Random Effect Models to Predict Operating Speed Distribution on Rural Two-Lane Highways

RANDOM EFFECT MODELS TO PREDICT OPERATING SPEED DISTRIBUTION ON RURAL TWO-LANE HIGHWAYS

Marco Bassani, Ph.D. (*corresponding author*)

Associate Professor, Politecnico di Torino, Department of Environment, Land and Infrastructures Engineering (DIATI)

24, corso Duca degli Abruzzi, Torino, I, 10129

Phone: +39 011 090 5635, E-mail: marco.bassani@polito.it

Cinzia Cirillo, Ph.D.

Associate Professor, University of Maryland, Department of Civil and Environmental Engineering

3250 Kim Building, 20742 College Park, MD, US

Phone: +1 301 405 6864, Fax: +1 301 405 2585, E-mail: ccirillo@umd.edu

Serena Molinari

MS student, Politecnico di Torino, Department of Environment, Land and Infrastructures Engineering (DIATI)

24, corso Duca degli Abruzzi, Torino, I, 10129

E-mail: serena.molinari@studenti.polito.it

Jean-Michel Tremblay

PhD student, University of Maryland, Department of Civil and Environmental Engineering

3250 Kim Building, 20742 College Park, MD, US

E-mail: jeanmi.tremblay@gmail.com

ABSTRACT

This paper presents the results obtained from the estimation of free-flow speed on two-lane rural highways. The data used for the analysis were collected in the Northwest of Italy using video cameras and a laser speed gun. The model structure adopted separates the estimate of the central tendency of speeds from the typical deviations of individual speeds. Hence, in the model the same set of variables can be used to determine both the mean value and the standard deviation of the speed distribution; the desired speed percentile is then calculated by considering the associated standard normal random variable (Z). Random effects (RE) were included in the model to account for the variability in time and space of the data that contain multiple measurements for the same road/section/direction and to remove any dependency between estimation errors from individual observations.

INTRODUCTION

Speed is a fundamental variable in the geometric design of highways and streets. Transportation engineers normally refer to the design speed when calculating the characteristics of geometric elements of the road, and to the operating speed when assessing the consistency of the adopted design values along the designed road alignment. Hence, operating speed models are fundamental to road design since they can anticipate the speeds that drivers will adopt.

Operating speeds reflect the speed behavior of drivers who are affected by the horizontal and vertical alignments as well as the cross section. Therefore, operating speed data are collected from isolated vehicles moving in free-flow conditions.

Free-flow speeds are, generally, normally distributed as indicated in many contributions (Transportation Research Circular 2011, Bassani et al. 2014). Usually, to assess if speed design consistency has been achieved, engineers calculate the 85^{th} percentile of the distribution (V_{85}), (Hassan 2004) which separates the population of prudent drivers from the small group of more aggressive drivers. Despite this widespread approach, some commentators contend that a good knowledge of the parameters describing the entire distribution is more powerful and useful for applications and inferences (Bonneson 2001, Transportation Research Circular 2011).

Figueroa Medina and Tarko (2005) illustrated this tenet with the following example. Two different distributions, the first with a low mean speed (50 km/h) and high standard deviation (15 km/h), and the second with a high mean speed (60 km/h) and low standard deviation (5 km/h), have the same V_{85} (65 km/h). Hence, the V_{85} alone is not able to provide a comprehensive interpretation of the effect of road geometrics on operating speeds and their distribution.

Several variables influence operating speed. If on urban streets cross sectional and environmental variables seem to be more significant in modeling (Wang et al. 2006, Bassani and Mutani 2013), on rural highways the characteristics describing horizontal and vertical alignments are deemed to be significant in a number of scientific and technical contributions (Transportation Research Circular 2011, National Cooperative Highway Research Program 2003). Despite the fact that a considerable amount of research on operating speeds has been produced in the last twenty years, more recent works have aimed at improving model predictions and at extending the spectrum of road typologies to those types that have not been fully investigated.

In July 2011 the Operational Effects of Geometrics Committee of the Transportation Research Board sponsored the publication of the E-C 151 Circular (Transportation Research Circular 2011). This document included some important remarks regarding the current level of research on the topic, and some criticisms and suggestions were made to improve model applicability and speed predictability.

Most models available in literature for cars can predict the V_{85} on horizontal curves, assuming they (the cars) maintain a constant speed throughout. Some can estimate speed variations approaching and exiting curves (Misaghi and Hassan 2005, Castro et al. 2011), and a few consider the possibility of variation in speeds within a curve (Figueroa Medina and Tarko 2007). In literature, only two models allow for the evaluation of the speed distribution. Figueroa Medina and Tarko (2005), under the assumption of normally distributed speeds, proposed a unique equation able to estimate any percentile of the mean (m) speed and the standard deviation (σ) in this form:

$$V_{ip} = m_i + Z_p \cdot \sigma_i + \varepsilon_i \qquad \text{(eq. 1)}$$

in which i represents the generic observed speed at a certain percentile p, and Z_p is the standardized normal variable corresponding to a selected percentile (p).

Himes et al. (2011) proposed the use of three simultaneous equations: one to model the posted speed limit (PSL) to be included afterwards in two other linear equations, which estimate the mean (m) speed and the standard deviation (σ) of speed respectively:

$$m = \alpha_m + \beta_m \cdot X_m + \gamma_s \cdot PSL + \varepsilon_m \qquad \text{(eq. 2)}$$

$$\sigma = \alpha_\sigma + \beta_\sigma \cdot X_\sigma + \gamma_\sigma \cdot PSL + \theta_\sigma \cdot m + \varepsilon_\sigma \qquad \text{(eq. 3)}$$

where α is a constant; X are vectors of exogenous geometric and traffic variables for mean speed (X_m) and speed deviation (X_σ); β, γ, and θ are vectors of estimable regression parameters for exogenous variables (X) and endogenous ones (PSL, mean speed and speed deviation respectively); and ε is the random disturbance term. The authors found that although the standard deviation of speed is affected by mean speed, the converse is not true; hence, they included the mean in eq. 3.

In the E-C 151 Circular, some remarks were also made on variables: many models contain the parameters describing the horizontal curvature, while few include variables related to the vertical alignment (i.e., grade), tangent, horizontal–vertical combinations, cross-section elements (i.e., width of lanes), available sight distance, and posted speeds.

Starting from this state of the art, our research proposes to calibrate linear regression models that are able to predict any speed percentile. The model formulation also takes into account possible random effects caused by the fact that multiple observations were taken at sections randomly selected from roads forming part of the two-lane highway network in the Northwest of Italy.

SPEED DATABASE

The speed database was compiled at various stages between 2005 and 2011 with data from several road sections in the provinces of Turin, Vercelli and Alessandria (Italy). A series of observational surveys were carried out on several typologies of rural roads including freeways, multilane, and two-lane highways on curved and tangent sections. In this investigation, only the sub-database of two-lane rural roads was considered for model calibration. Speed data were collected from a total of 13 roads and 37 sections, with the final database containing 6,567 free-flow speed observations used to calibrate the speed models. During the surveys, no data were measured in one of these lanes, thus bringing the total number of individual lanes investigated to 73. Table 1 lists the identification code for roads and sections, the name, the length of the road, and the main geometric characteristics of the sections. In particular, the table reports the lane width, the radius of the centerline (the symbol ∞ denotes tangent sections), and the average longitudinal grade across the sections. All the geometric characteristics were derived from regional GIS databases.

Table 1 also reports the minimum, the maximum, the V_{85} of speeds and the posted speed limit (PSL); furthermore, it includes the number of speed data in both directions present in the database.

