

# Uso de datos pasivos obtenidos mediante dispositivos inerciales para inferir la condición del pavimento en ciclovías

## Use of passive data obtained by inertial sensors to infer pavement conditions on bike lanes

M. Valle \*, J.C. Herrera <sup>1\*</sup>

\* Pontificia Universidad Católica de Chile, Santiago, CHILE

Fecha de Recepción: 08/10/2018

Fecha de Aceptación: 16/01/2019

PAG 33-44

### Abstract

The surface condition of bike lanes is one of the main factors that cyclists consider when choosing their route, because it affects their travelling comfort, inasmuch as it is related to the vibrations they experience while riding. The existing alternatives to determine this condition either do not correspond to the cyclist's reality or their application is complex. A methodology based on a logistic regression is proposed, which is objectively and specifically aimed at detecting bike-lane pavement failures and thus inferring its condition in the analyzed section. As a proof of concept of the proposed methodology, a field experiment was designed, which emulates two specific defects: pavement upheavals and potholes. Data collection was carried out by attaching inertial sensors to the test bicycle. The application of the proposed methodology allowed identifying the necessary inertial variables to assess the considered failures. Among the latter, the main ones are rotations in the three axes and vertical acceleration. Four models that could correctly identify pavement issues were generated with these variables. In the future, the results thereof will allow building an indicator to infer the surface condition based on the vibrations felt by cyclists, and thus establish a level of service associated to the bike-lane pavement through these indicators.

*Keywords:* Bike lanes, pavement condition, vibrations, logistic regression, IMU

### Resumen

Uno de los principales factores que consideran los ciclistas al escoger su ruta es la condición superficial de las ciclovías, ya que, al estar relacionado con las vibraciones experimentadas, afecta la comodidad del viaje. Las alternativas existentes para determinar esta condición no responden a la realidad del ciclista o son de compleja aplicación. Se propone una metodología basada en una regresión logística, que sea objetiva y específica para detectar irregularidades en el pavimento de ciclovías y así inferir la condición del mismo en el tramo analizado. Como prueba de concepto de la metodología propuesta, se diseñó un experimento en terreno que emula dos deterioros específicos: levantamientos del pavimento y baches. La recolección de estos se realizó mediante dispositivos inerciales adosados a una bicicleta de prueba. Al aplicar la metodología propuesta se identificaron las variables inerciales necesarias para identificar las irregularidades consideradas. Entre ellas destacaron las rotaciones en los tres ejes y la aceleración vertical. Con estas variables, se generaron cuatro modelos capaces de identificar correctamente los problemas del pavimento. A futuro, con los resultados obtenidos se puede construir un indicador capaz de inferir el estado de la superficie a partir de las vibraciones experimentadas por los ciclistas y así establecer un nivel de servicio asociado al pavimento de una ciclovía mediante estos indicadores.

*Palabras clave:* Ciclovías, estado del pavimento, vibraciones, regresión logística, IMU

## 1. Introduction

Different studies worldwide have demonstrated that exclusive bicycle infrastructures in good conditions favor their use (Garrard et al., 2008; Pucher et al., 2011; Pucher et al., 2010). A highly relevant aspect of the bicycle infrastructure is the bike-lane pavement condition. This is one of the main factors that bike riders consider when choosing their route, because the state of the road is directly related to the vibrations, which affect their travelling experience and comfort (Landis et al., 1997; Jensen, 2007; Lépine et al., 2014; Lépine et al., 2011; Martens, 2011).

A procedure used for assessing bike-lane pavements is the use of scales based on visual inspections. One of the biggest problems with this kind of evaluation is their subjectivity, because it partly depends on the criterion of the observer (Sprinkle Consulting, 2007; Landis, 1994).

The International Roughness Index (IRI) is another technique used for determining the surface condition of the pavement (Barbudo et al., 2015; MINVU, 2015). For example, in Australia, the ARRB uses a walking profilometer to collect the data needed for calculating the IRI (Cairney and King, 2003). In Belgium, Martens (2011) attached a third wheel to the bikes, which operated as a bike trailer and recorded the vibrations for calculating IRI as the bike travelled along the road. In Chile, the "Bike Lane Construction Manual: Technical Standard" (Ministry of Housing and Urban Planning (MINVU), 2015) indicates that a completed bike-lane construction should have an IRI of 4 m/km. Moreover, it

<sup>1</sup> Corresponding author:

Departamento de Ingeniería de Transporte y Logística.  
Pontificia Universidad Católica de Chile, Santiago, CHILE.  
E-mail: jch@ing.puc.cl



recommends controlling the surface with the Merlin device, by collecting 200 measurements at regular intervals of 200 meters. In other words, one per meter, which is costly in terms of time and labor.

In addition to the restrictions mentioned above, the IRI has to be modified in order to use it on bike lanes. In the first place, IRI was designed for motor vehicles circulating at high speeds, which cyclists cannot reach. Furthermore, the mass, size and geometry of the bicycle is quite different from those of a car, which directly affects the proposed index (Kohler, 2015; Parkin, 2009; Martínez et al., 2011). These aspects are significant when evaluating a pavement surface, because its dynamic response is highly sensitive to the speed and geometry of the vehicle used (Fuentes et al., 2010). Therefore, it implies changing the model accordingly, once the bike-lane profile is available.

