

# LA INFLUENCIA DE ELEMENTOS DE LA CARRETERA EN VELOCIDADES OPERACIONALES A LO LARGO DE ZONAS DE TRANSICIÓN USANDO EL MÉTODO DE REGRESIÓN BASADA EN EL ÁRBOL JERÁRQUICO<sup>1</sup>

Yesenia Cruz-Cantillo<sup>2</sup> e Ivette Cruzado<sup>3</sup>

Resumen: Los segmentos de carreteras que unen las zonas de alta velocidad con las zonas de baja velocidad son conocidos como zonas de transición; es en estas zonas que los conductores deben ajustar sus velocidades operacionales. La seguridad en las zonas de transición estará comprometida si los conductores no reducen la velocidad antes de entrar a la zona de baja velocidad. Un estudio fue llevado a cabo con el propósito de identificar las características geométricas de la carretera que influyen el comportamiento de los conductores según datos de velocidades antes, durante y después de diversas zonas de transición. Datos de velocidades y detalles geométricos de la carretera fueron tomados en zonas de transición de carreteras rurales de dos carriles en Pennsylvania. Los datos tomados fueron analizados con el modelo estadístico Regresión Basada en Árbol Jerárquico (RBAJ) y los resultados indican que la velocidad máxima postulada de 25mph es la variable que mayor predice las velocidades a lo largo de las zonas de transición.

**Palabras clave:** seguridad, regresión basada en árbol jerárquico, zonas de transición.

## THE INFLUENCE OF HIGHWAY ELEMENTS IN OPERATIONAL SPEEDS THROUGHOUT TRANSITION ZONES USING A REGRESSION METHOD BASED ON HIERARCHICAL TREE

Abstract: The highway segments located between high-speed rural areas and low-speed urban environments are known as transition zones; along these zones drivers should adjust their operating speeds. The safety of transition zones is compromised when drivers fail to reduce their speeds before entering the low-speed zone. A study was performed in order to identify the highway geometric characteristics that influence driver behavior based on speed data collected before, along, and after several transition zone sites. Data on both speed and highway geometric characteristics were collected at several transition zones along two-lane rural highways in Pennsylvania. The Hierarchical Tree-Based Regression (HTBR) method was used to analyze the data and the results identified a posted speed limit of 25mph as the best predictor of operating speeds along transition zones.

**Key words:** hierarchical tree-based regression, safety, transition zones.

### INTRODUCCIÓN

Aproximadamente el 77% de las carreteras de Estados Unidos son carreteras rurales (FHWA, 2004). Aunque el volumen de tráfico en estas carreteras es relativamente bajo, el 60% de las fatalidades en Estados Unidos ocurren en carreteras rurales. Estas carreteras proveen conexión entre pueblos y las velocidades postuladas por lo general se encuentran en el rango de 40 a 55 mph. Sin embargo, las áreas desarrolladas (comunidades rurales) requieren

---

<sup>1</sup> Artículo recibido el 4 de diciembre de 2012 y aceptado para publicación el 6 de febrero de 2012.

<sup>2</sup> Estudiante Doctoral, Departamento de Ingeniería Civil y Agrimensura, Universidad de Puerto Rico, Mayagüez, PR 00681-9000. E-mail: yesenia.cruz3@upr.edu

<sup>3</sup> Catedrático Auxiliar, Departamento de Ingeniería Civil y Agrimensura, Universidad de Puerto Rico, Mayagüez, PR 00681-9000. E-mail: ivette.cruzado@upr.edu

velocidades operacionales menores (entre 25 a 35 mph). Los segmentos de carretera que conectan las áreas rurales de alta velocidad con las áreas desarrolladas (urbanas) de baja velocidad son conocidos como zonas de transición. Los conductores son avisados que deben reducir la velocidad antes de entrar a la zona urbana por medio de rótulos.

La Asociación Americana de Oficiales de Carreteras Estatales y Transportación (AASHTO, por sus siglas en inglés) provee guías y criterios de diseño para todo tipo de carreteras en su publicación Normas para el Diseño Geométrico de Calles y Carreteras (2004). Sin embargo, aunque dicha publicación provee guías de diseño para carreteras de dos carriles, ya sean rurales o urbanas, no existen guías de diseño para las zonas de transición entre estas dos áreas. Además, variaciones mínimas en velocidades son indicativas de un diseño consistente. En el caso de zonas de transición, una reducción en velocidades es deseada; esto representa un desafío para el desarrollo de guías de criterio para estas zonas. Por esta razón, el Programa Nacional Cooperativo de Investigación en Carreteras (NCHRP, por sus siglas en inglés) ha indicado la necesidad de llevar a cabo estudios de investigación en zonas de transición para desarrollar guías de diseño que puedan ayudar a ingenieros a conseguir consistencia en el diseño de carreteras.

La identificación de elementos geométricos que influyen en las velocidades a lo largo de zonas de transición ha sido previamente estudiada utilizando modelos de regresión lineal y multiniveles (Cruzado y Donnell, 2010). El estudio discutido en este informe utiliza los mismos datos pero aplica el método de Regresión Basada en Árbol Jerárquico para analizar la misma. La aplicación de varios modelos estadísticos a la misma data es una práctica común entre ingenieros de transportación para la identificación del mejor modelo que pueda explicar la relación entre las variables de interés.

## REVISIÓN DE LITERATURA

Varios estudios se concentran tanto en analizar las velocidades operacionales en carreteras rurales de dos carriles así como en carreteras urbanas. Sin embargo, la literatura en zonas de transición entre estas dos áreas es limitada. El considerar las características geométricas que influyen las velocidades en zonas rurales de alta velocidad así como en zonas urbanas de baja velocidad, puede ayudar a identificar los elementos geométricos que están relacionados a las velocidades entre estas dos áreas. Es por esto que la revisión de literatura de este estudio se concentró en estudios de velocidades en zonas rurales de alta velocidad, en zonas urbanas de baja velocidad y en zonas de transición.

Se han llevado a cabo varios estudios de velocidades en carreteras rurales de alta velocidad. El elemento geométrico de las carreteras que ha sido identificado constantemente como aquel que influye mayormente las velocidades operacionales es el radio de curva horizontal (o grado de curvatura). Varios estudios han identificado una relación negativa entre la percentil 85 de velocidades y grado de curvatura; según aumenta el grado de curvatura, las velocidades operacionales disminuyen (Glennon et al., 1985; Lamm y Chouciri, 1987; Krammes et al., 1995; Passetti y Pambro, 1999; Fitzpatrick et al., 2000). Todos estos estudios utilizaron regresión lineal para modelar las velocidades operacionales. Estudios llevados a cabo para modelar la reducción de velocidades entre segmentos tangenciales y curvas también han identificado una mayor reducción de velocidad asociada con aumento en grado de curvatura (McFadden y Elefteriadou, 2000; Misaghi y Hassan, 2005). Otras variables asociadas con reducciones de velocidades son longitud de tangente, velocidades altas en las tangentes, curvas verticales inclinadas, presencia de intersección y aumento en ángulo de deflexión. Además, Misaghi y Hassan identificaron una relación positiva entre ancho de paseo y reducción de velocidad entre tangentes y curvas (2005).

