

Optimización de factores en la estimación de velocidad operacional en curvas horizontales

Optimization of factors for estimation of operational speed in horizontal curves

Victor Renan Castillo-Ñañez ^{1a}, Carlos Aníbal Calero-Valenzuela ^{1b}, Carlos Andrés Gaviria-Mendoza ²

¹ Grupo de investigación en Ingeniería de Tránsito, Facultad de Ingeniería Civil, Universidad del Cauca, Colombia. Orcid: [0000-0002-7934-189X](https://orcid.org/0000-0002-7934-189X), [0000-0003-2920-4130](https://orcid.org/0000-0003-2920-4130). Correos electrónicos: renancastillo@unicauca.edu.co, ccalero@unicauca.edu.co

² Universidad Militar Nueva Granda, Colombia. Orcid: [0000-0002-3741-43253](https://orcid.org/0000-0002-3741-43253). Correo electrónico: carlos.gaviria@unimilitar.edu.co

Recibido: 9 septiembre, 2024. Aceptado: 22 mayo, 2025. Versión final: 20 junio, 2025

Resumen

La consistencia de diseño geométrico permite valorar el trazado de una carretera a partir de las expectativas de los conductores. En la literatura se identificaron metodologías (Pollus y Mattar-Habib, 2004; Castro et al, 2006; Echaveguren, 2014) que definen el nivel de la consistencia a partir de la velocidad operacional (V85), para la estimación de ésta se utilizan modelos lineales en los que se incluye los elementos de la geometría y los parámetros operacionales como variables independientes; la hipótesis de este estudio es que los modelos no lineales generan una estimación adecuada de la V85, ya que tienen en cuenta la dependencia de los parámetros operacionales a través de la curva. El objetivo de este trabajo es comparar la estimación de la V85 en vías rurales de dos carriles utilizando el modelo de Datos en Panel y comparándolo con un modelo lineal. La muestra incluye 22,020 vehículos a flujo libre medidos en cinco puntos significativos de la curva en vías rurales de dos carriles en Puerto Rico. Para el análisis se incluyeron 23 variables cuantitativas y 46 variables categóricas, seleccionando las significativas con el método de Best subsets e integrándolas al modelo de Datos en Panel. Al compararlos, se observa similitud en los valores de R², dando como resultado que el modelo de Datos en Panel es adecuado para la estimación de la V85 para un tramo de vía tomando 5 puntos por segmento vial, teniendo estadísticos de error con valores menores a los obtenidos por los otros modelos.

Palabras clave: Datos en Panel; Filtro de Kalman; Best Subsets; Velocidad operacional; Diseño geométrico; Seguridad vial.

Abstract

The consistency of geometric design allows for evaluating a road layout based on drivers' expectations. In the literature, methodologies (Pollus and Mattar-Habib, 2004; Castro et al., 2006; Echaveguren, 2014) have been identified that define the level of consistency based on operational speed (V85). For estimating this, linear models are used, which include geometric elements and operational parameters as independent variables. The hypothesis of this study is that nonlinear models provide an adequate estimation of V85, as they account for the dependency of operational parameters through the curve. The objective of this work is to compare the estimation of V85 on rural two-lane roads using the Panel Data model and comparing it with a linear model. The sample includes 22,020 free-flow vehicles measured at

five significant points of the curve on rural two-lane roads in Puerto Rico. For the analysis, 23 quantitative variables and 46 categorical variables were included, with significant ones being selected using the Best Subsets method and integrated into the Panel Data model. When compared, similarity in R^2 values is observed, resulting in the conclusion that the Panel Data model is suitable for estimating V85 for a road section, taking five points per road segment, with error statistics having lower values than those obtained by the other models.

Keywords: Panel Data; Kalman Filter; Best Subsets; operational speed; geometric design consistency; road safety.

1. Introducción

La seguridad vial es un tema de importancia a nivel mundial; cada año mueren cerca de 1.19 millones de personas en todo el mundo en accidentes de tránsito, involucrando a personas que están en edades productivas [1]; por lo tanto, es necesario diseñar y construir carreteras seguras, de manera que los usuarios transiten de forma cómoda e inequívoca.

La consistencia de diseño de una carretera es un indicador de qué tan seguro es transitar por una vía, al estar estrechamente relacionada con la seguridad vial; [2] [3] [4]; por lo tanto, si una vía es segura, tiene consistencia buena y al disminuir la seguridad esta disminuye; la relación con la seguridad de una vía es directamente proporcional. La Figura 1 detalla las consideraciones que inciden en este aspecto de diseño geométrico.

1.1. Problema de investigación

Dado que la mortalidad en las vías es un problema que se sigue presentando, se requiere obtener modelos que pronostiquen de manera certera la velocidad de operación (V_{85}), dado que es la variable dependiente en la consistencia de diseño, y esta es la que controla la calidad del trazado, un trazado inseguro o de consistencia pobre aumenta las probabilidades de accidentes de tránsito. En virtud de que varios modelos planteados se han determinado con regresiones lineales, se proyecta utilizar herramientas estadísticas no lineales para la obtención de ecuaciones precisas, de tal manera que las metodologías para la evaluación de la consistencia del diseño se apliquen con confianza y se obtengan resultados ajustados a la realidad, como identificar elementos o segmentos de la infraestructura vial que tienen una alta probabilidad de generar riesgo de accidente en los usuarios y así tomar las respectivas medidas de corrección, en carreteras existentes o futuras.

1.2. Justificación

La consistencia de diseño geométrico se ha estudiado desde los años 70; autores como Leisch, J. E. y Leisch J.P. (1977), trabajaron en un procedimiento para determinar la consistencia del trazado teniendo en cuenta

la velocidad de operación en vías de Estados Unidos; Lamm, R. (1988), estudiaron la relación entre los accidentes, la seguridad vial, la velocidad de diseño y la velocidad de operación, en 1995 trabajaron la relación entre la consistencia de diseño y el coeficiente de fricción de demanda y disponible, basado en un estudio previo realizado por Lamm y Choueiri en 1991 donde se relacionaba la velocidad de operación con los coeficientes mencionados.

Polus y Mattar Habib (2004), trabajaron en el desarrollo de un modelo de consistencia basado en el perfil de velocidades, el cual determina la variación de la velocidad a través del recorrido de una vía. Considerando la consistencia de manera global y no individual, Garach et al (2014), utiliza la metodología propuesta por Polus y Mattar Habib (2004) y evalúa varios tramos de carreteras en Granada, España identificando que dicha metodología tiende a ser conservadora, calificando tramos como pobres a pesar de que tienen el parámetro Ra aceptable.

Sin embargo, a pesar de que las vías se hacen bajo la normatividad vigente, al año 2007, se registran 12 víctimas fatales por cada cien mil habitantes en el mundo [6]. Esto infiere, que es necesario establecer metodologías que evalúen la consistencia de diseño, dando herramientas adicionales para brindar vías más seguras a los usuarios.