Consequently, some studies have proposed developing other indexes. Denmark developed the Bicycle Profile Index (BPI), based on the methodology used to obtain IRI (Kohler, 2015). The calculation of this index uses a small motor vehicle equipped with accelerometers and laser profilometers that collect data as the vehicle circulates on the bike lane. The details of the algorithm for determining the BPI are subject to trade secret; the only aspect that is known in relation to the index is that it considers a longitudinal profile every 2.5 cm. Colombia created the Bike Lane Condition Index (ICC in Spanish) (Martínez et al., 2011). This index includes a series of factors such as pavement material, geometric characteristics, surface roughness and special features of road failures. In order to collect this information, field evaluations are carried out every two years.

In any of the cases mentioned above, the resources required for a proper data collection hinder the subsequent control and/or monitoring that allow an adequate management of the network over time. An alternative to avoid this problem is using electronic devices, which collect data of the movement, either attached to the bike or worn by the user. In Concepción (Chile), Echaveguren et al. (2015) installed accelerometers and satellite navigation systems (GPS) on the bikes, with the aim of determining the surface condition of the city bike lanes. These devices recorded the vertical acceleration and their location when travelling along the bike lane. With this information, bike lanes were classified into three categories according to the acceleration level only. In this way, the purpose was to establish a global level of service, because accelerations are directly related to the comfort of users. However, driving maneuvers reflecting discomfort, which are not related to the vertical acceleration only, are expected.

The objective of this research is to develop a methodology that can identify points in the bike lane where the pavement presents defects or deteriorations, by using data collected from people's usual behavior, without the need for them to do any specific action (that is, from passive data).

This methodology will allow deducing the pavement condition in the analyzed section and thus improve the infrastructure management, enhance users' experience when travelling along and encouraging its use. The main hypothesis of this study is that cyclists behave differently based on the condition of the bike-lane pavement, and that these behaviors can be inferred from the data collected during their trips. This means to identify the data needed to detect pavement defects and, therefore, to build the models. Particularly, the aim is to determine the feasibility of using inertial measurement units (IMU) to obtain data that allow generating models able to identify the location of road defects. One of the advantages of using this type of technology is that they enable an automatic data collection, which in turn makes the application of this methodology easier. A further advantage of IMU is their small size and easy accessibility, which makes scaling easier. However, this technology requires an effort from the user to make data available, which can affect the number of users reporting their data.

The rest of the paper is composed of three chapters. Chapter 2 introduces the methodology proposed to meet the main objective. Chapter 3 presents the main results obtained after applying the technology. Finally, chapter 4 indicates the main conclusions of the study and future lines of research.

## 2. Methodology

The proposed methodology considers the use of models whose dependent variable is related to the bike-lane pavement condition and the explanatory variables correspond to data provided by inertial sensors. This requires the calibration and validation of the models proposed. These tasks require data and, therefore, the following subsection describes the experiment carried out to collect the data. The second part of the present chapter describes the type of models considered and the indicators used for comparing their results. It is important to highlight that the data collected will allow doing a proof of concept of the methodology. That is, the experimental design does not aim at determining the most appropriate model, but verifying the potential of the idea behind the proposed methodology.

### 2.1 Experimental design

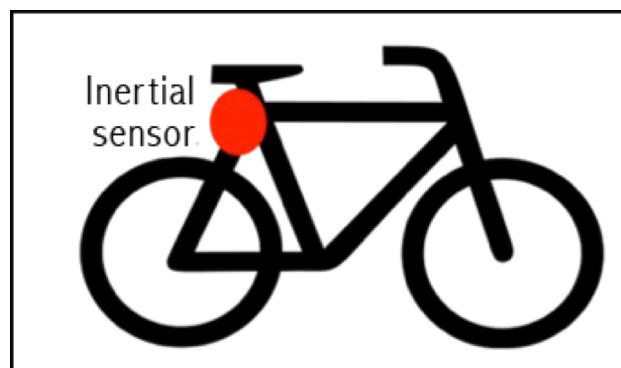
In order to perform the experiment, a 50-meter section of a bike lane inside the San Joaquín Campus of the Pontifical Catholic University of Chile was chosen. The location of this section is such that it allows the cyclist to reach the average circulation speed at its very beginning. This bike-lane wearing course is made of asphalt; it is two-way and isolated by physical dividers from the vehicle traffic flow. Figure 1 shows an aerial view of the location of the bike lane and a photograph showing its characteristics.



**Figure 1.** Map of the bike-lane location and photograph of its characteristics  
Source: Google Maps and self-preparation

Following the decision about the place for collecting the data, the instruments to be used were defined. A single person of 1.74 m high and 74 kg drove the 26" Moonstone Oxford bicycle with front shock absorber. An inertial sensor

was installed under the seat of the bike (Figure 2). The fact of using the same bicycle and driver allowed isolating the effect of other factors, which was a desirable aspect in this stage of the study.



