowmik et al. (2019) and Himes et al. (2013) estimated speed on segments, they focused on the proportion of vehicle speed and the mean speed, respectively. In addition, Lobo et al. (2018) adopted continuous speed data. However, models from continuous speed data cannot predict the operating speeds in different lands or roads: the model would require a new calibration based on local speed surveys, because of the difference in driver behavior and vehicle performances (Yan et al., 2017). Indeed, these models provided a potential for a more accurate investigation of driver's behavior as opposed to an estimation of operating speed.

Second, all studies provided separate models for a road type (e.g., two-lane, four-lane or multilane roads) and frequently the type of terrain (e.g., flat, rolling, and mountainous) which is the foundation reference for data collection and predictions. Nevertheless, the inclusion of more road categories may help improve the results thanks to greater data variability. Moreover, including the type of terrain as a variable, and geometric characteristics common to different roads' categories only may help setting an 'average' model suitable for each road. This 'average' model may contribute to estimate the operating speed in a variety of roads configurations and may result straightforward to be used for road authorities.

Third, several authors argued on the need to evaluate the effect of further variables on the speed prediction (e.g., Polus et al., 2000, Sil et al., 2020a, 2020b; Singh et al., 2012). This lack was mainly observed for (i) roadside characteristics, because no studies considered the presence of some elements (e.g., wall, curb or carriage margin delineator); (ii) traffic data, as the ratios between the number of cars in free-flow conditions and the passing total traffic flow, and total traffic flow and the road capacity were not investigated yet; (iii) the rate of forbidden overtaking, which may be considered as a driver for sight distance and; (iv) the visibility of roadside marking, which may affect vehicles speed and contribute to compliance with speed limits. Indeed, it expected that a clear visibility of markings might generate a positive effect on vehicles speed and encourage drivers to maintain credible speeds owing to higher perception of safety.

This paper aims to cover the former gaps.

## Material and method

In this section, details on research context, data collection, data cleaning and data analysis methods are provided.

### Research context

This research was performed in the province of Brescia (hereafter Brescia), which is on the eastern side of the Lombardy Region (Italy). Brescia has approximately 1 250 million inhabitants distributed among 205 municipalities and represents the largest province in Lombardy, with an area of approximately 4 784 km$^2$ and a density of about 264 inhabitants per km$^2$. Moreover, Brescia is the second province of the Region and the fifth in Italy per inhabitants.

Brescia represents one of the most important industrial, commercial, and social hubs in Italy, so that it originates/attracts major vehicular traffic daily (RL, 2016). However, its county road network is undersized if compared to the traffic volumes and the territorial extension, so that the accessibility to the whole province is critical (Faccin et al., 2011). This fact is also confirmed by the relevant number of road crashes, which yearly occur in this area: in 2019, 3 356 crashes were registered, which represented the 10% of the total number of road crashes and 65% of fatalities in Lombardy. Particularly, speeding caused about 9% of all crashes (Polis, 2020). To address this challenge, the Local Road Authority (i.e., the Provincia of Brescia) was motivated to estimate the operating speed to check the road design consistency, and possibly adjust speed limits for existing (and not monitored) county roads.

### Road selection and segmentation

The most critical county road sections were selected by the Road Authority, especially those presenting geometric characteristics that do not comply with the Italian Regulation (MIT, 2001). Therefore, selected roads enabled to obtain an exten-

ARTICLE IN PRESS

*V. Martinelli, R. Ventura, M. Bonera et al.*                                    *International Journal of Transportation Science and Technology xxx (xxxx) xxx*

sive representations of infrastructure geometry, traffic flow and type of terrain. Three types of county roads have been considered (Faccin et al., 2011):

- Main rural roads, that enable the traffic distribution from the metropolitan area to the secondary road network.
- Secondary rural roads, that enable the traffic penetration into a specific territorial area. These roads are connected to the extra-provincial network.
- Local rural roads, that enable the traffic access to specific localities and municipalities.

Furthermore, these roads cross heterogeneous territories. The southern area of Brescia is mainly flat, and industrial activities are located close to the road network. The eastern part of the province is characterized by rolling land, and several farmlands. Finally, the northern part is mainly mountainous, characterized by farmlands and it is an important touristic location. The surveyed roads are shown in Fig. 1.

Next, each road has been divided into homogeneous segments. Unlike Himes et al. (2013), Lobo et al. (2018) and Bhowmik et al. (2019), in this study a homogeneous segment is defined as a portion of road where shoulder (i.e., lane and shoulder width and type) and roadside elements (e.g., guardrail, counter-wall, boundary-wall, rock-wall) are kept consistent along it. This definition could be considered as a new subdivision for the selected roads. This choice aims to highlight if also the geometric cross-section characteristics can affect the vehicles' operating speed across the segments. Indeed, in each type of terrain, besides the road axis characteristics, both the roadside configuration and the roadside elements may change as well. Indeed, the road axis is predominantly straight for the roads located at flat (or rolling) terrain. For these roads, a prediction model on tangents or curves would not enable to obtain a considerable number of road sections on which surveys should be performed, as compared to model on segments.

The surveyed segments are located outside urban areas, and cross either rural settings or zones with scattered buildings. Segments were surveyed from September 2019 to September 2020, without being discontinued during Covid-19 pandemic, to have a very interesting free-flow traffic condition. Most of the segments were surveyed[3] on weekdays, during daylight, under good weather conditions and in dry pavement conditions. A total of 143 locations on 139.67 km of road network was surveyed. Road characteristics and survey locations are summarized in Table 2, which is self-explicative.

*Data collection and pre-processing*

Since the estimation of $V_{85}$ requires several components i.e., spot speed measurements and road characteristics for each segment, different data types are required, that are gathered from different sources.

As for the spot speed measurements, these were conducted by using a laser traffic counter provided by a reputable manufacturer. This device is based on the laser beams emission and reception that perpendicularly crosses the road axis. It was located on the roads' edge parallel to the travel direction, at about 1 m from the road surface, and at midpoint of the selected segment. Moreover, the device was hidden to avoid changes in driver's behavior owing to its presence. The device recorded time (i.e., date, hour, minutes, and seconds), spot speed [km/h], vehicle length [m] and travelling direction (i.e., ascending and descending) for each passing vehicle. Surveys lasted at least one hour to have a representative data sample during not peak period. The device enabled speeds registration in both directions (for secondary and local rural roads) or single direction (for main rural roads). It is noteworthy that speed data collection was carried out only in specific environmental and traffic conditions where drivers can reach best driving performance (i.e., dry roads, free flow conditions, daylight hours, and good weather conditions). Particularly, the surveys were conducted excluding 08:00 am–10:00 am and 05:00 pm–07:00 pm, being these time periods uncongested (Provincia di Brescia, 2011).

As for the road characteristics, firstly drawing on the previous literature, road axis, cross section, roadside, marking roads, traffic data and land crossed were identified as possible factors affecting the operating speeds. Next, data on these factors (and sub factors) were collected for each segment by using several sources. Table 3 presents the explanatory factors and data sources, and provides the self-explanatory descriptive statistics (i.e., the minimum, maximum, mean, and standard deviation values at a segment level). In addition, for the sake of clarity, factors and related subfactors were reported.

Although most of factors are easily collected and computed, some others deserve a few of clarifications.

The curvature change rate of homogeneous segment ($CRR_M$) is the integral mean of the point curvature along the homogeneous segment length on the horizontal alignment.

Let: $P_j = (x_j, y_j)$ be the j<sup>th</sup> point of the polygonal that discretizes the segment; $w$ be the total number of points of the polygonal that discretizes the segment; $k_j$ be the curvature at point $P_j$; s be the curvilinear abscissa (i.e., the distance from the first point measured along the polygonal); $\Delta s_j$ be the length of $j^{th}$ polygonal segment; $\theta_j$ be the angle between two vectors tangent to polygonal by points $P_j = (x_j, y_j)$ and $P_{j+1} = (x_{j+1}, y_{j+1})$. $CCR_M$ is numerically computed as follows:

$$\Delta x_j = x_{j+1} - x_j \tag{1}$$

---

[3] Segments that are not investigated had the following characteristics: they were close to intersections or survey operating conditions were unsafe. More precisely, segments containing intersections or roundabouts were excluded from the analysis, ensuring that all the vehicles travelling the segments have right-of-way.