For certain roads, several cross sections were selected when differences in the geometric characteristics and/or margin treatments were observed. The minimum distance between the closest sections of the same road was set equal to 2 km, and some surveys were carried out at different times.

Speed data were collected using two different techniques (Fig. 1): longitudinal measurement by means of a laser speed gun (Fig. 1A), and cross-sectional measurement by means of a digital video camera positioned perpendicularly to the road axis (Fig. 1B). In the case of longitudinal measurement, the laser speed gun was located close to the shoulder in order to decrease by as much as possible the angle α (Fig. 1A), thus minimizing the underestimation of speed due to the cosine effect. In the case of transversal measurement, the video camera (A) was placed at a sufficient distance from the shoulder to contain the two references B and C in the sequence of frames. The two lengths L_1 and L_2 were calculated by direct measurement of the distances L, H, H_1, and H_2 ($L_1=L \cdot H_1/H$, $L_2=L \cdot H_2/H$). The time necessary to cover the two lengths (Δt_1 and Δt_2) was measured from the video recording ($\Delta t_1 = t'_1 - t_1$, $\Delta t_2 = t'_2 - t_2$). The average speeds along the measurement section were determined simply by the ratio between distances and times ($v_1 = L_1/\Delta t_1$ and $v_2 = L_2/\Delta t_2$). The type of technique used depended on the characteristics of the survey site, and the observation points were carefully selected so as to minimize any disturbance to traffic and to avoid any change in driver behavior.

To merge the data coming from the longitudinal and transversal surveys into the same dataset, a preliminary comparison of the speeds detected with the two devices was carried out. In particular, differences in speed distribution ascribable to the precision and accuracy of the two measurement systems were carefully checked. A comparison of the speeds detected for the same vehicle by the two methods is reported in Fig. 2. All the data fall between the ± 10% lines across the equality line. The high coefficient of determination and the small standard error confirmed the possibility of using the two datasets jointly to calibrate the same speed model.

Speed data were collected under free-flow conditions assuming a minimum headway of 6 s, and at points where drivers assumed constant speeds, hence along tangents in sections far from curves, and in sections located at the center of curves. Each observation included in the database was associated firstly with the corresponding percentile p and secondly with the standardized normal variable Z_p, derived from the average and standard deviation calculated on the sub-dataset of the lane ($Z_p = 0$ when p = 50%, and $Z_p = 1.036$ when p = 85%).

Field inspections were carried out to collect information on the geometric characteristics of the cross section (i.e., lane and shoulder width, posted speed limit, as well as the presence of driveways, retaining walls, ramps and barriers). Such investigations were supported by the collection of the same information from aerial views on Google Earth®. Regarding the curvature (corresponding to the inverse of the radius) and the longitudinal grade, data were obtained from GIS databases thanks to the cooperation of the public agencies that manage the road networks.

Each set of measurements was subjected to the chi-square (χ^2) and the Kolmorogov-Smirnov tests to assess whether data were normally distributed. In all cases, the tests were successful. It should be noted that fifteen sections pertaining to roads SP230-VC and SP177-TO had less than 100 speed observations, which is typically the minimum number of observations recommended in speed surveys. This was due to the very low traffic volumes encountered during the surveys. In these cases, the goodness of fit tests demonstrated that in each direction the number of speed data were sufficient to form a normal distribution. Moreover, since the minimum and maximum speed ranges were sufficiently broad, it was reasonable to consider the recorded speed data representative of the speed conditions for those sections.

The posted speed limit was between 50 and 90 km/h (30 and 55 mi/h), and the maximum grade was ± 8.5%. Fourteen of thirty-seven sections were on curves with radii ranging between 150 and 8,351.25 m. The lane width was between 3 and 3.8 m. The road characteristics in Table 1 are representative of typical Italian two-lane rural highways, where the V_{85} is frequently observed to be above the posted speed limit (in the dataset this happens 32 times out of a total of

37 observations). In particular, the average values of V_{85} were higher than the PSL by 20.1 km/h at 50 km/h, by 19.4 km/h at 70 km/h, and by 5.7 km/h at 90 km/h.

Data were differentiated according to the direction of driving, since the same road feature may produce different effects depending on the side occupied with respect to the driving trajectory (right and closer, left and farther). Hence, all the elements located on the roadside were considered two times for the two driving directions.

Table 2 contains the list of the thirty-three variables considered in the investigation; raw statistics (minimum and maximum values, average – μ – and standard deviation – σ) are also provided which indicate the variability of each parameter, and consequently the field of validity for the models. Furthermore, the frequency, which counts the number of times a variable has a non-zero value in the database, has also been reported. The variables included in Table 2 are a mix of numerical continuous, numerical discrete and Boolean, and are denoted in the table by the symbols NC, ND and B respectively. In the case of Boolean variables, 0 indicates that the element is absent, while 1 indicates that the element is present.

Some variables have been considered twice in an effort to understand if either their presence or density affects driver behavior. All the variables characterized by units expressed in No./km were estimated by summing the number of elements (i.e., ramps, driveways, intersections, and pedestrian crossings) in a section of 1 km across each investigated cross section.

Throughout the surveys, the authors dedicated particular attention to the roadside characteristics, since these are normally not taken into account in operating speed investigations on two-lane rural roads (Transportation Research Circular 2011).

For both the right (R) and left (L) sides, the presence and width of shoulders (SR, SL, SRW, and SLW respectively), the presence and density of ramps (RLS, RRS, TRDLS, and TRDRS), the presence and the density of driveways (DLS, DRS, DDLS, and DDRS), sidewalks (SLS and SRS), parking lanes (PKLLS and PKLRS), safety barriers (SBLS and SBRS), and retaining walls

(WLS and WRS) were carefully noted as well as the presence and density of pedestrian crossings (Ped and PedD). The presence along the margins of lay-bys, paved areas at the side of a highway designated for use as emergency parking, has been taken into account with the two variables LBSR and LBSL.

Finally, the authors decided to include the posted speed limit (PSL) in the model thus assuming it to be an exogenous variable similar to that indicated in Himes et al. (2013). It should be noted that the current practice of Italian road Agencies is to omit the PSL as a function of the observed operating speed and, thereby, of the geometric design of the road.

MODELING APPROACH

In order to evaluate the operating speed, the authors adopted the model structure proposed by Figueroa Medina and Tarko (2005) and used by the former in a previous work to estimate the operating speed in an urban environment (Bassani et al. 2014). This model structure separates the estimate of the central tendency of speeds from the typical deviation of individual speeds, which is a function of the driving skills and decisions of individual drivers. In the speed dataset, each speed data is a speed percentile (p) for a single direction (d), on a specific section (s) and on a specific road (r), thus the sample dataset consists of data randomly extracted from sections and roads along the road network.

Random Effect Model

The data collected according to the methodology previously described contains multiple measurements from different sections of the same road (i.e., road #3, 5, 6, and 7 of Table 1). Multiple observations were, in fact, available for the same road, the same section and for both directions. Hence, random effects (RE), which correctly account for the sampling design, were included in the model to remove any dependency between estimation errors from individual observations.