2. Metodología

La base de datos se creó en en dos etapas, primero se obtuvo la información geométrica y posteriormente se recopiló los datos de la variable dependiente. En la revisión bibliográfica se establecieron los modelos empleados para determinar la velocidad de operación, arrojando las variables más utilizadas en la literatura existente. El siguiente punto de análisis de normalidad de datos se comprueba que los datos utilizados tengan distribución normal, por lo cual, son factibles de desarrollar con ellos modelos para la estimación de la velocidad de operación, como Datos en Panel, cuya premisa es que los datos tengan dicha distribución para que las estimaciones se ajusten a la realidad.

En la selección del valor representativo, se determina un valor por punto de la curva el cual constituye a todos los

registros. Luego se realiza una selección de las variables más importantes determinadas en la literatura. Posteriormente se realiza la selección de variables con un método llamado Best Subsets o los mejores subgrupos, lo cual complementa las variables determinadas por la literatura, luego se determina el modelo de Datos en Panel con ayuda del software Stata V14; con el modelo encontrado, se realiza una depuración de variables utilizando el principio de parsimonia.

Una vez depurado el modelo, se comparó con un modelo lineal creado a partir de las variables encontradas y se analiza el comportamiento de ambos, para realizar dicho analisis, fue necesario utilizar el 30% de los datos para validación; después se pasa el modelo de Datos en Panel por el Filtro de Kalman, analizando la conducta del nuevo modelo y comparándolo con los modelos anteriores, para esta actividad fue necesario utilizar el software Matlab, el cual contiene dentro de sus herramientas este método estadístico.

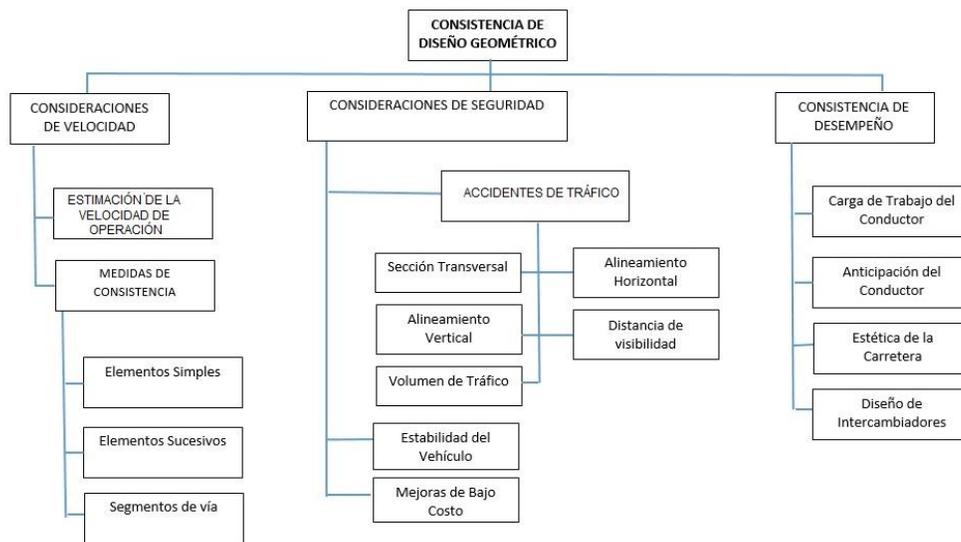


Figura 1. Principales clasificaciones de la consistencia de diseño geométrico. Fuente: Modificada de State of the art of highway geometric design consistency [5].

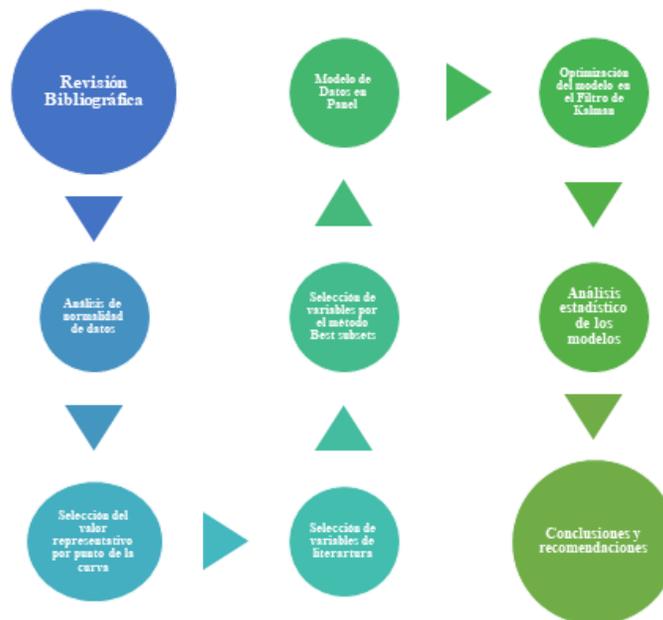


Figura 2. Metodología de la investigación. Fuente: elaboración propia.

Por último, se hace un análisis de los modelos creados a fin de determinar la superioridad o no, de cualquiera de los modelos. En la Figura 2 se muestra el esquema de la metodología utilizada para esta investigación.

3. Datos a analizar

Los datos de velocidad a flujo libre se tomaron en 5 puntos de las curvas horizontales, en la mitad de la tangente de entrada (1/2TE), en el PC, MC, PT y en la mitad de la tangente de salida (1/2TS), las curvas pertenecen a 10 tramos viales de Puerto Rico y la longitud total es de 20.969 Km. Se identificaron dichos segmentos en vías rurales utilizando criterios como TPD (500-21,000), ubicación rural, límite de velocidad (≥ 35 MPH; 56.3 km/h aprox.), tipo de terreno (plano, ondulado, montañoso), condiciones del pavimento, ausencia de mediana y dispositivos de control de velocidad a más de 0.5 millas de distancia. Se recopilaban características de la sección transversal, alineamiento horizontal y vertical, y datos generales del segmento.

Se utilizaron equipos como odómetros, pistolas láser y niveles electrónicos para medir dimensiones como ancho de carril, distancia de visibilidad, pendiente, radio de curva, entre otros; además, se instalaron equipos de clasificación de vehículos en los cinco puntos de las curvas horizontales, estos dispositivos registraron velocidad, tipo de vehículo, espaciamiento y dirección. Una vez registradas las velocidades, se aplicó la técnica de Boxplot para eliminar observaciones con cambios de velocidad abruptos o velocidades inusuales (muy altas o bajas). Esto garantizó que los datos representaran velocidades a flujo libre. Finalmente, se desarrolló un algoritmo en Matlab para unir las variables geométricas y de velocidad en una base de datos final, asegurando que cada observación estuviera correctamente asociada a su punto de medición.