**Fig. 1.** Surveyed roads of Province of Brescia.

**Table 2**
Road characteristics and survey locations.

| Road name | Road type | Road length [km] | Carriage-way [n°] | Carriage-way lane [n°] | Type of Terrain | Posted[6] speed limit [km/h] | Homogeneous segments [n°] | Survey locations [n°] |
|---|---|---|---|---|---|---|---|---|
| Ring road | Main rural road | 6.85 | 2 | 2 | flat | 90 | 11 | 12[7] |
| SPBS 510 | Main rural road | 5.12 | 2 | 2 | flat | 90 | 10 | 9 |
| SP VII | Secondary rural road | 4.49 | 1 | 2 | flat | 70; 90 | 9 | 7 |
| SPBS 236 | Secondary rural road | 13.3 | 1 | 2 | flat | 70; 90 | 16 | 11 |
| SP BS 668 | Secondary rural road | 8.64 | 1 | 2 | flat | 50; 70; 90 | 16 | 12 |
| SP BS 510 | Secondary rural road | 1.55 | 1 | 2 | flat | 70; 90 | 3 | 1 |
| SP BS 510 | Secondary rural road | 19.73 | 1 | 2 | rolling | 70; 90 | 26 | 15 |
| SP XI | Local rural road | 10.34 | 1 | 2 | rolling | 50; 70; 90 | 29 | 11 |
| SP 16 | Local rural road | 11.3 | 1 | 2 | flat | 50, 70, 90 | 22 | 10 |
| SPBS 237 | Secondary rural road | 10 | 1 | 2 | mountain | 50; 90 | 27 | 18 |
| SP 48 | Secondary rural road | 12.25 | 1 | 2 | mountain | 70 | 24 | 15 |
| SP III | Local rural road | 21.1 | 1 | 2 | mountain | 90 | 36 | 9 |
| SP 5 | Local rural road | 15 | 1 | 2 | mountain | 90 | 15 | 8 |

[6] More than one posted limit may be set for some roads owing to different road characteristics.
[7] Speeds' measurement is doubled in a single segment because the device registers speeds along single direction for separate carriageways.

$$\Delta y_j = y_{j+1} - y_j \tag{2}$$

$$\Delta s_j = \sqrt{\Delta x_j^2 + \Delta y_j^2} \tag{3}$$

$$k_j = \frac{\theta_j}{\Delta s_j} \tag{4}$$

$$CCR_m = \frac{\sum_{j=1}^{w-1} k_j \Delta s_j}{\sum_{j=1}^{w-1} \Delta s_j} = \frac{\sum_{j=1}^{w-1} \theta_j}{\sum_{j=1}^{w-1} \Delta s_j} \tag{5}$$

**Table 3**
Summary characteristics for each factor (and sub-factors).

| Response Variable | Symbol | Unit of measure | Type | Data source | Minimum | Maximum | Mean | Standard Deviation |
|---|---|---|---|---|---|---|---|---|
| The 85th percentile of the distribution of the free-flow operating speed | $V_{85}$ | [km/h] | Continuous | DM | 38.00 | 108.00 | 73.91 | 14.13 |
| **Explanatory Factor** | **Symbol** | **Unit of measure** | **Type** | **Data source** | **Minimum** | **Maximum** | **Mean** | **Standard Deviation** |
| Horizontal and vertical alignment | | | | | | | | |
| Curvature change rate | $CCR_m$ | [rad/m] | Continuous | AR | $2.90 * 10^{-5}$ | $1.73 * 10^{-2}$ | $4.32 * 10^{-3}$ | $4.52 * 10^{-3}$ |
| Homogeneous segments length | $L$ | [m] | Continuous | DC | 125.00 | 2150.00 | 688.51 | 413.56 |
| Longitudinal slope | $SL$ | [-] | Continuous | DC | −0.08 | 0.08 | $-2.77 * 10^{-5}$ | 0.03 |
| Cross section | | | | | | | | |
| Lane width | $LW$ | [m] | Continuous | DM | 2.43 | 4.60 | 3.39 | 0.41 |
| Right shoulder width[8] | $RSW$ | [m] | Continuous | DM | 0.20 | 4.50 | 0.85 | 0.66 |
| Paved right shoulder | $PRS$ | [-] | Binary | OFM | 0 | 1 | 0.59 | 0.49 |
| Number of lanes | NL | [n°] | Discrete | OFM | 1 | 2 | 1.08 | 0.28 |
| Roadside | | | | | | | | |
| Number of lateral access | $NA$ | [n°] | Discrete | DC | 0 | 9 | 3 | 2 |
| Guardrail | $G$ | [-] | Binary | OFM | 0 | 1 | 0.38 | 0.49 |
| Counter-wall | $CW$ | [-] | Binary | OFM | 0 | 1 | 0.10 | 0.30 |
| Boundary-wall | $BW$ | [-] | Binary | OFM | 0 | 1 | 0.02 | 0.14 |
| Rock-wall | $RW$ | [-] | Binary | OFM | 0 | 1 | 0.08 | 0.28 |
| Curb | $C$ | [-] | Binary | OFM | 0 | 1 | 0.14 | 0.35 |
| Carriage margin delineator | $MD$ | [-] | Binary | OFM | 0 | 1 | 0.34 | 0.48 |
| Road Marking and Sign | | | | | | | | |
| Rate of forbidden overtaking | FO | [%] | Continuous | DC | 0.00 | 100.00 | 82.41 | 32.64 |
| Visible median marking | $MMv$ | [-] | Binary | OFM | 0 | 1 | 0.78 | 0.42 |
| Visible external marking | $EMv$ | [-] | Binary | OFM | 0 | 1 | 0.87 | 0.34 |
| Posted speed limit value | $PSL$ | [km/h] | Continuous | OFM | 50.00 | 90.00 | 81.04 | 12.36 |
| Posted speed limit sign | $VS$ | [-] | Binary | OFM | 0 | 1 | 0.56 | 0.50 |
| Traffic | | | | | | | | |
| Annual average daily traffic | $AADT$ | [veh/d] | Continuous | TP | 345 | 36,000 | 6868 | 5722 |
| Percentage of cars with respect to total passing flow | CAR/PF | [%] | Continuous | DM | 9.87 | 100.00 | 43.87 | 19.57 |
| Ratio between the total passing flow and road capacity | PF/MC | [-] | Continuous | DM | 0.02 | 0.97 | 0.34 | 0.28 |
| Percentage of trucks | $TR$ | [%] | Continuous | DM | 0.00 | 40.40 | 9.24 | 8.22 |
| Land crossed | | | | | | | | |
| Mountainous terrain | $MT$ | [-] | Binary | OFM | 0 | 1 | 0.40 | 0.49 |
| Flat terrain | $FT$ | [-] | Binary | OFM | 0 | 1 | 0.43 | 0.50 |

AR = Automatic Routine; DC = Digital Cartography; DM = Direct Measure; TP = Third Part; OFM = On the field measurement.

[8] RSW represents the distance between the right lane external limit and any fixed object at the roadside.

Data on the number of lateral accesses (*NA*) differ for road type: *NA* represents the number of acceleration and deceleration lanes for main rural road, the number of intersections or industrial access points for secondary rural roads, and the number of residential driveways or driveways towards main attractors (i.e., hospitals, industries, municipalities) for local rural road, respectively. These data were inferred manually by digital cartography.

The annual average daily traffic (*AADT*) was inferred by the transportation model provided by Lombardy Region. AADT was calculated from peak hour flow value, and it was referred to as equivalent cars traffic flow. The percentage of cars with respect to total passing flow (*CAR/PF*) is computed by the ratio between the number of cars (*CAR*) in free-flow conditions and the total passing flows (*PF*) recorded during the survey. Conversely, the ratio between the total passing flow and road capacity is computed by dividing the total passing flows passing along the segment and the road capacity, which is computed according to HCM (2016).

The rate of forbidden overtaking (*FO*) is referred to overtaking sight distance and is a function of road design speed (MIT, 2001). For ease, it was inferred by referring to the length of the central road marking line that is continuous. Next, it is divided by the total segment length (*L*) multiplied by 100.

Afterwards, for all binary variables, the specific characteristics were observed on the field, and recorded as input 1 if detected, and 0 otherwise. For the marking visibility, the codification has been made by setting not visible marking as the base condition. Likewise, for the land crossed (e.g., mountainous and flat terrain codification by setting rolling terrain as the base condition).

*Data cleaning*

Spot speed data were automatically collected. Therefore, some data cleaning was applied before calibrating the speed model. More precisely, a total of 2 863 between bicycles and motorcycles and 16 167 commercial vehicles were excluded

(according to a length-based criterion), because the main goal was to provide a model for cars that represent the highest quota of traffic. Preprocessed data returned 60 175 car spot speed records. Next, a time headway of 5 s was used to identify free-flow conditions (e.g., Polus et al., 2000; Medina and Tarko, 2005; Semeida, 2013; Russo et al., 2016). The $3\sigma$ statistical criterion was used to remove outliers (e.g., Esposito et al., 2011; Yan et al., 2017). Moreover, we ensured at least 100 observations in free-flow conditions at each location for each direction of passing vehicles (e.g., Polus et al., 2000; Medina and Tarko, 2005; Russo et al., 2016; Yan et al., 2017; Maji et al., 2018). Finally, data cleaning returned 27 052 car spot speed records.