RE evaluate the existence of any difference between the speed predictions for all directions/sections/roads and the corresponding predictions for a specific direction/section/road. They are considered to be normally distributed as per the following:

$$\alpha \sim N\left(0, \sigma^2\right) \quad \text{(eq. 4)}$$

The dependent variable ($V_{rsd,i}$), which represents the generic observed speed (i) at a certain percentile (p) in a direction (d), section (s) and road (r), is then derived from an RE model as follows:

$$V_{rsd,i} = \beta_o + \sum_{k=1}^{K} \beta_k^C \cdot X_{ki} + \sum_{j=1}^{J} \beta_j^D \cdot \left(Z_p \cdot X_{ji}\right) + \sum_n \alpha + \varepsilon_{rsd,i} \quad \text{(eq. 5)}$$

in which β_0 is the general model intercept, β_k^C and β_j^D are calibration parameters for the variables affecting the estimated mean (X_k), and the estimated standard deviation (X_j) respectively, Z_p is the standardized normal variable, $\sum_n \alpha$ is the sum of normally distributed random effects (α), and $\varepsilon_{rsd,i}$ is the error associated with each measurement. In eq. 5, the second term represents the central tendency term, while the third represents the dispersion term.

If random effects are excluded from eq. 5, the model becomes a fixed effect model, which is widely used in literature to model operating speeds. In contrast to the RE model, the FE model does not correctly account for the sampling design of this investigation when multiple measurements are included in the database. Such observations share unknown effects, so it is not possible to assume independence of errors for individual observations.

In eq. 5 the subscript k indicates the number of significant variables affecting the central tendency (X_k), and j indicates the number of significant variables affecting the deviation from the mean (X_j).

Variable Selection

The Bayesian Information Criterion (BIC) postulated by Schwarz (1978) was used to select the variables significantly affecting driver speed behavior from all the possible covariates. The model with the lowest BIC function value (f_{BIC}), calculated according to eq. 6, is given preference:

$$f_{BIC} = -2 \cdot \hat{L} + k \cdot \ln(n) \qquad \text{(eq. 6)}$$

In this equation, $\hat{L}$ is the maximized value of the log-Likelihood function, n the number of observations, and k is the number of parameters included in the model. According to the structure of eq. 5, the number of parameters is equal to the sum of the size of coefficients β_0, β_k^C and β_j^D. Only the variables that contributed to the minimization of the BIC function were selected and included in the model.

RESULTS AND DISCUSSION

Two RE models have been calibrated; the synthesis of results from model calibration can be found in Table 3. Analyses have been carried out with the R-software version 3.0.2; in particular, the REML algorithm running the *lme4* package (R Development Core Team 2005) was used for model estimation.

Model #1 was calibrated using all the variables of Table 2 selected according to the BIC criterion. Model #2 excludes the variables with low frequency (variables PKLLS, PKLRS, WLS, and WRS in Table 2), and all the Boolean variables that have a corresponding numerical one (SR, SL, RRS, RLS, DLS, DRS, ILS, IRS, and Ped in Table 2), with a total of twenty variables considered in the modeling process as compared to the thirty-three indicated in Table 2.

It can be noted that only two variables (i.e., curvature and pedestrian crossing density) affect the central tendency of speed distribution, while thirty-two variables affect the dispersion.

With nineteen variables in total, model #2 is more parsimonious since it requires the use of twelve fewer variables (19 instead of 31) than model #1. Model #2 is expressed in equation form as follows:

$$V_i = 79.33 - 7.59 \cdot PedD - 1946.0 \cdot \frac{1}{R} + \\ + Z_p \cdot \begin{bmatrix} 20.54 - 274.1 \cdot \left(\frac{1}{R}\right) - 3.46 \cdot SLS - 1.56 \cdot SBLS - 0.83 \cdot IDRS - 2.88 \cdot SRS + \\ - 0.64 \cdot SBRS - 13.38 \cdot SRW - 13.85 \cdot SLW - 0.062 \cdot PSL + 0.42 \cdot DDRS + \\ - 0.29 \cdot TRDLS - 3.52 \cdot LW - 0.43 \cdot IDLS + 0.069 \cdot LG + 0.072 \cdot \Delta PSL + \\ - 1.62 \cdot PedD + 3.46 \cdot TRDRS + 1.16 \cdot LBRS + 1.31 \cdot LBLS \end{bmatrix} \quad \text{(eq. 7)}$$

where the meaning of symbols used to identify the variables is available in Table 2.

Significant variables

Model #2 is henceforth regarded as a reference in any discussion on the most significant variables because its interpretation is more straightforward. The results reported in Table 3 demonstrate that the central tendency is only affected by the pedestrian crossing density (PedD) and curvature (1/R). This finding confirms once again the strong influence of the local tortuosity of horizontal alignment on speeds (Transportation Research Circular 2011), and the effect of pedestrian activities that may result in physical obstacles along the travelled way in the case of rural roads. The negative sign of the estimated coefficients indicates that an increase in these variables results in a decrease in the observed average speed.

Variables that are normally included in models for the estimation of the 85^{th} percentile of speed distribution, are included here as terms that affect the speed dispersion around the mean. The variables that have been found to be significant can be divided into two categories:

- those which contribute to a reduction in operating speed dispersion when they increase, i.e. the curvature (1/R), the presence of sidewalks (SLS and SRS) and safety barriers (SBLS and SBRS) on both sides, the intersection density (IDLS and IDRS), the lane width (LW), the shoulder width on the right side (SRW), and finally the ramp density on the left side (TRDLS); and
- those which contribute to an increase in operating speed dispersion when their values increase, i.e. the posted speed limit (PSL) and its variation (ΔPSL), the driveway density on the right side (DDRS, the longitudinal grade (LG), pedestrian (pedD) and right side ramp density (TRDRS), and the presence of a lay-by on both sides (LBLS and LBRS).

The first category groups those variables that have the ability to concentrate the different percentiles around the mean speed, while the second includes those that tend to increase the dispersion.

Variables in the first category induce a more conservative behavior in all the drivers, who consider the presence of physical obstacles (i.e., safety barriers and sidewalks) or pedestrian activities along the roadsides a source of danger. Similarly, the presence of intersections and ramps on the left side of the travelled way are viewed as a potential source of conflicts with opposing vehicles. A comparison with literature results is hereafter proposed to facilitate an appreciation of the similarities and differences between our results and those previously found by other authors. Unfortunately, this is possible only for a limited number of variables such as the lane width (LW), the longitudinal grade (LG), and the posted speed limit (PSL).

The negative sign for LW in Table 3 is partially unexpected and goes against the few indications that can be found in the literature. For example, SETRA (1986) and Lamm and Choueiri (1987) found that an increase in LW corresponds to an increase in the V_{85}. It is worth noting that more recent contributions do not explicitly consider LW as a significant parameter affecting the operating speed of rural highways (Misaghi and Hassan 2005, Transportation Research Circular 2011). Figueroa Medina and Tarko (2005) introduced LW through the variable named “pavement width” (PAV), that includes the travelled way width and the two paved shoulders. They found that an increase in PAV results in an increase in mean speed and a decrease in speed deviation along tangent segments. PAV did not prove to be significant in the case of curved segments.

As regards to the effect of longitudinal grade (LG), the results in literature are quite controversial. Figueroa Medina and Tarko (2005) found that when LG increases the mean speed decreases, while the deviation increases. These findings are similar to the ones observed in this research. Conversely, Himes et al. (2011, 2013) observed that when LG increases both the mean speed and deviation decrease. The results obtained here on the PSL confirm what was previously found by Himes et al. (2011, 2013).

With reference to Table 3, it is interesting to note the opposite effects on data dispersion caused by the two shoulder widths (SRW and SLW) and the two ramp densities (TRDRS and TRDLS), whose coefficients have opposite signs.

Regarding the effects on speed dispersion in the case of model #1, it can be noted that some variables that have been included in terms of presence (Boolean variables) and density (numerical variables) present opposite signs (i.e., driveways and ramp presence and density, shoulder presence and width). This result leads to a difficult interpretation of the effects of such variables in the model, which, in turn, encouraged the authors to introduce model #2.