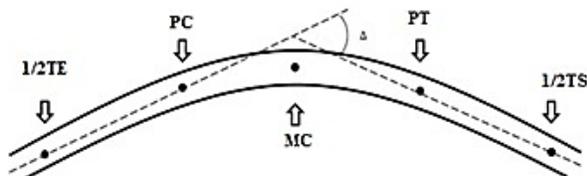


Figura 3. Puntos de la curva. Fuente: modificada de [7].

Como se mencionó, estos datos de velocidad de cada punto van acompañados de información geométrica y características propias de la vía, para un total de 68 variables adicionales, 23 cuantitativas y 45 variables categóricas.

Finalmente, se agruparon los datos en una base obteniendo la velocidad de 20020 sujetos (veh)

observados a flujo libre, en los puntos especificados, tangente de entrada (TE), (PC), (MC), (PT), y tangente de salida (TS) como se muestra en la figura 3; además, se presenta en el Anexo 1 un cuadro detallado de los estadísticos descriptivos de las principales variables numéricas.

Teniendo esa base de datos, se procede a verificar la normalidad de los datos, de tal manera que se tenga certeza de la procedencia de los mismos, así, proseguir con el análisis en el modelo deseado, ya que esta es una de las premisas del modelo de Datos en Panel, los datos utilizados deben tener distribución normal; para este análisis se creó un algoritmo en el software Matlab el cual arrojó resultados positivos al analizar los datos por el test de normalidad de Kolmogorov-Smirnov.

Teniendo en cuenta que el registro de las velocidades se hizo en puntos específicos de la curva horizontal, fue necesario calcular el percentil 85 en cada punto de la curva (velocidad de operación V85), así se obtiene un valor representativo de cada punto, acompañado de valores constantes de elementos geométricos de la curva. Para obtener la velocidad de operación (V85) en los 5 puntos de las curvas de todos los tramos, fue necesario desarrollar un algoritmo que permita hacer esta tarea de manera automática, este se realizó en el software Matlab.

4. Conclusiones

4.1. Variables seleccionadas

Con el fin de obtener un modelo que estime la velocidad de operación de una vía, este estudio muestra las variables que más han sido utilizadas en la literatura.

En la revisión de literatura se identificaron modelos utilizados para la estimación de la velocidad de operación (V85) (Lamm, R. y Choueiri, E. (1987), Lamm, R., et al. (1990), Kanellaidis, G., et al. (1990), Morrall, J. y Talarico, R. (1994), Islam, M. y Seneviratne, P. (1994), Krammes, R., et al. (1995), Voigt, A. (1996), McFadden, J. y Elefteriadou, L. (1997), Cardoso, J. et al. (1998), Ottesen, J. y Krammes, R. (2000), McFadden, J. y Elefteriadou, L. (2000), Gibreel, G., et al. (2001), Polus, A., Fitzpatrick, K., and Fambro, D. (2000), Misaghi, P. y s.

Hassan, Y. (2005), Castro, M. et al. (2006), R. D. K. Shallama, M. Ali Ahmedb. (2016), Anitha Jacob, Dhanya R, Anjaneyulu, M.V.L.R (2013), Fitzpatrick, K., et al. (1999).

En la Figura 4 las variables más utilizadas para la estimación de la velocidad de operación, entre estas la

mayor es el radio (42%), seguido de la longitud de la curva (16%), la deflexión (17%), la longitud de la curva (16%), el grado de curvatura (14%); en menor proporción se presentan las siguientes variables: el percentil 85 de la velocidad de operación en la tangente de entrada (V_t) (5%), el peralte (3%) y la tasa de cambio de la curvatura (CCR) (3%).

Según M. Castro, J. F. Sánchez, J. A. Sánchez, y L. Iglesias [8] los modelos para determinar la velocidad de operación deben tener variables que no dependan entre sí, como es el caso del radio con la longitud de curva y la deflexión con el grado de curvatura; teniendo esto en cuenta, es necesario realizar un análisis con las variables que permitan obtener las más importantes.

Para determinar otras variables se utilizó Best Subsets o los Mejores Subgrupos, la cual se soporta en el Software Minitab 17, ésta permite determinar una combinación de variables para un modelo que mejor se ajuste [9], o sea, bajo la variable dependiente, se hace una serie de regresiones lineales simultaneas con las combinaciones de variables disponibles, la que mejor tengan el valor del coeficiente Cp-Mallows, clasifican como posibles variables a entrar en el modelo de Datos en Panel.

De esta manera las variables más representativas que inciden en el comportamiento de la velocidad de operación según la técnica estadística Best Subsets son las siguientes:

1. Ancho de paseo pavimentado IZQ (m)
2. Peralte promedio IZQ (%)
3. Longitud (m)
4. Velocidad de diseño de curva horizontal (km/h)
5. Velocidad de diseño inferida (km/h)
6. Radio de curva (m)
7. Delta o deflexión ($^{\circ}$)
8. Coeficiente Fricción (ad.)
9. Código carril (ad.)
10. Código Calzada (ad.)
11. Longitud a acceso en tangente menor a 60.96m. (m)
12. Línea central marcada (ad.)
13. Límite de velocidad (km/h)
14. Tipo de terreno (ad.)
15. Zanja izquierda (ad.)
16. Bandas sonoras centro o calzada (ad.)
17. Uso de suelo Derecha (ad.)

Estas variables se ponen en un archivo de Excel, el cual se analiza en el programa Stata V14, con la técnica estadística de Datos en Panel.

Variables utilizadas en ecuaciones de estimación de V_{85}

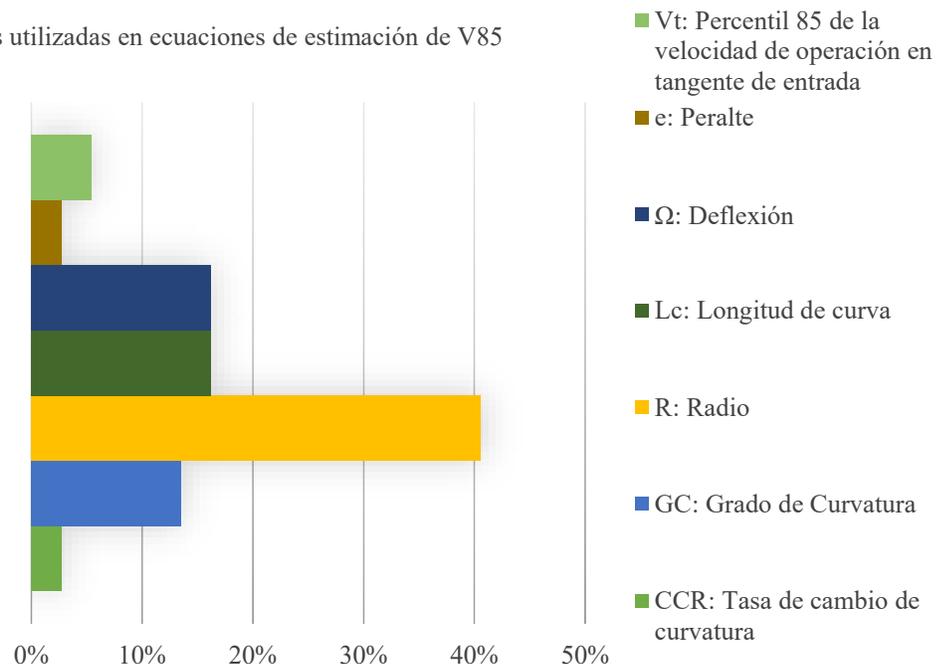


Figura 1. Variables utilizadas en ecuaciones de estimación de la V_{85} . Fuente: elaboración propia.