As for road characteristics, the $3\sigma$ statistical criterion was used to remove outliers from the matrix of variables. However, no value was removed and a total of 6 748 elementary data were returned.

*Data analysis method*

The data analysis followed these tasks: (a) Operational speed computation. (b) Correlation analysis among explanatory $V_{85}$ variables. (c) Inferential analysis (Full regression and stepwise regression analysis for model refinement).

Task (a) is required to compute the observed operational speed (i.e., $V_{85,i,obs}$) for each road section investigated. It was performed as follows. Let:

- $N$ be the set of all surveyed locations and $i \in N$ an individual location.
- $V$ be the set spot speed and $v \in V$ an individual spot speed.
- $Z$ be the set of frequency bin of spot speed and $z \in Z$ an individual bin. This set is explained by this example: if we consider a frequency bin of each selected spot speed $v$, the corresponding $v_i$ in this bin is a member of it.

$V_{85,i,obs}$ is computed as follows: First, the absolute frequency of spot speed observations $q_{z,i}$ for a given frequency bin $z \in Z$ and for a given surveyed location $i \in N$ is computed as follows:

$$q_{z,i} = Count(v_i)_z \forall z \in Z, \forall i \in N, \forall v \in V \tag{6}$$

Second, frequency bins are sorted in ascending order. Third, the absolute cumulative frequency $Q_{z,i}$ ($\forall z \in Z, \forall i \in N$), which represents the total number of spot speed belonging to each frequency bin not higher than the considered frequency bin $z$, is computed as follows:

$$Q_{1,i} = q_{1,i} \tag{7}$$

$$Q_{2,i} = Q_{1,i} + q_{2,i} \tag{8}$$

$$Q_{|z|,i} = Q_{|z|-1,i} + q_{|z|,i} = Q_i \tag{9}$$

Fourth, the relative cumulative frequency $F_{z,i}$ ($\forall z \in Z, \forall i \in N$) is built. It represents the relative number of the observed spot speed elements which belong to each frequency bin not higher than the considered frequency bin $z$. More precisely, $F_{z,i}$ is calculated as follows:

$$F_{1,i} = \frac{Q_{1,i}}{Q_i} \tag{10}$$

$$F_{2,i} = \frac{Q_{2,i}}{Q_i} \tag{11}$$

$$F_{|z|,i} = \frac{Q_{|z|,i}}{Q_i} = 1 \tag{12}$$

Fifth, the relative cumulative frequency is plotted on a diagram to build the cumulative distribution function. The X-axis reports the spot speed bins and the Y-axis the value of the percentiles of this distribution. Sixth, the corresponding operational speed is measured by fixing the 85th-percentile of this distribution. This can be performed graphically as follows: (i) Move vertically along the segment representing the 85th percentile. (ii) Move horizontally to intercept a value returned by the cumulative distribution function of spot speeds. (3) Move vertically to intercept the segment representing the corresponding spot speed frequency bin. The value of the estimated $V_{85,i,obs}$ is the X coordinate at the interception.

Task (b) reports the result of a correlation analysis to detect the mutual statistical dependence between couples of the variables investigated. Thus, a preliminarily check to investigate the occurrence of multi-collinearity problems is performed. However, it is worth noting that correlation indexes do not necessarily disenable some (correlated) variables from appearing together in the model: they just highlight if there are multi-collinearity problems. The correlation moves towards the possible reduction of explanatory variables to have a straightforward (relatively) model to manage.

Task (c) provides the outcomes of the main factors that can affect $V_{85}$, by using statistical models. Specifically, the response variable $V_{85}$ is continuous and follows a normal distribution. Hence, the multiple linear regression (MLR) model

ARTICLE IN PRESS

*V. Martinelli, R. Ventura, M. Bonera et al.*      *International Journal of Transportation Science and Technology xxx (xxxx) xxx*

was employed. This technique is easier to use and understand, simple to interpret and straightforward to assess using basic statistics (e.g., Sil et al., 2020b). Indeed, the value of the coefficient shows the relative 'importance' of each factor in explaining the response variable, while keeping all other variables constant at their means, being the partial derivative of the response variable with respect to the considered explanatory variable. For instance, the higher the coefficient for an explanatory variable, the higher is the dependence of the response variable. The sign of parameters is also important. A negative sign implies a reduction in the speed prediction for each increase in the corresponding explanatory variable and vice versa.

More formally, let: $\widehat{V}_{85,i}$ be the predicted operating speed at location $i \in N$; $X_{ik}$ be the value of the $k^{th}$ predictor at location $i \in N$; $b_k$ be the regression coefficient associated with $k^{th}$ variable; $C$ be the constant of the regression (hyperplane intercept); $m$ be the total number of explanatory variables considered in the MLR model. The MLR model for operating speed prediction is given by:

$$\widehat{V}_{85,i} = \sum_{k=1}^{m} b_k X_{ik} + C \forall i \in N \tag{13}$$

The ordinary least squares method was used to estimate the best possible coefficients of the multiple regression models as follows. First, a full MLR model (i.e., with all the variables) was run. Next, the principle of parsimony must be considered to obtain an improved and reduced model through the selection of variables in the MRL model. Different procedures could be adopted to select the variables. For instance, the literature indicated that several variables could be identified for the purposes of their selection. However, because this paper also investigates variables that were disregarded in the literature, no specific MLR model could be exploited for this purpose. Therefore, other methods for deriving significant variables were considered such as the stepwise regression. It was applied using forward selection and backward elimination. Forward selection starts with no model terms and tests the addition of each predictor using a chosen model fit criterion. Next, predictors were incorporated, one at a time, if the inclusion gives the most statistically significant improvement of the fit. Backward elimination operates exactly in the reverse and starts with all model terms. Next, it tests the deletion of a predictor and deletes predictors, one at time, if the loss gives the most statistically insignificant deterioration of the fit. Based on the highest adjusted R-squared value (denoted by $R^2_{adj}$), F value and its corresponding p-value of the resulting model, backward or forward stepwise technique was preferred. To summarise, these techniques automatically individuate the independent variables that best contribute to observed data explanation. Moreover, they can help remove redundant and, possibly, high correlated variables from the set of explanatory variables, and improve the fit of the model.

Once the model has been estimated, it has been evaluated by the following goodness-of-fit statistics: the $R^2_{adj}$ and the highest linear correlation between predictors and the response variable, selected by global F-test and corresponding significance value. The sign of the coefficients and their significance were also evaluated. Next, the residuals analysis was performed by: (i) the predicted zero sum of residual (i.e., the random disturbance term across all predictors and observations); (ii) the absence of heteroskedasticity; and (iii) the normal distribution of residuals.

More precisely let $e_i$ be the residual of the predicted operating speed at site $i \in N$, the predicted sum is computed as follows:

$$\sum_{i \in N} e_i = \sum_{i \in N} \left( V_{85,i,obs} - \widehat{V}_{85,i} \right) = 0 \tag{14}$$

Finally, the mean absolute deviation (*MAE*) and the root mean squared error (*RMSE*) are computed (i.e., the average magnitude of the errors in a set of forecasts, without considering their direction) to better evaluate the prediction performance of the model. Since they are negatively oriented scores, lower values are better. The coefficient of variation (*CoV*) was also computed. More formally, the statistical indicators were estimated as follows.

$$MAE = \frac{1}{|N|} \sum_{i \in N} \left| V_{85,i,obs} - \widehat{V}_{85,i} \right| \tag{15}$$

$$RMSE = \sqrt{\frac{\sum_{i \in N} \left| V_{85,i,obs} - \widehat{V}_{85,i} \right|^2}{|N|}} \tag{16}$$

$$CoV = \frac{RMSE}{\frac{\left( \sum_{i \in N} \widehat{V}_{85,i} \right)}{|N|}} \tag{17}$$

All procedures were applied using both GenStat and PhStat software, but the results were reported by the software PhStat, due to its more user-friendly output.