Aunque el uso de regresión lineal en estudios de velocidades es común, estudios recientes se han enfocado en aplicar diferentes modelos estadísticos para modelar las velocidades a lo largo de carreteras rurales de dos carriles. Park y Saccomanno desarrollaron modelos para predecir la percentil 85 de la reducción de velocidades entre tangente y curvas utilizando data agregada y desagregada. Los autores aplicaron un modelo multinivel el cual identificó una relación negativa entre radio de curva y reducción de velocidad. Los resultados del estudio también indicaron que el uso de regresión lineal subestima los parámetros de coeficiente, así como la desviación estándar de la variable independiente (radio de curva). El estudio realizado por Park y Saccomanno concluyen que el uso de modelos jerárquicos es más apropiado cuando se considera la asociación entre comportamiento del conductor (velocidades operacionales) y las características de la carretera ya que estos modelos eliminan la falacia ecológica que resulta al agregar los datos de velocidades, como es el caso de regresión lineal (2006).

Varios estudios investigan las velocidades operacionales en carreteras urbanas, incluyendo estudios que no utilizan regresión lineal. Un estudio llevado a cabo por Tarris et al. aplicaron modelos de data en panel para modelar

las velocidades operaciones en calles urbanas de baja velocidad (1996). Los resultados del estudio indican que un aumento en grado de curvatura está asociado con una reducción en percentil 85 de velocidades operaciones. Un estudio realizado por Poe y Mason llegaron a la misma conclusión; los autores usaron modelos mixtos para establecer la relación entre características geométricas y velocidades en carreteras urbanas (2000). Además de grado de curvatura, una variable asociada con índice de peligro (debido a la distancia de obstáculos cercanos a la carretera) fue identificada como influyente en velocidades operacionales. Wang et al. también utilizan modelos mixtos a data de velocidades en zonas urbanas; las variables independientes asociadas con cambios en velocidades fueron: presencia de encintado, obstáculos a la orilla de la carretera, presencia de acera, estacionamiento a la orilla de la carretera y presencia de intersecciones (2006).

Estudios enfocados a zonas de transición son escasos y, aunque guías de diseño están disponibles para carreteras rurales y urbanas, no hay guías de diseño para la transición entre estas dos zonas. Stamatiadis et al. realizaron un estudio debido a la necesidad de guías de diseño para las zonas de transición (Stamatiadis et al., 2004; Stamatiadis et al., 2006). Los autores indicaron que las carreteras rurales, al pasar por un área desarrollada, deben tener diferentes velocidades postuladas así como diferentes secciones transversales para influenciar al conductor en ajustar su velocidad operacional de acuerdo al área. Métodos de flexibilidad en el diseño fueron aplicados a 22 lugares de estudio en once estados; se recopiló data de accidentes antes y después de un cambio físico en el lugar de estudio. Ejemplos de flexibilidad en el diseño incluyeron aplicación de encintado e implementación de pacificadores de tráfico, entre otros. Los resultados del análisis indicaron una reducción en la frecuencia de accidentes durante el periodo “después” pero no hubo cambios significativos en las velocidades operacionales.

Cambios en velocidades a lo largo de 20 zonas de transición en Pennsylvania fueron analizados usando regresión lineal y modelos multinivel (Cruzado y Donnell, 2010). Los resultados indicaron que las características geométricas de la carreteras asociadas con reducciones en velocidad fueron reducciones en ancho de carril, paseo y distancia lateral libre de obstáculos, aumento en número de accesos, presencia de encintado, presencia de curva horizontal y aumento en distancias de zona de transición. Los resultados de ambos modelos fueron similares en términos de los coeficientes pero el modelo multinivel resultó en errores estándares mayores que los errores obtenidos con el modelo de regresión lineal.

Estudios recientes sugieren el uso de data desagregada para modelar velocidades, tales como data en panel, modelos mixtos y modelos multinivel, los mismos identifican las desventajas de utilizar data agregada, como lo es regresión lineal (Misaghi y Hassan, 2005; Park y Saccomanno, 2006). Estudios de transportación los cuales utilizan regresiones basadas en árboles jerárquicos son limitados; algunos de sus aplicaciones son frecuencia de accidentes, severidad de accidentes y emisiones vehiculares (Xuedong et al., 2010; Chang y Wang, 2006; Fomunung et al., 1999).

## RECOPIACIÓN DE DATOS

Para este estudio se tomo datos de la geometría de las carreteras en varias zonas de transición en Pennsylvania y velocidades a lo largo de estos segmentos para así poder determinar las características que están asociadas con velocidades operacionales. Se utilizaron los datos tomados por Cruzado y Donnell (2009).

### Datos de Características de la Carretera

Para poder determinar posibles lugares de estudio, se utilizó el sistema de vídeo del Departamento de Transportación de Pennsylvania (PennDOT). Todos los lugares de estudio estaban localizados a lo largo de carreteras rurales de dos carriles en la cual una reducción de velocidad operacional era requerida debido a la cercanía de una comunidad rural. La reducción de velocidad era indicada por la presencia de un rótulo de Reducción de Velocidad Adelante, el cual delineaba el comienzo de la zona de transición. Un requisito esencial era que el rótulo de Reducción de Velocidad Adelante estuviera acompañado por un rótulo de Velocidad Máxima Postulada indicando la nueva y reducida velocidad reglamentada; este segundo rótulo delineaba el final de la zona de transición. Solo se escogieron lugares de estudio que incluían la versión del rótulo de Reducción de Velocidad Adelante según especificado en la versión 2000 del Manual de Dispositivos Uniformes de Control de Tráfico (MUTCD, por sus siglas en inglés) lo cuales indican la velocidad reducida requerida (FHWA y DOT, 2000). La versión 2003 del MUTCD especifica una versión distinta de este rótulo, pero ésta no se utiliza frecuentemente en las

carreteras rurales de Pennsylvania. La Figura 1 muestra el comienzo de una zona de transición indicada por la localización del rótulo que especifica la velocidad reducida requerida en la zona de baja velocidad.



Figura 1: Ejemplo del Comienzo de una Zona de Transición.