**Figure 2.** Diagram showing the location of the inertial sensor

The inertial sensor was a Smartphone with an accelerometer, a gyroscope, a compass and a navigation and localization satellite system (GPS). The SensorLog application (Thomas, 2017) was used for collecting the necessary data. The data collection frequency was 30 observations per second, where each observation is composed of a data series. Working with a high data-collection frequency was important to capture the type of desired behavior, since the expected maneuvers are usually quite fast. For example, if data is

collected every 3 seconds, a maneuver for dodging or passing over a defect may not be captured.

Two specific, usually annoying problems for users were chosen: pavement upheavals (generally caused by the roots of trees) and the presence of potholes in the road. The latter tends to be more annoying under rainy weathers; when it accumulates water, it will most probably wet the cyclist when passing through (Martínez et al., 2011). When confronted to these issues, users tend to dodge or reduce

their speed to deal with road defects and to avoid the resulting inconveniences.

For recreating the first problem, wooden speed bumps were built in two different heights: six and four centimeters. The initial and final part of the speed bump was softened by a slope, which resembles real conditions. As for the pothole, cyclists tend to dodge this type of failures in the pavement; therefore, cones were used to recreate it. In this way, the

cyclist would be forced to make the maneuver that the study sought to analyze.

Five scenarios were built, where the first did not present alterations. Figure 3 shows the detail of the elements used and their location in the four other scenarios. One repetition was undertaken for the base scenario without alterations, while two repetitions were performed for the other scenarios with alterations. This allowed a total of nine runs throughout the analysis section with different elements.

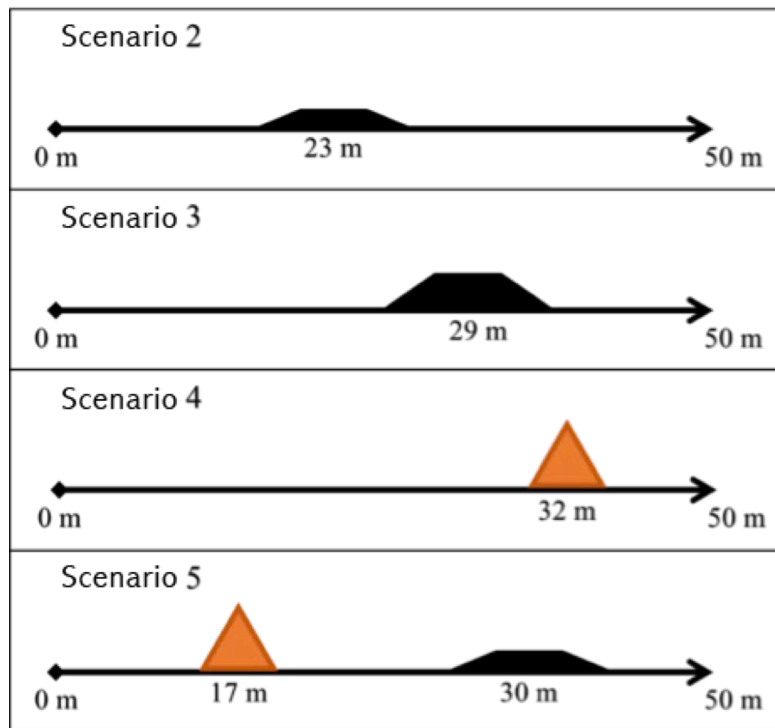


Figure 3. Experiment scenarios

### 2.1.1 Data Description and Processing

The SensorLog application provides a series of data for each observation, from which data presenting changes during the experiment and not related to each other were chosen. The following data were selected: instantaneous speed (m/s),

geographical location in coordinates, true north or horizontal orientation in azimuth (360 degrees), rotation in the three axes (rad/s) and real acceleration in the three axes expressed in G ( $G = 9.8 \text{ m/s}^2$ ). Figure 4 shows the orientation of each axis in relation to the bike's movement.

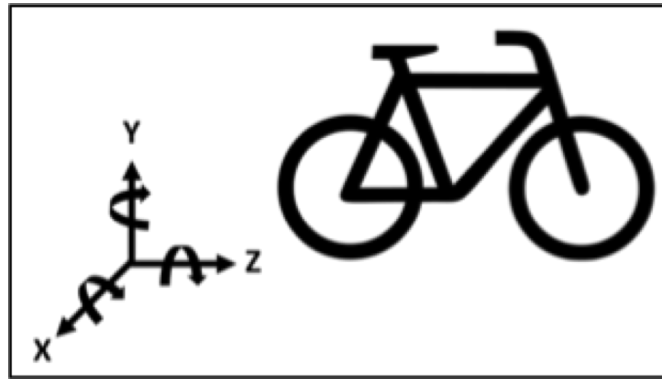


Figure 4. Diagram of the coordinates of the axes in relation to the bike's movement

The geographical coordinates of each observation were used to define their distance from the starting point (in meters). To that effect, the distance was calculated with Vincenty's formulas.