## Results and discussion

According to task (a), the speed cumulative distribution function was empirically determined for each $i \in N$ and the speed value associated to the 85th percentile was assumed as $V_{85,i,obs}$ by applying eqns. from (6) to (12). This task had generated 251 values of $V_{85,i,obs}$ over 143 locations. Fig. 2a, 2b and 2c (self-explanatory) show the results of $V_{85,i,obs}$ for some locations, according to the different roads considered. $V_{85,i,obs}$ was estimated by applying eqns. from (6) to (12).
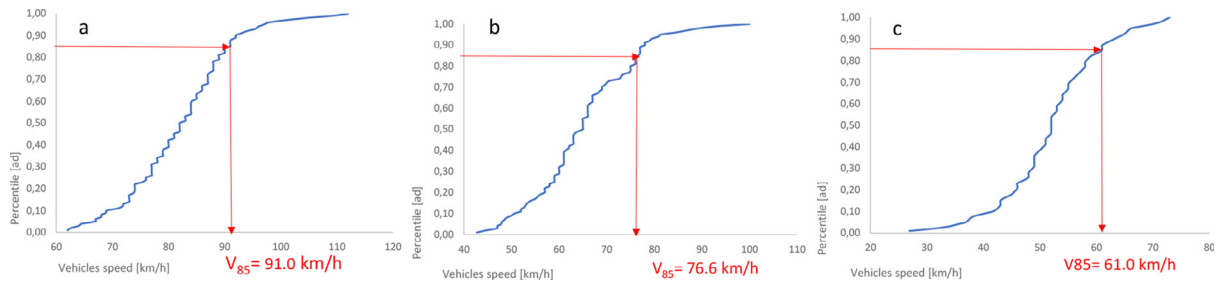
ARTICLE IN PRESS

V. Martinelli, R. Ventura, M. Bonera et al.                                    International Journal of Transportation Science and Technology xxx (xxxx) xxx

**Fig. 2.** a. Example of $V_{85}$ determination in a specific site of main road. Fig. 2b. Example of $V_{85}$ determination in a specific site of secondary road. Fig. 2c. Example of $V_{85}$ determination in a specific site of local road.

According to task (b), a correlation analysis was performed to detect if correlations between the factors exist. The results are shown in Table 4, which reports values and signs of correlation indexes between a pair-comparison of all factors.

A bold font represents correlation indexes greater than 0.5, which, to our knowledge, can be considered as a threshold separating strong from possible weak correlations. It is worth noting that most of the correlation indexes have rather small values, thus the risk of potential multi-collinearity problems appears to be low. However, for some of the $25 \times 25$ pairs of variables, the pair-wise correlation indexes have values greater than 0.5 and deserve some attention as multi-collinearity problems could occur. Particularly, a strong and negative correlation (-0.78) was found between the two variables "*Rate of forbidden overtaking*" (*FO*) and "*Number of lanes*" (*NL*). This is a somewhat expected result: indeed, the presence of multiple lanes implies a greater opportunity of overtaking and therefore a lower probability to find road sections where overtaking is prohibited. Likewise, the results show a strong positive correlation between "*Annual average daily traffic*" (*AADT*) and "*Number of lanes*" (*NL*) variables. This is an expected result, since the busiest roads are more likely to have a greater number of lanes. Similar considerations can be made for the other pairs of strongly correlated variables.

According to task (c) by applying eqn. (14), the inferential analysis was carried out. Specifically, $V_{85}$ is used as a response variable to look for a better estimate of the operating speed. Two models were estimated encompassing different road categories. Model 1 was estimated by using data collected on secondary and local roads only (i.e., two-lane county roads). Model 2 adds data of the main roads (i.e., the four-lane county roads). It is worth noting that some geometric elements for the main roads (e.g., presence of median, left shoulder width) were excluded from the analysis, because these attributes were not available on secondary and local roads. Conversely, the rate of forbidden overtaking is considered in all models because this value is zero in the case of main roads. Therefore, different roads' categories within the same model may be included.

Table 5 and Table 6 show the results at the best fits of Model 1 and Model 2, respectively. Moreover, Tables 5 and 6 report both the full and the final model, for each model. The coefficients (*Estimate*) and their significance (*p-value*), which is bold when <0.001, are shown for each model. Finally, the last part of both tables reports the summary statistics.

Generally speaking, Model 1 and Model 2 properly fit data[4]. Indeed, the statistical test on F returns a small p-value as a goodness-of-fit (<0.001). Therefore, there seems to be strong evidence for a regression effect here (i.e., not all the $b_i$ are zero). In addition, Model 1 and Model 2 explain at least 70% of $V_{85}$'s variance by the selected predictors. However, as shown in Tables 5 and 6, two models show different results. First, the number of significant variables differs (i.e., 9 for secondary and local roads, 11 for all roads). In addition, $R^2_{adj}$ and F values are smaller when $V_{85}$ is estimated excluding data relating to main rural roads. Nevertheless, we can affirm a better fit when all roads are incorporated into the regression model.

Since Model 2 contains all significant predictors of Model 1, but fits better data, it is discussed in what follows.

As for $R^2$, the results improve previous works (i.e., $R^2 = 0.68 \div 0.84$ in Polus et al., 2000; $R^2 = 0.52$ in Zuriaga et al., 2010; $R^2 = 0.70$ in Russo et al., 2016). In addition, the $R^2_{adj}$ explains 83% of $V_{85}$'s variance (using 11 predictors instead of 25) and it is consistent with some existing models estimated on tangents or segments (i.e., Medina and Tarko, 2005; Maji et al., 2018). As for SE (where a smaller value ensures better precision of line of best fit) an acceptable value is returned. As for F-test, the results ensure high significance of the selected model (*p*-value < 0.001). Therefore, we can reject the null hypothesis that all the $b_k$ are zero. As for the predictors, the result confirms that they are very significant (*p*-value < 0.001 for most of these) showing a strong regression effect. This result is also confirmed by setting a 95% confidence interval. No generic coefficient $b_k = 0$ is included in the defined interval (i.e., $b_k = 0$ value is not included in the columns "Lower 95%" and "Upper 95%"). Therefore, this result confirms that the insignificant inclination of each predictor can be rejected.

The physical meaning and the sign of the variables are discussed in what follows.

---

[4] After the application of stepwise regression, the original (final) models did not contain the variable "total passing flow" and "road capacity" (i.e.,PF/MC). A Reviewer recommended us to also include this variable, because was significant in the original (full) models. As a result of this choice, the new (final) models were significant and returned a goodness of fit, which is lower than the original (final) ones (F = 124.53 *vs* 100.40). However, this slightly lesser fit is compensated by the higher significance of other explanatory variables, such as the right shoulder width; thus, it increases the explanatory capacity of the models. Therefore, these models were preferred and shown in this study.

**Table 4**

Correlation matrix for the factors. The pairs of high correlated variables (correlation coefficient greater than 0.50 o lesser than −0.50 are bold edited).