4.2. Datos en Panel

Es un método estadístico utilizado en las ciencias económicas, es una combinación de datos de series de tiempo y de corte transversal. Un modelo de regresión de datos en panel tiene la siguiente forma [10]:

$$Y_{it} = \alpha_{it} + X_{it}\beta + U_{it} \quad (1)$$

Donde:

i: Individuo o unidad de estudio (sujeto).

t: tiempo

α : vector de interceptos con n parámetros

β : es un vector de K parámetros y

Xit: es la i-ésima observación al momento t para las variables K explicativas.

La técnica estadística tiene dos variantes llamadas método de efectos fijos y método de efectos aleatorios, el primero considera que cada individuo tiene un término constante diferente, de tal manera que las unidades de corte transversal son afectadas por las variables explicativas y estas a su vez se distinguen por cualidades propias medidas por medio del intercepto. El modelo de efectos aleatorios considera que las características propias de cada individuo tienen igual relevancia que las variables explicativas, por ello considera que no son independientes entre sí los efectos individuales, sino que están alrededor de un valor distribuidos de manera aleatoria [11].

Para este estudio se utilizara el software Stata V14, este es utilizado por investigadores y personal académico alrededor del mundo con grandes aportes. Uno de los supuestos del modelo de Datos en Panel es que las variables son independientes entre sí, para comprobar que cumplen con la hipótesis anterior es necesario hacer una matriz de correlación para saber cuales son las que violan la premisa y así poder desestimarlas.

Las variables correlacionadas se compilan en la Tabla 1.

Tabla 1. Variables correlacionadas

Variable 1	Variable 2	Variable 3
Vdisch (4)	Veldisinf (5)	
CodCal (10)	Anchopaspavizq (1)	
Coefric (8)	Delta (7)	
Vdisch (4)	Lim Vel (13)	Veldisinf (5)
CodCal (10)	CodCarr (9)	

Nota: Numeración tomada de las variables representativas determinadas con el método Best Subsets.

Para empezar se ejecuta el Panel por el método de efectos fijos y efectos aleatorios, además, se hace la prueba de

Hausman para verificar cuál de los métodos es el mejor [11], de ahí en adelante se analizan los datos por el método elegido, como se muestra en la figura 5.

Como el *Pvalor* es mayor a 0.05 se debe admitir la hipótesis nula de igualdad de condiciones, por lo cual el mejor método es el de efectos aleatorios [11], el cual se presenta en la figura 6. Una vez determinado el método, se empieza con la sustracción de las variables analizando el comportamiento del R^2 .

El modelo de Datos en Panel con las variables resultantes tienen un valor de R^2 alto, un *Pvalor* de cada variable menor a 0.05 e intervalos de confianza coherentes.

Las variables seleccionadas para el modelo son:

- Veldisinf: Velocidad de diseño inferida
- Radio
- Delta
- Codcal: Código calzada
- TipoTer: Tipo de terreno
- Zanzizq: zanja izquierda

Aunque el modelo presente 6 variables, 3 categóricas y 3 cuantitativas, y un R^2 alto de 0.8558, como se muestra en la figura 7, se aplica el principio de parsimonia, el cual dice que los métodos simples son destacados ante los [13], por lo cual, examinando la literatura, se eligen las variables que son más utilizadas en modelos para estimar la velocidad de operación V_{85} , estas son Radio y Delta; caso contrario están las demás variables: Veldisinf, Codcal, Tipoter y Zanzizq.

La aplicación de dicho principio se evalúa de la misma forma, analizando el comportamiento del valor de R^2 , por lo cual, el modelo se define con las siguientes variables, con un valor de R^2 de 0.83:

- Veldisinf: velocidad de diseño inferida.
- Radio
- Delta
- Constante.

Para determinar la eficiencia del modelo se realiza una regresión lineal, obteniendo las variables de la misma manera que se establecieron las del método de Datos en Panel; esta arrojó las variables Vdisch (Velocidad de diseño de curvas horizontales) [km/h], Veldisinf (Velocidad de diseño inferido) [Km/h], y Delta (Deflexión) [radianes] y una constante, los valores de los coeficientes, y un R^2 como se muestra en la figura 8.

El valor de R^2 es 0.82, similar al de la regresión con Datos en Panel, siendo este valor ajustado a los obtenidos en otros estudios, los modelos se presentan en la Tabla 2.

Al obtener los modelos se procede a compararlos utilizando el 30% de los datos, en este procedimiento se analizó el comportamiento de los modelos, como se muestra en las Figuras 9 a 13.

Tabla 2. Variables correlacionadas

Modelo	Variables			
	Veldisinf	Radio	Delta	Constante
Datos en Panel	0.6137267	0.0034777	-0.1087	35.79596
Lineal	Veldisinf	Vdisch	Delta	Constante
	0.2538001	0.4630048	-0.0701	23.30297

	Coefficients			
	(b) modelo_fe	(B) modelo_re	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
Anchopaspavizq	-2.682655	-.493007	-2.189648	1.25117
epromizq	.104409	.0934282	.0109767	.0461075
Long	-.0053841	-.0007632	-.0046209	.0017212
radio	.0030865	.0027175	.000369	.0011015
Delta	-.1191671	-.112387	-.00678	.0117892
coefric	-6.479485	-5.939739	-.5397458	2.176261
Codcar	-1.10711	-.7871443	-.3199659	.2360397
Lacmen60_96	.0814438	.0882813	-.0068374	.3342792
zanizq	-1.239538	-1.583061	.3435235	.3972743

b = consistent under Ho and Ha; obtained from xtreg
 B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(6) = (b-B)'[(V_b-V_B)^(-1)](b-B) = 12.13
 Prob>chi2 = 0.1454
 (V_b-V_B is not positive definite)

Figura 5. Test de Haussman. Fuente: elaboración propia.

Source	SS	df	MS	Number of obs	F(3, 376)	Prob > F	R-squared	Adj R-squared	Root MSE
Model	46160.326	3	15386.7753	380	566.58	0.0000	0.8189		
Residual	10211.0661	376	27.1570907			0.8189			
Total	56371.3922	379	148.737182						5.2112

V85	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Vdisch	.4630048	.0397448	11.65	0.000	.3848549 .5411547
veldisinf	.2538001	.0323621	7.84	0.000	.1901667 .3174335
Delta	-.0701078	.0110262	-6.36	0.000	-.0917885 -.048427
_cons	23.30297	1.663922	14.00	0.000	20.03121 26.57473

Figura 8. Regresión lineal con las variables seleccionadas. Fuente: elaboración propia.

