

Modelo de clasificación del riesgo en intersecciones rurales en T y validación del tiempo de evasión como medición alternativa de la seguridad de tránsito en intersecciones

Risk classification model in rural T-form intersections and time to evasion evaluation as surrogate safety measure

Alejandro Torres Flores*¹, Francisco Torres Avilés*, José M^a Pardillo Mayora**

*Universidad de Santiago de Chile. CHILE

**Universidad Politécnica de Madrid. ESPAÑA

Fecha de recepción: 01/ 06/ 2010

Fecha de aceptación: 31/ 08/ 2010

PAG. 353 - 370

Resumen

En los cruces o intersecciones, los conductores afrontan una multitud de opciones relacionadas con el camino, velocidad, y trayectoria que, en combinación con numerosos movimientos del tránsito, complican la tarea del conductor y aumentan considerablemente la potencial ocurrencia de un accidente. En este artículo se utilizan datos de conflictos de tránsito, tomados en cruces de la Provincia de Talagante en la Región Metropolitana de Santiago, Chile. Usando la metodología de componentes principales, se propone un índice que represente el nivel de riesgo de cada conflicto de tránsito en una intersección, determinándose que los valores más altos del índice están relacionados a un mayor riesgo de que un conflicto termine en accidente, mientras que valores bajos indican un menor riesgo. Posteriormente, se desarrolla un modelo de clasificación del riesgo en intersecciones, aplicando la teoría de árboles de decisión. Por último, se valida el Tiempo de Evasión (TE) como variable cuantificadora de la severidad de los conflictos de tránsito, tomando como base variables ya estudiadas a nivel mundial, como el Tiempo hasta la Colisión (TC) y el Tiempo de Post-invasión (TPI).

Palabras Clave: Análisis de componentes principales, árboles de clasificación, conflicto de tránsito, tiempo a la colisión, tiempo de evasión

Abstract

In the intersections, the drivers confront many options related to the way, speed and trajectory, which ones in combination to many options of traffic movements, they make difficult the task of the conductor and considerably increase the potential occurrence of an accident. In this paper, data of traffic conflicts are used, taken in rural T-form intersections from Talagante district in the Metropolitana region, Chile. Using the principal component analysis, an index are proposed that represents the risk level of each traffic conflict in an intersection, determining that the highest values of the index are related to a greater risk of than a traffic conflict it finished in accident, whereas low values indicate a smaller risk. After that, a model of classification of the risk in intersections is developed, applying the theory of decision trees. Finally, Time to Evasion is validated like quantifying variable of the traffic conflicts, taking as it bases studied variables at world-wide level, like Time to Collision and Post - encroachment Time.

Keywords: Principal component analysis, classification trees, time to collision, time to evasion, traffic conflicts

1. Introducción

Uno de los objetivos que se persigue al diseñar una intersección es minimizar el número de potenciales eventos que pudieran traducirse en accidentes, a la vez de facilitar la comprensión y movimientos a través de ella por parte de los usuarios, con los estándares más altos de seguridad. Es en este punto donde radica la importancia de realizar estudios rápidos conducentes a evaluar la seguridad en intersecciones, así como también ir evaluando de manera pronta las mejoras efectuadas.

1. Introduction

One of the main objectives intended to be achieved when designing an intersection is to minimize the events leading to potential accidents, as well as to facilitate the users to understand the displacement through it by employing the highest safety standards. The importance of developing expedite studies leading to asses intersection safety lays in this factor, as well as the prompt evaluation of improvements conducted.

¹ Autor de correspondencia / Corresponding author:

E-mail: alejandro.torres@usach.cl



Para determinar si una intersección es segura, se han empleado diversas metodologías como los modelos de predicción de accidentes, modelos de simulación de tránsito y técnicas que no necesitan grandes bases de datos para determinar el potencial riesgo de accidentes, como el análisis de conflictos de tránsito y las auditorías de seguridad vial o las llamadas mediciones alternativas, las que fundamentan esta investigación.

Las mediciones alternativas de seguridad vial podrían ser un indicador para cuantificar la seguridad, y ayudar a decidir entre diferentes alternativas, sin tener que recurrir a los costosos estudios de accidentalidad o a los no menos costosos proyectos de reconstrucción; una opción es complementar este estudio con otro tipo de medidas para determinar la seguridad en los diferentes elementos de la carretera; precisamente en este punto es donde pueden tener un amplio campo de aplicación las mediciones alternativas de seguridad en intersecciones.

2. La técnica del conflicto de tránsito

Perkins y Harris (1968) identificaron los patrones de conflictos de tráfico que relacionaron con algún tipo de accidente, la técnica fue considerada como una medida del potencial de accidentes. Glauz y Migletz (1982), aportan las siguientes consideraciones,

- a) Relación con la seguridad: los conflictos deben estar relacionados con los accidentes.
- b) Relación con el lugar: los conflictos de tránsito pueden ser muy útiles para diagnosticar problemas en un determinado lugar o como una medida de la eficiencia de la mejora de algún sitio de la vía.
- c) Consistencia: la definición de un conflicto debe tener una variación mínima entre diferentes observadores que registran el mismo evento.
- d) Repetibilidad: la definición debería tener un nivel aceptable de variación en repetidas observaciones realizadas por el mismo observador, en el mismo sitio y bajo condiciones idénticas.
- e) Economía: los datos obtenidos deben ser fiables y obtenidos en un tiempo y a unos costos razonables.

En el año 1987 en la Universidad de Lund, Suecia el profesor Christer Hydén, introdujo un concepto que denominó Tiempo hasta el Accidente, dentro de lo que se conoce como Técnica Sueca de Conflictos de Tráfico, la cual se caracterizó principalmente por tres aspectos.

In order to determine whether an intersection is safe or not, diverse methodologies have been used such as accident prediction models, transit simulation models and techniques not requiring greater data bases to achieve an accident potential risk, as traffic conflict analysis and safety road audits or the so called surrogate measurement, which are the basis of the current research.

The surrogate measurement of road safety might be a useful indicator to quantify safety and help to choose among different alternatives, not needing to conduct expensive accident studies or quite expensive reconstruction projects. An option is to complement such studies with other type of measurements in order to determine the safety of different elements existing in a highway, which is precisely a way to obtain a wide application field from safety surrogate measurements at intersections.

2. Traffic conflict technique

Perkins and Harris (1968) identified the traffic conflict patterns related to some kind of accident. Such technique was considered as accidents potential measurement. Glauz and Migletz (1982), provide the following considerations.

- a) Relation with safety: conflicts must be related with accidents.
- b) Relation with the place: traffic conflicts may be useful to diagnose problems at a given place or as an efficiency measurement of the improvement conditions at a given point of the road.
- c) Soundness: the definition of a conflict must have a minimum variation among different observers registering the same event.
- d) Repetition: the definition should have an acceptable variation level at repeated observations conducted by the same observer, at the same place and under identical conditions.
- e) Economy: obtained data must be trustworthy and obtained at reasonable time and costs.

In 1987 at Lund, Sweden, Professor Christer Hydén introduced a concept named Time to Collision, as part of the Swedish Traffic Conflicts Technique, which was mainly characterized by three aspects.

- i) Definió medidas objetivas para medir la severidad de los conflictos: Tiempo al Accidente (TA) y Velocidad Conflictiva (VC).
- ii) Solamente estudia los conflictos considerados como severos.
- iii) La información se recopila de manera subjetiva por medio de observadores.

Los investigadores Almqvist y Hydén (1994) creen que esta técnica es la forma más directa de considerar una intersección con un alto riesgo de accidentes. Gettman y Head (2003) definieron los conflictos de tránsito como situaciones observables que pueden indicar la seguridad relativa que puedan tener diferentes diseños de intersecciones. Estos eventos de conflictos ocurren entre dos vehículos que están sobre una trayectoria de colisión, pero no ocurre debido a alguna acción evasiva; estos conflictos ocurren en un sitio particular en un tiempo y espacio, denominado punto de conflicto.

3. Mediciones alternativas o sustitutas de seguridad vial

Las mediciones alternativas de seguridad son medidas de tiempo, los cuales se registran durante la ocurrencia de un conflicto de tránsito. Estas mediciones evalúan cuán cerca estuvo un conflicto entre vehículos que circulan a través de una intersección, de terminar en una colisión y del nivel de severidad que hubiese tenido dicha colisión.