The observations were grouped every 50 centimeters, which defined 101 records per run. The observations collected between 25 cm upstream and 25 cm downstream were considered for each point. With regard to the speed and horizontal orientation, the arithmetic mean was calculated for each point, while the root mean square was calculated for the acceleration and rotation in each axis, since these values fluctuated around zero. This generated 909 records in total. The first 10 meters of each row presented problems in relation to the geographical location; therefore, the decision was made to eliminate these records. Following this purge, the total number of records was reduced to 729.

A new variable was created for each record, representing the presence of problems in the pavement ( $y$ ), and corresponding to the dependent variable that this work expects to predict. This variable adopts values of one or zero, where the first indicates that the respective record

corresponds to a position where a problem or deterioration on the pavement exists (and has an influence on cyclist behavior) and the second represents the absence of problems or deterioration.

In order to establish the values of variable  $y$ , the location of the elements in the simulated scenarios was used and the length affecting each element was determined. In relation to speed bumps, 1.5 meters around the problem was considered enough. Regarding the cones, the maneuver was expected to show a longer effect, because in order to dodge the cones it was necessary to change the behavior before and after the element itself. Therefore, a length of 2.5 meters was considered with the cones located at the center.

Table 1 shows an extract of the database used, where  $y$  is the estimated dependent variable. The course column indicates the run to which it belongs, thereby providing the simulated scenario. Dist (m) is the distance from the start point to the data recording point. The rotation and the acceleration around the 'i' axis are expressed by Rot<sub>i</sub> and Acc<sub>i</sub>, respectively. Figure 4 shows different rotations around each axis.

Table 1. Database Extract

$y$	Course	Dist. (m)	Speed (m/s)	Rot <sub>x</sub> (rad/s)	Rot <sub>y</sub> (rad/s)	Rot <sub>z</sub> (rad/s)	Acc <sub>x</sub> (G)	Acc <sub>y</sub> (G)	Acc <sub>z</sub> (G)
0	2	13,5	3,57	0,09	0,10	0,19	0,20	0,23	0,17
0	2	16	3,57	0,08	0,08	0,00	0,07	0,09	0,17
1	3	24,5	3,51	1,33	0,27	0,25	0,52	1,13	0,92

## 2.2 Model Definition and Construction

In order to meet the proposed objective, a logistic regression was used, which allows relating a dependent variable to a set of independent variables, in a non-linear

way. In these regressions, the dependent variable is characterized by adopting values ranging from zero to one, which can be interpreted as a probability. Maximum



likelihood estimation allow us to estimate this type of models (Hosmer and Lemeshow, 2000).

The probability of occurrence of the object of study is estimated according to Equation 1.

$$P_i = \frac{1}{1 + e^{-f(\bar{x}_i)}} \quad (1)$$

In this case,  $P_i$  represents the probability of the analyzed point  $i$  to present a problem on the pavement. This probability is determined based on the characteristics gathered in point  $i$ , which correspond to a subset of the data presented for each row of Table 1, represented by vector  $\bar{x}_i$ . Thus,  $f(\bar{x}_i)$  is the functional form in which these independent variables gathered in point  $i$  relate to each other.

Consequently, the generation of the model consists in determining the function  $f(\bar{x}_i)$  that best allows deducing the presence of defects on the bike lane. This means to define the variables that are part of the function and, therefore, to identify those best capable of capturing the cyclist's maneuvers that are related to the presence of deterioration in the pavement. Thus, when determining the bike lane points that will most likely present problems or defects, it is possible to infer the pavement condition in that section.

A correlation analysis is carried out in order to define the independent variables included in  $f(\bar{x}_i)$  (Table 2). The analysis considers that one variable is related to another if their correlation values are higher than 0.5, and they have a close relation if this value is higher than 0.7 (Devijver and Kittler, 1982).

**Table 2.** Correlation between variables

	<b>Rot_x</b>	<b>Rot_y</b>	<b>Rot_z</b>	<b>Acc_x</b>	<b>Acc_y</b>	<b>Acc_z</b>	<b>Speed</b>
<b>Rot_x</b>	1						
<b>Rot_y</b>	0,12	1					
<b>Rot_z</b>	0,07	<b>0,57</b>	1				
<b>Acc_x</b>	<b>0,53</b>	0,40	0,27	1			
<b>Acc_y</b>	<b>0,77</b>	0,10	0,09	<b>0,56</b>	1		
<b>Acc_z</b>	<b>0,61</b>	0,08	0,05	<b>0,51</b>	<b>0,72</b>	1	
<b>Speed</b>	-0,10	-0,01	-0,01	-0,05	-0,05	-0,06	1

Since several rates are close to the cutoff value (0.5), the decision was made to further analyze the three closely correlated variables: Rot\_x, Acc\_y, and Acc\_z. Rot\_x and Acc\_y are related when the cyclist encounters a speed bump, because the acceleration in the 'y' direction and the rotation around the 'x' axis are expected to increase. These two variables are also related to Acc\_z, because the cyclist instinctively reduces the speed (changes in Acc\_z) when encountering a problem in the bike lane. Therefore, none of the combinations of these variables was included in the same model. Following this consideration, different combination of independent variables were tested.