Synthesis of statistics and error analysis

Table 4 contains a synthesis of the statistical analyses obtained from the two RE models. These statistics confirm that the data fit very well the random effect models, although the presence of random effects in the data is entirely dependent on the sampling design. From the same table, it is important to note that the random effects associated with the road are not significant, which attests that RE models are more transferable than those obtained with FEs; REs associated with the section are higher than REs associated with the direction. Furthermore, from a comparison of models #1 and #2, it is evident that the latter has very good statistics albeit with the use of a lower number of variables. The use of simpler models is always an attractive option; in this specific case, statistics confirm that the prediction capability of model #2 is very similar to that of model #1 notwithstanding the fact that the former uses twelve fewer variables.

The quality of the results obtained can be appreciated in Fig. 3, where a comparison of observed vs. predicted speed values is presented. All sets of points appear to be distributed randomly along the identity line (y=x) which is required by the regression. The predicted values of the two RE models (Fig. 3A and 3B) are closer to the line, which indicates a very good fit.

Finally, Fig. 4 reports all the error term (residuals) distributions for the two models. The plots are in pairs: the first for the roads and the second for the sections for each of the two models. The expected results in this scenario would be for all 13 boxes of road residuals and all 37 boxes of section residuals to have approximately the same height and shape, indicating that errors are centered around zero for all roads/sections and the variance of residuals for each road/section are the same.

Models #1 and #2 address most of the regression problems. In fact, the errors for all roads and sections are centered around zero, hence the fitted values are not biased for any road or any section.

Model validation

For each combination of road/section/direction, approximately 90% of observations for the training sample (the one used for estimation of random effects and coefficients) and 10% of the observations for the validation sample were kept. In total, the training sample consists of 5,877 observations while the validation sample includes 690 observations. Observations were assigned to one sample randomly to avoid any possible bias arising from the order in the original sample.

Although random effects are, by definition random, their values from the training sample have been estimated. Plot A in Fig. 5 compares the predicted values to the observed ones for all 690 validation observations. The identity line was determined with the expectation that points would be distributed along it. The location of points far from the identity line would indicate a systematic error or bias in the application. Overall, the analysis proved satisfactory with points distributed along the line with relatively similar variance for all values of observed speeds.

Since the road effect is not significant, plots B, C, D and E in Fig. 5 show the predicted values for sections 13, 23, 25, 36 respectively. These sections have been chosen at random for the purpose of cross validation. Plots include values for both directions within the section. Points on the plots generally show the same patterns derived for the complete validation sample. However, in the case of section 23 prediction errors are more noticeable: this is possibly due to this particular case being based only on a handful of observations.

CONCLUSIONS

In this paper random effect (RE) linear models have been applied to predict a full range of percentile values for the operating speed along tangents or curves of two-lane rural highways. The model has been calibrated with data collected from surveys of isolated vehicles on a number of road sections in the Northwest of Italy. A total of 6,567 observations have been collected from 37 sections randomly selected from 13 roads. In total, 33 geometric and environmental variables have been taken into account for model estimation. The structure of the model separates the central tendency from the dispersion of speed data allowing the estimation of not only the 85th percentile (which is usually regarded as a reference measure for operating speeds) but also the evaluation of any percentile through the standardized normal variable Z_p.

The structure of the model makes it possible to distinguish between the effects on the mean and those on the deviation, for the benefit of those Road Agencies interested in the implementation of speed management actions.

The free-flow speed data collected in this study violate the assumption of ordinary least squares regression; therefore, FE models would have produced biased estimates and large dispersion of the residuals. The RE model, which correctly accounts for the sampling design, proved to produce a very low level of errors and to not suffer from the presence of outliers. The two RE models estimated, highlight once again (Transportation Research Circular 2011) the effects of the curvature on the central tendency of speed distribution. The density of pedestrian crossings is the only other variable that significantly affects the mean value of free-flow speeds. According to the same model several cross-sectional elements located along the road margins contribute to the dispersion of speed data around the central value. While each element on its own has a negligible effect on driver behavior, the combined sum of these effects has a significant impact (on driver behavior). This demonstrates that speeds are affected by a number of cross-sectional characteristics normally neglected in currently operating speed models for rural roads.

When analyzing the single RE model, it is possible to conclude that road effects are negligible, that most of the errors are associated with a road section (in particular road #7 – Table 1 – which is also the one for which a lower number of speed data were available) and, to a lesser extent, with the direction effect, and, finally, that the residuals have low standard deviation.

In the paper, an attempt to reduce the number of variables used (in the model) has been made. In particular, all the variables affected by low frequency in the data (PKLLS, PKLRS, WLS, and WRS), and all the Boolean variables that have a corresponding numerical one (SR, SL, RRS, RLS, DLS, DRS, ILS, IRS, and Ped) were discarded in the calibration of model #2. Results from model #2 demonstrate that, when compared to model #1, which includes 31 variables, the statistical fit is significantly improved in spite of a lower number of variables used in model calibration (19 in total).

In summary, the RE model incorporates the sampling design, has a very low level of errors, it does not show outliers, and the behavior of errors remains constant even for extreme values of predicted speeds. This is very positive since many models lose strength at the edge of the sampling space. In other words, minor errors were not observed for low predicted speeds nor major errors for high predicted speeds.

Future investigation will seek to validate the model presented here with speed data relative to different driving environments and collected with different measurement technologies; furthermore, comparisons with other modeling techniques aimed at the estimation of the entire speed distribution, which are ongoing in the context of operating speeds studies, will be pursued as well.

ACKNOWLEDGEMENTS

The research work included in this paper has been made possible thanks to the joint project between the Politecnico di Torino (Italy) and the University of Maryland (US) titled: "$(MS)^3$ - Advanced Models, Methods and Materials for Sustainable, Safe and Smart Infrastructures". The authors acknowledge funding received from the *Compagnia di San Paolo* (Italy) and from the *Politecnico di Torino* under the grant: "*Bando per il Finanziamento di Progetti di Internazionalizzazione della Ricerca*", approved with the Rectoral Decree n. 208 of the 24th of May, 2013.

The *Città Metropolitana di Torino,* and the Provinces of *Vercelli* and *Alessandria* are greatly acknowledged for the access to the GIS databases used to collect roadway inventory information.

REFERENCES

Bassani, M., Dalmazzo, D., Marinelli, G., and C. Cirillo (2014). “The Effects of Road Geometrics and Traffic Regulations on Driver-Preferred Speeds in Northern Italy. An Exploratory Analysis”. *Transp. Res. Part F*, 25, 10-26.

Bassani, M. and G. Mutani (2013). “Effects of Environmental Lighting Conditions on Operating Speeds on Urban Arterials”. *Transp. Res. Rec.*, 2298, 78-87.

Bonneson, J. A. (2001). “Controls for Horizontal Curve Design”. *Transp. Res. Rec.*, 1751, 82-89.

Castro, M., Sánchez, J.F., Sánchez, J.A., and Iglesias, L. (2011). “Operating Speed and Speed Differential for Highway Design Consistency”. J. of Transp. Eng., 137 (11), 10.1061/(ASCE)TE.1943-5436.0000309, 837-840.

Figueroa Medina, A.M. and A. P. Tarko. (2005). “Speed Factors on Two-Lane Rural Highways in Free-Flow Conditions”. *Transp. Res. Rec.*, 1912, 39-46.