La medida alternativa de severidad de un conflicto más aceptada es el tiempo hasta la colisión (TC), aunque otras medidas (tiempo posterior a la invasión, TPI; o tasa de deceleración, TD) han sido propuestas como medidas de otras características de situaciones de conflicto (Gettman y Head, 2003). La dificultad radica en definir una serie de medidas secundarias que tengan una conexión razonable con la valoración de la seguridad de mejoras en las vías, además de poder registrar de manera sencilla estas medidas en estudios de campo.

Al revisar la bibliografía y los principales estudios desarrollados, se puede observar claramente que los autores reconocen el TC como una de las principales medidas de severidad de los conflictos (Perkins y Harris, 1968; Hayward, 1972; Allen et al., 1978; Glauz y Migletz, 1982; Hydén, 1987; Brown y Cooper, 1990; Van der Horst, 1990; Gettman y Head, 2003; Archer, 2005), cuyo umbral para considerar un conflicto severo lo fijan en $TC < 1.5$ segundos.

- i) Defined objective measurement to evaluate severity conflicts: Time to Collision (TC) Controversial Speed (CS).
- ii) Only severe conflicts were considered.
- iii) Information is collected in a subjective way by means of observers.

The researchers Almqvist and Hydén (1994) believe this technique is the most direct method to evaluate an intersection with high accident risks. Gettman and Head (2003) defined traffic conflicts as observable situations that may indicate the relative safety different intersection designs may have. Such conflict events take place between two vehicles running on a collision trajectory, but not due to an evasive action. These conflicts take place at a particular place and space known as conflict point.

3. Road safety alternative or surrogate measurement

Surrogate safety measurements are time measurement, which are registered while the traffic conflict is in progress. Such measurements evaluate how close the conflict between vehicles facing an intersection may end up in a collision and the severity level the collision might have.

The most accepted surrogate measurement of conflict severity is Time to Collision (TC), even though some others (post-encroachment time, PET; or deceleration rate, DR) have been proposed as measurement for different conflict situations (Gettman and Head, 2003). The problem is to define a series of secondary measurements having a reasonable connection with the assessment of improvements developed to road safety. Besides, it is necessary to register, as easily as possible, such measurements in field studies.

When reviewing bibliography and the main studies conducted, it is clearly noticeable that authors consider TC as one of the most important measurement of conflict severity (Perkins and Harris, 1968; Hayward, 1972; Allen et al., 1978; Glauz and Migletz, 1982; Hydén, 1987; Brown and Cooper, 1990; Van der Horst, 1990; Gettman and Head, 2003; Archer, 2005), which severe conflict threshold is fixed at $TC < 1.5$ seconds.



Para un punto de conflicto, la medida del TC se define como la diferencia entre el tiempo final de encuentro del vehículo que gira y el tiempo estimado en que llegará al punto de conflicto el vehículo que se traslada en línea recta y que posee la prioridad de paso, si éste continuara con la misma trayectoria y con la misma velocidad que poseía antes de comenzar a frenar para evitar la colisión.

Para un punto de conflicto, la medida del tiempo posterior a la invasión (TPI) se define como la diferencia de tiempo entre la salida del punto de conflicto de la zona de encuentro y la llegada a ese punto de conflicto del vehículo que se traslada en línea recta y que posee la prioridad de paso. En la revisión documental se establece que mientras más cercano a cero sea el valor de TPI, más cerca estuvo de producirse la colisión (Allen et al., 1978).

En la gráfica de la Figura 1, se muestra la ubicación tiempo-espacio de los conceptos punto de conflicto, TC y TPI.

En el presente estudio se introducirá la medida Tiempo de Evasión (TE), definida como el tiempo medido en segundos que transcurre al momento de realizar una maniobra evasiva (frenado) hasta que el parachoques delantero del vehículo llega al punto de conflicto definido. Este nuevo concepto, está asociado a la percepción del conductor al acercarse a la intersección, el cual entrega el tiempo que tiene para realizar una maniobra en caso de que se produzca un evento inesperado. Su cálculo queda definido por el cociente entre la Distancia de Frenado (DF) y la Velocidad de Aproximación (VA). Esta nueva medición alternativa de seguridad (TE), se correlacionará con las medidas ya estudiadas, como son TC y TPI, que serán las 3 mediciones alternativas de seguridad que se utilizarán en este estudio.

For a conflict point, TC measurement is defined as the difference between the final encounter time for a turning vehicle and the estimated time to reach the conflict by another vehicle displacing in a straight line, which has the right of way; provided that the latter continued the same trajectory and kept the same speed before starting to restrain itself in order to avoid collision.

For a conflict point, the measurement of post-encroachment time (PET) is defined as the difference of time between the exit of conflict point in the encounter zone and the arrival to such conflict point by a vehicle displacing in straight line, which has the right of way. Documentary review establishes that closest to zero value the PET was the closest probability of collision was (Allen et al., 1978).

Graph in Figure 1, indicates time-space allocation of conflict point concepts, TC and PET.

The current study will introduce the measurement Time to Evasion (TE), defined as time expressed in seconds, from the moment an evasive movement is made (braking) until the front end bumper reaches the defined point of conflict. This new concept is associated to the driver's perception when approaching the intersection, which provides time to make an evasive movement in case an unexpected event occurs. Calculation is defined by the quotient between Breaking Distance (BD) and Approaching Speed (AS). This new surrogate safety measurement (TE) will be correlated with measurement already studied, such as TC and PET, thus becoming 3 surrogate measurements used by this study.

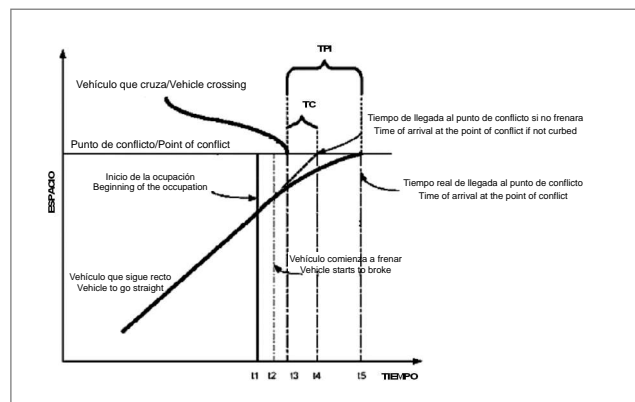


Figura 1. Ubicación tiempo-espacio de un punto de conflicto, tiempo hasta la colisión y tiempo posterior a la invasión

Figure 1. Ubication of time-space conflict point, time before collision and time after invasion

A su vez, se define la Distancia de Frenado (DF) como la longitud medida en metros, desde el punto donde es accionado el freno del vehículo (lo que da inicio a una maniobra evasiva), hasta el punto de conflicto definido en la intersección estudiada. La Velocidad de Aproximación (VA), velocidad medida 150 metros antes del punto de conflicto definido.

4. Estrategia metodológica

Está basada en los métodos multivariantes de la estadística clásica, los cuales son aplicados con el fin de analizar la información disponible. En esta investigación cada unidad experimental se define por los conflictos de tránsito.

Para el análisis se considera,

- a) El muestreo de información de las variables originales (TC, TE, TPI, DF y VA) para cada uno de los conflictos identificados en las intersecciones, incluyendo un análisis correlacional de la información.
- b) Resumir la información disponible a través de una nueva variable o índice, utilizando la metodología conocida como Componentes Principales (Peña, 2002). Esto permitirá construir un índice cuantitativo del riesgo para cada intersección incluida en el estudio, con las variables originales estandarizadas de forma multivariada.
- c) Finalmente, es de interés estudiar el comportamiento del índice propuesto con el fin de determinar posibles grupos en los datos que representen a los niveles de riesgo.

Se proponen reglas de clasificación basadas en la metodología propuesta por Hothorn et al. (2006), cuyo procedimiento general se enfoca en construir un árbol de regresión/clasificación midiendo sucesivamente la asociación entre la variable respuesta y las particiones de las covariables aplicando múltiples tests de permutación (Strasser y Weber, 1999), hasta que no se detecte asociación significativa, por lo que el proceso recursivo se detiene. Para cumplir con este objetivo se procede en tres fases.

1. Se construye un árbol de regresión independiente para cada intersección, cuya variable respuesta corresponderá al índice de riesgo cuantitativo obtenido del análisis de componentes principales. Esta teoría de regresión busca relacionar el conjunto de variables originales (DF, VA, TC, TPI y TE) con una variable respuesta, que en este caso corresponde al índice de riesgo de la intersección.