Finally, four models were evaluated in the experiment presented in this study. That is, four different formulations of  $f(\bar{x}_i)$  were evaluated. Considering these four models, Equation 2 established the general formulation of function  $f(\bar{x}_i)$ .

$$f(\bar{x}_i) = \alpha + \beta_y \cdot \text{Acc}_y + \gamma_x \cdot \text{Rot}_x + \gamma_y \cdot \text{Rot}_y + \gamma_z \cdot \text{Rot}_z \quad (2)$$

Depending on the variables to be included in each model,  $\alpha, \beta_y, \gamma_x, \gamma_y, \gamma_z$  are the parameters to be estimated for each of them.

With regard to each model proposed, each parameter's statistical significance was defined by the Z-test for a 95% confidence interval. For models in general, a typical goodness-of-fit test was run, which analyzed whether the proposed model was statistically different from the model considering just the constant. Moreover, the value of pseudo R was considered, which is related to the model's predictive capacity.

Additionally, the cross-validation technique was used with the aim of ensuring that the resulting parameters were independent from the database division (Jung and Hu, 2015). Therefore, the 729 records were randomly divided in five files. Iteratively, one of them was used for the validation of the models, while the other four were used in the calibration stage.

### 2.2.1 Analyses

A sensitivity analysis was carried out for each variable of each model. That is, we analyzed how the value of the dependent variable changes as the value of the studied independent variable changes, thereby keeping all others constant (and equal to their average value in the sample).

With the purpose of comparing the proposed models, a detailed study was undertaken dealing with the number of



ENGLISH VERSION:

mistakes made and coincidences achieved according to the True Positives (TP), False Positives (FP), True Negatives (TN)

and False Negatives (FN). Figure 5 details each concept according to the actual data and the results of the models.

<b>Model</b>		<b>Reality</b>	
		0	1
	0	TN	FN
	1	FP	TP

**Figure 5.** Results of a Diagnosis Test

Errors type I (FP), type II (FN) and True Positives will be studied specifically. The first are usually less risky in this case, since they can be considered as a preventive measure. However, they imply a field-review expenditure when it is not required. Type II errors are concerning, because, in this study, it means that a deterioration in the pavement has been overlooked. True Positives are equal to the coincidences of actual problems, so they are highly important.

In order to calculate these values, the mistakes made by the models will be averaged in the five cross-validation files. Regarding the True Positives, the average coincidence of each model will be calculated.

Finally, the models' TP and FN are analyzed in detail according to the simulated element: high speed bump, low

speed bump and cones. The aim is to associate specific models to certain failures in the pavement and to identify the most relevant variables in each element.

### 3. Results

Table 3 shows the average parameters obtained for the four selected models, the values for maximum likelihood, the pseudo R and the classification of the models according to their variables. For comparison purposes, the table also includes the results for the model considering constants only (model E).

**Table 3.** Average values for each model

<b>Model</b>	$\alpha$	$\beta_y$ (Acc_y)	$\gamma_x$ (Rot_x)	$\gamma_y$ (Rot_y)	$\gamma_z$ (Rot_z)	<b>Likelihood</b>	<b>Pseudo R<sup>2</sup></b>	<b>Classification</b>
A	-7,41	-	13,35	2,36	5,60	-43,42	64%	Complex-Rotation
B	-7,64	8,43	-	3,96	4,27	-46,21	61%	Complex-Acceleration
C	-7,17	-	14,25	-	7,10	-46,12	61%	Simple- Rotation
D	-6,71	7,85	-	-	6,48	-53,74	55%	Simple- Acceleration
E	-3,06	-	-	-	-	-106,31	-	Only constants

Thus, as an example, model B is defined by Equation 3, where  $f(x)$  is as follows:

$$f(x) = -7,64 + 8,43 \cdot \text{Acc}_y + 3,96 \cdot \text{Rot}_y + 4,27 \cdot \text{Rot}_z \quad (3)$$

The rotation around the 'z' axis ( $\text{Rot}_z$ ), associated to the natural inclination when maneuvering the bike, is observed in all four models. The rotations when dodging cones or facing speed bumps are very important. In these situations, the cyclist usually loses balance, and rotations are more relevant than accelerations.

Models A and C include the rotation around the 'x' axis and models B and D present the acceleration in the direction of the 'y' axis. Therefore, the first group will be

called rotational models and the second group, acceleration models. The rotation around the 'y' axis is present in models A and B, which implies that, in models C and D, a deviation of the axis through which the cyclist travels, is only explained by the rotation around the 'z' axis. Since models A and B have an additional variable, they were considered complex models, and models C and D were called simple models.

In relation to the sensitivity analysis of the variables, Figure 6 shows the results obtained with the different models. In each graph, the vertical axis represents the probability that the pavement shows deterioration as each variable changes. The variables of each model are in the horizontal axis. The rotations are expressed in radians per second and the acceleration in G.