El objetivo del estudio era identificar las características de la carretera que influyen las velocidades operacionales en la misma. Para poder minimizar la probabilidad de que el conductor estuviera influenciado por otros factores, zonas de transición con las siguientes características no fueron considerados en el estudio:

1. Presencia de una intersección controlada por semáforos o rótulos
2. Porcentaje alto de vehículos pesados (más de 10%)
3. Alto volumen vehicular los cuales minimizan la probabilidad de tomar datos de vehículos de flujo libre
4. Superficies y líneas de pavimentos en condiciones pobres

Register for free at <https://www.scipedia.com> to download the version without the watermark

Visitas al campo confirmaron si las localizaciones inicialmente identificadas con el uso del sistema de video de PennDOT eran apropiadas para la investigación. Usando el criterio de selección de lugar antes descrito, 20 lugares de estudio fueron seleccionados. Los cambios en límites de velocidad así como el largo de zona de transición para los lugares de estudio se encuentran listados en la Tabla 1, junto con otra información relevante para la identificación y localización de estas zonas.

Tabla 1: Descripción de Lugares de Estudio.

ID del Lugar	Pueblo / Comunidad	Condado	Ruta	Segmento	Reducción en Límite de Velocidad (mph)	Largo de Zona de Transición (pies)
1	Alverda	Indiana	553WB	0160-0170	55 → 35	535
2	Brush Valley	Indiana	56WB	0420	55 → 35	690
3	Corsica	Jefferson	322 WB	0020-0030	55 → 35	725
4	Cross Keys	Juniata	35 NB	0050-0060	55 → 40	540
5	Cross Keys	Juniata	35 SB	0070-0080	55 → 40	375
6	Curwensville	Clearfield	453 NB	0390-0410	45 → 25	750
7	Curwensville	Clearfield	879 EB	0100-0110	45 → 25	500
8	Ernest	Indiana	110 EB	0070-0080	55 → 35	1065
9	Fousetown	Huntingdon	655 SB	0540-0550	55 → 35	750
10	Freeport	Butler	356 NB	0110	55 → 40	690

ID del Lugar	Pueblo / Comunidad	Condado	Ruta	Segmento	Reducción en Límite de Velocidad (mph)	Largo de Zona de Transición (pies)
11	Homer City	Indiana	3035 NB	0010-0030	55 → 35	1020
12	Madisonburg	Centre	192 EB	0210-0220	55 → 40	800
13	Mifflintown	Juniata	35 NB	0510-0520	45 → 25	570
14	Orbisonia	Huntingdon	522 NB	0210	40 → 25	925
15	Osceola Mills	Clearfield	53 NB	0480-0490	45 → 25	460
16	Shirleysburg	Huntingdon	522 SB	0310-0320	55 → 35	700
17	Spruce Creek	Huntingdon	45WB	0080-0090	55 → 35	675
18	Unionville	Centre	3040 NB	0360-0370	45 → 35	665
19	Warriors Mark	Huntingdon	550SB	0110-0120	55 → 35	410
20	Zion	Centre	550 NB	0520-0540	55 → 40	860

Además de los cambios en el límite de velocidad y la longitud de la zona de transición en cada lugar de estudio, las características geométricas de la carretera se midieron en cada uno de los lugares donde se colocaron los sensores de velocidad, tales como ancho de carril, distancia lateral, y pendiente vertical, entre otros. Datos de información de otras características de la carretera también se recolectaron, tales como la presencia de vallas de seguridad, presencia de curva horizontal y el tipo y número de otros rótulos de carretera. La Tabla 2 muestra un resumen de las características de la carretera en los 20 lugares de estudio.

Tabla 2: Resumen de Características Descriptivas de las Características de las Carreteras.

Característica Medida	Promedio	Desviación Estándar	Mínimo	Máximo
Ancho de Carril (pies)	10.65	0.570	9.7	13
Ancho de Paseo Pavimentado (pies)	3.43	1.911	0	8.6
Ancho de Paseo No Pavimentado (pies)	1.27	1.807	0	12
Ancho de Carretera Pavimentada (pies)	28.33	4.493	23	41
Zona Libre de Obstáculos (pies)	8.88	5.244	0	30
Pendiente (%0)	-0.50	2.926	-9.2	5.6
Número de Accesos (ambos lados)	2.02	1.787	0	7
Característica Contada	Proporción en la Muestra (%)			
Presencia of Curva Horizontal con Rótulo de Advertencia*	20.2			
Presencia of Curva Horizontal sin Rótulo de Advertencia*	29.2			
Presencia de Encintado	17.9			
Presencia de Valla de Seguridad	21.7			
Presencia de Rótulo de Advertencia de Intersección	9.3			
Presencia de Rótulo de Advertencia de Zona Escolar	10.2			
Presencia de Rótulo de Advertencia de Curva Adelante	11.4			
*Fue teorizado que la presencia de una curva horizontal acompañada con un rótulo de advertencia era percibida como una curva más cerrada y por lo tanto influenciando a los conductores a reducir la velocidad a un grado mayor que una curva que no tuviese dicho rótulo.				

### Datos de Velocidades

La toma de velocidades fue llevada a cabo utilizando los sensores Hi-Star de la compañía Numetrics los cuales emplean tecnología de imagen magnética. Estos sensores son considerados no intrusivos, eliminando así la

posibilidad de que los conductores ajustaran sus velocidades debido al equipo visible y el personal de recolección de datos. El sensor se coloca en el centro del carril y una cubierta de goma se coloca encima del sensor, pareciendo ser parte de la superficie del pavimento; esta cubierta reduce la atención que pueda ocasionar el sensor minimizando así la probabilidad de que los conductores reduzcan la velocidad debido al equipo. El sensor es capaz de registrar los datos de velocidad, temperatura del pavimento, la condición del pavimento (seco o húmedo), la longitud del vehículo y la hora de cada observación.

En cada lugar de estudio se colocaron cuatro sensores: dos en los límites de la zona de transición, identificados por las localizaciones de los rótulos de velocidad reducida y límite de velocidad, y dos sensores fueron colocados 500 pies antes y después de los límites de la zona de transición. Estos dos sensores adicionales se colocaron con el fin de recoger datos de velocidad de los conductores que pudieran estar ajustando sus velocidades operacionales antes de la zona de transición y más allá del final de ésta. La Figura 2 muestra la localización de los cuatro sensores Hi-Star ubicados en cada lugar de estudio y su número de identificación para fines de análisis de datos.