Random-effects GLS regression

Group variable: id

R-sq: within = 0.2736, between = 0.9466, overall = 0.8793

Number of obs = 265, Number of groups = 53

Obs per group: min = 5, avg = 5.0, max = 5

corr(u_i, X) = 0 (assumed)

Wald chi2(17) = 837.83, Prob > chi2 = 0.0000

V85	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Anchopaspavizq	-.493007	.5436684	-0.91	0.365	-1.558577 .5725635
epromizq	.0934282	.0798365	1.17	0.242	-.0630485 .2499049
Long	-.0007632	.0055607	-0.14	0.891	-.0116619 .0101355
Vdisch	.2536447	.0933505	2.72	0.007	.0706812 .4366083
veldisinf	.4420457	.1153723	3.83	0.000	.2159202 .6681711
radio	.0027175	.0018961	1.43	0.152	-.0009987 .0064337
Delta	-.112387	.0229665	-4.89	0.000	-.1574005 -.0673735
coefric	-5.939739	3.796211	-1.56	0.118	-13.38018 1.500698
Codcar	-.7871443	.577486	-1.36	0.173	-1.918996 .3447075
Codcal	5.242749	2.376826	2.21	0.027	.5842554 9.901243
Lacmen60_96	.0882813	.6084975	0.15	0.885	-1.104352 1.280914
Lincenmarc	-6.516329	3.187235	-2.04	0.041	-12.7632 .2694635
lim_vel	-.184058	.2931637	-0.63	0.530	-.7586483 .3905324
tipoter	-1.038895	.7528875	-1.38	0.168	-2.514528 .436737
zanizq	-1.583061	.8039752	-1.97	0.049	-3.158823 -.0072985
Bandson	-4.036746	4.569266	-0.88	0.377	-12.99234 4.91885
Uso_suelo_der	-1.578528	1.575456	-1.00	0.316	-4.666366 1.50931
_cons	49.2324	16.84156	2.92	0.003	16.22355 82.24124

sigma_u = 2.5597897, sigma_e = 3.9092257, rho = .30009859 (fraction of variance due to u_i)

Figura 6. Datos en Panel por el método de efectos aleatorios. Fuente: elaboración propia.

Random-effects GLS regression

Group variable: id

R-sq: within = 0.2485, between = 0.9232, overall = 0.8558

Number of obs = 265, Number of groups = 53

Obs per group: min = 5, avg = 5.0, max = 5

corr(u_i, X) = 0 (assumed)

Wald chi2(6) = 640.21, Prob > chi2 = 0.0000

V85	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
veldisinf	.5366508	.0395694	13.56	0.000	.4590962 .6142053
radio	.0028594	.0012297	2.33	0.020	.0004491 .0052696
Delta	-.1025699	.0119089	-8.61	0.000	-.1259109 -.079229
Codcal	3.105748	1.349039	2.30	0.021	.4616808 5.749815
tipoter	-2.229442	.663044	-3.36	0.001	-3.529494 -.9293888
zanizq	-1.980025	.7769083	-2.55	0.011	-3.502737 -.4573126
_cons	44.00336	2.710699	16.23	0.000	38.69049 49.31624

sigma_u = 3.1006982, sigma_e = 3.9461097, rho = .3817314 (fraction of variance due to u_i)

Figura 7. Variables seleccionadas por el método de Datos en Panel con efectos aleatorios. Fuente: elaboración propia.

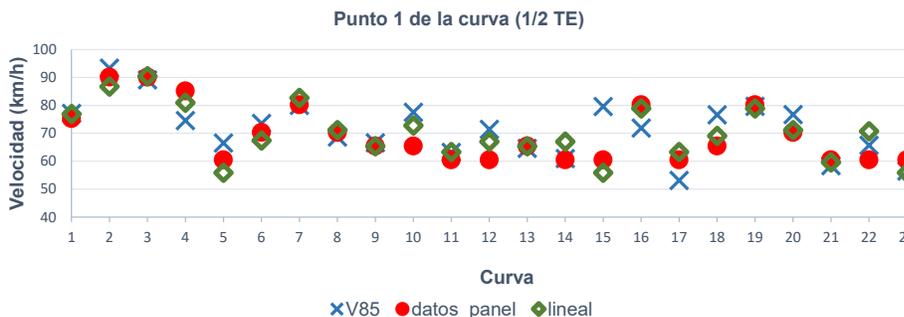


Figura 9. Comparación de los modelos en el punto 1 de la curva (1/2 TE). Fuente: elaboración propia.

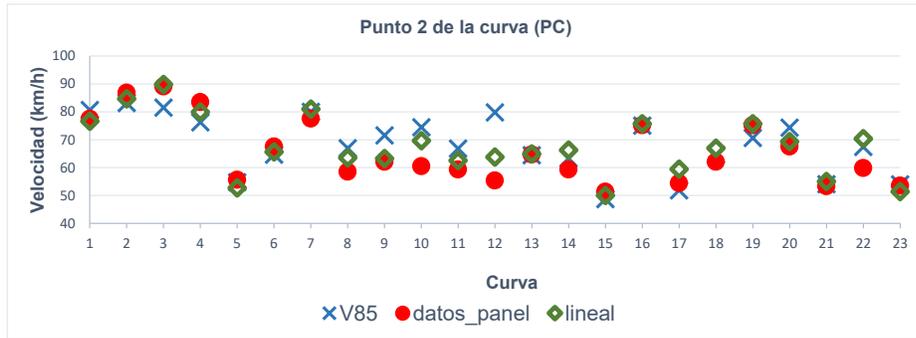


Figura 10. Comparación de los modelos en el punto 2 de la curva (PC). Fuente: elaboración propia.

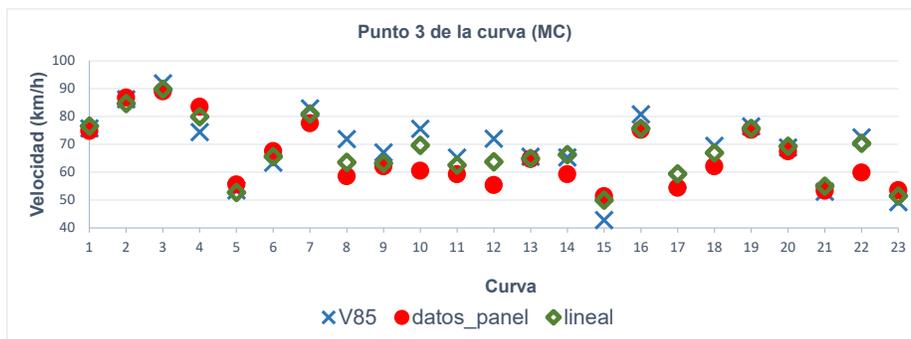


Figura 11. Comparación de los modelos en el punto 3 de la curva (MC). Fuente: elaboración propia.