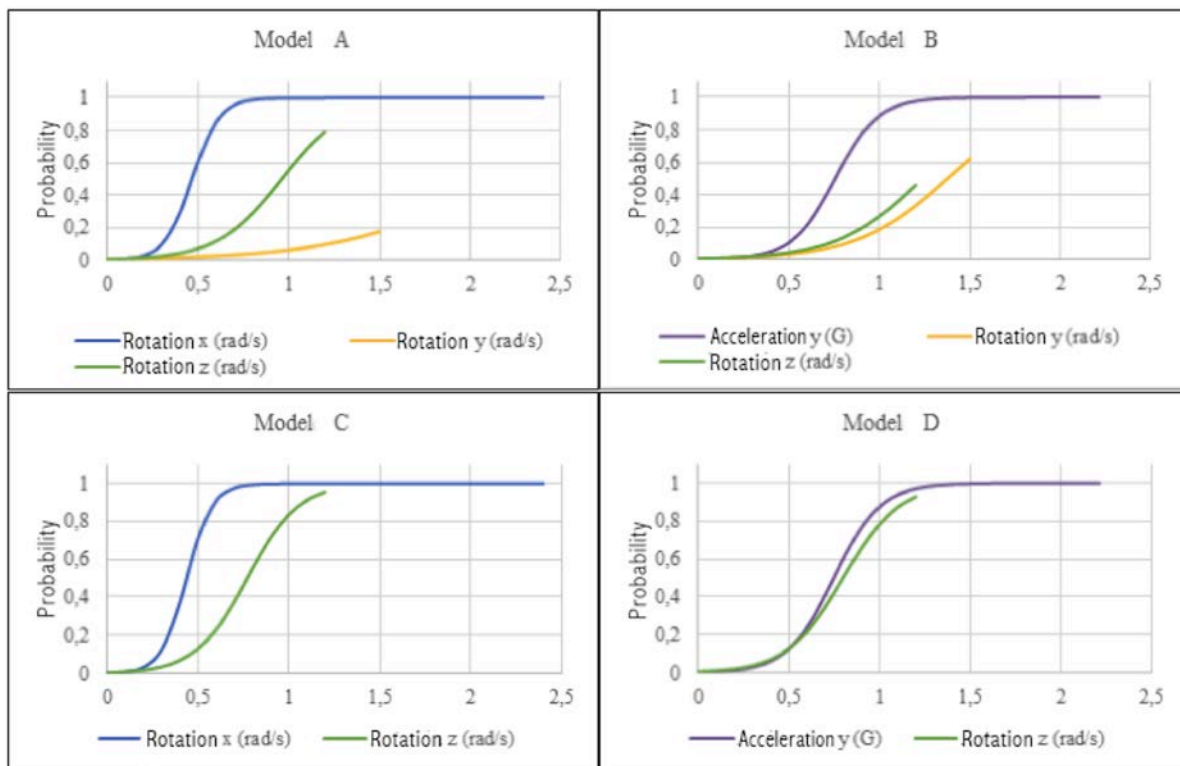


Figure 6. Sensitivity of variables for each model

The shape of the curves of variables  $\text{Rot}_x$  and  $\text{Acc}_y$  reveals their importance when determining a deterioration in the pavement. Both variables have a robust threshold for change, because it remains pretty much the same in both complex (A and B) and simple models (C and D). For  $\text{Rot}_x$ ,

the limit is approximately 0.5 radians per second, while for  $\text{Acc}_y$  it has to exceed 0.8 G.

When analyzing the complex models,  $\text{Rot}_y$  stands out for its little influence on Model A. In the model with acceleration, the variable has an influence on the probability



ENGLISH VERSION.....

of occurrence of the condition when its values are close to the higher limit.

The *Rot\_z* variable (which is included in all models) is highly relevant in simple models (C and D). The curve of this variable gets sharper, similar to the curve of *Rot\_x* and *Acc\_y*. In both models, the threshold for change is equal to 0.8 radians per second, which demonstrates the robustness of this variable in these models.

In order to compare these models with the validation data, it was necessary to calculate the average of how many times (considering the five different subsets used as validation data) these models made the error type I and the error type II, as well as the average percentage of coincidences (Table 4).

When analyzing errors type I, model C makes more mistakes, while model A makes an error only 1.6 times, on average. With regard to errors type II, acceleration models make more mistakes than rotational models. In both cases, the complex version of the model makes more mistakes than the simple version. Brief, model A is preferred, despite the fact that, marginally speaking, it makes more errors type II than model C. As for the percentage of coincidences, once again, rotational models obtain better results than those with acceleration, and simple models can identify more failures than complex models (within their category).

**Table 4.** Errors and True Positives by model, with validation data

<b>Model</b>	<b>Error Type I</b>	<b>Error Type II</b>	<b>True Positives</b>
A	1,6	3,4	55%
B	1,8	3,8	50%
C	2,6	3,2	58%
D	1,8	3,6	53%

A differentiated analysis for all three elements simulated with the validation data is presented below.

High Speed Bump

Table 5 shows the coincidences on the total number

of cases with  $\gamma=1$  for all four models and for each of the three points where  $\gamma=1$ . All four models detect the failure with the records obtained in the exact location of the high speed bump and the next location. The records obtained in the location prior to the element show an average performance.