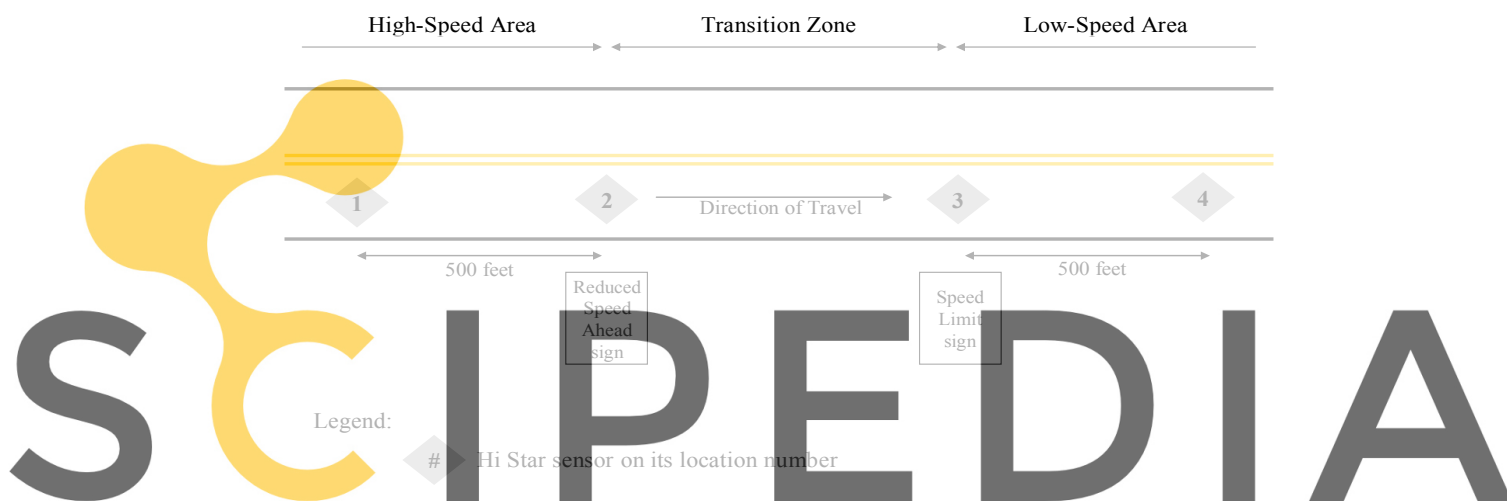


Figura 2: Localización de Sensores.

Register for free at <https://www.scipedia.com> to download the version without the watermark

Los datos de velocidad se tomaron durante el día y en condiciones meteorológicas favorables. Es común tomar datos durante condiciones de luz diurna y seco pavimento en las fases iniciales de un estudio de velocidades para investigar el comportamiento del conductor en "condiciones ideales" ya que los conductores tienden a conducir con mayor precaución durante la noche o en condiciones de pavimento húmedo. Sólo los datos de la velocidad de los vehículos de flujo libre se consideraron ya que estos vehículos sólo están influenciados por las características de las carreteras. Un vehículo de flujo libre es definido como aquel que tiene una separación entre vehículos de al menos 5 segundos (McFadden y Elefteriadou, 2000). Además, debido a que la disposición de los sensores permite que un vehículo pueda ser "rastreado", sólo se consideró los datos de aquellos vehículos cuyas velocidades fueron tomadas en los cuatro sensores. La Tabla 3 muestra el promedio de las velocidades y la desviación estándar para los cuatro sensores en cada lugar de estudio junto con el número total de vehículos por lugar y el número total de observaciones.

## METODOLOGÍA

El método de Regresión Basada en Árbol Jerárquico (RBAJ) puede ser aplicado cuando es necesaria la categorización de las observaciones y se puede utilizar para identificar patrones en conjuntos de datos cuantiosos a través de la partición recursiva de predictores siguiendo tres pasos básicos.

El modelo RBAJ es un método de selección de variables de tipo "paso-a-paso" similar a la regresión paso-a-paso (Breiman et al., 1984). El método RBAJ se basa en un algoritmo que avanza según dos preguntas son formuladas y contestadas de forma iterativa. La primera pregunta cuestiona cual variable de todas las consideradas puede producir la máxima reducción en la variabilidad de la variable dependiente. La segunda pregunta determina el valor de la variable elegida (ya sea discreta o continua) que resulta en la reducción máxima en la variabilidad de la

variable dependiente. El algoritmo termina cuando la condición final deseada se cumple y un modelo de árbol es estimado (Washington et al., 1996).

Una de las ventajas del método RBAJ es que no requiere que los datos sigan una distribución en particular, a diferencia de un análisis de regresión tradicional, ya que es un método no paramétrico. Otra ventaja es que, debido a que se producen fracturas en los valores atípicos, esto lo hace resistente a los efectos de estos valores extremos (Roberts et al., 1999). Además, puesto que RBAJ elige sólo las variables independientes más importantes y valores de dichas variables que dan como resultado la reducción máxima en desviación, una identificación preliminar de las variables independientes no es necesaria. Finalmente, los resultados obtenidos con el método RBAJ son constantes, libremente de la transformación de las variables independientes, eliminando así la necesidad de probar un sinnúmero de transformaciones con el fin de estimar el mejor modelo (Hallmark et al., 2001).

**Tabla 3: Promedio y Desviación Estándar de Velocidades.**

ID del Lugar	Tamaño de Muestra	Promedio y Desviación Estándar de Velocidades por Sensor (mph)			
		1	2	3	4
1	124	47.9 (7.24)	49.6 (7.20)	50.3 (6.07)	47.6 (7.12)
2	68	52.8 (9.50)	52.4 (7.71)	44.2 (8.47)	43.1 (8.04)
3	98	51.3 (5.44)	49.9 (5.72)	46.3 (6.17)	43.1 (6.15)
4	104	57.6 (7.97)	53.9 (7.69)	52.6 (6.67)	48.2 (6.75)
5	231	58.2 (6.78)	52.3 (7.16)	49.6 (6.79)	45.5 (6.81)
6	99	42.6 (7.18)	41.6 (6.00)	35.8 (7.37)	28.7 (6.26)
7	159	52.0 (6.11)	47.0 (5.62)	44.4 (6.91)	37.4 (6.19)
8	149	57.1 (6.20)	53.0 (7.02)	49.5 (7.50)	46.7 (6.48)
9	478	58.4 (6.40)	53.1 (5.92)	48.3 (7.02)	47.4 (7.65)
10	148	51.7 (6.08)	51.0 (5.82)	49.6 (5.90)	49.0 (6.14)
11	141	43.3 (6.88)	41.4 (6.11)	36.6 (5.30)	36.9 (5.96)
12	73	54.5 (6.20)	52.6 (5.96)	48.8 (7.09)	38.9 (9.57)
13	130	43.8 (6.89)	41.7 (5.02)	28.4 (4.08)	30.1 (4.42)
14	112	53.4 (7.03)	49.2 (6.17)	39.2 (5.81)	36.1 (5.97)
15	81	46.7 (5.98)	41.7 (5.02)	41.7 (5.42)	36.6 (5.89)
16	122	54.0 (6.90)	50.8 (5.51)	45.7 (6.20)	36.7 (5.66)
17	164	58.2 (6.25)	55.5 (6.08)	50.4 (6.04)	46.3 (6.02)
18	52	58.1 (7.54)	53.3 (7.06)	52.0 (6.31)	50.8 (6.08)
19	178	50.2 (5.75)	45.5 (5.31)	49.5 (6.44)	42.5 (6.01)
20	148	53.3 (5.56)	52.0 (6.01)	47.8 (5.58)	43.9 (6.26)
Total:	2,859 conductores	11,436 observaciones de velocidades			

El modelo estimado utilizando el método RBAJ utiliza varios algoritmos, tales como la Detección Automática de Interacción Chi-Cuadrado (CHAID, por sus siglas en inglés), CHAID Exhaustivo, Árboles de Clasificación y Regresión (CART, por sus siglas en inglés), y Rápida, Imparcial, Eficiente, Árbol de Estadística (QUEST, por sus siglas en inglés), entre otros.

