



MODELOS DE REGRESIÓN PARAMÉTRICOS POLINOMIALES Y NO PARAMÉTRICOS B-SPLINES. UNA APLICACIÓN EN INGENIERÍA.

Polynomial parametric and non-parametric B-Splines regression models. An Engineering Application.

 ¹Byron Toalombo Rojas*

 ²Antonio Meneses Freire

 ³Lourdes Zúñiga Lema

 ¹Ricardo Espín Guerrero

¹Universidad Técnica de Ambato, Facultad de Sistemas, Electrónica e Industrial, Ambato, Ecuador.

²Universidad Nacional de Chimborazo, Facultad de Ingeniería, Carrera de Ingeniería en Telecomunicaciones, Riobamba, Ecuador.

³Escuela Superior Politécnica de Chimborazo, Facultad de Informática y Electrónica, Carrera Ingeniería en Telecomunicaciones, Riobamba, Ecuador.

*byronmtr@gmail.com

RESUMEN

Se realizó un estudio de los modelos de regresión paramétricos polinomiales y no paramétricos B-splines aplicados a una simulación de impacto de un automóvil contra una carrocería de autobús. Se estableció un diseño no experimental, transversal y correlacional. Se utilizó el software R y se consideraron los criterios: rechazo de la nulidad de los coeficientes de los modelos mediante la prueba de hipótesis t de Student, validez de los modelos mediante la prueba F de Snedecor de la tabla ANOVA, bondad de ajuste, intervalos de confianza al 95%, y cumplimiento de los supuestos de distribución normal, no autocorrelación y homocedasticidad de los residuos. Para la selección del modelo de regresión idóneo se aplicó la prueba no paramétrica de Wilcoxon a partir de las longitudes de los intervalos de confianza. Los modelos de regresión paramétricos polinomiales se ajustaron a las curvas con forma parabólica o que presentan una curvatura sin cambios abruptos, adaptándose mejor a las relaciones de la simulación de impacto vehicular, cuya variable explicativa es la velocidad del vehículo que impacta. Mientras los modelos de regresión no paramétricos B-splines brindaron un mejor ajuste a las curvas con forma de campana y con cambios de curvatura más abruptos.

Palabras claves: *Simulación de impacto vehicular, modelos de regresión polinomiales, modelos de regresión B-splines, bondad de ajuste, intervalo de confianza, normalidad.*

ABSTRACT

A study of the polynomial parametric and nonparametric B-splines regression models applied to a simulation of an impact of a car against a bus body was performed. A non-experimental design, cross-sectional, and correlational was established. The R software was used and the following criteria were considered: the rejection of the nullity of the coefficients of the models through the Student's t-hypothesis test, the validity of the models using the Snedecor F-test of the ANOVA table, the goodness of fit, 95% confidence intervals, and compliance with the assumptions of normal distribution, non-autocorrelation and homoscedasticity of the residuals. The Wilcoxon nonparametric test was applied to select the appropriate regression model based on the lengths of the confidence intervals. The parametric polynomial regression models were fitted to curves with parabolic shape or curvature without abrupt changes, fitting better to the relationships of the vehicle impact simulation, whose explanatory variable is the speed of the impacting vehicle. While the nonparametric B-splines regression models provided a better fit to bell-shaped curves with more abrupt curvature changes.

Keywords: *Vehicle crash simulation, polynomial regression models, B-Splines regression models, goodness of fit, confidence interval, normality.*

Fecha de recepción: 04-12-2022

Fecha de aceptación: 15-06-2022

Fecha de publicación: 11-10-2022

I. INTRODUCCIÓN

Los accidentes de tránsito representan un problema que atenta contra la salud de las personas, los principales factores que inciden en la ocurrencia de este tipo de acontecimientos son entre otros la impericia de los conductores y las condiciones de las vías (1).

Ante tal situación los fabricantes de automotores y de autobuses no pueden intervenir directamente, sin embargo, están en capacidad de minimizar la afectación de los accidentes en las personas involucradas, a través de la fabricación de componentes que brinden buena resistencia al impacto y a la deformación.

Esto constituye un aspecto inherente a la ingeniería mecánica automotriz, siendo que en el diseño de los vehículos se consideran aspectos como la forma y los materiales de los componentes automotrices (2).

Bajo el contexto indicado, los fabricantes de vehículos y de autobuses suelen implementar herramientas tecnológicas para el análisis estático y dinámico de las colisiones y choques.

La simulación de impacto es una de las herramientas indispensables, que consiste en la utilización de un software de Elementos de Análisis Computarizado (CAE) para representar un choque real. De este modo se conoce el efecto que las cargas de impacto provocan en las estructuras de los automotores (estados de tensiones y deformaciones) (3).

En el análisis estático y dinámico de una colisión de vehículos o de un choque intervienen las magnitudes físicas, tales como: velocidad lineal, fuerza, esfuerzo, tiempo de impacto, factor de seguridad de diseño (FDS), entre otras. El FDS es el cociente calculado entre el esfuerzo de cedencia de un material y el esfuerzo dado por las condiciones a las que está sometido el diseño (4).

De entre las magnitudes señaladas, la velocidad de circulación de los vehículos es la que condiciona el comportamiento del fenómeno, pues no se experimenta el mismo resultado en un siniestro de tránsito a distintas velocidades (5). En este sentido, el presente estudio considera el análisis del efecto que produce la velocidad de circulación de un automóvil en las magnitudes fuerza de impacto, tiempo de duración del

impacto, deformación y factor de seguridad (FDS) durante un evento de choque contra la parte posterior de la carrocería de un autobús.