| | CCRm | L | SL | LW | RSW | PRS | NL | NA | G | CW | BW | RW | C | MD | FO | MMv | EMv | PSL | VS | AADT | CAR/PF | PF/MC | TR | MT | FT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CCRm | 1,00 | | | | | | | | | | | | | | | | | | | | | | | | |
| L | −0,22 | 1,00 | | | | | | | | | | | | | | | | | | | | | | | |
| SL | 0,00 | 0,00 | 1,00 | | | | | | | | | | | | | | | | | | | | | | |
| LW | −0,19 | 0,17 | 0,02 | 1,00 | | | | | | | | | | | | | | | | | | | | | |
| RSW | −0,28 | 0,00 | 0,06 | 0,31 | 1,00 | | | | | | | | | | | | | | | | | | | | |
| PRS | 0,05 | 0,19 | −0,03 | 0,29 | 0,04 | 1,00 | | | | | | | | | | | | | | | | | | | |
| NL | −0,17 | −0,05 | 0,00 | 0,09 | −0,06 | −0,14 | 1,00 | | | | | | | | | | | | | | | | | | |
| NA | 0,17 | −0,22 | −0,01 | −0,34 | −0,18 | −0,12 | −0,15 | 1,00 | | | | | | | | | | | | | | | | | |
| G | −0,01 | −0,12 | −0,11 | 0,18 | 0,07 | 0,14 | 0,31 | −0,21 | 1,00 | | | | | | | | | | | | | | | | |
| CW | 0,22 | −0,07 | 0,14 | −0,10 | −0,05 | 0,26 | −0,10 | 0,08 | −0,25 | 1,00 | | | | | | | | | | | | | | | |
| BW | −0,01 | −0,11 | 0,03 | −0,03 | 0,08 | 0,00 | −0,04 | 0,08 | −0,05 | −0,05 | 1,00 | | | | | | | | | | | | | | |
| RW | 0,31 | 0,00 | 0,04 | 0,01 | −0,21 | 0,09 | −0,09 | −0,02 | −0,22 | 0,10 | −0,04 | 1,00 | | | | | | | | | | | | | |
| C | 0,03 | 0,12 | −0,01 | 0,25 | −0,20 | 0,33 | −0,14 | 0,00 | −0,33 | −0,11 | −0,07 | 0,25 | 1,00 | | | | | | | | | | | | |
| MD | −0,04 | −0,04 | −0,01 | −0,17 | 0,04 | −0,49 | −0,12 | 0,05 | **−0,50** | −0,20 | −0,04 | 0,11 | −0,10 | 1,00 | | | | | | | | | | | |
| FO | 0,33 | 0,02 | 0,00 | −0,09 | −0,01 | 0,23 | **−0,78** | 0,22 | −0,30 | 0,15 | 0,03 | 0,15 | 0,21 | 0,04 | 1,00 | | | | | | | | | | |
| MMv | −0,20 | −0,10 | 0,00 | **0,51** | 0,18 | 0,02 | 0,10 | −0,11 | 0,12 | −0,07 | 0,05 | −0,09 | 0,05 | 0,02 | −0,18 | 1,00 | | | | | | | | | |
| EMv | −0,12 | −0,03 | −0,07 | 0,03 | 0,02 | 0,18 | 0,05 | 0,03 | 0,14 | 0,06 | −0,14 | −0,11 | 0,02 | −0,26 | −0,07 | 0,24 | 1,00 | | | | | | | | |
| PSL | −0,14 | 0,17 | 0,00 | −0,09 | 0,01 | −0,15 | 0,22 | −0,02 | 0,04 | −0,03 | 0,01 | 0,00 | −0,19 | 0,10 | −0,23 | 0,05 | −0,13 | 1,00 | | | | | | | |
| VS | −0,25 | 0,06 | 0,06 | 0,34 | 0,18 | 0,10 | 0,26 | −0,14 | 0,10 | −0,08 | −0,05 | −0,17 | 0,10 | −0,03 | −0,25 | 0,19 | 0,21 | −0,37 | 1,00 | | | | | | |
| AADT | −0,18 | 0,03 | −0,05 | 0,29 | 0,09 | −0,09 | **0,73** | −0,31 | 0,34 | −0,13 | −0,01 | −0,03 | −0,05 | −0,15 | **−0,58** | 0,20 | 0,03 | 0,12 | 0,28 | 1,00 | | | | | |
| CAR/PF | **0,50** | −0,08 | 0,00 | −0,43 | −0,38 | 0,05 | −0,37 | 0,44 | −0,20 | 0,15 | 0,00 | 0,19 | 0,01 | 0,04 | 0,40 | −0,31 | −0,13 | 0,06 | −0,44 | **−0,52** | 1,00 | | | | |
| PF/MC | −0,48 | 0,23 | 0,00 | 0,41 | 0,24 | 0,06 | 0,13 | −0,34 | 0,11 | −0,19 | −0,01 | −0,22 | 0,12 | −0,06 | −0,23 | 0,30 | 0,14 | 0,05 | 0,41 | 0,33 | **−0,71** | 1,00 | | | |
| TR | −0,38 | 0,13 | 0,00 | 0,43 | 0,44 | 0,10 | 0,19 | −0,43 | 0,12 | −0,20 | −0,03 | −0,18 | 0,01 | 0,07 | −0,23 | 0,23 | 0,02 | −0,06 | 0,41 | 0,36 | **−0,57** | 0,44 | 1,00 | | |
| MT | **0,64** | −0,14 | 0,00 | −0,22 | −0,32 | 0,08 | −0,24 | 0,22 | −0,04 | 0,30 | 0,01 | 0,37 | 0,00 | 0,04 | 0,39 | −0,22 | −0,23 | 0,01 | −0,37 | −0,27 | **0,66** | **−0,63** | −0,47 | 1,00 | |
| FT | −0,41 | 0,02 | 0,00 | 0,05 | 0,24 | −0,16 | 0,36 | −0,21 | 0,11 | −0,27 | −0,12 | −0,25 | −0,19 | 0,05 | **−0,54** | 0,07 | 0,15 | −0,03 | 0,29 | 0,26 | −0,34 | 0,18 | 0,48 | **−0,66** | 1,00 |

ARTICLE IN PRESS

V. Martinelli, R. Ventura, M. Bonera et al.                                    International Journal of Transportation Science and Technology xxx (xxxx) xxx

**Table 5**

Regression results for secondary and local rural roads – Model 1.

| Explanatory Factors | Symbol | Full model | | | | Final model | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Estimate | p-value | Lower 95% | Upper 95% | Estimate | p-value | Lowe 95% r | Upper 95% |
| Constant | C | 85.06 | <0.001 | 67.44 | 102.68 | 79.41 | <0.001 | 72.06 | 86.77 |
| Curvature change rate | $CCR_m$ | −1624.35 | **<0.001** | −1929.48 | −1319.23 | −1615.81 | **<0.001** | −1846.32 | −1385.30 |
| Homogeneous segments length | L | −0.0004 | 0.73 | −0.003 | 0.002 | | | | |
| Longitudinal slope | SL | −20.04 | 0.20 | −50.67 | 10.59 | | | | |
| Lane width | LW | −2.96 | 0.11 | −6.64 | 0.72 | | | | |
| Right shoulder width | RSW | 1.06 | 0.27 | −0.82 | 2.94 | 1.47 | 0.02 | 0.20 | 2.74 |
| Paved right shoulder | PRS | 2.20 | 0.15 | −0.83 | 5.23 | | | | |
| Number of lateral accesses | NA | −1.24 | **<0.001** | −1.74 | −0.74 | −1.50 | **<0.001** | −1.87 | −1.14 |
| Guardrail | G | −1.54 | 0.43 | −5.37 | 2.29 | | | | |
| Counter-wall | CW | −0.04 | 0.99 | −4.99 | 4.91 | | | | |
| Boundary-wall | BW | 4.11 | 0.27 | −3.26 | 11.47 | | | | |
| Rock-wall | RW | 0.13 | 0.95 | −4.06 | 4.32 | | | | |
| Curb | C | −0.41 | 0.84 | −4.29 | 3.47 | | | | |
| Carriage margin delineator | MD | −0.24 | 0.89 | −3.63 | 3.15 | | | | |
| Rate of forbidden overtaking | FO | −0.06 | 0.04 | −0.12 | −0.001 | −0.07 | **<0.001** | −0.11 | −0.03 |
| Visible median marking | MMv | 4.37 | 0.04 | 0.24 | 8.50 | 4.86 | 0.001 | 2.09 | 7.63 |
| Visible external marking | EMv | −4.34 | 0.18 | −10.68 | 1.99 | −6.55 | 0.005 | −11.10 | −2.00 |
| Posted speed limit value | PSL | 0.02 | 0.76 | −0.09 | 0.12 | | | | |
| Posted speed limit sign | VS | −0.46 | 0.74 | −3.15 | 2.23 | | | | |
| Annual average daily traffic | AADT | 0.0004 | 0.12 | −0.0001 | 0.0008 | | | | |
| Percentage of cars with respect to total passing flow | CAR/PF | 0.15 | 0.004 | 0.05 | 0.25 | 0.23 | **<0.001** | 0.16 | 0.30 |
| Ratio between the total passing flow and road capacity | PF/MC | 8.47 | 0.003 | 2.89 | 14.05 | 8.18 | **<0.001** | 4.41 | 11.94 |
| Percentage of trucks | TR | −0.04 | 0.66 | −0.23 | 0.15 | | | | |
| Mountainous terrain | MT | −3.19 | 0.17 | −7.71 | 1.34 | −5.34 | **<0.001** | −7.90 | −2.79 |
| Flat terrain | FT | 2.83 | 0.14 | −0.90 | 6.55 | | | | |

| Regression statistics | DF | SS | MS | DF | SS | MS |
|---|---|---|---|---|---|---|
| Regression | 24 | 27964.99 | 1165.21 | 9 | 25886.72 | 2876.30 |
| Residual | 205 | 11671.95 | 56.94 | 198 | 6518.96 | 32.92 |
| Total | 229 | 39636.94 | | 207 | 32405.68 | |
| $R^2$ | 0.71 | | | | 0.80 | |
| $R^2_{adj}$ | 0.67 | | | | 0.79 | |
| Standard Error | 7.55 | | | | 5.74 | |
| Observations | 230 | | | | 208 | |
| F | 20.47 | | | | 87.36 | |
| Significance F | $5.21 * 10^{-42}$ | | | | $4.65 * 10^{-64}$ | |

DF = Degree of freedom; SS = Sum of squares; MS = Mean of squares.

As for the roadside axis geometry, the results shows that a 1 rad/m increase in the curvature change rate ($CCR_M$) strongly reduces $V_{85}$, while keeping all other variables constant at their means, and confirming previous research (e.g., Zuriaga et al., 2010; Esposito et al., 2011; Russo et al., 2016; Yan et al., 2017). This result is realistic because $CCR_M$ is an index of the major or minor tortuosity of a homogeneous segment. Moreover, it suggests that tortuosity of a segment strongly affect the driving behavior in terms of $V_{85}$.