**Table 5.** Coincidences for high speed bump by model and depending on the location of the record in relation to the failure

<b>Model</b>	<b>PRIOR</b>	<b>EXACT</b>	<b>AFTER</b>	<b>TOTAL</b>
A	1/2	2/2	2/2	<b>83%</b>
B	1/2	2/2	2/2	<b>83%</b>
C	1/2	2/2	2/2	<b>83%</b>
D	2/2	2/2	2/2	<b>100%</b>

### Low Speed Bump

Table 6 shows the same information as Table 5, but in relation to the low speed bump. The records obtained in the location prior to the low speed bump do not allow detecting the failure easily, regardless of the model. However, in the

same way as the high speed bump, all four models detect the failure with the records obtained in the exact location and the next one. Once again, simple models (C and D) make a better prediction than complex models in relation to this element.

**Table 6.** Coincidences for low speed bump by model and depending on the location of the record in relation to the failure

<b>Model</b>	<b>PRIOR</b>	<b>EXACT</b>	<b>AFTER</b>	<b>TOTAL</b>
A	1/4	3/4	4/4	<b>67%</b>
B	1/4	4/4	3/4	<b>67%</b>
C	2/4	3/4	4/4	<b>75%</b>
D	1/4	4/4	4/4	<b>75%</b>

### Potholes

Table 7 is similar to the two tables above, but in this case the effect of the failure ( $y=1$ ) is considered in five points. The records obtained in the far end locations practically do not allow identifying the pothole, regardless of the model. The records obtained in the center locations allow all four models to partially identify the failure. If all the five locations are

considered, models A and C (rotational) yield better results. However, they do not reach a 50% performance. If just the three center locations are analyzed, models A and C exceed a 50% performance. This suggests that the fact of considering the records immediately before and after the exact location of the pothole would improve the performance of the models.

**Table 7.** Coincidences for potholes by model and depending on the location of the record in relation to the failure

<b>Modelo</b>	<b>INITIAL</b>	<b>PRIOR</b>	<b>EXACT</b>	<b>AFTER</b>	<b>FINAL</b>	<b>SUBTOTAL</b>	<b>TOTAL</b>
A	0/4	2/4	3/4	2/4	1/4	<b>58%</b>	<b>40%</b>
B	0/4	2/4	2/4	1/4	1/4	<b>42%</b>	<b>30%</b>
C	1/4	3/4	2/4	2/4	0/4	<b>58%</b>	<b>40%</b>
D	1/4	1/4	2/4	1/4	0/4	<b>33%</b>	<b>25%</b>

As a summary of the analysis by elements, and in relation to speed bumps, all four models behave in a similar way, where the most simple ones are more accurate. With regard to the pothole, rotational models show a better response than those with accelerations. Consequently, it is preferable to choose model A or C.

#### 4. Final remarks

The proposed methodology is able to detect failures in the bike-lane pavement in an objective and specific way. Therefore, when determining the points in the bike lane that will most probably present failures, it is possible to deduce the pavement condition in that section.

The logistic regressions did not pose problems for building the models with the characteristics of the sample and the dependent variable. Despite the fact that the experiment was very controlled, the designed methodology is easy to calibrate and apply at a large scale and, consequently, to real scenarios.

With regard to the experiment carried out, all four models can objectively, autonomously and correctly identify the problems studied on the pavement of the bike-lane. Although all four models show similar behaviors, model A slightly stands out due to the values obtained in the experiment presented herein. The construction of these models allowed identifying the data needed to evaluate the condition of the pavement, where the rotations in the three axes and the vertical acceleration are the most important.

An advantage of the proposed methodology is the use of inertial sensors, since it simplifies the data collection process. Due to the technological progress and the increasing offer of these devices, they are now easily accessible for anyone. Likewise, it is not an invasive methodology for obtaining data; therefore, in order to determine the condition of the pavement, it is not necessary to make interventions nor close the bike lane. All you need is bikes travelling with the sensor installed on them.

An important aspect to be considered when scaling this methodology is the bias recorded in the data, due to the driving style of the cyclist and the type of bicycle used

(Lépine et al., 2011). In order to validate the data, a great number of cyclists are required to circulate several times on the same section of analysis. Additionally, it is also necessary to collect data referred to other failures that might appear in the pavement. Future works should deal with all these aspects, with the aim of obtaining a representative model for other situations.

Relying on a larger amount of data would allow refining the models developed in this work. In this perspective, new problems on bike-lane pavements could be identified, and large-scale models could be calibrated on site. Depending on the amount of data, it would be interesting to analyze other methodologies for developing these models. For example, the use of properly calibrated learning machines, which take into account the great number of potential problems, similar to the approach adopted by Catalan et al. (2018) in the context of driving public-transport vehicles.

Given the data requirement mentioned above, the simplicity of this technology to obtain data gains more importance. For example, public bikes can be equipped with inertial sensors and GPS. In this way, the data needed to apply the proposed methodology could be collected on a continuous basis and on a large scale and thus identify the condition of the pavement in the bike lanes. In this case, it is even possible to calibrate the device and the model based on the own characteristics of public bikes, in addition to study other difficulties that cyclists must face or other types of sections (with curves, inclination, etc.).