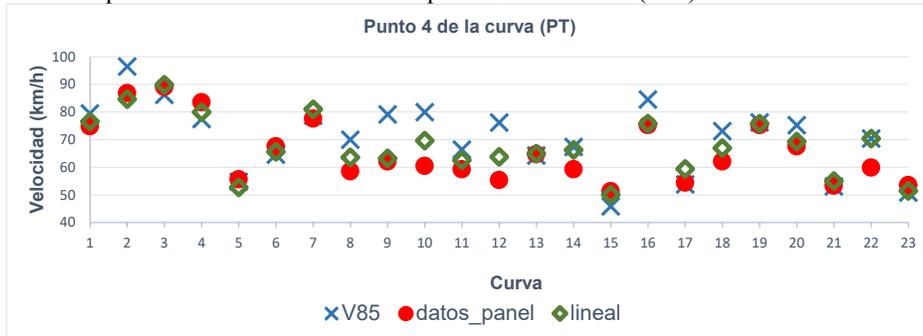


Figura 12. Comparación de los modelos en el punto 4 de la curva (PT). Fuente: elaboración propia.

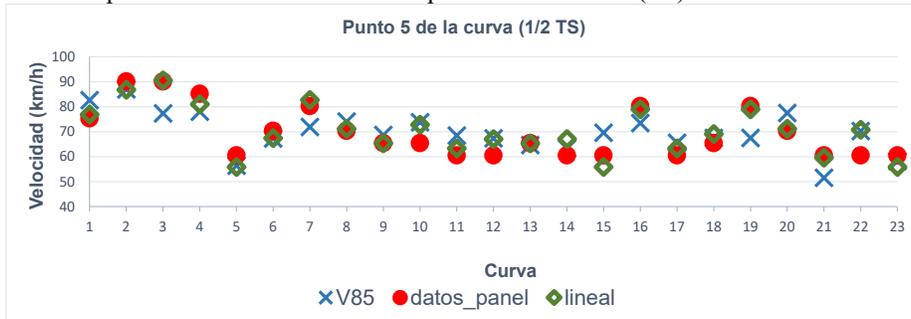


Figura 13. Comparación de los modelos en el punto 5 de la curva (1/2 TS). Fuente: elaboración propia.

Como se puede apreciar en las Figuras 9 a 13 el modelo de Datos en Panel genera resultados similares al lineal, ahora, aplicamos el filtro de Kalman con el modelo de Datos en Panel para ver cómo se comportan las estimaciones.

4.3. El Filtro de Kalman

El filtro de Kalman es uno de los métodos estadísticos que realizan estimación de sistemas dinámicos estado-espacio [12], la forma matemática del filtro es la siguiente:

$$X_t = AX_{t-1} + w_{t-1}$$

Donde la fase $X \in \mathcal{R}^n$, se intenta estimar dentro de un proceso regulado en tiempo discreto con forma lineal estocástica, con una ecuación de medición, la cual es:

$$Z_t = HX_t + v_t$$

Donde, $Z \in \mathcal{R}^n$.

Los supuestos del filtro consisten en que w_t y v_t son variables aleatorias, las cuales son ruido y además independientes entre sí con media igual a cero y distribución normal [12].

$$\begin{aligned} p(w) &\cong N(0, Q) \\ p(v) &\cong N(0, R) \end{aligned}$$

Donde Q y R son matrices de covarianza de la perturbación del proceso y de la medida, respectivamente.

El filtro de Kalman basa su funcionamiento en la retroalimentación del proceso, es decir, estima el proceso en un instante de tiempo determinado y logra la retroalimentación por medio de la información tomada. Así, se considera que el filtro se divide en dos partes: la primera contiene las ecuaciones de estimación del estado y la segunda las de corrección o actualización, convirtiéndose este en un algoritmo de predicción-corrección, la Figura 3 describe el proceso del filtro de kalman.

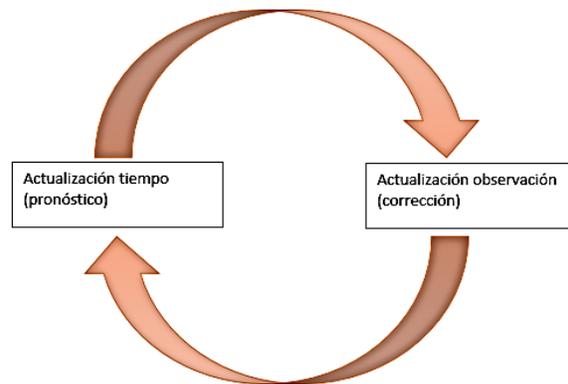


Figura 14. Ciclo del Filtro de Kalman. Fuente: Modificada de [11].

El filtro de Kalman es un algoritmo que provee una solución recursiva óptima, por el método de mínimos cuadrados, para un sistema dinámico lineal; de esta manera, las iteraciones del filtro dependen del tamaño de la muestra. El esquema del filtro de Kalman utilizado se muestra en la siguiente Figura 15:

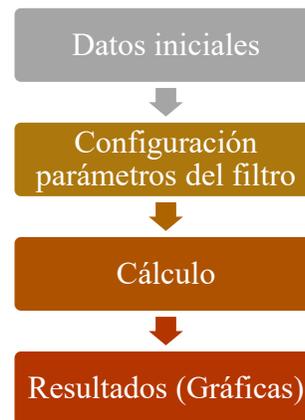


Figura 15. Proceso del filtro de Kalman. Fuente: elaboración propia.

El uso del filtro de Kalman se realizó con ayuda del software Matlab, que tiene esta herramienta estadística a modo de Toolbox. Al suministrarle los datos al programa arrojó lo que se puede apreciar en las Figuras 16 a 20.

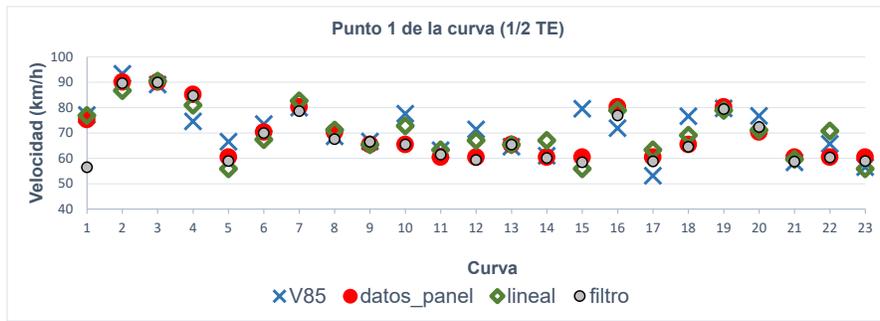


Figura 16. Comparación de los modelos en el punto 1 de la curva (1/2 TE). Fuente: elaboración propia.