As for geometry cross section, the results show that the lane width does not significantly influence $V_{85}$. Although it might seem counterintuitive, this outcome is consistent with the findings of other previous works (i.e., Polus et al., 2000; Maji et al., 2018a). Probably, the strong influence of other factors might overshadow the significance of lane width factor. Conversely, a 1 m increase in the right shoulder width (RSW) induces a slightly increase in $V_{85}$. This outcome indicates that the existence of a wide right shoulder induces a safety perception in the drivers, resulting in a higher driving speed. Coherently, the presence of paved right shoulder (PRS) increases $V_{85}$. This variable only occurs for Model 2 and confirms previous studies (Singh et al., 2012). Hence, this result suggests that drivers could consider paved shoulder the safest, and hence the driving speed would be higher. Nevertheless, in more critical segments, one may experiment a colored pavement to alert driver of a possible hazard due to the cross-section characteristics. Therefore, contrary to expectations, in defining a homogeneous segment, none of roadside elements (e.g., guardrail, counter-wall, boundary-wall, rock-wall) significantly affects the speed across the segment, apart the shoulder type. Furthermore, unlike e.g., Semeida (2013), Yan et al. (2017) and Bhowmik et al. (2019), a 1-unit increase of number of lanes (NL) increases $V_{85}$. This result is acceptable because drivers could move to the left lane (if any) where speeds are often higher than the right lane.

As for roadside configuration, the results show that a 1-unit increase of later access (NA) reduces $V_{85}$, confirming previous studies (e.g., Medina and Tarko, 2005; Semeida, 2013; Russo et al., 2016). This result suggests that cautious drivers can adapt

**Table 6**
Regression results for main, secondary, and local rural roads – Model 2.

| Explanatory Factors | Symbol | Full model | | | | Final model | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Estimate | p-value | Lower 95% | Upper 95% | Estimate | p-value | Lower 95% | Upper 95% |
| Constant | C | 76.64 | <0.001 | 57.40 | 95.89 | 69.06 | <0.001 | 59.00 | 79.11 |
| Curvature change rate | $CCR_m$ | −1605.11 | **<0.001** | −1895.87 | −1314.35 | −1491.22 | **<0.001** | −1707.51 | −1274.93 |
| Homogeneous segments length | L | −0.0002 | 0.84 | −0.003 | −0.002 | | | | |
| Longitudinal slope | SL | −20.28 | 0.19 | −50.37 | 9.81 | | | | |
| Lane width | LW | −2.32 | 0.20 | −5.87 | 1.23 | | | | |
| Right shoulder width | RSW | 1.10 | 0.24 | −0.72 | 2.93 | 1.54 | 0.01 | 0.33 | 2.75 |
| Paved right shoulder | PRS | 3.25 | 0.02 | 0.45 | 6.05 | 2.25 | **0.006** | 0.66 | 3.84 |
| Number lanes | NL | 8.45 | 0.02 | 1.29 | 15.61 | 11.63 | **<0.001** | 7.32 | 15.95 |
| Number of lateral accesses | NA | −1.18 | **<0.001** | −1.66 | −0.69 | −1.45 | **<0.001** | −1.81 | −1.10 |
| Guardrail | G | −1.14 | 0.55 | −4.86 | 2.58 | | | | |
| Counter-wall | CW | −0.37 | 0.88 | −5.20 | 4.45 | | | | |
| Boundary-wall | BW | 4.09 | 0.27 | −3.16 | 11.34 | | | | |
| Rock-wall | RW | 0.55 | 0.79 | −3.55 | 4.64 | | | | |
| Curb | C | −0.75 | 0.70 | −4.54 | 3.04 | | | | |
| Carriage margin delineator | MD | 0.10 | 0.95 | −3.22 | 3.43 | | | | |
| Rate of forbidden overtaking | FO | −0.07 | 0.02 | −0.12 | −0.01 | −1.45 | **<0.001** | −0.11 | −0.03 |
| Visible median marking | $MM_v$ | 4.38 | 0.03 | 0.32 | 8.45 | 5.06 | **<0.001** | 2.40 | 7.72 |
| Visible external marking | $EM_v$ | −4.99 | 0.12 | −11.21 | 1.24 | −8.17 | **<0.001** | −12.62 | −3.72 |
| Posted speed limit value | PSL | 0.01 | 0.80 | −0.09 | 0.11 | | | | |
| Posted speed limit sign | VS | −0.52 | 0.70 | −3.16 | 2.12 | | | | |
| Annual average daily traffic | AADT | 0.0001 | 0.31 | −0.0001 | 0.0004 | | | | |
| Percentage of cars with respect to total passing flow | CAR/PF | 0.15 | 0.004 | 0.05 | 0.25 | 0.22 | **<0.001** | 0.15 | 0.29 |
| Ratio between the total passing flow and road capacity | PF/MC | 7.87 | 0.005 | 2.46 | 13.28 | 5.5 | **<0.001** | 1.92 | 9.07 |
| Percentage of trucks | TR | −0.03 | 0.70 | −0.21 | 0.14 | | | | |
| Mountainous terrain | MT | −3.99 | 0.08 | −8.39 | 0.41 | −7.05 | **<0.001** | −9.53 | −4.57 |
| Flat terrain | FT | 2.33 | 0.20 | −1.28 | 5.94 | | | | |
| **Regression statistics** | DF | SS | | MS | | DF | SS | | MS |
| Regression | 25 | 36599.06 | | 1463.96 | | 11 | 33393.54 | | 3035.78 |
| Residual | 225 | 12476.85 | | 55.45 | | 214 | 6470.75 | | 30.24 |
| Total | 250 | 49075.91 | | | | 225 | 39864.29 | | |
| $R^2$ | 0.75 | | | | | 0.84 | | | |
| $R^2_{adj}$ | 0.72 | | | | | 0.83 | | | |
| Standard Error | 7.45 | | | | | 5.50 | | | |
| Observations | 251 | | | | | 226 | | | |
| F | 26.40 | | | | | 100.40 | | | |
| Significance F | $2.29 * 10^{-53}$ | | | | | $4.26 * 10^{-78}$ | | | |

DF = Degree of freedom; SS = Sum of squares; MS = Mean of squares.

their driving style (i.e., they could slow down the speed) when approaching any access considered (i.e., acceleration and deceleration lanes, intersections or industrial drive-ways or residential drive-ways or drive-ways towards main attractors).

As for road marking and signs, the results showed that the rate of forbidden overtaking ($FO$) affected negatively $V_{85}$. This is a new result. Since $FO$ is detected by median markings, our results suggest that a greater percentage of it could ensure a reduction in speed. In this case, the application of lane markers may further alert the driver. Furthermore, the presence of a visible median marking ($MM_V$) increases $V_{85}$. In contrast, the presence of a visible external marking ($EM_V$) decreases $V_{85}$. These are novel results, because the presence of vertical signs or marking have not been much investigated in the literature, to our knowledge. This result suggests that road marking plays an important role in driver behavior and may foster a greater perceived safety, relating to median markings. The opposite effect of external markings on $V_{85}$ could be attributed to drivers' perception of cross section narrowing. Indeed, the presence of visible external marking ensures a visual distinction between lane and right shoulder. In general, an effective marking drives the correct position of the vehicle on the road and avoid potential collisions with vehicles passing in opposite direction. Furthermore, markings result particularly effective in conditions of poor visibility (e.g., in the nighttime or in foggy and rainy weather) and in the critical points of the road. Conversely, unlike Himes (2013), posted speed limits did not result significantly influence $V_{85}$. However, even if this result might seem counterintuitive, it confirms other literature findings (e.g., Bhowmik et al., 2019; Maji et al., 2018a) and the reasons of this outcome could be various. First, the strong influence of other factors (such as vertical and horizontal alignment) could overshadow the significance of posted speed limits. Second, the visibility of speed limits signals is often poor (in the context of the study), and this could have induced a lack of awareness in the users. Third, the non-observance of speed limits is unfortunately a widespread issue in Italy and this cloud have influenced the regression outcomes.
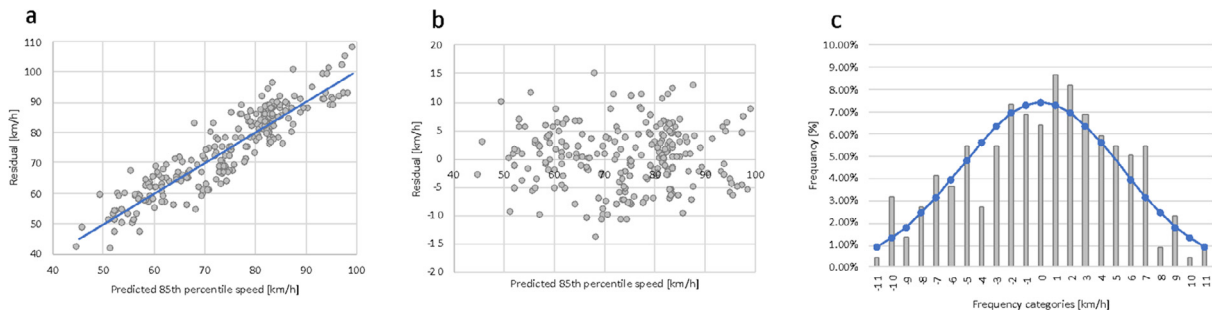
**Fig. 3.** a. Scattered plots for performance evaluation of developed model. Fig. 3b. Residuals distribution. Fig. 3c. Frequency distribution of residuals.