At the same time, Breaking Distance (BD) is defined as the length expressed in meters, from the point the vehicle break is pushed (which starts an evasive movement) up to the conflict point defined at the intersection studied. The Approaching Speed (AS) is measured 150 meters before the conflict point selected.

4. Methodological strategy

It is based on multivariate methods of classic statistics, which are used in order to analyze the information available. In this research each experimental unit is defined by traffic conflicts.

Such analysis considers:

- a) Information sampling of original variables (TC, TE, PET, BD and AS) for each conflict identified at intersections, including a correlation analysis of conflicts.
- b) Summary of available information by means of a new variable or index, using the methodology known as Principal Components (Peña, 2002). Besides, it will lead to a development of a quantitative risk index for each intersection comprised by this study, using a multivariate standardization of the original variables.
- c) Eventually, it is quite important to study the proposed index behavior in order to determine possible groups among data, which will represent the risk levels.

Proposed classification rules are created using the methodology presented by Hothorn et al. (2006), which general procedure is based on the construction of a regression/classification tree by successively measuring the association between the response variable and the covariate partitions by means of multiple permutation tests (Strasser and Weber, 1999), until no significant association is detected, consequently the recursive process ends. So as to reach this objective three phases were developed.

1. An independent regression tree is designed for each intersection, which response variable shall correspond to the quantitative risk index obtained from the analysis of main components. This regression theory is intended to relate a set of original variables (BD, AS, TC, PET and TE) with a response variable that in this case, corresponds to the intersection risk index.

2. La validación de cada árbol-modelo se realizará usando los modelos ajustados para cada intersección, y se seleccionará aquel que clasifica mejor, usando como medida de discrepancia el promedio de los cuadrados de las diferencias entre los valores observados y los predichos o error cuadrático medio (ECM) estimado.
3. Una vez seleccionado el modelo de árbol de regresión con el mínimo ECM y a partir del índice cuantitativo, se propone una regla de clasificación basada en un árbol cualitativo de clasificación, que indique la existencia o no existencia del riesgo, con el fin de dar una respuesta más parsimoniosa al problema de determinar si un conflicto que se observa en intersección es considerado en “riesgo alto de accidente” o “riesgo bajo de accidente”.

Los cálculos y análisis del presente trabajo se realizarán usando el software estadístico “R” y considerando un nivel de significación del 5%.

5. Recopilación y análisis de la información

Características de la Información

Las intersecciones seleccionadas para la recolección de datos en campo, están ubicadas en una zona de carácter rural de la provincia de Talagante, Región Metropolitana de Santiago, Chile. Las intersecciones corresponden a la ruta G-78 con G-374 (I1), la ruta G-30 con G-374 (I2) y la ruta G-40 con G-46 (I3), las cuales se muestran en la Figura 2.

2. The assessment of each tree model will be done by employing the adjusted models for each intersection and, the most suitable of them will be selected using as discrepancy measurement the average resulting from the difference between observed and predicted square values or the estimated mean square error (MSE).
3. Once the regression tree model is selected with a minimum MSE, and by using a quantitative index, a classification rule is proposed. Such rule is based on a qualitative classification tree, which indicates the presence or absence of risk thus providing a thorough response to evaluate when an observed intersection conflict is considered as an “accident high risk” or “accident low risk”.

Calculations and analysis in the current research will be carried out by using the “R” statistic software and considering a 5% significance level.

5. Information collection and analysis

Information Features

The intersections selected for information collection at field sites, are located in a rural zone in Talagante District, Santiago Metropolitan Region, Chile. Such intersections are G-78 highway intersecting G-374 (I1), G-30 highway intersecting G-374 (I2) and G-40 highway intersecting G-46 (I3), which are shown in Figure 2.



Figura 2. Intersecciones seleccionadas para el estudio: a) ruta G-78 con G-374; b) ruta G-30 con G-374; c) ruta G-40 con G-46 Fuente: Google Earth
Figure 2. Selected intersections for the study: a) road G-78 and G-374; b) road G-30 and G-374; c) road G-40 and G-46 Source: Google Earth

Los tres cruces están conformados por el encuentro de 2 vías bidireccionales, su prioridad de paso está regulado por una señal PARE. Las mediciones se efectuaron en época y horarios laborales (de 10:00 a 17:00 horas), durante 3 días para cada intersección, de las cuales se obtuvieron 635, 463 y 591 conflictos medidos para las intersecciones I1, I2 e I3, respectivamente. Estas intersecciones tienen características similares, así se esperaría que no hubiera factores externos que marquen diferencias en los resultados.

The three crossroads are shaped up by the encounter of two 2-ways roads and there is a STOP signal regulating the right of way. Measurements were conducted during labor periods and working hours (from 10 am to 5 pm), during three days for each intersection, from which 635, 463 and 591 measured conflicts were obtained for intersections I1, I2 and I3, respectively. Such intersections have similar features, so it is expected that no external factors would make a difference in results.

Análisis de Asociación

En la Tabla 1, la matriz triangular superior muestra el nivel de asociación lineal que existe entre las variables involucradas en el estudio, mientras que la triangular inferior muestra los p-valores asociados a la hipótesis de correlación igual cero. Note que el nivel de asociación entre variables es alto, reafirmados por los p-valores inferiores a 0.05, para las variables relacionadas con los tiempos de medición en las 3 intersecciones propuestas.

Association Analysis

In Table 1, the upper triangular matrix shows the linear association level existing among the involved variables in this study, while lower triangular matrix shows p-values associated with the correlation hypothesis equal zero. Notice that the association level among variables is high, confirmed by p-values lower than 0.05 for variables related to measure times in the 3 proposed intersections.

Tabla 1. Coeficiente de correlación y p-valores correspondiente a las intersecciones 1, 2 y 3
Table 1. Correlation coefficient and p-values corresponding to intersections 1, 2 and 3

	Intersección/Intersection	DF	VA	TE	TC	TPI
DF	I1:G78 G374					
	I2:G30 G374		-0.063 0.002	0.740 0.768	0.734 0.733	0.596 0.631
	I3:G40 G46		0.047	0.689	0.619	0.480
VA	I1:G78 G374					
	I2:G30 G374	0.111 0.974		-0.668 -0.583	-0.665 -0.616	-0.574 -0.485
	I3:G40 G46	0.251		-0.635	-0.628	-0.527
TE	I1:G78 G374					
	I2:G30 G374	<0.001 <0.001	<0.001 <0.001		0.996 0.989	0.812 0.727
	I3:G40 G46	<0.001	<0.001		0.878	0.663
TC	I1:G78 G374					
	I2:G30 G374	<0.001 <0.001	<0.001 <0.001	<0.001 <0.001		0.811 0.736
	I3:G40 G46	<0.001	<0.001	<0.001		0.789
TPI	I1:G78 G374					
	I2:G30 G374	<0.001 <0.001	<0.001 <0.001	<0.001 <0.001	<0.001 <0.001	
	I3:G40 G46	<0.001	<0.001	<0.001	<0.001	

Los valores negativos existentes en la triangular superior de la tabla, muestran consistencias respecto a lo observado en terreno, pues se espera que a mayor VA menores sean los tiempos asociados a la llegada del vehículo a la intersección. Cabe destacar que entre DF y VA existen valores cercanos a cero (p -valores >0.05), lo que implica que DF estaría no correlacionada linealmente con la velocidad del vehículo antes de realizar la maniobra evasiva. Tiene sentido esta conclusión, puesto que para este cálculo se usó el coeficiente de correlación de Pearson, el cual detecta sólo asociación lineal.

Modelo de clasificación cuantitativo del riesgo de las intersecciones

Las componentes principales son nuevas variables que sintetizan la información contenida en las originales. Estas nuevas variables, son no correlacionadas y se encuentran en orden decreciente según orden de importancia. Se presentan sólo las dos primeras componentes, debido a que el porcentaje de variabilidad acumulado explicado por ambas dimensiones supera el 70% en cada intersección.

Para el análisis de componentes principales, utilizado para construir el índice cuantitativo de riesgo, se usará la matriz de correlaciones estimada en cada intersección. Las componentes principales y sus variabilidades son presentadas en la Tabla 2.