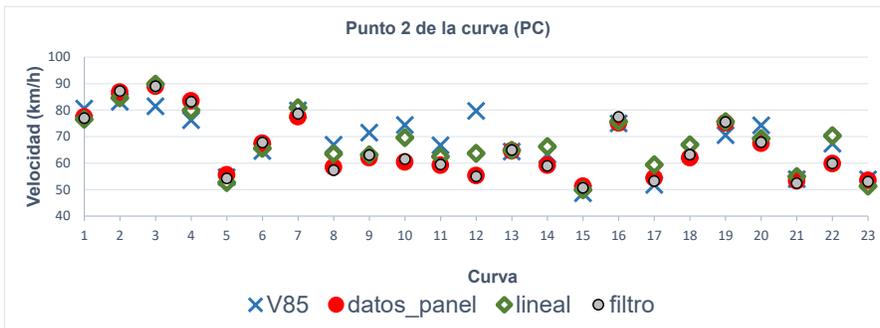


Figura 17. Comparación de los modelos en el punto 2 de la curva (PC). Fuente: elaboración propia.

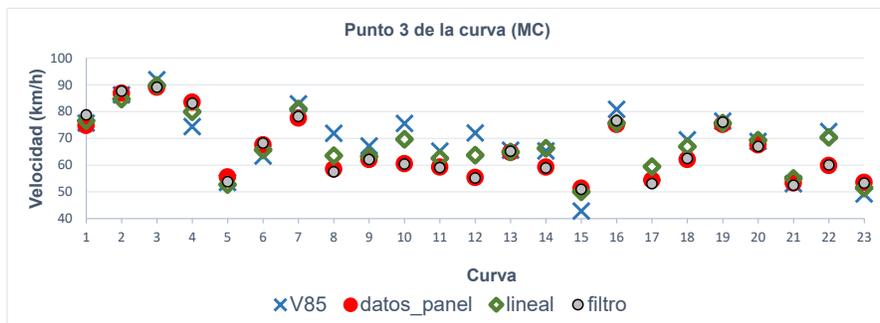


Figura 18. Comparación de los modelos en el punto 3 de la curva (MC). Fuente: elaboración propia.

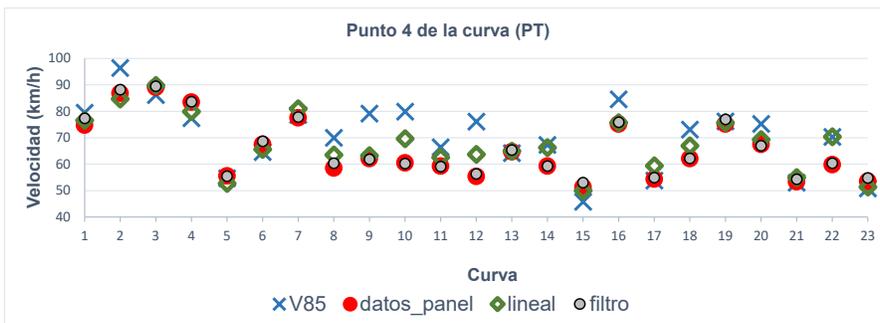


Figura 19. Comparación de los modelos en el punto 4 de la curva (PT). Fuente: elaboración propia.

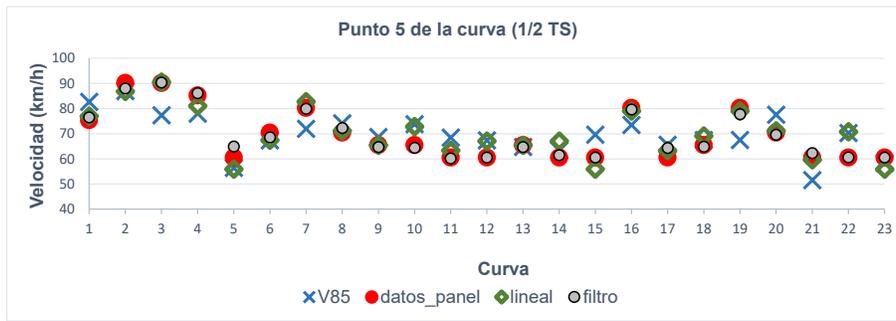


Figura 20. Comparación de los modelos en el punto 5 de la curva (1/2 TS). Fuente: elaboración propia.

Los datos arrojados por el filtro se analizaron con ayuda del software Minitab 17.

4.4. Análisis estadístico de los modelos

A los modelos elaborados en los capítulos anteriores se realiza un análisis de varianza ANOVA por sus siglas en inglés, este procedimiento se hace para cada punto de la curva, sabiendo de antemano que la velocidad se obtiene en 5 puntos diferentes de cada curva horizontal.

El análisis realizado en el software Minitab 17 compara los tres tratamientos en una sola corrida y arrojó que los modelos tienen el mismo comportamiento para todos los puntos de la curva, como se muestra en las siguientes Figuras 22, 23, 24, 25.

La comparación de los valores en cada uno de los puntos arroja que ninguno de los modelos es mejor, todos tienen el comportamiento similar, obteniendo los valores de la media muy parecidos.

Fuente	GL	SC	Ajust.	MC	Ajust.	Valor F	Valor p
Factor	3	96.00	32.00	0.31	0.818		
Error	84	8665.15	103.16				
Total	87	8761.15					

Figura 21. ANOVA de los modelos en el punto 1. Fuente: elaboración propia.

Fuente	GL	SC	Ajust.	MC	Ajust.	Valor F	Valor p
Factor	3	120.6	40.20	0.33	0.801		
Error	84	10124.3	120.53				
Total	87	10244.9					

Figura 22. ANOVA de los modelos en el punto 5. Fuente: elaboración propia.

Fuente	GL	SC	Ajust.	MC	Ajust.	Valor F	Valor p
Factor	3	188.2	62.74	0.48	0.700		
Error	84	11069.6	131.78				
Total	87	11257.8					

Figura 23. ANOVA de los modelos en el punto 5. Fuente: elaboración propia.

Fuente	GL	SC	Ajust.	MC	Ajust.	Valor F	Valor p
Factor	3	340.7	113.6	0.85	0.468		
Error	84	11172.1	133.0				
Total	87	11512.8					

Figura 24. ANOVA de los modelos en el punto 4. Fuente: elaboración propia.

Fuente	GL	SC	Ajust.	MC	Ajust.	Valor F	Valor p
Factor	3	340.7	113.6	0.85	0.468		
Error	84	11172.1	133.0				
Total	87	11512.8					

Figura 25. ANOVA de los modelos en el punto 5. Fuente: elaboración propia.