Figueroa Medina, A.M., and Tarko, A.P. (2007). “Speed Changes in the Vicinity of Horizontal Curves on Two-Lane Rural Roads”. J. of Transp. Eng., 133 (4), 10.1061/(ASCE)0733-947X(2007)133:4(215), 215-222.

Fitzpatrick, K., Miaou, S-P., Brewer, M., Carlson, P., and Wooldridge, M.D. (2005). “Exploration of the Relationships between Operating Speed and Roadway Features on Tangent Sections”. J. of Transp. Eng., 131 (4), 10.1061/(ASCE)0733-947X(2005)131:4(261), 261-269.

Hassan, Y. (2004). “Highway Design Consistency: Refining the State of Knowledge and Practice”. *Transp. Res. Rec.*, 1881, 63-71.

Himes, S.C., Donnell, E.T., and Porter, R.J. (2011). New Insights on Evaluations of Design Consistency for Two-Lane Highways. *Transp. Res. Rec.*, 2262, 31-41.

Himes, S.C., Donnell, E.T., and Porter, R.J. (2013). Posted speed limit: To include or not to include in operating speed models. *Transp. Res. Part A*, 52, 23–33.

Jacob, A., and Anjaneyulu, M.V.L.R. (2013). "Operating Speed of Different Classes of Vehicles at Horizontal Curves on Two-Lane Rural Highways". J. of Transp. Eng., 139 (3), 10.1061/(ASCE)TE.1943-5436.0000503, 287-294.

Lamm, R., and Choueiri, E. M. (1987). "Recommendations for evaluating horizontal design consistency based on investigations in the State of New York." *Transp. Res. Rec.*, 1122, 68-78.

Misaghi, P., and Hassan, Y. (2005). "Modeling operating speed and speed differential on two-lane rural roads." J. Transp. Eng., 131 (6), 10.1061/(ASCE)0733-947X(2005)131:6(408), 408-418.

NCHRP (National Cooperative Highway Research Program). (2003). "Design Speed, Operating Speed, and Posted Speed Practices". Report No. 504, Transportation Research Board of the National Academies, Washington, D.C., ISBN 0-309-08767-8.

R Development Core Team. (2005). "R: A Language and Environment for Statistical Computing". <www.R-project.org> (Feb. 2, 2015).

Schwarz, G. (1978). "Estimating the Dimension of a Model". *Ann. of Stat.*, 6, 461-464.

SETRA (1986). "Vitesses pratiquées et geometric de la Route." Note d'information série Circulation Sécurité Équipement Exploitation (CSEE) n° 10. Ministère de l' Equipment, du Logement, de l'Aménagement du Territoire et des Transports, France.

Transportation Research Circular (2011). "Modeling Operating Speed". E-C151 Synthesis Report, Transportation Research Board of the National Academies, Washington, D.C.

Wang, J., Dixon, K.K., Li, H. and M. Hunter (2006). "Operating Speed Model for Low-Speed Urban Tangent Streets Based on In-Vehicle Global Positioning System Data". *Transp. Res. Rec.*, 1961, 24-33.

Table 1. Geometric and operative characteristics of the selected road sections

Road #	Section #	Road name	Road length (km)	Lane width (m)	Radius (m)	Grade (%)	V_{min} (km/h)	V_{max} (km/h)	V_{85} (km/h)	PSL (km/h)	# data
1	1	SP70-VC	3.830	3.60	∞	± 1.50	39.0	97.0	76.0	70	429
2	2	SP8-VC	7.790	3.70	178.47	± 3.00	45.0	132.0	75.0	90	618
3	3	SP299-VC	57.250	3.75	∞	0.00	32.0	157.0	85.0	50	972
	4			3.60	334.57	± 0.50	45.0	128.0	87.0	90	799
	5			3.70	∞	± 1.00	38.0	130.0	89.0	90	669
4	6	SS460-TO	61.757	3.75	1000.00	± 1.44	57.0	122.6	93.2	70	192
5	7	SS565-TO	18.180	3.75	304.00	± 5.14	24.0	114.0	94.0	70	312
	8			3.75	∞	± 2.09	57.0	124.0	98.0	70	101
	9			3.75	3226.00	± 4.69	52.0	114.0	91.0	70	101
	10			3.25	150.00	± 8.50	46.0	77.0	70.1	50	87
6	11	SP55-AL	13.978	3.00	∞	± 1.50	42.6	129.4	90.3	70	107
	12			3.00	∞	0.00	46.7	128.6	94.3	70	120
	13			3.00	∞	0.00	50.0	114.0	90.4	70	108
	14			3.00	∞	± 2.00	49.6	129.4	101.7	70	125
	15			3.00	∞	0.00	49.0	127.8	95.9	70	127
	16			3.00	∞	0.00	46.3	128.3	98.9	70	138
	17			3.00	∞	0.00	49.4	149.6	95.3	70	128
7	18	SP230-VC	39.466	3.50	∞	± 0.50	42.0	115.2	88.5	90	41
	19			3.50	∞	± 0.50	73.0	130.0	106.6	90	29
	20			3.50	2250.00	± 0.50	70.0	118.0	107.0	90	26
	21			3.50	∞	± 0.50	58.0	127.0	112.9	90	28
	22			3.50	∞	0.00	55.0	130.0	100.3	90	30
	23			3.50	1502.00	± 1.00	32.0	72.0	57.4	50	32
	24			3.50	∞	0.00	59.0	110.0	92.6	70	27
	25			3.50	909.00	± 1.00	44.0	96.0	78.0	70	37
	26			3.50	8351.25	0.00	47.0	98.0	85.5	70	38
	27			3.50	452.00	0.00	44.0	70.0	57.5	50	36
	28			3.50	1503.00	± 1.00	56.0	130.0	97.0	90	43
	29			3.50	∞	± 1.00	65.0	119.5	107.9	90	38
	30			3.50	∞	0.00	54.7	108.1	85.1	70	42
	31			3.50	∞	0.00	44.3	120.2	83.6	70	140
8	32	SP177-TO	10.927	3.20	300.00	± 2.50	27.6	93.3	75.8	70	116
9	33	SP176-TO	4.961	3.80	∞	0.00	42.6	106.6	75.3	90	67
10	34	SP267-TO	8.839	3.50	∞	0.00	34.7	105.4	68.5	50	154
11	35	SP220-TO	4.045	3.80	∞	0.00	44.2	89.1	79.8	70	100
12	36	SP183-TO	2.156	3.50	∞	0.00	36.9	134.8	82.5	50	161
13	37	SP23-TO	92.223	3.50	550.00	± 0.50	36.4	192.1	102.1	90	249