El algoritmo CHAID es uno de los métodos más antiguos de clasificación del árbol originalmente propuesto por Kass (1980), y se basa en las pruebas de significancia ajustada (pruebas Bonferroni). CHAID se utilizó en este estudio debido a que es un algoritmo que permite múltiples divisiones (particiones) de un nodo, lo cual es muy adecuado para el análisis de conjuntos de datos extremos (Antipov y Pokryshevkaya, 2009). El procedimiento CHAID se basa en tres pasos:

1. Fusión: hacer referencia de fusión de categorías no significativas para cada variable independiente. Cada categoría final de la variable independiente resulta en un nodo secundario si esta variable independiente se utiliza para dividir el nodo. Al final, el valor ajustado del error (valor de p), que se va a utilizar en el paso de división, se calcula.
2. División: encontrar la mejor división para cada variable independiente. Se elige cuál variable independiente puede dividir el nodo de la manera más eficaz y la selección se realiza mediante la comparación del valor ajustado de p que está relacionado con cada variable independiente.
3. Parada: es el proceso de verificación que determina si los pasos para el desarrollo del árbol debe ser finalizado de acuerdo con las reglas de detención.

El análisis CHAID abarca una o más variables de predicción (continua, ordinal categórica o nominal categórica), una variable de destino (valores continuos, nominales u ordinales), y se pueden hacer varios ajustes a los diversos parámetros. Estos ajustes incluyen niveles de significancia utilizados en fusión y división y un criterio para detener el proceso de separación (SPSS, 1999). En esta investigación, los modelos RBAJ se estimaron utilizando el programa SPSS (versión 17.0).

### MODELO ESTIMADO Y DISCUSIÓN DE RESULTADOS

El modelo de Regresión Basada en Árbol Jerárquico (RBAJ) se utilizó para predecir las velocidades operaciones a lo largo zonas de transición de carreteras rurales de dos carriles usando los datos de velocidades recopilados por los cuatro sensores Hi-Star. El modelo que se muestra en la Figura 3 ilustra la estructura final del árbol para velocidades operacionales que implican cinco variables de división: límite de velocidad 25 mph, límite de velocidad 55 mph, ancho de paseo pavimentado, ancho de carretera pavimentada y ancho de carril. La Tabla 4 incluye las estadísticas descriptivas de cada nodo del árbol.

La primera división óptima para velocidades de operación a lo largo de zonas de transición de carreteras rurales de dos carriles (nodo 0) es función de la Velocidad Límite Postulada 25 mph, que clasifica a la velocidad de operación en dos grupos: si el límite de velocidad no es de 25 mph (nodo 1), el árbol predice una velocidad operacional promedio de 49.301 mph, y si el límite de velocidad es de 25 mph (nodo 2), la velocidad operacional promedio es 35.935 mph.

En el segundo nivel del árbol, Límite de Velocidad 55 mph y Ancho de Carretera Pavimentada conduce a dos divisiones adicionales. Límite de Velocidad 55 mph divide los datos en dos subgrupos: si el límite de velocidad no es de 55 mph y si el límite de velocidad es de 55 mph. La velocidad de funcionamiento más alta (52.780 mph) se produce cuando el límite de velocidad es 55 mph. Ancho de Carretera Pavimentada divide los datos en seis subgrupos. La velocidad operacional más alta (44.384 mph) se produce cuando las carretera pavimentadas es mayor de 31.7 pies y menor de 35 pies (nodo 9).

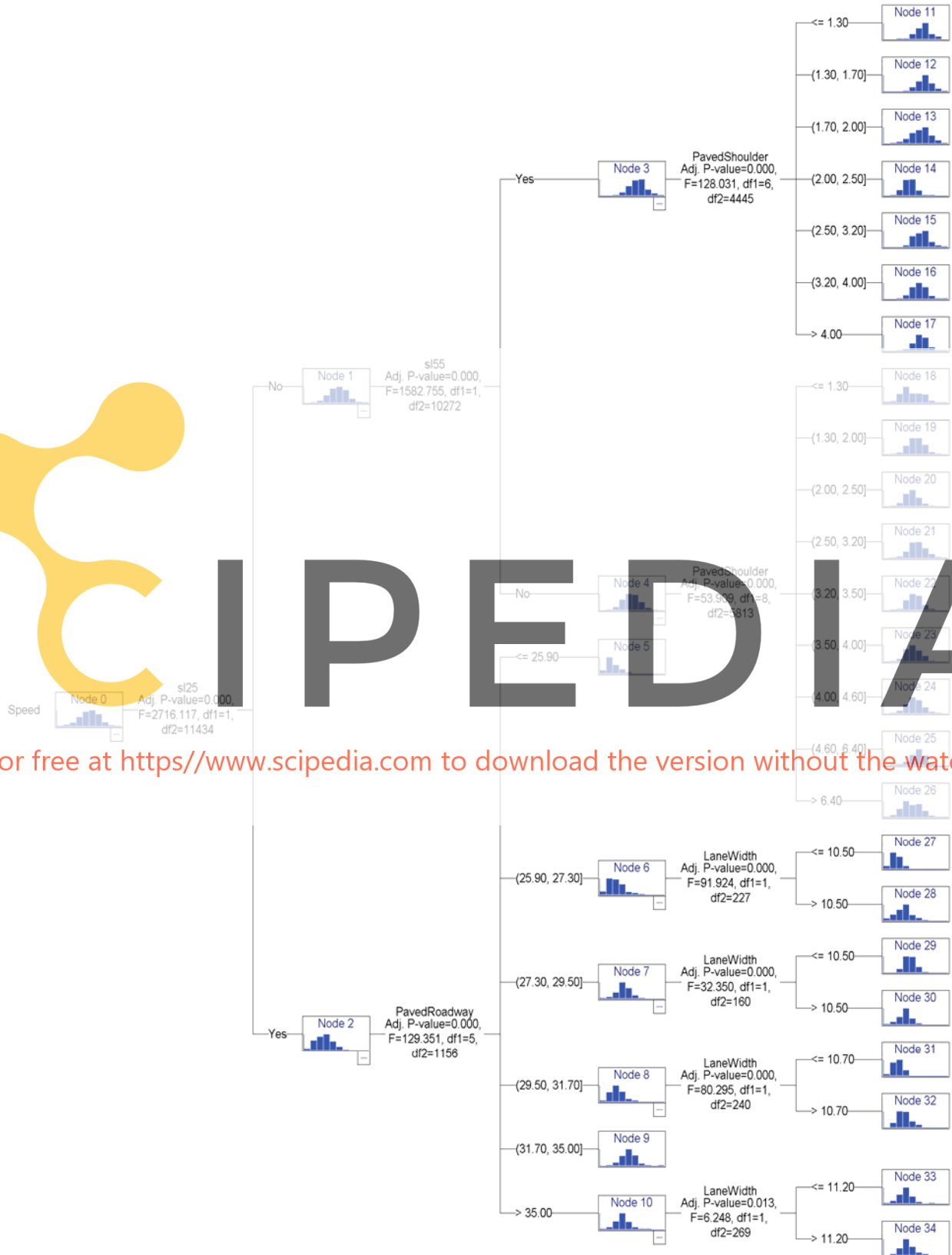
En el tercer nivel del árbol, Ancho de Paseo Pavimentado y Ancho de Carril conducen a las divisiones finales. Ancho de Paseo Pavimentado conduce a la división en los nodos 3 y 4, partiendo los datos en 7 subgrupos cuando la velocidad límite es 55mph: ancho de paseo pavimentado menor o igual a 1.30 pies, entre 1.30 y 1.70 pies, entre 1.7 y 2.0 pies, entre 2.0 y 2.5 pies, entre 2.5 y 3.20 pies, entre 3.20 y 4.0 pies, y mayor de 4 pies. Según el modelo, la velocidad operacional mas alta (56.203 mph) ocurre cuando el ancho del paseo pavimentado se encuentra en el rango entre 1.30 y 1.70 pies. La variable de Ancho de Carril conduce a la división de los nodos 6, 7, 8 y 10. Para carreteras entre 27.3 y 29.5 pies (Nodo 7), Ancho de Carril divide la cada en dos grupos: ancho de carril menor o igual a 10.5 pies y mayor de 10.5 pies. El árbol indica que, para este grupo, la velocidad operacional mayor (41.691 mph) ocurre cuando el ancho de carril es igual o menor a 0.5 pies. Divisiones similares para Ancho de Carril ocurren en otros grupos de acuerdo al ancho de la carretera pavimentada.