Posteriormente, fue necesario comparar los modelos obtenidos entre sí y, además, frente a tres modelos establecidos en la literatura, los cuales tienen variables en común y forma lineal; para ello fue necesario tomar 9 curvas al azar y hacer el análisis de estadísticos de error; como el error cuadrático medio MSE (por sus siglas en inglés), desviación media absoluta MAD (por sus siglas en inglés) y la raíz del error cuadrático medio RMSE (por sus siglas en inglés). Los modelos son los siguientes:

$$V_{85} = 102.44 - \frac{2471.81}{R} + 0.012Lc - 0.10\Omega \quad \text{KRAMMES, R., et al. (1995)}$$

$$V_{85} = 99.61 - \frac{2951.37}{R} + 0.014Lc - 0.13\Omega - 71.82e \quad \text{VOIGT, A. (1996)}$$

$$V_{85} = 41.62 - 1.29GC + 0.049Lc - 0.12\Omega + 0.95Vt$$

McFADDEN, J. y
ELEFTERIADOU, L.
(1997)

La tabla 3 presenta los resultados de la evaluación realizada a las curvas 9 curvas seleccionadas. Es claro desde estos resultados, que los modelos que más se ajustan son los propuestos en este estudio (es decir, Datos en Panel, Lineal y Filtro de Kalman).

Tabla 3. Comparación estadística de los modelos

Estadístico	Modelos					
	1	2	3	4	5	6
MSE	43.1	26.1	222.4	173.6	1541.4	43.2
RMSE	6.6	5.1	14.9	13.2	39.3	6.6
MAD	5.1	4.3	13.2	13.8	38.9	5.0

1: Datos de Panel; 2: Lineal; 3: Krammes; 4: Voigt; 5: McFaden; 6: Filtro de Kalman

5. Conclusiones

De las 23 variables cuantitativas y 46 variables cualitativas, se seleccionaron las significativas con el método de Best Subsets, se analizaron con el software Stata V14 cuales son las mejores para el modelo de Datos en Panel, las cuales fueron: velocidad de diseño inferida (km/h), radio (m) y ángulo de deflexión (radianes). De esta misma manera, se analizaron cuáles eran las mejores variables para generar un modelo lineal, el cual arrojó las siguientes variables: Vdisch (Velocidad de diseño de curvas horizontales) [km/h], Veldisinf (Velocidad de diseño inferido) [km/h], y Delta (ángulo de deflexión) [radianes], las cuales son variables muy poco utilizadas en la literatura relacionada.

Al comparar el modelo de Datos en Panel con el lineal, se observa similitud en los valores de R², sin embargo, el modelo lineal tiene menores valores en los estadísticos MSE, RMSE y MAD, por lo cual, se puede concluir que el modelo lineal, estima mejor la velocidad de operación en curvas horizontales

El filtro de Kalman se adapta muy bien al modelo de Datos en Panel generando un proceso de suavización de la curva, lo que se traduce en una mejoría del modelo, por lo que obtuvo valores intermedios entre los modelos preestablecidos comportándose de manera dinámica en las estimaciones de la velocidad de operación (V85).

El modelo de datos en Panel estima con gran precisión los valores de velocidad operacional en curvas con radios menores a 50 m y mayores a 200 m.

Agradecimientos

A la Universidad del Cauca por la oportunidad de realizar este trabajo.

Financiación

No aplica.

Contribución de los autores

V. R. Castillo-Ñañez: Conceptualización, Metodología, Software, Validación, Análisis formal, Manuscrito inicial. C. A. Calero-Valenzuela: Conceptualización, Metodología, Análisis formal, Investigación, Revisión y Edición del Manuscrito. C. A. Gaviria- Mendoza: Validación, Investigación, Recursos, Manuscrito inicial, Supervisión.

Todos los autores han leído y están de acuerdo con la versión publicada del artículo.

Conflicto de interés

Declaramos que no existe ningún conflicto de intereses en relación con este artículo. Los autores no tienen intereses financieros, personales, ni profesionales que puedan influir en los resultados o en la interpretación de los hallazgos encontrados.

Declaración de la Junta de Revisión Institucional

No aplica.

Declaración de consentimiento informado

No aplica.

Referencias

- [1] Organización Mundial de la Salud, "Informe Sobre La Situación Mundial De La Seguridad Vial 2023," Ginebra, Suiza, ISBN 978-92-4-008651-7.
- [2] A. Polus and C. Mattar-Habib, "New Consistency Model for Rural Highways and Its Relationship to Safety," *J. Transp. Eng.*, vol. 130, no. 3, pp. 286–293, 2004, doi: [https://doi.org/10.1061/\(ASCE\)0733-947X\(2004\)130:3\(286\)](https://doi.org/10.1061/(ASCE)0733-947X(2004)130:3(286))
- [3] P. Misaghi, Y. Hassan, "Modeling Operating Speed and Speed Differential on Two-Lane Rural Roads," *J. Transp. Eng.*, vol. 131, no. 6, pp. 408–418, 2005, doi: [https://doi.org/10.1061/\(ASCE\)0733-947X\(2004\)130:3\(286\)](https://doi.org/10.1061/(ASCE)0733-947X(2004)130:3(286))

[4] M. D. Wooldridge, K. Fitzpatrick, D. W. Harwood, I. B. Potts, L. Elefteriadou, and J. D. Torbic, *Geometric design consistency on high-speed rural two-lane roadways*. 2003.

[5] Y. H. G. M. Gibreel, S. M. Easa, Member, ASCE and I. A. El-Dimeery, “state of the art of highway geometric design consistency,” *Journal of Transportation Engineering*, vol. 125, no. August, pp. 305–313, 1999, [https://doi.org/10.1061/\(ASCE\)0733-947X\(1999\)125:4\(305\)](https://doi.org/10.1061/(ASCE)0733-947X(1999)125:4(305))

[6] T. Toroyan, “Global status report on road safety,” World Heal. *Organisation*, p. 318, 2015.

[7] C. A. Calero Valenzuela, “Metodología para la Evaluación de la Consistencia de Diseño de Carreteras Rurales de Dos Carriles,” Puerto Rico University, 2015.

[8] M. Castro, J. F. Sánchez, J. A. Sánchez, and L. Iglesias, “Operating Speed and Speed Differential for Highway Design Consistency,” *J. Transp. Eng.*, vol. 137, no. 11, pp. 837–840, 2011, doi: [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000309](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000309)

[9] C. Gil Martínez, “Técnicas de regularización y selección del mejor modelo,” 2018.

[10] M. Mayorga and E. Muñoz, “La técnica de datos de panel. Una guía para su uso e interpretación,” Banco Cent. Costa Rica, Dep. Investig. económicas, p. 18, 2000.

[11] R. Montero, Efectos fijos o aleatorios: test de especificación. Documentos de Trabajo en Economía Aplicada. Universidad de Granada. España. 2011.

[12] Á. Solera Ramírez, “El filtro de kalman,” Nota técnica, Dep. Investig. económicas, Banco Cent. Costa Rica, pp. 1–32, 2003.

[13] J. E. Hanke, and D. W. Wichern, *Pronosticos en los Negocios*. Pearson Educación, 2010.