The existing negative values in the upper triangular matrix shown in the table, are consistent with field observations, since it is expected that at higher AS shorter the associated time would be for the vehicle to reach the intersection. It is important to mention that between BF and AS values are close to zero (p -values >0.05), which means that BD would not be linearly correlated with the vehicle speed before executing an evasive movement. This conclusion makes sense, since this calculation used Pearson coefficient correlation, which only detects linear association.

Quantitative intersection risk classification model

Principal components are defined as new variables summarizing the information contained in the original ones. Such new variables are not correlated and they are placed in decreasing order in accordance to importance level. Only the two first components are indicated, since the accumulated variability percentage expressed by both dimensions exceeds the 70% in each intersection.

For the analysis of main components, employed to build up the risk quantitative index, an estimated correlation matrix will be used for each intersection. The principal components and their variables are shown in Table 2.

Tabla 2. Componentes principales correspondientes a las intersecciones en estudio
Table 2. Main components corresponding to intersections under study

	Intersección/Intersection G78 G374		Intersección/Intersection 2 G30 G374		Intersección/Intersection 3 G40 G46	
	Comp. 1	Comp. 2	Comp. 1	Comp. 2	Comp. 1	Comp. 2
-						
DF	0.382	0.675	0.406	0.620	0.361	0.710
VA	-0.353	0.737	-0.322	0.784	-0.361	0.701
TE	0.510	0.015	0.515	0.017	0.510	0.048
TC	0.509	0.012	0.515	-0.031	0.519	-0.008
TPI	0.459	-0.026	0.449	0.016	0.458	-0.053
Varianza/Variance	3.740	0.938	3.595	1.003	3.425	1.051
% Variabilidad Explicada/Variability Explained	74.80	18.75	71.91	20.06	68.50	21.02

Además de verificar que las componentes escogidas acumularan una variabilidad explicada mayor a un 60%, se exigió que las componentes tuvieran una interpretación lógica relacionada con el riesgo de la intersección. Así, se justifica consistentemente la construcción del índice.

Debido a que la primera componente principal, cuyos coeficientes son los asociados a la combinación lineal del conjunto de variables originales que formará el índice, explica el valor umbral de porcentaje acumulado de variabilidad exigido (> 65%) en las tres intersecciones, se puede decir que estas nuevas variables resumen adecuadamente la información de las variables originales asociadas a las mediciones alternativas de seguridad, luego, las primeras componentes son potenciales candidatas a representar un índice de riesgo.

Considerando las primeras componentes para cada intersección, los índices propuestos para cada una de las intersecciones involucradas estarán dadas por las fórmulas (1) - (3).

Índice de I1/Index for I1

$$I_1 = -0.382 DF + 0.353 VA - 0.510 TE - 0.509 TC - 0.459 TPI \quad (1)$$

Índice de I2/Index for I2

$$I_2 = -0.406 DF + 0.322 VA - 0.515 TE - 0.509 TC - 0.449 TPI \quad (2)$$

Índice de I3/Index for I3

$$I_3 = -0.361 DF + 0.360 VA - 0.510 TE - 0.519 TC - 0.458 TPI \quad (3)$$

Es posible verificar que la interpretación de la primera componente principal se asocia a un índice de riesgo de la intersección, pues, los coeficientes de esta componente son negativos para las variables que de tener valores pequeños reflejan un mayor riesgo de la intersección, como son las variables DF, TC, TPI y TE. Por otra parte, el coeficiente de la variable VA es positivo, por lo que al presentar valores altos se obtiene también un mayor riesgo de la intersección. En síntesis, valores bajos del índice representan un bajo riesgo de la intersección y valores altos representan un alto riesgo.

Los coeficientes que acompañan a cada una de las variables originales, representan el peso que tienen éstas sobre la variable respuesta. En los índices obtenidos, las variables TE y TC son las de mayor peso, lo que sugiere una mayor importancia de estas mediciones. En la Tabla 3 se observan las contribuciones en términos porcentuales. Con las variables estandarizadas, se controla el peso que puedan entregar las variabilidades marginales. Luego de este análisis, se descarta el uso del segundo componente, puesto que explica menor variabilidad que el primero.

In addition to make sure that selected components accumulate an explained variability higher than 60%, it was demanded that components had a logical interpretation related to the intersection risk. Consequently the construction of index is consistently justified.

Since the first principal component, which coefficients are the ones associated with the linear combination of the original variable set, to be comprised by the index, expresses the threshold of demanded accumulated percentage value (> 65%) for the three intersections. Therefore, it can be observed that these new variables properly summarize the information from original variables associated with surrogate safety measurement, then, the first components are potential candidates to represent a risk index.

Considering the first components for each intersection, the proposed indexes for every involved intersection are given by formulas (1) - (3).

It is possible to conclude that the interpretation of the first principal component is associated with an intersection risk index, since the coefficients of these components are negative for the variables. If the latter had smaller values, it would reflect a greater intersection risk, such as variables BD, TC, PET and TE. On the other side the coefficient of SA variable is positive, so when yielding greater values, a greater intersection risk is also obtained. Summarizing, lower index values indicate a lower intersection risk and higher values represent a higher risk.

The coefficients accompanying each one of these original variables represent the importance they have on the response variables. In the obtained indexes, the TE and TC variables are the most important ones, which suggest a greater importance of these measurements. Table 3 shows the contribution on percentage terms. By using standardized variables, the importance provided by marginal variability can be controlled. Concluded this analysis, the use of a second component is discarded, since it expresses a lower variability than the first one.



Tabla 3. Contribuciones de las variables originales en los índices de riesgo propuestos
 Table 3. Contribution of original variables on proposed risk indexes

	Contribuciones por índice de riesgo/Index of risk contributions (%)		
	Intersección/Intersection G78 G374	Intersección/Intersection 2 G30 G374	Intersección/Intersection 3 G40 G46
DF	17.26	18.40	16.35
VA	15.95	14.59	16.30
TE	23.05	23.33	23.10
TC	23.00	23.33	23.51
TPI	20.74	20.34	20.74

La distribución de los valores estimados de cada índice se muestra en la Tabla 4. Lo más relevante que se puede observar es la posibilidad de proponer además un punto de corte evidente, pues, el 5% de los vehículos de la muestra presentan valores de índice mayores a 2.529 para la I1, 3.073 para la I2 y 3.248 para la I3; clara evidencia que un índice cercano a 3.0 puede ser un valor referencial que permita distinguir si un vehículo está en un riesgo inminente de accidente. Además, el 5% de los vehículos de la muestra que presentan estos valores, coincide con el valor de $TC < 1.5$ seg, umbral para considerar un conflicto severo respaldado en la bibliografía consultada, sobre todo en las experiencias de Allen et al. (1978), Hydén (1987), Van der Horst (1990) y Archer (2005).

Una vez determinado el índice para cada intersección, se ajustó un árbol por cada una de ellas, etapa considerada como de entrenamiento, donde se ajustaron tres modelos cuya respuesta era el valor del índice calculado a partir de las componentes principales. El árbol procederá a utilizar a las variables originales como factores para clasificar a la variable respuesta o índice (Hothorn et al., 2006). Una vez que se dispuso de los tres modelos de árbol, se consideró a las otras dos intersecciones como muestras independientes y se procedió a ajustarlas con los distintos modelos obtenidos en la etapa de entrenamiento (etapa de validación). Finalmente se calculan los errores cuadráticos medios asociados a cada índice predicho por los tres modelos de la etapa de entrenamiento. En la Tabla 5, se observan los errores cuadráticos de "predicción", los cuales se pueden considerar bajos, dado que el índice propuesto toma valores en el rango (-5.0, 5.0). Aunque la información asociada a la I1 entrega errores muy parecidos a la I2, el modelo y datos de esta última intersección será el utilizado para el proceso de construcción del árbol de decisión del riesgo de carácter cualitativo.

The distribution of estimated values for each index is shown in Table 4. The most relevant fact is the possibility of proposing a clear cut off point, since the 5% of vehicles in the sample, yield values higher than 2.529 for I1; 3.073 for I2 and 3.248 for I3; which is a clear evidence that an index close to 3.0 could be a reference value to distinguish whether a vehicle is under imminent risk of accident. Furthermore, the 5% of vehicles in the sample yielding such values, matches the $TC < 1.5$ sec, which is the threshold to be considered as severe, according to searched bibliography, as well as from experiences by Allen et al. (1978), Hydén (1987), Van der Horst (1990) and Archer (2005).