Register for free at <https://www.scipedia.com> to download the version without the watermark





Register for free at <https://www.scipedia.com> to download the version without the watermark



**Figura 3: Modelo RBAJ de Predicción de Velocidades Operacionales en Zonas de Transición en Carreteras Rurales de Dos Carriles.**

**Tabla 4: Modelo RBAJ – Variables Descriptivas y Resultados por Nodo.**

Node	Mean	Std. Deviation	N	Percent	Predicted Mean	Parent Node	Primary Independent Variable					
							Variable	Sig.(a)	F	df1	df2	Split Values
0	47.94	9.218	11,436	100.0%	47.94							
1	49.30	8.327	10,274	89.8%	49.30	0	sl25	0.000	2,716.117	1	11,434	No
2	35.93	7.922	1,162	10.2%	35.93	0	sl25	0.000	2,716.117	1	11,434	Yes
3	52.78	7.717	4,452	38.9%	52.78	1	sl55	0.000	1,582.755	1	10,272	Yes
4	46.64	7.778	5,822	50.9%	46.64	1	sl55	0.000	1,582.755	1	10,272	No
5	28.71	6.257	99	0.9%	28.71	2	PavedRoadway	0.000	129.351	5	1,156	<=
6	31.59	6.783	229	2.0%	31.59	2	PavedRoadway	0.000	129.351	5	1,156	(25.89999999999999999999, 27.30000000000000000000]
7	39.16	6.190	162	1.4%	39.16	2	PavedRoadway	0.000	129.351	5	1,156	(27.30000000000000000000, 29.5]
8	32.84	5.988	242	2.1%	32.84	2	PavedRoadway	0.000	129.351	5	1,156	(29.5, 31.69999999999999999999]
9	44.38	6.910	159	1.4%	44.38	2	PavedRoadway	0.000	129.351	5	1,156	(31.69999999999999999999, 35]
10	38.12	6.097	271	2.4%	38.12	2	PavedRoadway	0.000	129.351	5	1,156	> 35
11	54.86	6.780	268	2.3%	54.86	3	PavedShoulder	0.000	128.031	6	4,445	<= 1.3
12	56.20	6.794	1,187	10.4%	56.20	3	PavedShoulder	0.000	128.031	6	4,445	(1.3, 1.7]
13	52.68	8.967	640	5.6%	52.68	3	PavedShoulder	0.000	128.031	6	4,445	(1.7, 2]
14	41.39	6.112	141	1.2%	41.39	3	PavedShoulder	0.000	128.031	6	4,445	(2, 2.5]
15	53.01	7.024	149	1.3%	53.01	3	PavedShoulder	0.000	128.031	6	4,445	(2.5, 3.20000000000000000000]
16	50.57	7.037	1,095	9.6%	50.57	3	PavedShoulder	0.000	128.031	6	4,445	(3.20000000000000000000, 4]
17	52.19	6.488	972	8.5%	52.19	3	PavedShoulder	0.000	128.031	6	4,445	> 4
18	43.58	9.988	245	2.1%	43.58	4	PavedShoulder	0.000	53.909	8	5,813	<= 1.3
19	47.80	7.159	1,306	11.4%	47.80	4	PavedShoulder	0.000	53.909	8	5,813	(1.3, 2]
20	43.70	6.465	382	3.3%	43.70	4	PavedShoulder	0.000	53.909	8	5,813	(2, 2.5]
21	49.04	7.786	785	6.9%	49.04	4	PavedShoulder	0.000	53.909	8	5,813	(2.5, 3.20000000000000000000]
22	45.21	7.098	444	3.9%	45.21	4	PavedShoulder	0.000	53.909	8	5,813	(3.20000000000000000000, 3.5]
23	45.21	7.098	444	3.9%	45.21	4	PavedShoulder	0.000	53.909	8	5,813	(3.5, 4]
24	46.29	7.357	1,038	9.1%	46.29	4	PavedShoulder	0.000	53.909	8	5,813	(4, 4.59999999999999999999]
25	49.70	6.271	536	4.7%	49.70	4	PavedShoulder	0.000	53.909	8	5,813	(4.59999999999999999999, 6.40000000000000000000]
26	43.45	9.306	623	5.4%	43.45	4	PavedShoulder	0.000	53.909	8	5,813	> 6.40000000000000000000
27	28.42	4.077	130	1.1%	28.42	6	LaneWidth	0.000	91.924	1	227	<= 10.5
28	35.76	7.369	99	0.9%	35.76	6	LaneWidth	0.000	91.924	1	227	> 10.5
29	41.69	5.424	81	0.7%	41.69	7	LaneWidth	0.000	32.350	1	160	<= 10.5
30	36.63	5.894	81	0.7%	36.63	7	LaneWidth	0.000	32.350	1	160	> 10.5
31	30.06	4.419	130	1.1%	30.06	8	LaneWidth	0.000	80.295	1	240	<=
32	36.06	5.970	112	1.0%	36.06	8	LaneWidth	0.000	80.295	1	240	>
33	37.35	6.194	159	1.4%	37.35	10	LaneWidth	0.013	6.248	1	269	<=
34	39.21	5.810	112	1.0%	39.21	10	LaneWidth	0.013	6.248	1	269	>

Growing Method: CHAID

a. Bonferroni adjusted

La estimación de riesgo es un indicador del rendimiento del modelo. La Tabla 5 muestra los estimados de riesgo, desviación estándar, varianza, la proporción de la varianza debido al error (varianza inexplicada) y la proporción de la varianza explicada por el modelo desarrollado.