Table 2. Summarized raw statistics of considered variables

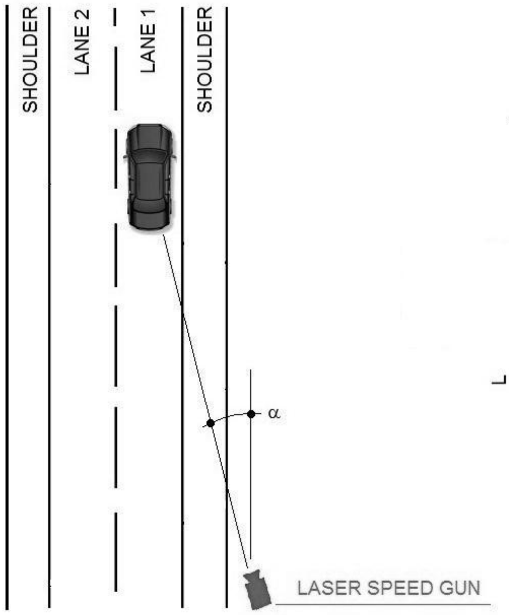
#	Variable	symbol	type	min.	max.	μ	σ	Frequency	
1	Posted speed limit (km/h)	PSL	ND	50	90	73.0	13.610	73	100%
2	Posted speed limit variation (km/h)	ΔPSL	ND	0	20	0.5	3.287	2	3%
3	Lane width (m)	LW	NC	3.0	3.8	3.5	0.260	73	100%
4	Curvature (m^{-1})	1/R	NC	0	$6.67 \cdot 10^{-3}$	$8.28 \cdot 10^{-4}$	$1.61 \cdot 10^{-3}$	28	38%
5	Longitudinal grade (%)	LG	NC	-8.85	8.50	-0.012	2.110	41	56%
6	Shoulder right	SR	B	0	1	-	-	69	95%
7	Shoulder left	SL	B	0	1	-	-	69	95%
8	Shoulder right width (m)	SRW	NC	0.0	1.5	0.9	0.437	69	95%
9	Shoulder left width (m)	SLW	NC	0.0	1.5	0.9	0.437	69	95%
10	Ramp left side	RLS	B	0	1	-	-	5	7%
11	Ramp right side	RRS	B	0	1	-	-	5	7%
12	Ramp density left side (No./km)	TRDLS	NC	0.0	2.0	0.1	0.323	5	7%
13	Ramp density right side (No./km)	TRDRS	NC	0.0	2.0	0.1	0.323	5	7%
14	Driveways left side	DLS	B	0	1	-	-	55	75%
15	Driveways right side	DRS	B	0	1	-	-	55	75%
16	Driveway density left side (No./km)	DDLS	NC	0.0	8.0	2.3	2.321	55	75%
17	Driveway density right side (No./km)	DDRS	NC	0.0	8.0	2.3	2.325	55	75%
18	Intersections left side	ILS	B	0	1	-	-	39	53%
19	Intersections right side	IRS	B	0	1	-	-	39	53%
20	Intersection density left side (No./km)	IDLS	NC	0.0	5.0	1.1	1.362	39	53%
21	Intersection density right side (No./km)	IDRS	NC	0.0	5.0	1.1	1.362	39	53%
22	Lay-by left side	LBLS	B	0	1	-	-	25	34%
23	Lay-by right side	LBRS	B	0	1	-	-	27	37%
24	Sidewalk left side	SLS	B	0	1	-	-	13	18%
25	Sidewalk right side	SRS	B	0	1	-	-	13	18%
26	Pedestrian crossing	Ped	B	0	1	-	-	12	16%
27	Pedestrian crossing density (No./km)	PedD	NC	0.0	4.0	0.3	0.845	12	16%
28	Parking lanes left side	PKLLS	B	0	1	-	-	1	1%
29	Parking lanes right side	PKLRS	B	0	1	-	-	1	1%
30	Safety barrier left side	SBLS	B	0	1	-	-	19	26%
31	Safety barrier right side	SBRS	B	0	1	-	-	20	27%
32	(Retaining) Wall left side	WLS	B	0	1	-	-	1	1%
33	(Retaining) Wall right side	WRS	B	0	1	-	-	1	1%

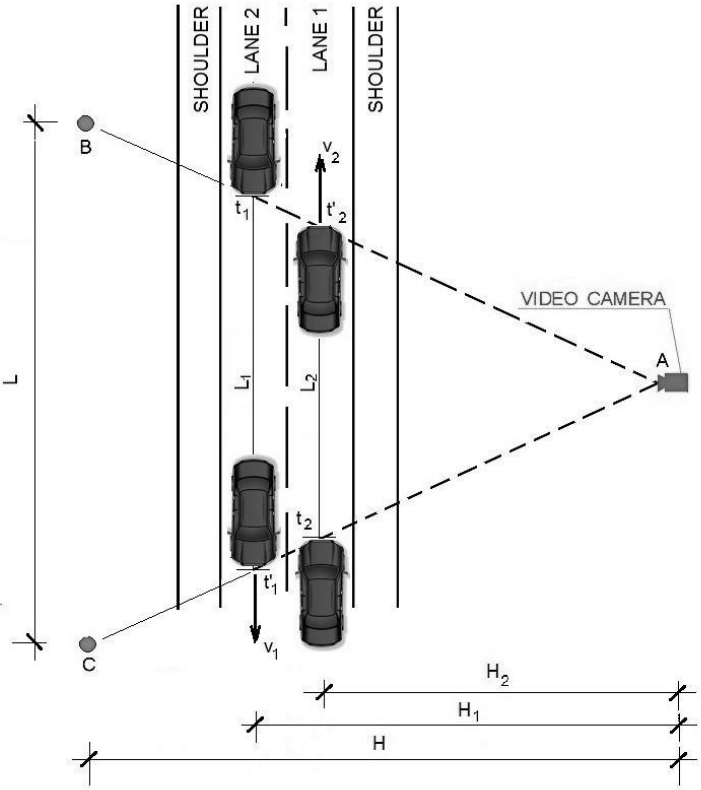
Table 3. Model coefficients and significant variables for random effect (RE) models