Once the index is determined for each intersection, a tree was fitted for each one of them, which is considered as a training stage, where three models were fitted and which response corresponds to the index value computed from principal component analysis. The tree will proceed to use the original variables as classification factors of the response or index (Hothorn et al., 2006). After elaborating the three tree models, the other two intersections were considered as independent samples and were fitted with the different models obtained from the training stage (validation stage). Finally the mean square errors associated with each predicted index, for three models in the training stage, are calculated. Table 5 indicates the "predictive" square errors, which may be considered as low, provided that the proposed index takes values in the range of (-5.0, 5.0). Although the information associated to I1 provides errors quite similar to I2, the model and data from the latter intersection will be used in the process of building a decision tree of qualitative risk.

Tabla 4. Medidas de distribución de los índices de riesgo propuestos
 Table 4. Distribution measurements of proposed risk indexes

Intersección/Intersection	Mínimo/Minimum	1er. Cuartil/1st Quartile	Mediana/Median	Media	3er. Cuartil/3rd Quartile	Percentil/Percentile 95	Máximo/Maximum
G78 G374	-7,919	-1,524	0,557	-1,1e-15	1,455	2,529	3,771
G30 G374	-5,715	-1,281	-0,354	8,1e-16	1,601	3,073	4,569
G40 G46	-5,663	-1,263	-0,384	-2,5e-15	1,673	3,248	4,580

Tabla 5. Error cuadrático medio del índice de riesgo, según la base de datos ajustada
 Table 5. Average quadratic error of risk index, as per adjusted data base

Error cuadrático medio/Mean square error	Predicción/Prediction			
	Observado/Observed			
Intersección/Intersection		G78 G374	G30 G374	G40 G46
G78 G374		0.040	0.134	0.340
G30 G374		0.137	0.046	0.290
G40 G46		0.148	0.163	0.089

El árbol de clasificación seleccionado para clasificar el índice de riesgo se muestra en la Figura 3. Note la importancia que presenta la variable TC en la clasificación del riesgo del índice. Las otras variables que permiten determinar el nivel cuantitativo de riesgo son TE y TPI. El efecto “a mayor valor del índice, mayor es el riesgo en la intersección” se puede observar inicialmente en el árbol de I2 (Figura 3), pues cuando $TC \leq 3.25$ [s], $TC \leq 2.12$ [s] y $TC \leq 1.47$ [s], el índice toma valores mayores a 3, que equivale al 5.2% del total de conflictos en la muestra de entrenamiento. Si $TC > 1.47$ [s], los valores del índice muestran un riesgo relativamente alto, pues toma valores entre 2.0 y 3.0. En esta etapa, TE y TPI son variables potenciales para clasificar índices que no explican un “muy alto riesgo”.

The classification tree selected to classify the index risk is shown in Figure 3. Notice the importance represented by TC variable in the classification of risk index. Other variables that facilitate the determination of quantitative risk level are TE and PET. The effect “the greater index value, the greater risk at intersection” can be initially observed in I2 tree (Figure 3) provided that when $TC \leq 3.25$ [sec], $TC \leq 2.12$ [sec] and $TC \leq 1.47$ [sec], the index reaches values higher than 3, which correspond to 5.2% of total conflicts in the training sample. If $TC > 1.47$ [sec], the index values show a relatively high risk, because it reaches values between 2.0 and 3.0. In this stage, TE and PET are potential variables to classify indexes that do not express a “very high risk”.

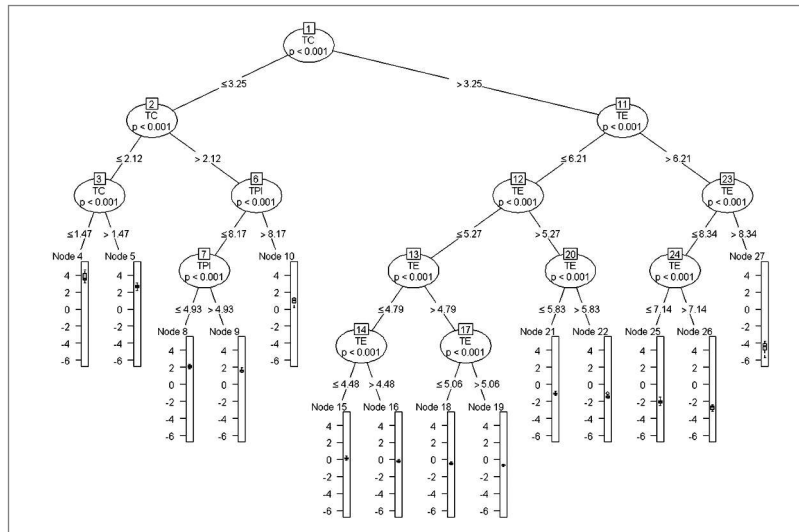


Figura 3. Árbol de clasificación índice cuantitativo del riesgo Intersección I2
 Figure 3. Classification tree of risk quantitative index for intersection I2

Modelo de clasificación cualitativo del riesgo de las intersecciones

En este punto se clasifica el riesgo de las intersecciones de una manera más simple y parsimoniosa. De los 14 grupos resultantes en el modelo cuantitativo (Figura 3), se procedió a construir una nueva variable respuesta asociada a sólo 2 grupos. El criterio fue categorizar el índice anterior considerando los diversos resultados obtenidos durante el desarrollo del presente trabajo y la bibliografía existente. La categorización se realizó en base a la clasificación obtenida para el grupo de más alto riesgo y se aplicó la metodología de Hothorn et al. (2006) para construir un árbol, denominado en este caso de Clasificación, puesto que la respuesta es cualitativa. Es así como los valores de Índice ≥ 3 y TC ≤ 1.47 [s], son de referencia para clasificar un evento en la intersección como de riesgo alto de accidente. Esta categorización es respaldada por investigaciones anteriores, pues se señala que el valor umbral de TC que indica un riesgo elevado de ocurrencia de una colisión es de 1.5 segundos.

Se establecieron las categorías “riesgo alto de accidente” (RA), correspondiente a un Índice ≥ 3 y TC ≤ 1.47 [s], y “riesgo bajo de accidente” (RB), correspondiente a un Índice < 3 y TC > 1.47 [s], dando como resultado el ajuste de un árbol de clasificación cualitativa con respuesta binaria (Figura 4).

Classification model of qualitative risk at intersections

In this point the risk at intersections is classified in a thorough and simpler manner. From 14 out of resulting groups in the quantitative model (Figure 3), a new response variable associated only to two groups was built. The criterion employed was to categorize the previous index, by considering diverse results obtained during the current study and available bibliography. The categorization was done based on the classification obtained from the highest risk group and the methodology by Hothorn et al. (2006) was applied to construct a tree, in this case called Classification Tree, because its response is of a qualitative nature. Therefore, Index values ≥ 3 and TC ≤ 1.47 [sec], are useful reference values to classify an intersection event as a high accident risk. This categorization is supported by previous researches, because it states that a high TC threshold, 1.5 sec, indicates a high collision risk.

“High accident risk” categories were established (HR), corresponding to an Index ≥ 3 and TC ≤ 1.47 [sec], and “Low accident risk” (LR), corresponding to an Index < 3 and TC > 1.47 [sec], providing as a result the adjustment of a qualitative classification tree with a binary response (Figure 4).

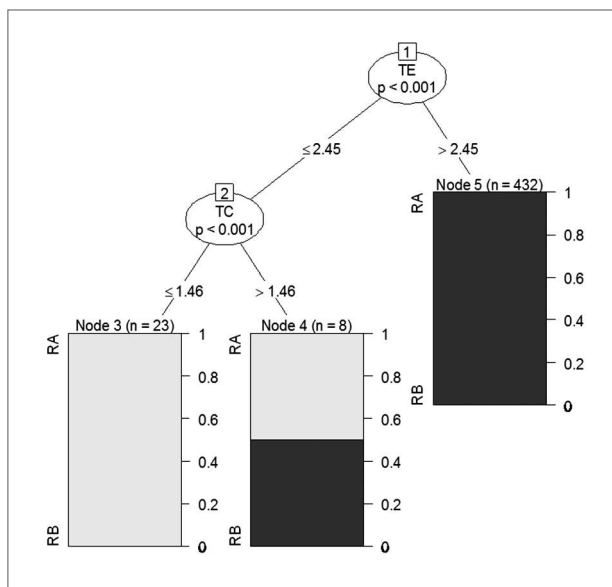


Figura 4. Árbol de clasificación cualitativo del riesgo para la intersección I2

Figure 4. Classification tree of risk qualitative risk for Intersection I2

Del árbol de respuesta binaria resultante, se puede comentar que sólo las variables TC y TE forman parte del modelo. Si se tienen valores de $TE > 2.45$ [s] el riesgo de accidente es bajo. Por el contrario, para $TE \leq 2.45$ [s] y $TC \leq 1.46$ [s] el riesgo de accidente es alto. Para $TE \leq 2.45$ [s] y $TC > 1.46$ [s] existe un grupo no asociado directamente a las clasificaciones RA o RB y que representa una categoría de incertidumbre, pues el 50% de ellas presentarían alto riesgo de accidente. Un porcentaje de estas observaciones son candidatas a ser mal clasificadas.