**Tabla 5: Proporción de Varianza Explicada por el Modelo.**

	<b>Estimado de Riesgo</b>	<b>Desviación Estándar</b>	<b>Varianza</b>	<b>Proporción de Varianza debido al Error (Varianza Inexplicada)</b>	<b>Proporción de Varianza Explicada por el Modelo</b>
<b>Modelo con Todos los Sensores</b>	51.946	9.218	84.972	0.611	38.87%

Además, otros cuatro árboles, correspondientes a cada ubicación del sensor, fueron desarrollados. Los cuatro modelos RBAJ se construyeron para predecir respectivamente: las velocidades de operación en la zona de alta velocidad (500 pies antes del rótulo de Reducción de Velocidad Adelante) correspondientes al sensor 1; las velocidades de operación en los rótulos de Reducción de Velocidad Adelante (comienzo de la zona de transición) correspondiente al sensor 2; velocidades operacionales en el rótulo de Velocidad Máxima Postulada (final de la zona de transición) correspondiente al sensor 3 y; velocidades de operación en la zona de baja velocidad (500 pies después del rótulo de Velocidad Máxima Postulada) correspondientes al sensor 4.

Para propósitos de la entrada de datos, los cuatro modelos tienen la misma variable dependiente y diferentes variables independientes, que representan los datos característicos de las carreteras y los factores de control de tráfico. Ha de hacerse notar que, considerando los árboles desarrollados por separado, esto resulta en la pérdida de información (pérdida de variables y observaciones). Los resultados se muestran en la Tabla 6.

**Tabla 6: Resultados de Otros Modelos RBAJ.**

		<b>Sensor 1</b>	<b>Sensor 2</b>	<b>Sensor 3</b>	<b>Sensor 4</b>
<b>División</b>	<b># 1: Óptimo</b>	Ancho de Paseo No Pavimentado	Numero Total de Entradas	Pendiente	Ancho de Carril
	<b># 2: Segundo Nivel</b>	Pendiente	Pendiente	Velocidad Limite	Ancho de Paseo Pavimentado
		Ancho de Carril	Ancho de Carril	Ancho De Carretera Pavimentada	Velocidad Limite
				Ancho de Paseo No Pavimentado	Zona de Espacio Libre Lateral
<b># 3: Tercer Nivel</b>	Ancho de Paseo Pavimentado				
		Velocidad Limite			
<b>Estadísticas</b>	<b>Estimado de Riesgo</b>	43.207	38.458	41.470	43.367
	<b>Desviación Estándar</b>	8.321	7.537	8.635	8.754
	<b>Varianza</b>	69.239	56.806	74.563	76.633
	<b>Proporción de Varianza debido al Error (Varianza Inexplicada)</b>	0.624	0.677	0.556	0.566
	<b>Proporción de Varianza Explicada por el Modelo</b>	37.60%	32.30%	44.38%	43.41%

El valor estimado del riesgo es la varianza en el nodo. La proporción de la varianza explicada por el modelo es mayor en el modelo 3 (44.38%), el cual predice la velocidad operacional en el sensor 3, el cual corresponde al final de la zona de transición.

Los resultados de los cuatro árboles independientemente estimados indican que el mejor predictor de las velocidades operacionales en el área de alta velocidad (sensor 1) es el ancho del paseo no pavimentado, seguido por pendiente, ancho de carril, ancho de paseo pavimentados y límite de velocidad postulada. Al comienzo de la zona de transición (sensor 2), el número de accesos es la variable principal para predicciones de velocidades, seguido por la pendiente y ancho de carril. Al final de la zona de transición (sensor 3) la pendiente es la variable más influyente en las velocidades de operación, seguido por límite de velocidad postulada, ancho de carretera pavimentada, y ancho de paseo pavimentado. Finalmente, en la zona de baja velocidad (sensor 4), las variables identificadas como influyentes en las velocidades de operación fueron ancho de carril, ancho de paseo pavimentado, límite de velocidad postulada, y ancho de la zona lateral libre de obstáculos.

## CONCLUSIONES Y RECOMENDACIONES

En esta investigación el método de Regresión Basada en Árbol Jerárquico (RBAJ) se utilizó para predecir las velocidades de operación a lo largo de zonas de transición en carreteras rurales de dos carriles empleando datos de 20 lugares de estudio en Pensilvania. Los resultados indican que las velocidades operacionales dependen de cinco factores: límite de velocidad 25 mph, límite de velocidad de 55 mph, ancho de paseo pavimentado, ancho de carretera pavimentada y ancho de carril; la variable de Límite de Velocidad 25 mph fue el mejor predictor.

Además, el método RBAJ se utilizó para predecir velocidades de funcionamiento en cuatro lugares a lo largo de las zonas de transición: área de alta velocidad, inicio y final de la zona de transición, y la zona de baja velocidad. Los resultados muestran que los factores que son identificados como significativos en modelos de predicción de velocidades varían dependiendo de la zona donde los sensores estaban colocados.

La velocidad operacional en las zonas de alta velocidad (500 pies antes del rótulo de Reducción de Velocidad Adelante) depende de cinco factores: ancho de paseo no pavimentado estabilizado, pendiente, ancho de carril, ancho de paseo pavimentado y velocidad límite postulada. La velocidad operacional al comienzo de la zona de transición depende de tres variables: número total de accesos, pendiente y ancho de carril. La velocidad operacional al final de la zona de transición depende de cuatro factores: pendiente, límite de velocidad postulada, ancho de carretera pavimentada y el ancho de paseo no pavimentado estabilizado. Las variables que influyen en la velocidad operacional en la zona de baja velocidad son: ancho de carril, ancho de paseo pavimentado, límite de velocidad postulada y la distancia lateral libre de obstáculos.

El ancho de paseo no pavimentado resultó ser el predictor más importante de velocidades operacionales en las zonas de alta velocidad áreas situadas a 500 pies antes del rótulo de Reducción de Velocidad Adelante. En la zona de transición, (entre los sensores 2 y 3) los cambios en velocidades están relacionados con el número total de accesos. En las zonas de baja velocidad, el ancho de carril resultó el mejor predictor de velocidades en dicha área.