#	Variable	Model #1				Model #2			
		Estimate	Std.error	t-value	p-value	Estimate	Std.error	t-value	p-value
-	Intercept (β_0)	79.34	1.82	43.41	$<1\cdot10^{-4}$	79.33	1.82	43.49	$<1\cdot10^{-4}$
1	PedD	-7.61	1.94	-3.91	$<1\cdot10^{-4}$	-7.59	1.94	-3.91	$<1\cdot10^{-4}$
2	1/R	-1949.0	1063.0	-1.83	0.067	-1946.0	1061.0	-1.83	0.067
3	Z	10.99	0.69	16.01	$<1\cdot10^{-4}$	20.54	0.60	33.96	$<1\cdot10^{-4}$
4	Z · 1/R	-134.1	11.69	-11.48	$<1\cdot10^{-4}$	-274.1	13.62	-20.13	$<1\cdot10^{-4}$
5	Z · SLS	-2.81	0.12	-23.49	$<1\cdot10^{-4}$	-3.46	0.11	-30.94	$<1\cdot10^{-4}$
6	Z · SBLS	-1.27	0.11	-11.11	$<1\cdot10^{-4}$	-1.56	0.10	-15.06	$<1\cdot10^{-4}$
7	Z · IDRS	-0.13	0.048	-2.62	0.009	-0.83	0.030	-28.15	$<1\cdot10^{-4}$
8	Z · DLS	-2.34	0.13	-17.76	$<1\cdot10^{-4}$	-	-	-	-
9	Z · Ped	1.38	0.20	6.82	$<1\cdot10^{-4}$	-	-	-	-
10	Z · SRS	-3.14	0.12	-26.89	$<1\cdot10^{-4}$	-2.88	0.11	-26.55	$<1\cdot10^{-4}$
11	Z · SBRS	-1.44	0.11	-12.91	$<1\cdot10^{-4}$	-0.64	0.092	-6.97	$<1\cdot10^{-4}$
12	Z · SRW	-14.40	0.19	-76.52	$<1\cdot10^{-4}$	-13.38	0.24	-54.72	$<1\cdot10^{-4}$
13	Z · SLW	11.94	0.19	63.25	$<1\cdot10^{-4}$	13.85	0.24	56.80	$<1\cdot10^{-4}$
14	Z · ILS	-1.37	0.11	-12.54	$<1\cdot10^{-4}$	-	-	-	-
15	Z · PSL	0.15	0.005	27.41	$<1\cdot10^{-4}$	0.062	0.0003	19.77	$<1\cdot10^{-4}$
16	Z · DDRS	1.00	0.022	45.32	$<1\cdot10^{-4}$	0.42	0.021	20.44	$<1\cdot10^{-4}$
17	Z · DRS	-2.80	0.12	-23.52	$<1\cdot10^{-4}$	-	-	-	-
18	Z · IRS	-2.40	0.11	-22.42	$<1\cdot10^{-4}$	-	-	-	-
19	Z · SR	3.99	0.17	23.82	$<1\cdot10^{-4}$	-	-	-	-
20	Z · PKLLS	-3.29	0.28	-11.54	$<1\cdot10^{-4}$	-	-	-	-
21	Z · TRDLS	-4.21	0.21	-19.66	$<1\cdot10^{-4}$	-0.29	0.09	-3.25	0.001
22	Z · RLS	5.58	-0.34	16.27	$<1\cdot10^{-4}$	-	-	-	-
23	Z · LW	-1.83	0.24	-7.48	$<1\cdot10^{-4}$	-3.52	0.23	-15.02	$<1\cdot10^{-4}$
24	Z · WRS	-3.38	0.24	-13.77	$<1\cdot10^{-4}$	-	-	-	-
25	Z · IDLS	0.22	0.05	4.51	$<1\cdot10^{-4}$	-0.43	0.030	-14.26	$<1\cdot10^{-4}$
26	Z · LG	0.09	0.01	8.36	$<1\cdot10^{-4}$	0.069	0.014	5.04	$<1\cdot10^{-4}$
27	Z · DDLS	0.11	0.02	5.10	$<1\cdot10^{-4}$	-	-	-	-
28	Z · ΔPSL	-0.11	0.01	-8.36	$<1\cdot10^{-4}$	0.072	0.006	11.35	$<1\cdot10^{-4}$
29	Z · WLS	-1.99	0.23	-8.59	$<1\cdot10^{-4}$	-	-	-	-
30	Z · PedD	0.73	0.08	9.13	$<1\cdot10^{-4}$	1.62	0.064	25.22	$<1\cdot10^{-4}$
31	Z · PKLRS	1.63	0.29	5.68	$<1\cdot10^{-4}$	-	-	-	-
32	Z · TRDRS	-2.53	0.21	-11.81	$<1\cdot10^{-4}$	0.41	0.085	4.75	$<1\cdot10^{-4}$
33	Z · RRS	3.74	0.33	11.18	$<1\cdot10^{-4}$	-	-	-	-
34	Z · LBRS	0.40	0.09	4.21	$<1\cdot10^{-4}$	1.16	0.087	13.21	$<1\cdot10^{-4}$
35	Z ·LBLS	-	-	-	-	1.31	0.074	17.77	$<1\cdot10^{-4}$

Table 4. Synthesis of statistical analysis for RE models #1 and #2

	Model #1		**Model #2**	
BIC function at convergence	36774.62		27820.38	
Multiple R^2	0.9332		0.9840	
Random effects	Variance	Std.Dev.	Variance	Std.Dev.
Direction:(Section:Road)	19.45	4.41	19.10	4.37
Section:Road	85.96	9.27	85.75	9.26
Road	$1.16\cdot10^{-9}$	$3.41\cdot10^{-5}$	$1.87\cdot10^{-9}$	$4.32\cdot10^{-5}$
Residual	3.86	1.96	3.63	1.90





(A) **(B)**

Fig. 1. Speed surveying techniques used in the investigation

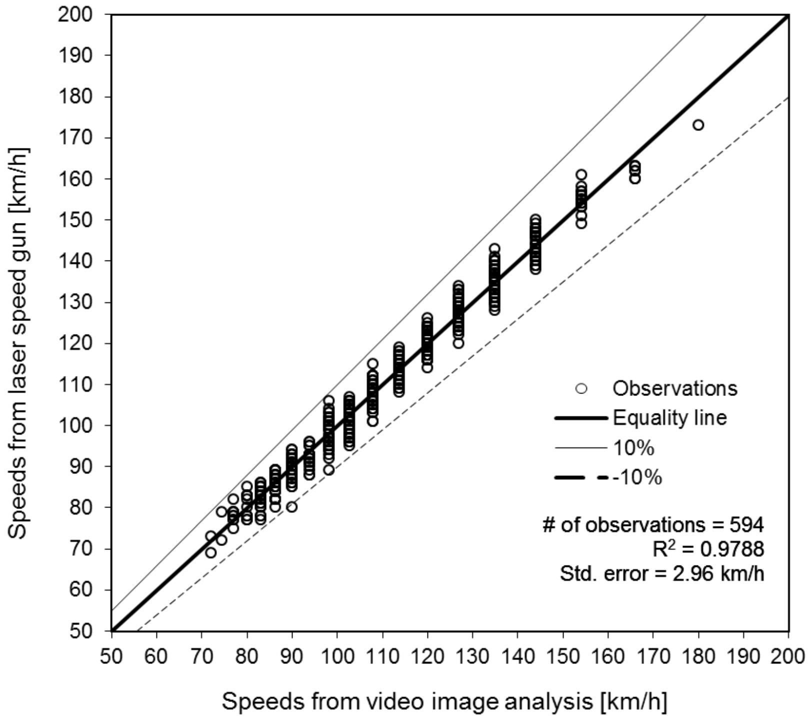


Fig. 2. Comparison of operating speeds detected for the same vehicle from video image analysis and laser speed gun

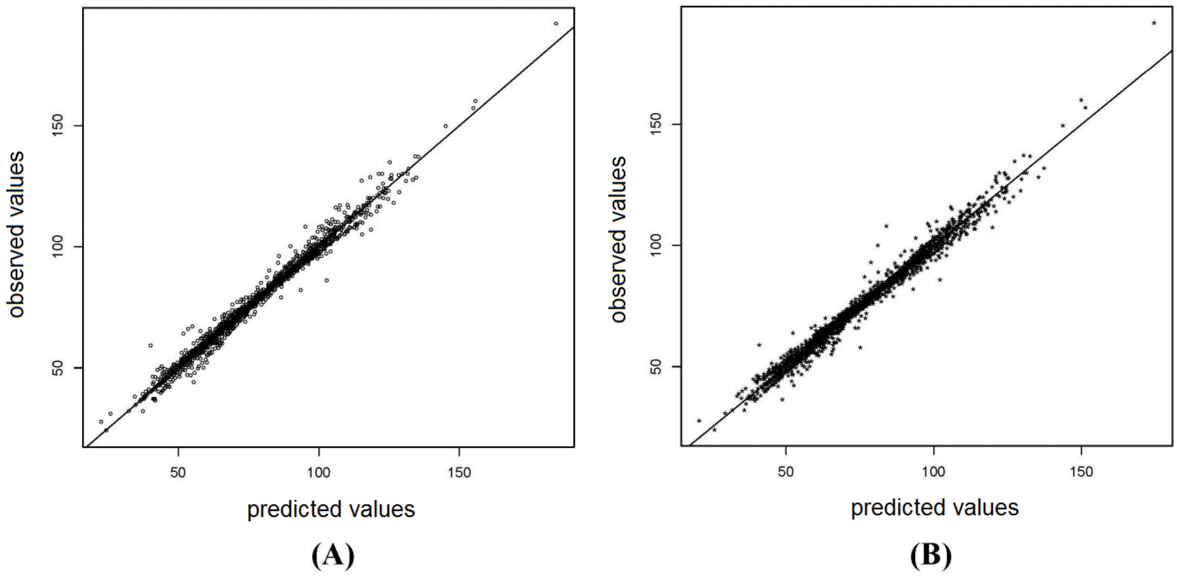


Fig. 3. Comparison of predicted and observed operating speed data. Data distribution for (A) model #1, and (B) model #2