Técnicamente, del total de conflictos observados con las clasificaciones RA y RB, se calculó una tasa de clasificación incorrecta para cada intersección. La Tabla 6 muestra los porcentajes de clasificación incorrecta estimados al clasificar los datos de cada intersección, alcanzando a lo más un 2.5%, producida en la I1, del total de observaciones registradas en cada uno de los cruces en estudio, siendo un valor razonable para este árbol de respuesta binaria.

Para una mayor comprensión de los grupos creados por la categorización del modelo de clasificación cualitativo, se presentan los gráficos de dispersión TC v/s Índice de Riesgo y TE v/s Índice de Riesgo para la I2 en la Figura 5. En ella se observan los dos grupos de riesgo generados por los puntos de corte del árbol de clasificación cualitativo, teniendo en consideración que existe un 0.87% de observaciones mal clasificadas (Tabla 6).

The resulting binary tree shows that only TC and TE variables are comprised in the model. If values are $TE > 2.45$ [sec], the accident risk is low. In contrast, if $TE \leq 2.45$ [sec] and $TC \leq 1.46$ [sec], the accident risk is high. For $TE \leq 2.45$ [sec] and $TC > 1.46$ [sec], there is a group which is not directly associated to classifications HR and LR and represents an uncertainty group, since the 50% of them would represent a higher accident risk. A percentage of these observations are subject to misclassifications.

Technically, from total studied conflicts under HR and LR classifications, a missclassification rate was computed for each intersection. Table 6 shows the estimated missclassification percentages when classifying data at each intersection, reaching a 2.5% at the most for I1, from the total observations registered in each one crossroads in study, which is a reasonable value for this binary response tree.

For a better understanding of selected groups for the categorization of qualitative classification model, the graphs of TC dispersion v/s Risk Index and TE v/s Risk Index for I2 are shown in Figure 5. It shows two risk groups generated by cut off points of the qualitative classification tree, taking into account that there are a 0.87% of misclassified observations (Table 6).

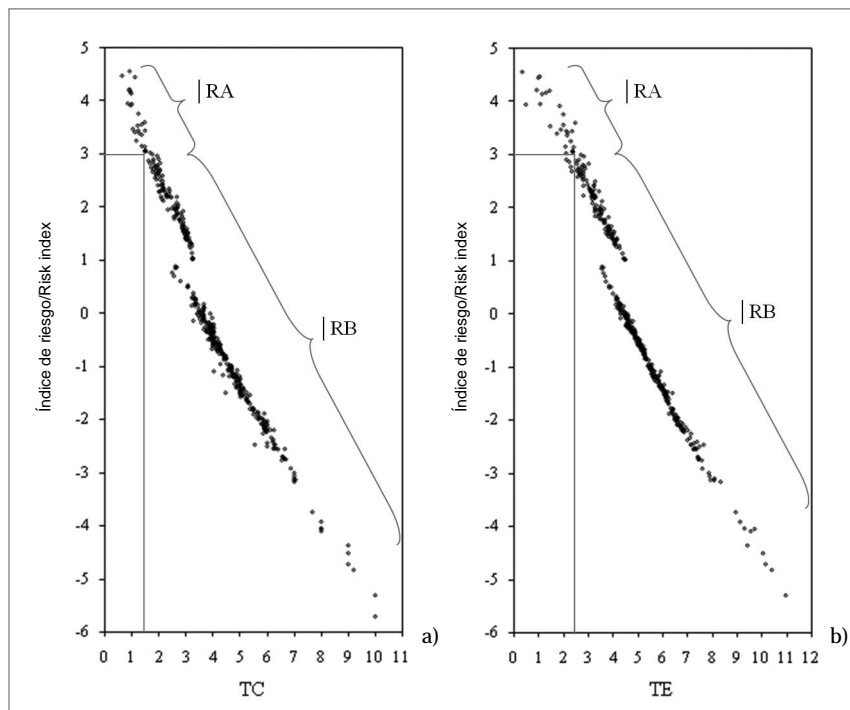


Figura 5. Gráficos de dispersión Índice de Riesgo versus a) Tiempo hasta la Colisión (TC) , b) Tiempo de Evasión (TE)
 Figure 5. Graphs for dispersion risk index versus a) Time to Collision (TC), b) Time to Evasion (TE)

Tabla 6. Porcentaje de clasificación incorrecta de los datos de cada intersección con el árbol de respuesta binaria
 Table 6. Percentage of data base incorrect classification for each intersection with binary response tree

Intersección/Intersection	Base Predicha/Predicted Base		
	G78 G374	G30 G374	G40 G46
Total observaciones/Total observation	635	463	591
N° mal clasificadas/N° misclassified	16 (RB: 16; RA: 0)	4 (RB: 4; RA: 0)	13 (RB: 2; RA: 11)
Mal clasificadas/Misclassified (%)	2,5	0,87	2,2

6. Discusión

La metodología propuesta para la construcción del Índice de Riesgo de la Intersección, resulta ser una herramienta útil para resumir el conjunto de mediciones obtenidas usando la Técnica de Conflictos de Tránsito. Esto permitió el diagnóstico de la seguridad vial en cada intersección. Su mayor utilidad está relacionada con la interpretación de la nueva variable indicadora, pues en este estudio, se pudo asociar la primera componente principal al riesgo de la intersección, donde cada medición alternativa de seguridad que forma parte del índice, tiene un peso representativo en relación a su importancia en el riesgo de la intersección.

Para cada uno de los índices de riesgo de las intersecciones, se concluye que las variables originales más influyentes son las mediciones de tiempo, especialmente TC y TE. Esto se refleja en los coeficientes o pesos que presentan estas variables en los índices de cada intersección.

A partir de la importancia de TE dentro de los índices de riesgo, fue posible probar que ésta es una medición alternativa de seguridad que puede ser utilizada para diagnosticar el riesgo de una intersección, siendo una característica significativa que introduce la percepción de la seguridad de un conductor al acercarse a un cruce. Así, se pudo determinar que los valores más altos del índice de riesgo están relacionados a un mayor riesgo de que un conflicto termine en accidente, mientras que valores bajos indican un menor riesgo.

El modelo de clasificación permitió encontrar la relación existente entre las mediciones alternativas de seguridad (DF, VA, TE, TC y TPI) y el índice de riesgo construido, generando distintos grupos de riesgo a partir de criterios de decisión asociados a las variables originales.

6. Discussion

The proposed methodology for the construction of Intersection Risk Index becomes a useful tool to summarize the set of measurement obtained by employing the Traffic Conflict Technique. Such methodology facilitated the diagnosis of road safety for each intersection. Its greater benefit is the interpretation of a new indicator variable, since this study permitted the association of the first principal component with intersection risks, where each surrogate safety measurement in the index has a representative importance in relation to its intersection risk importance.

It is concluded that, for each intersections risk index, the most influent original variables are time measurement, especially TC and TE. It is reflected by the coefficients or weights associated to those variables included in each intersection index.

From TE importance among risk indexes, it was possible to prove that this is an useful surrogate safety measurement to diagnose an intersection risk, thus being a significant feature that introduces a driver's safety perception when approaching a cross road. So it was possible to determine that higher risk index values are related to higher risk probabilities for a conflict to become an accident, meanwhile lower values indicate a lower risk.

The classification model allowed finding the existing relation between surrogate safety measurements (BD, AS, TE, TC and PET) and the constructed risk index, generating different risk groups from decision criteria associated with original variables.