Finally, the methodology developed in this work can be used in various ways. In the short run, it provides the necessary information to develop a pavement management system specifically for bike lanes, because it allows identifying the problems on the road and making the corresponding maintenance works. In the long run, this methodology can generate a pavement quality index for bike lanes, based on the vibrations experienced by cyclists, so that a level of service can be associated to them. The level of service would enable the development of a pavement management system to improve users' experience, provide information for building cycling route selection models, and it could be useful for calculating the capacity of bike lanes.

#### 5. References

- Barbudo A., Jiménez R. R., Ledesma E. F., Sierra M. J. (2015), Regularidad superficial y adherencia en vías ciclistas - recomendaciones de diseño disponibles. *Informes de La Construcción*, 67(540), 124-135. <https://doi.org/10.3989/ic.15.010>
- Cairney P., King K. (2003), *Development of a performance based specification for a major bicycle facility*. ARRB Transport Research. Vermont, Victoria, Australia.
- Catalán H., Löbel H., Herrera J.C. (2018), Classifying Drivers' Behavior in Public Transport Using Inertial Measurement Units and Decision Trees. *The 21st IEEE International Conference on Intelligent Transportation Systems*. 4-7 de Noviembre. Maui, Hawaii, USA.
- Devijver P. A., Kittler, J. (1982), *Pattern Recognition: A Statistical Approach*. Prentice-Hall, Londres, Inglaterra.
- Echaveguren T., Montoya C., Tudela A. (2015), Evaluación de Condición Superficial de Ciclovías usando Aceleraciones Verticales Obtenidas con GPS. *XVII Congreso Chileno de Ingeniería de Transporte*, 1, 1-14. Concepción, Chile.
- Fuentes L., Gunaratne M., Hess D. (2010), Evaluation of the effect of pavement roughness on skid resistance. *Journal of Transportation Engineering*, 136(7), 640-653. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000118](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000118)
- Garrard J., Rose G., Lo S. K. (2008), Promoting transportation cycling for women: The role of bicycle infrastructure. *Preventive Medicine*, 46(1), 55-59. <https://doi.org/10.1016/j.ypmed.2007.07.010>
- Hosmer D. W., Lemeshow S. (2000), *Applied Logistic Regression*. (2da Ed.). New York: Chichester, Wiley.
- Jensen S. U. (2007), Pedestrian and Bicyclist Level of Service on Roadway Segments. *Transportation Research Record*, 2031(1), 43-51. <https://doi.org/10.3141/2031-06>



- Jung Y., Hu J. (2015)**, AK-fold averaging cross-validation procedure. *Journal of nonparametric statistics*, 27(2), 167-179. <http://dx.doi.org/10.1080/10485252.2015.1010532>
- Kohler E. (2015)**, Medición de Irregularidad Superficial en Pavimentos de Aeropuertos, Viales y Ciclovías. *Ingeniería de Transporte*, 18(2), 21-24.
- Landis B. (1994)**, Bicycle Interaction Hazard Score: a Theoretical Model. *Transportation Research Record*, (1438), 3-8.
- Landis B., Vattikuti V., Brannick M. (1997)**, Real-Time Human Perceptions: Toward a Bicycle Level of Service. *Transportation Research Record*, 1578(1), 119-126. <https://doi.org/10.3141/1578-15>.
- Lépine J., Champoux Y., & Drouet, J. M. (2011)**, Excitation techniques for testing bike vibration transmission in the laboratory. En *Society for Experimental Mechanics* (Vol. 6, pp. 263-289). Québec, Canada.
- Lépine J., Champoux Y., & Drouet J. M. (2014)**, Road bike comfort: on the measurement of vibrations induced to cyclist. *Sports Engineering*, 17, 113-122. <https://doi.org/10.1007/s12283-013-0145-8>
- Martens, S. (2011)**. Bicycles measure the quality of cycle paths in Belgium. *Eltis the urban mobility portal*. Recuperado de <http://www.eltis.org>.
- Martínez G., Fuentes L. G., Torregroza L. M. (2011)**, Revisión del sistema de gestión de pavimentos de la red ciclorrutas de Bogotá. *Revista ingeniería de construcción*, 26(2), 150-170.
- MINVU (2015)**. Construcción de ciclovías: Estándar técnico. Espacios públicos urbanos. Gobierno de Chile. Santiago, Chile.
- Parkin J. (2009)**, The humps and the bumps: objective measurement using an instrumented bicycle. *Research and Innovation Conference*, 1, 3.
- Pucher J., Dill J., & Handy S. (2010)**, Infrastructure, programs, and policies to increase bicycling: an international review. *Preventive medicine*, 50, S106-S125. <https://doi.org/10.1016/j.ypmed.2009.07.028>
- Pucher J., Buehler R., Seinen M. (2011)**, Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. *Transportation Research Part A: Policy and Practice*, 45(6), 451-475. <https://doi.org/10.1016/j.tra.2011.03.001>
- Sprinkle Consulting Inc. (2007)**, Bicycle Level Of Service: Applies Model. Tampa, Florida, Estados Unidos.
- Thomas B. (2017)**, SensorLog (Versión 1.9.2) [Software de Aplicación Móvil].