As for traffic data, the results show that a 1% increase of the percentage of cars with respect to total traffic flow (*CAR/PF*) increases $V_{85}$, that is a novel result. This result suggests that car drivers can maintain a larger speed (clearly in free-flow conditions) if their behavior is not conditioned by other traffic components (e.g., trucks). Moreover, an increase in the ratio between the total passing flow and road capacity (*PF/MC*) seems to induce an increase in $V_{85}$. Although this (novel) outcome can appear counterintuitive, it is noteworthy that model was fitted only considering free-flow speed data, since traffic congestion states were excluded during the data cleaning phase (see section 3.3). This implied a mean observed ratio (i.e., mean *PF/MC*) of 0.34 and then validity of the obtained model is guaranteed around this value. Thus, further research is recommended.

Finally, as for land crossed, the type of terrain (*MT*) affected negatively the $V_{85}$. This is a novel result. More precisely, drivers tend to reduce their speeds when the MT is in a mountainous terrain. Effectively, this result suggests that these roadway characteristics do not enable drivers to maintain high speeds, albeit the traffic may be lower than that of rolling or flat context. Lower speeds could be due to the major tortuosity of the road alignment or the presence of tight bends and ravines adjacent to the road. These factors cause drivers to give more attention when driving.

Once the influence of predictors on operating speed has been analyzed, some statistics are reported to further evaluate the performance of the estimated model. More precisely the observed vs predicted $V_{85}$, the analysis of residuals, and the prediction capacity of this model are discussed.

From the available dataset, the scattered plots show that predicted $V_{85}$ values are uniformly distributed along the expected $V_{85}$ values and are close to the ideal situation (Fig. 3a).

As for the residual analysis, the application of Eq. (14) returns a value of $2.24 * 10^{-14}$, which is effectively close to zero. Furthermore, the absence of heteroskedasticity is verified. That means that residuals vary by the same amount with higher or lower values of explanatory variables. The points cloud is not characterized by a well-defined pattern (Fig. 3b). Since the points on the plot distribute on both side of x and y axes, the residual values are randomly located between $\pm 2\sigma$. Thus, the regression equation can better reflect the rules (Yan et al., 2017). Most of residuals are within the interval $\pm 10.72$ (i.e., two standard deviations of the mean value known as residual range $\mu \pm 2\sigma$), and more than 94 % of the sites had residuals within the residual range.

Fig. 3c shows that the frequency distribution of residuals overlaps the normal distribution, for the most. Therefore, the histogram of the residuals of the model is a good fit to the normal distribution curve.

Moreover, we verified the prediction performance (capacity) of the model by computing *MAE*, *RMSE* and *CoV*. The statistical indicators were estimated using Eq. (15), (16) and (17), and were compared with the results on tangents, curves, and segments (related to speed prediction model for passenger cars) of other studies[5] as shown in Table 7.

Specifically, Table 7 shows that values of statistical indicators are consistent with past research. Moreover, earlier studies indicated that the models may be acceptable when *MAE*, *RMSE* and *CoV* are within 15.00, 19.20 and 0.20, respectively (Sil et al., 2020a). Therefore, we can conclude that Model 2 is reliable in predicting $V_{85}$ (as well as Model 1), because low *MAE*, *RMSE* and *CoV* were obtained that are also within the abovementioned limits.

Finally, to check further the viability of the developed model as an aiding support tool in road speed diagnosis, a validation procedure has been performed, by comparing observed and estimated operating speed and by analyzing the residuals. The procedure considered different roads from those used during calibration phase (i.e., road segments where the detector was not available). These roads were selected from the Provincia di Brescia (the funder of this research) that also requested for this validation. The investigated roads presented an average length of 10 km. For this validation, the authors sampled a number of locations that corresponds to about 10% of the original number of observations for fitting the model, according to previous studies (i.e., Cvitanić and Maljković, 2017; Russo et al., 2016). The results of *MAE*, *RMSE* and *CoV* (3.47 km/h, 3.79 km/h and 0.04, respectively) further confirmed the transferability of the developed model for road segments where the traffic surveys were not available.

---

[5] A scrupulous comparison could not be conducted, because only two similar studies have been performed (*e.g.*, Himes et al., 2013; Lobo et al., 2018). Therefore, all studies containing previous indicators have been reported.

ARTICLE IN PRESS

V. Martinelli, R. Ventura, M. Bonera et al.                                    International Journal of Transportation Science and Technology xxx (xxxx) xxx

**Table 7**
Comparison results of statistical indicators.

| Authors | $MAE[km/h]$ | $RMSE[km/h]$ | $CoV[ad]$ |
|---|---|---|---|
| Zuriaga et al., 2010 | N/A | 3.92 ÷ 5.91*, 8.80** | N/A |
| Esposito et al., 2011 | 4.44 ÷ 9.48*, 4.26 ÷ 8.82** | 5.57 ÷ 11.53*, 5.30 ÷ 10.86** | 0.07 ÷ 0.15*; 0.06 ÷ 0.13** |
| Himes et al., 2013 | N/A | 1.45 ÷ 5.44**** | N/A |
| Semeida, 2013 | N/A | 3.11 ÷ 10.32** | N/A |
| Russo et al., 2016 | 8*, 7**, 7*** | 10.02*, 8.59**, 8.66*** | 0.14*, 0.10**, 0.02*** |
| Eboli et al., 2017 | N/A | 4.33*, 6.27** | N/A |
| Yan et al., 2017 | 2.69* | 3.54* | N/A |
| Lobo et al., 2018 | 3.4**** | 5.67**** | N/A |
| Majii et al., 2018a | N/A | 4.37* | N/A |
| Maji et al., 2018b | 2.15 ÷ 2.31** | 2.50 ÷ 2.95** | 0.05 ÷ 0.06** |
| Sil et al., 2020a | 7.08* | 7.72* | 0.1* |
| This paper - Model 1 | 4.48**** | 5.60**** | 0.08**** |
| This paper - Model 2 | **4.30****** | **5.35****** | **0.07****** |

(*) curve; (**) tangent; (***) curve and tangent; (****) segment.

## Conclusions and research perspectives

Vehicle operating speed is a crucial factor of road safety, that may severely affect the occurrences and severity of crashes on rural roads. The inconsistency of roads' geometric design combined with drivers' inappropriate behavior (i.e., speeding) could be potential causes for risk collisions. Therefore, checking the consistency of successive road elements and providing a possible support to adjust speed limits, can make an important contribution to achieve road safety targets. This is particularly relevant for existing roads that were built without a specific and solid normative. Usually, the 85th percentile of the operating speed (i.e., $V_{85}$) distributions in free-flow traffic is considered for this scope. Although the computation of this percentile is quite simple, many road authorities cannot collect speed data for each road, thus, they might not be able to evaluate roads' operating conditions relating to consistency and safety. Therefore, the use of prediction models could be a useful tool to define the relationship between the speed on roadway facilities and several predictors (e.g., geometric characteristics, traffic data, adjacent land, and driver characteristics).

Although the literature provided relevant models for this purpose, this paper contributes to the field in a threefold manner, as follows.

- A partition of county roads in homogeneous segments is considered for the $V_{85}$ estimation. This partition is based on the change in cross-section characteristics for the selected roads. This choice enables to highlight if also the cross-section characteristics can affect the vehicles speed across the segments. Conversely to expectations, only the presence of a wider or paved shoulder and the number of lanes affected positively the operating speed $V_{85}$.
- The inclusion of all road categories in a single prevision 'average' model, set up by a new variable, *i.e.,* type of terrain, and the inclusion of variables common to different roads. The main advantage of this choice is that it enables to involve the high degree of variability that may characterize a territory and the road in a unique prevision model. Consequently, it may result in an improved fitting performance if compared with those of single specific models fitted for each road category.
- The introduction of new predictors, in addition to the abovementioned, i.e., the ratio between the number of cars in free-flow conditions and all passing traffic flow, the ratio between the total passing flow and road capacity, the visibility of road markings (median and external) and the rate of forbidden overtaking. The new variables enlarged the results of previous studies, because they helped refine the $V_{85}$ estimation.

Two models have been specified, calibrated and tested using 60 000+ real data on the Province of Brescia (Italy). The main results showed the $V_{85}$ increases when a wider or paved shoulder and visible median road markings occur, and when the number of lanes and the percentage of cars with respect to total passing flow increase as well. Conversely, $V_{85}$ decreases as the curvature of road axis, the number of lateral accesses and the rate of forbidden overtaking increase. In addition, the visibility of external road markings and the road's path on a mountainous terrain are factors that decrease the speed. These models could be a useful tool to the Province to identify inconsistent and unsafe county road segments and, possibly adjust safe speed limits on its overall county road network.