Del modelo resultante (Figura 3) se observa que sólo las mediciones de tiempo TE, TC y TPI influyen en la clasificación del índice para nuevos datos, esto se explica dado que estas variables reúnen implícitamente la información asociada a DF y VA del vehículo en un conflicto. Con lo cual el algoritmo de creación del modelo elimina las variables DF y VA, considerándola información redundante. A partir de este modelo fue posible identificar los distintos grupos de riesgo generados. En el primer nodo, se dividen hacia la izquierda aquellas observaciones que presentan índices de riesgo positivos y hacia la derecha negativos, asociándose a grupos de mayor y menor riesgo, respectivamente. Por ende, la partición inicial genera una adecuada clasificación en términos de simplificar la identificación de dos grandes grupos de riesgo.

A partir de las dos ramas principales se generaron nuevas particiones, que finalmente terminan en 14 grupos de clasificación. Esta gran cantidad de grupos dificulta el asociarlos a un nivel de riesgo específico, ya que impide una interpretación práctica de cada uno de ellos. Sin embargo, su importancia radica en que a través de este análisis fue posible identificar el grupo de “más alto riesgo” dentro de los datos y las condiciones que deben cumplir los conflictos para formar parte de éste, Índice ≥ 3 y TC ≤ 1.47 [s], siendo esta última variable la que mejor clasifica a los conflictos con más altos índices de riesgo. El porcentaje de casos que son considerados dentro de este grupo asciende al 5.2% de todos los conflictos.

El modelo de clasificación cualitativo construido a partir de un índice de respuesta binaria, hace más práctica la clasificación encontrada con el árbol de clasificación cuantitativo, pues permite asociar cada conflicto a sólo dos grupos específicos de riesgo. De esta forma se identificó el grupo de los conflictos más graves como aquellos asociados a un alto riesgo “Riesgo Alto de accidente”, y aquel grupo cuya cualidad es presentar un bajo riesgo “Riesgo Bajo de accidente”. La primera categoría binaria (RA) corresponde al grupo de más alto índice en el modelo cuantitativo y la segunda categoría binaria (RB) reúne los trece grupos restantes.

Estos modelos facilitan la clasificación de un conflicto como de alto o bajo riesgo de accidente, dado que sólo requieren de las mediciones TE y TC para una adecuada categorización. Con esto, es posible corroborar la importancia de las variables TC y TE, ya que desde un comienzo del análisis presentaron una alta contribución en la construcción del índice y posteriormente se encontraron presentes en los árboles de clasificación cuantitativo y cualitativo, desplazando en este último a las mediciones DF, VA y TPI.

The resulting model (Figure 3) shows that only time measurements TE, TC and PET influence the classification index for new data, which is explained because these variables implicitly gather the information related to BD and AS of a vehicle during a conflict. Therefore the creation model algorithm eliminates BD and AS variables, considering them as redundant information. From this model it was possible to identify different risk groups generated. In the first node, those observations presenting positive risk indexes are placed at the left and the negative ones to the right, in relation to higher and lower risk groups, respectively. Consequently, the initial partition generates an adequate classification in terms of simplifying the identification of two greater risk groups.

From the two main branches new partitions were generated, that finally end up into 14 classification groups. This huge amount of groups makes the association with a specific risk level very difficult, since it obstructs a practical interpretation for each one of them. However, its importance lays in the fact that this analysis allowed the identification of the “highest risk group”, among data and conditions to be met by conflicts in order to become part of such group, Index ≥ 3 and TC ≤ 1.47 [sec], being this latter variable the one that allows the best classification of conflicts with the highest risk indexes. The percentage of cases considered by this group reaches the 5.2% of all conflicts.

The qualitative classification model built up from a binary response index makes the classification more practical against the quantitative classification tree, because it allows the association of each conflict with only two specific risk groups. In this way, the group of most serious conflicts was identified as those associated with a “High Accident Risk”, and the group that represents a low risk was identified as “Low Accident Risk”. The first binary category (HR) corresponds to the group of highest risk in the quantitative model and, the second binary category (LR) gathers the thirteen remaining groups.

Such models facilitate the classification of a conflict into high or low accident risk, since they only require TE and TC measurements for an adequate categorization. So it is possible to verify the importance of TC and TE variables, since from the analysis beginning they showed a high contribution to the index construction and later they became part of quantitative and qualitative classification trees, then setting the qualitative classification tree aside for BD, AS and PET.

Dentro de las variables del modelo cualitativo, el TC es la más estudiada entre las mediciones alternativas de seguridad, siendo descrita como la principal medida de severidad de conflictos. Las investigaciones anteriores indican que para un valor de TC menor a 1.5 segundos existe un elevado riesgo de ocurrencia de accidente, valor semejante al establecido por ambos árboles de clasificación desarrollados en el presente estudio, que clasifican conflictos con riesgo alto de accidente como aquellos cuyos valores de $TC < 1.47$ [s] en el modelo cuantitativo y $TC < 1.46$ [s] en el cualitativo. Con esto, es posible concluir que el algoritmo utilizado en la creación de los modelos de clasificación presenta una acertada categorización de riesgo en las intersecciones estudiadas.

En cuanto a la variable TE que forma parte del modelo cualitativo, sus estudios realizados en memorias de título anteriores se basan en medirla y buscar su relación con parámetros asociados a la seguridad. Sin embargo, no se evalúa su efectividad como medición alternativa de seguridad. El presente trabajo tuvo por objetivo validar el TE como variable cuantificadora de los conflictos de tránsito, obteniéndose como resultado la importancia de dicha variable dentro del análisis realizado. Esto se observa en una primera instancia en la correlación casi lineal existente entre el TE y el TC (0.878 a 0.996), variable cuya importancia ha sido demostrada a nivel mundial, y en su alta contribución en el índice de riesgo de la intersección (que promedia un 26.2%) con respecto a las otras variables utilizadas. La validación de TE fue finalizada con el análisis del modelo de clasificación cualitativo, donde su importancia radicó en formar parte de dicho árbol y obtener un valor crítico de 2.45 [s], asociado al riesgo de accidente.

Se puede concluir a partir del modelo de clasificación cualitativo (Figura 4), que es posible determinar cada grupo, ya sea de “riesgo alto” o “riesgo bajo”, sólo con las condiciones $TC \leq 1.46$ [s] o $TE \leq 2.45$ [s] para un conflicto de alto riesgo y, por el contrario, $TC > 1.46$ [s] o $TE > 2.45$ [s] para un conflicto con riesgo bajo de accidente. Esto se explica, puesto que el grupo que cumple las condiciones $TE \leq 2.45$ [s] y $TC > 1.46$ [s] no puede ser asociado a un nivel de riesgo específico (RA o RB), pues está compuesto simultáneamente por conflictos de riesgo alto y bajo, es decir, este grupo es más bien un error del modelo. En definitiva, el error asociado al modelo cualitativo, de a lo más un 2.5%, es el error cometido en decir que el conflicto con un $TE \leq 2.45$ [s] equivale a la condición $TC \leq 1.46$ [s] para clasificar los conflictos en alto riesgo de accidente.

Among qualitative model variants, TC is the most studied surrogate safety measurement, being described as the main severity conflicts measurement. Previous researches indicate that for a TC minor than 1.5 seconds there is a high risk of accident probability, which is a similar value to the one indicated for both classification trees developed by the current study, classifying conflicts as high accident risk those whose values are $TC < 1.47$ [sec] for quantitative model and $TC < 1.46$ [sec] for qualitative model. Consequently, it is possible to conclude that the algorithm employed in the creation of classification models yields a correct risk categorization for studied intersections.

As far as TE variable is concerned, which is part of qualitative model, the studies conducted for previous degree thesis are focused in its measurement and finding its relation with parameters associated to safety. However, its effectiveness as a surrogate safety measurement is not assessed. The current research had as main objective the validation of TE as a quantifying variant of traffic conflicts, obtaining as result the importance of such variant in the conducted analysis. In first instance this is observed by the almost linear correlation between TE and TC (0.878 a 0.996), which importance has been worldwide demonstrated, and due to a high contribution to the intersection risk index (26.2% average) in regards to other employed variants. TE validation was completed with the qualitative classification model, where its importance laid in becoming part of such tree and obtaining a 2.45 [sec] critical value, associated with the accident risk.