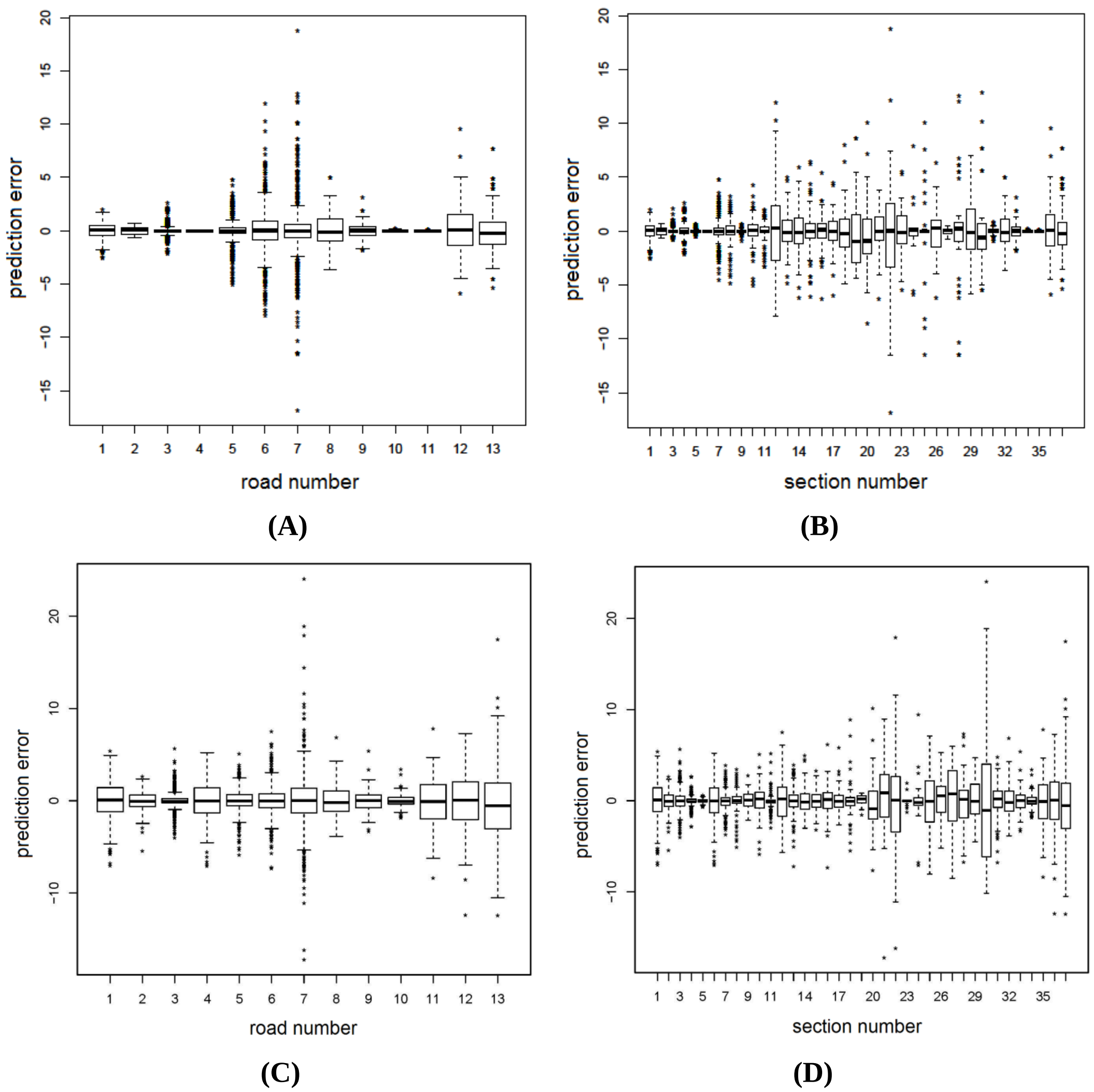
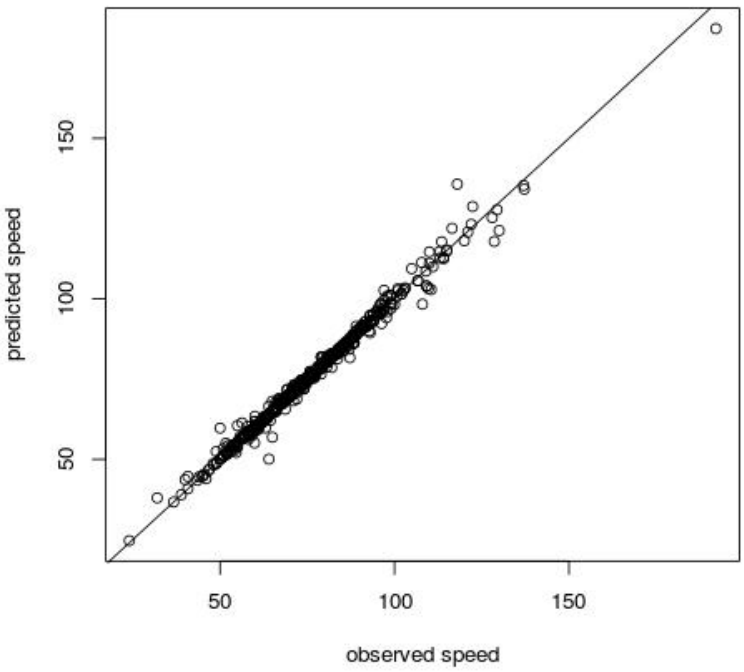
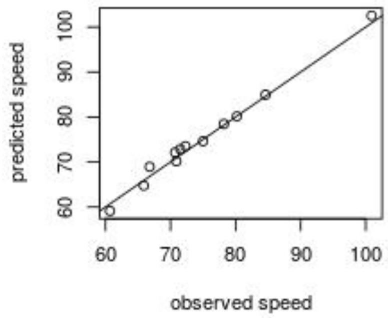


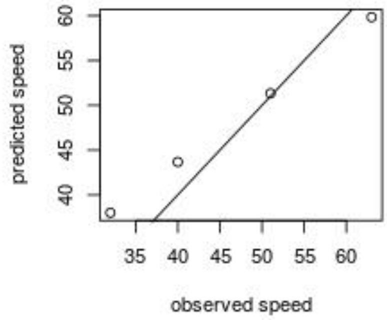
Fig. 4. Error term distributions for roads and sections for model #1 (A and B), and model #2 respectively (C and D)



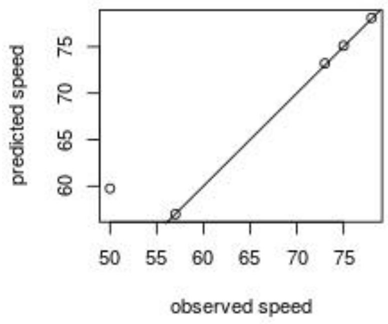
(A)



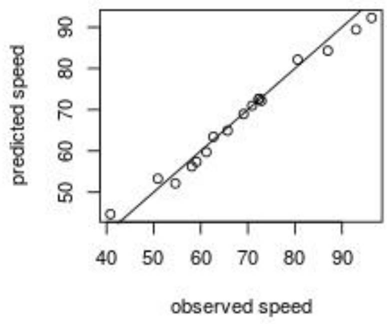
(B)



(C)



(D)



(E)

Fig. 5. Results from model validation