Los datos en este estudio son limitados debido a que no hay suficientes lugares con suficientes medidas de la misma variable en todas las localizaciones de los sensores. Por lo tanto, al desarrollar modelos de árboles por separados (de acuerdo a la localización del sensor) los resultados de los modelos no fueron constantes. Se recomienda el tomar datos adicionales y aplicar el método RBAJ con el fin de confirmar los resultados mostrados en este estudio.

## REFERENCIAS

American Association of State Highway and Transportation Officials – AASHTO (2004). *A Policy on Geometric Design of Highways and Streets*, Washington, D. C.

Antipov, E. y Pokryshevkaya, E. (2009). “Applying CHAID for Logistic Regression Diagnostics and Classification Accuracy Improvement”, MPRA Paper No. 21499.

Breiman, L., Friedman, J., Olshen, R. F. y Stone, C. (1984). *Classification and Regression Trees*, Wadsworth International Group, Belmont, California.

- Chang, L. y Wang, H. (2006). "Analysis of Traffic Injury Severity: An Application of Non-Parametric Classification Tree Techniques", *Accident Analysis and Prevention*, Vol. 38, No. 5, pp. 1019-1027.
- Cruzado, I. y Donnell, E. T. (2010). "Factors Affecting Driver Speed Choice Along Two-Lane Rural Highway Transition Zones", *Journal of Transportation Engineering (American Society of Civil Engineers – ASCE)*, Vol. 136, No. 8, 2010, pp. 755-764.
- Federal Highway Administration – FHWA (2004). "Highway Statistics 2004", Disponible en: <http://www.fhwa.dot.gov/policy/ohim/hs04/>, Accedido en: 4 de julio de 2011.
- Federal Highway Administration – FHWA y The United States Department of Transportation - DOT (2000). *Manual on Uniform Traffic Control Devices (MUTCD)*.
- Fitzpatrick, K., Elefteriadou, L., Harwood, D. W., Collins, J. M., McFadden, J., Anderson, I. B., Krammes, R. A., Irizarry, N., Parma, K. D., Bauer, K. M. y Passetti, K. (2000). "Speed Prediction for Two-Lane Rural Highways", Report FHWA-RD-99-171, Federal Highway Administration, McLean, Virginia.
- Fomunung, I., Washington, S., y Guensler, R. (1999). "A statistical model for estimating oxides of nitrogen emissions from light duty motor vehicles", *Transportation Research Part D: Transport and Environment*, Vol. 4, No. 5, pp. 333-352.
- Glennon, J. C., Neuman, T. R. y Leisch, J. E. (1985). "Safety and Operational Considerations for Design of Rural Highway Curves", Report FHWA/RD-86/035, Federal Highway Administration, McLean, Virginia.
- Hallmark, S., Guensler, R. y Fomunung, I. (2001). "Characterizing On-Road Variables that Affect Passenger Vehicle Modal Operation", *Transportation Research Part D: Transport and Environment*, Vol. 7, No. 2, pp. 81-98.
- Kass, G. V. (1980). "An Exploratory Technique for Investigating Large Quantities of Categorical Data", *Journal of Applied Statistics*, Vol. 29, No. 2, pp. 119-127.
- Krammes, R., Brackett, R. O., Shafer, M., Ottesen, J., Anderson, I., Fink, K., Collins, K., Pendleton, O. y Messer, C. (1995). "Horizontal Alignment Consistency for Rural Two-Lane Highways", Report FHWA-RD-94-034, Federal Highway Administration, Washington D.C.
- Lamm, R. y Choueiri, E. M. (1987). "Recommendations for Evaluating Horizontal Design Consistency Based on Investigations in the State of New York", *Transportation Research Record*, No. 1122, pp. 68-78.
- McFadden, J. y Elefteriadou, L. (2000). "Evaluating Horizontal Alignment Design Consistency of Two-Lane Rural Highways", *Transportation Research Record: Journal of the Transportation Research Board*, No. 1737, pp. 9-17.
- Misaghi, P. y Hassan, Y. (2005). "Modeling Operating Speed and Speed Differential on Two-Lane Rural Roads", *Journal of Transportation Engineering*, Vol. 131, No. 6, pp. 408-418.
- Park, Y-J. y Saccomanno, F. F. (2006). "Evaluating Speed Consistency Between Successive Elements of a Two-Lane Rural Highway", *Transportation Research Part A: Policy and Practice*, Vol. 40, No. 5, pp. 375-385.
- Passetti, K. A. y Pambro, D. B. (1999). "Operating Speeds on Curves with and without Spiral Transitions", *Transportation Research Record: Journal of the Transportation Research Board*, No. 1658, pp. 9-16.
- Poe, C. M. y Mason, J. M. (2000). "Analyzing Influence of Geometric Design on Operating Speeds Along Low-Speed Urban Streets", *Transportation Research Record*, No. 1737, pp.18-25.
- Roberts, C. A., Washington, S. y Leonard, J. (1999). "Forecasting Dynamic Vehicular Activity on Freeways: Bridging the Gap Between Travel Demand and Emerging Emissions Models". *Transportation Research Record*, No. 1664, pp. 31-39.

- Stamatiadis, N., Pigman, J. G. y Hartman, D. (2004). "Safety Consequences of Flexibility in Highway Design for Rural Communities", Final Report for National Cooperative Highway Research Program Project, pp.15-22.
- Stamatiadis, N., Pigman, J. G. y Hartman, D. (2006). "Safety Consequences from Design Flexibility in Rural to Urban Transitions", *Transportation Research Board 85<sup>th</sup> Annual Meeting*, p. 18.
- Statistical Product and Service Solutions – SPSS (1999). "AnswerTree Algorithm Summary", SPSS White Paper.
- Tarris, J., Poe, C., Mason, J. M. y Goulias, K. (1996). "Predicting Operating Speeds on Low-Speed Urban Streets: Regression and Panel Analysis Approaches", *Transportation Research Record*, No. 1523, pp. 46-54.
- Wang, J., Dixon, K. K., Li, H. y Hunter, M. (2006). "Operating-Speed Model for Low-Speed Urban Tangent Streets Based on In-Vehicle Global Positioning System Data", *Transportation Research Record*, No. 1961, pp. 24-33.
- Washington, S., Wolf, J. y Guensler, R. (1996). "A Binary Recursive Partitioning Method for Modeling Hot-Stabilized Emissions from Motor Vehicles", *Transportation Research Record*, No. 1587, pp. 96-105.
- Xuedong, Y., Richards, S., y Xiaogang, S. (2010). "Using Hierarchical Tree-Based Regression Model to Predict Train-Vehicle Crashes at Passive Highway-Rail Grade Crossings", *Accident Analysis and Prevention*, Vol. 4, No. 1, pp. 64-74.