It can be concluded from a qualitative classification model (Figure 4), that it is possible to determine each group, either “high risk” or “lower risk”, only under $TC \leq 1.46$ [sec] or $TE \leq 2.45$ [sec] conditions, for a high risk conflict. On the contrary, $TC > 1.46$ [sec] or $TE > 2.45$ [sec] for a conflict with lowers accident risks. This is explained because the group that meets the requirements $TE \leq 2.45$ [sec] and $TC > 1.46$ [sec] cannot be associated with a specific risk level (HR or LR), since it is simultaneously made of high and low risk conflicts, meaning that this group is quite an error in the model. Definitively, the error associated to the qualitative model, a 2.5% at the most, is the error made by indicating that a conflict with $TE \leq 2.45$ [sec] is equivalent to condition $TC \leq 1.46$ [sec] to classify conflicts in high accident risk.

La equivalencia entre los puntos de corte obtenidos para TE y TC, puede ser reafirmada mediante la observación de los gráficos de dispersión de la Figura 5, en los cuales se distingue la similitud de los grupos generados con un valor crítico de $TC = 1.46$ [s] y un $TE = 2.45$ [s], ya sea RA o RB, cuya discontinuidad coincide con un quiebre estructural de los datos. Por último, el validar el TE como medición alternativa de seguridad, permite simplificar la evaluación del riesgo de las intersecciones estudiadas, puesto que el cálculo de TE sólo requiere la medición de la VA y DF determinadas en terreno, sin la necesidad de realizar el análisis de una línea temporal de sucesos para cada conflicto de tránsito.

7. Conclusión

Los modelos de clasificación cuantitativo y cualitativo entregan criterios que permiten tomar decisiones del riesgo que presenta una intersección, en base a puntos de corte adecuados, y aplicar el mismo modelo a la información de las otras intersecciones con tasas de error pequeño. Debido a esto, se ratifica que el modelo de la Intersección I2 se ajusta también para las otras intersecciones estudiadas, por lo que este modelo y sus valores críticos son representativos para intersecciones en forma de T con características similares a las utilizadas en este estudio. De la misma forma, la metodología utilizada, puede ser aplicada en la creación de modelos de clasificación del riesgo en cualquier intersección que permita determinar las variables DF, VA, TE, TC y TPI en conflictos de tránsito. En trabajos posteriores y sobre la base de nuevas mediciones en otras intersecciones en "T", se efectuará el ajuste del índice propuesto en esta primera etapa exploratoria.

Una ventaja identificada para estos modelos, es la evaluación del riesgo de una intersección luego de la mejora en cuanto a infraestructura y/o dispositivos de control de tránsito, al igual que un análisis Antes - Después, pero sin realizar una comparación mediante la ocurrencia de accidentes, sino que considerando sólo conflictos de tránsito. Esto se realiza comparando la variación del porcentaje de conflictos que son clasificados con "Riesgo alto de accidente", luego de efectuada la mejora en la intersección.

The equivalence among the cut off points obtained for TE and TC, can be confirmed by observing the dispersion graphs in Figure 5, which clearly distinguishes the similarity of groups generated with a critic value $TC = 1.46$ [sec] and a $TE = 2.45$ [sec], either HR or LR, which discontinuity matches with a data structural breakdown. Finally, validating TE as a surrogate safety measurement, allows the simplification of risk evaluation for the studied intersections, since TE calculation only requires the SA and BF measurement from field sites, without needing to conduct an incident time line analysis for each traffic conflict.

7. Conclusion

Quantitative and Qualitative classification models provide criteria that allow making risk decisions on a given intersection, based on adequate cut off points and, applying the same model to information from other intersection at low error rates. Therefore, it is confirmed that Model of Intersection I2 matches other studied intersections, so this model and critical values are representative for those intersection in T shape with similar features to the ones employed in this research. In the same way, the employed methodology may be applied for the creation of risk classification models at any intersection enabling the assessment of BD, AS, TE, TC and PET variables in traffic conflicts. For later studies and based on new measurement from other "T" intersections, the adjustment of the proposed index in this exploratory stage will be adjusted.

An identified advantage for these models is the evaluation of an intersection risk after introducing infrastructure improvements and/or transit control devices, as well as a Before-After analysis, but not comparing from accidents occurrences, however only considering traffic conflicts. This is done by comparing the average variation of conflicts classified as "High Accident Risk", and after the intersection is already improved.

Aunque los modelos creados permitan la clasificación de los conflictos como alto o bajo riesgo de accidentes, sería necesario un estudio que pueda asociar la cantidad de conflictos pertenecientes al grupo “Riesgo alto de accidente”, con la interpretación de una intersección segura o insegura. Sin embargo, este estudio propone la posibilidad de evaluar la seguridad de las intersecciones en términos comparativos a través del modelo de clasificación cualitativo.

Para futuras investigaciones asociadas al análisis realizado en el presente trabajo, se recomienda, determinar la efectividad de los modelos en la evaluación de la seguridad, a través del Tiempo de Evasión medido en aquellos eventos que no corresponden a conflictos de tránsito, y estudiar la correspondencia entre el índice creado y los conflictos de alto riesgo, con los accidentes que se puedan registrar en la intersección. Por otro lado, es recomendable realizar mediciones en intersecciones de diferente tipología, en cuanto a geometría (en cruz) y dispositivos de control de tránsito (semaforizadas), y analizar el comportamiento del modelo propuesto.

Although elaborated models allow the classification of conflicts as high or low accident risks, it would be necessary to count with a study enabling the association of conflicts amount pertaining to “High accident risk” group, with the interpretation of a safe or unsafe intersection. However, this study proposes the possibility of evaluating the safety intersections level in comparative terms, by means of a qualitative classification model.

For future researches associated with the analysis conducted by the current study, it is recommended to determine the models effectiveness for safety evaluation by means of measurement of Time to Evasion for those events that do not correspond to traffic conflicts and; to study the connection between the elaborated index and high risk conflicts with accidents that may occur at the intersection. On the other side, it is recommended to develop some measurement at different typology intersections, as far as geometry is concerned (cross-shaped) traffic control devices (traffic lights) and to analyze the proposed model behavior.

8. Referencias / References

- Allen, Shin, y Cooper (1978), “Analysis of Traffic Conflicts and Collisions” Transportation Research Record 667, pags. 67 – 74. Transportation Research Board, Estados Unidos.
- Almqvist S., y Hydén C. (1994), “Métodos para valorar la seguridad en el tráfico en países en desarrollo” Building Issues Vol. 6 N° 1, págs. 3 – 21.
- Archer (2005), “Indicators for traffic safety assessment and prediction in micro-simulation modeling: a study of urban and suburban intersections.
- Brown G. y Cooper P. (1990), “Role of conflicts in traffic analysis”
- Gettman D., y Head L. (2003), Surrogate Safety Measures From Traffic Simulation Models, Final Report. Publication N° FHWA-RD-03-050. U.S. Department of Transportation, Federal Highway Administration. Virginia.
- Glauz W. y Migletz D. J. (1982), “Traffic Conflicts Techniques for use at intersections” Transportation Research Record 770, pags. 21 – 29. Transportation Research Board, Estados Unidos.
- Hayward J. C. (1972), “Near-Miss Determination Through Use of a Scale of Danger” Report TTSC-7115. University Park, The Pennsylvania State University.
- Hothorn T., Hornik K. y Zeileis A. (2006), Unbiased Recursive Partitioning: A Conditional Inference Framework. Journal of Computational and Graphical Statistics, 15(3), 651-674.
- Hydén C. (1987), “The Swedish Traffic Conflict Technique” Departamento de Ingeniería y Planificación del Tráfico, Universidad de Lund, Suecia.
- Perkins S. y Harris J. (1968), “Traffic Conflicts Characteristics – Accident Potencial at Intersections” General Motors Research Laboratories.
- Peña D. (2002), Análisis de Datos Multivariantes. McGraw Hill.
- Strasser y Weber (1999), On the Asymptotic Theory of Permutation Statistics. Mathematical Methods of Statistics, 8, 220-250.
- Van Der Horst R. (1990), “A time-based analysis of road user behaviour at intersections”
- Zeeger C. y Parker M. (1988a), “Traffic Conflict Techniques for safety and operations – engineers guide” Departamento de Transportes de los Estados Unidos (FHWA – IP – 88 – 026).
- Zeeger C. y Parker M. (1988b), “Traffic Conflict Techniques for safety and operations – observers manual” Departamento de Transportes de los Estados Unidos (FHWA – IP – 88 – 027).