Nevertheless, this research is not very large in scale and further data collection is recommended to confirm the results. However, this study is sufficiently large to contribute to the research evidence base on this topic, because we provide a more explicit indication of key determinants affecting $V_{85}$, especially if the high degree of context variability is considered. The findings can also be applied to similar contexts with comparable elements, e.g., the overall road length, terrain type, land use. Conversely, the model is straightforward and has general validity. Thus, providing new input data, it can be applied to any context.

Nevertheless, the inclusion of all road categories in a single prevision 'average' model could induce a high variability in the values assumed by the explanatory variables, which could lead to unstable model estimation. This disadvantage, that is largely compensated by the higher fitting performance of the obtained model, will be deeply investigated in future studies.

ARTICLE IN PRESS

*V. Martinelli, R. Ventura, M. Bonera et al.*                                    *International Journal of Transportation Science and Technology xxx (xxxx) xxx*

Besides the stated limitations, further research can be developed to improve the results by using more data and testing the influence of roadside elements even more. A further refinement of $V_{85}$ estimation may be searched by applying the Montecarlo simulation methods, owing to the numerous sources of uncertainties in data collection. Thus, a larger database may be implemented for those segments where few data were available. Finally, machine learning algorithms build models based on sample data, known as "training data", to make predictions or decisions without being explicitly programmed to do so. Therefore, the application of these algorithms may open rooms for a more accurate prediction of $V_{85}$.

## CRediT authorship contribution statement

**Valentina Martinelli:** Methodology, Formal analysis, Data Curation, Writing - original draft, Writing - review & editing, Visualization. **Roberto Ventura:** Methodology, Formal analysis, Data Curation, Writing - review & editing, Visualization. **Michela Bonera:** Writing - review & editing, Visualization. **Benedetto Barabino:** Conceptualization, Methodology, Data Curation, Writing - original draft, Writing - review & editing, Supervision, Project administration. **Giulio Maternini:** Conceptualization, Supervision, Funding acquisition.

## Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

## References

Aarts, L., van Nes, N., Wegman, F., Van Schagen, I.N.L.G., Louwerse, R., 2009. January. Safe speeds and credible speed limits (sacredspeed): a new vision for decision making on speed management. In: Compendium of Papers of the 88th TRB Annual Meeting, pp. 11–15.

ACI - Automobile Club d'Italia, 2019. Incidenti stradali, anno 2019. (In Italian).

Barabino, B., Di Francesco, M., Mozzoni, S., 2017. Time reliability measures in bus transport services from the accurate use of automatic vehicle location raw data. Quality Reliability Eng. Int. 33 (5), 969–978.

Bhowmik, T., Yasmin, S., Eluru, N., 2019. A multilevel generalized ordered probit fractional split model for analyzing vehicle speed. Anal. Methods Accid. Res. 21, 13–31.

Cvitanić, D., Maljković, B., 2017. Operating speed models of two-lane rural state roads developed on continuous speed data. Tehnicki vjesnik/Technical Gazette 24 (6).

Eboli, L., Guido, G., Mazzulla, G., Pungillo, G., 2017. Experimental relationships between operating speeds of successive road design elements in two-lane rural highways. Transport 32 (2), 138–145.

Esposito, T., Mauro, R., Russo, F., Dell'Acqua, G., 2011. Speed prediction models for sustainable road safety management. Procedia-social Behav. Sci. 20, 568–576.

ETCS - European Transport Safety Council, 2019. Reducing Speeding in Europe. PIN Flash report 36.

Faccin C., Zavanella L., Savoldi E., Boroni A., Rossini P., 2011. Provincia di Brescia. Piano del traffico della viabilità extraurbana. (In Italian).

Martinelli, V., Ventura, R., Bonera, M., Barabino, B., Maternini, G., 2022. Effects of urban road environment on vehicular speed. Evidence from Brescia (Italy). Transp. Res. Proc. 60, 592–599.

Medina, A.M.F., Tarko, A.P., 2005. Speed factors on two-lane rural highways in free-flow conditions. Transp. Res. Rec. 1912 (1), 39–46.

Forbes, G., Gardner, T., McGee, H. W., Srinivasan, R., 2012. Methods and practices for setting speed limits: An informational report (No. FHWA-SA-12-004). United States. Federal Highway Administration. Office of Safety.

Hashim, I.H., Abdel-Wahed, T.A., Moustafa, Y., 2016. Toward an operating speed profile model for rural two-lane roads in Egypt. J. Traffic Transp. Eng. (Engl. Ed.) 3 (1), 82–88.

HCM - Highway Capacity Manual, 2016. 6th Edition: A Guide for Multimodal Mobility Analysis. Transportation Research Board, Washington, D.C.

Himes, S.C., Donnell, E.T., Porter, R.J., 2013. Posted speed limit: to include or not to include in operating speed models. Transp. Res. A: Policy Pract. 52, 23–33.

Lamm, R., Psarianos, B., Mailaender, T., 1999. Highway design and traffic safety engineering handbook.

Lobo, A., Amorim, M., Rodrigues, C., Couto, A., 2018. Modelling the operating speed in segments of two-lane highways from probe vehicle data: a stochastic frontier approach. J. Adv. Transp. 2018.

Maji, A., Tyagi, A., 2018. Speed prediction models for car and sports utility vehicle at locations along four-lane median divided horizontal curves. J. Modern Transp. 26 (4), 278–284.

Maji, A., Sil, G., Tyagi, A., 2018. 85th and 98th percentile speed prediction models of car, light, and heavy commercial vehicles for four-lane divided rural highways. J. Transp. Eng., Part A: Syst. 144 (5), 04018009.

Maji, A., Singh, D., Agrawal, N., Zaman, M., 2018b. Operating speed prediction models for tangent sections of two-lane rural highways in Oklahoma State. Transp. Lett., 1–8.

MIT - Ministry of Infrastructure and Transport, 2001. Norme funzionali e geometriche per la costruzione delle strade. DM (6792/2001). SO alla GU n.3. Del 4.1.2002. (In Italian).

MIT - Ministry of Infrastructure and Transport, 2006. Direttiva sulla corretta ed uniforme applicazione delle norme del codice della strada in materia di segnaletica e criteri per l'installazione e la manutenzione. Prot. N. 777 (In Italian).

Polis – Regione Lombardia, 2020. Incidenti stradali a ciclisti regione Lombardia. (In Italian).

Polus, A., Fitzpatrick, K., Fambro, D.B., 2000. Predicting operating speeds on tangent sections of two-lane rural highways. Transp. Res. Rec. 1737 (1), 50–57.

Provincia di Brescia, 2011. Piano del Traffico della Viabilità Extraurbana.

Regione Lombardia – RL, 2016. Piano Regionale per la Mobilità e i Trasporti.

ARTICLE IN PRESS

*V. Martinelli, R. Ventura, M. Bonera et al.*                    *International Journal of Transportation Science and Technology xxx (xxxx) xxx*

Russo, F., Antonio Biancardo, S., Busiello, M., 2016. Operating speed as a key factor in studying the driver behaviour in a rural context. Transport 31 (2), 260–270.

Semeida, A.M., 2013. Impact of highway geometry and posted speed on operating speed at multi-lane highways in Egypt. J. Adv. Res. 4 (6), 515–523.

Sil, G., Nama, S., Maji, A., Maurya, A.K., 2020a. Modeling 85th percentile speed using spatially evaluated free-flow vehicles for consistency-based geometric design. J. Transp. Eng., Part A: Syst. 146 (2), 04019060.

Sil, G., Nama, S., Maji, A., Maurya, A.K., 2020b. Effect of horizontal curve geometry on vehicle speed distribution: a four-lane divided highway study. Transp. Lett. 12 (10), 713–722.

Singh, D., Zaman, M., White, L., 2012. Neural network modeling of 85th percentile speed for two-lane rural highways. Transp. Res. Rec. 2301 (1), 17–27.

Ventura, R., Barabino, B., Vetturi, D., Maternini, G., 2020. Bridge safety analysis based on the function of exceptional vehicle transit speed. Open Transp. J. 14 (1), 222–236.

Wang, B., Hallmark, S., Savolainen, P., Dong, J., 2018. Examining vehicle operating speeds on rural two-lane curves using naturalistic driving data. Accid. Anal. Prev. 118, 236–243.

Yan, Y., Li, G., Tang, J., Guo, Z., 2017. A novel approach for operating speed continuous predication based on alignment space comprehensive index. J. Adv. Transp. 2017.

Zuriaga, A.M.P., García, A.G., Torregrosa, F.J.C., D'Attoma, P., 2010. Modeling operating speed and deceleration on two-lane rural roads with global positioning system data. Transp. Res. Rec. 2171 (1), 11–20.

## Further Reading

AA.VV. (2018). Brescia. Documento unico di programmazione 2016 – 2018. (In Italian).

Institute of Transportation Engineers – ITE (1993). Speed Zoning Guidelines. A proposed recommended practice.