El trabajo se enfoca en la caracterización de los modelos teóricos estadísticos de regresión paramétricos polinomiales y no paramétricos B-splines. Los primeros son tradicionalmente conocidos y relativamente sencillos de utilizar, sin embargo, conllevan algunas restricciones para ser utilizados en cada caso particular (6).

Adicionalmente es necesario considerar la bondad de ajuste de los modelos, para tomar una decisión respecto al que se adapta de mejor manera a las condiciones de los problemas estudiados (7). El impacto de la investigación radica en el suministro de información útil para el desarrollo de tecnologías en la fabricación de carrocerías de autobuses.

Modelos de regresión paramétricos polinomiales.

La regresión polinomial es un caso especial de regresión múltiple, con solo una variable regresora o predictora X. El modelo de regresión polinomial de una variable explicativa o regresora se puede expresar de la siguiente manera:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_k x_i^k + \epsilon_i, \quad i = 1, \dots, n. \quad (1)$$

Donde:

n es el número de datos disponibles a ser utilizados para hallar el modelo.

k es el grado del polinomio (8).

Métricas de medición del error y bondad de ajuste.

Para la evaluación de un modelo de regresión se obtiene la tabla ANOVA a partir de los parámetros de la variación: regresión, error y total, como se muestra en la Tabla 1:

Fuente de variación	G. L.	Sumas cuadráticas	Medias cuadráticas	F_0	p-valor
Regresión	k	$SSR = \hat{\beta}_1 \cdot S_{xy}$	$MSR = \frac{SSR}{k}$	$\frac{MSR}{MSE}$	
Error	$n - p$	$SSE = SST - \hat{\beta}_1 \cdot S_{xy}$	$MSE = \frac{SSE}{n - p}$	-	-
Total	$n - 1$	SST	-	-	-

Tabla 1. Tabla análisis de varianza, ANOVA. Fuente: (9).

F_0 , es el estadístico de prueba F de Snedecor.
 MSR , media cuadrática de la regresión.
 MSE , media cuadrática del error.
 SSR , suma de cuadrados de la regresión.
 SSE , suma de cuadrados del error.
 n , es el número de datos disponibles a ser utilizados para hallar el modelo.
 k , es el grado del polinomio.
 $p = k+1$.

La bondad de ajuste de un modelo de regresión es el principal criterio que valora su capacidad de predicción de una variable respuesta. Para el efecto se mide la variabilidad explicada por el modelo sobre la variable a predecir y la forma de hacerlo es a través del uso del error estándar de los residuos (RSE), el test F de bondad de ajuste mostrado en la tabla ANOVA y el coeficiente de determinación ajustado R^2 ajustado (10).

Modelos de regresión no paramétricos B-splines.

Los modelos de regresión no paramétricos permiten una estimación flexible de efectos no lineales y no requieren ningún supuesto restrictivo con respecto a una determinada forma funcional paramétrica.

En el caso de una sola covariable continua x , el modelo estándar para la regresión no paramétrica se define de la siguiente manera:

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, \dots, n. \quad (2)$$

Para la variable de error ϵ_i , se supone que la función f presenta ciertas características de suavidad, como continuidad o diferenciabilidad, pero no se especifica una forma paramétrica específica. Una spline es una función que se construye por partes a partir de funciones polinomiales, el término proviene del "listón elástico" (11). Los splines "son curvas polinómicas por trozos continuamente diferenciables hasta un orden prescrito" (12). Entre los ejemplos más característicos se tienen C_0 , una spline lineal por trozos y C_1 un spline cúbico.

Función base B-spline

Una función B-spline es la función de base interpolativa de máxima diferenciación y representa una generalización de la curva de Bézier. Los B-splines se definen por su orden m y el número de nodos interiores N (hay dos puntos finales que son en sí mismos nodos, por lo que el número total de nodos será $N + 2$). El grado del

polinomio B-spline será el orden de spline $m-1$ (13).

Una B-spline de grado n (de orden spline $m = n+1$) es una curva paramétrica compuesta de una combinación lineal de B-splines de base $B_{i,n}(x)$ de grado n dada por:

$$B(x) = \sum_{i=0}^{N+n} \beta_i \cdot B_{i,n}(x), \quad x \in [t_0, t_{N+1}] \quad (3)$$

La función f tiene la siguiente estructura:

$$f(x) = a_1 \cdot B_1(x) + \dots + a_k \cdot B_k(x) \quad (4)$$

Donde:

k , número de bases.

a_1, \dots, a_k , parámetros desconocidos.

B_1, \dots, B_k , funciones polinomiales de base B-spline.

β_i , puntos de control o puntos de Boor.

La secuencia t se conoce como una secuencia de nodos, siendo un nodo el término individual de la secuencia (13).

La regresión basada en B-spline de orden p se considera de la estructura:

$$f(x) = a_0 + a_1x + \dots + a_px^p + B_1(x - C_1)^p + \dots + B_k(x - C_k)^p \quad (5)$$

Donde:

a_0, \dots, a_p y B_1, \dots, B_k , coeficientes a determinar.

El spline f puede ser expresado como una combinación lineal de la forma:

$$f(x) = a_0 \cdot B_0(x) + \dots + a_p \cdot B_p(x) + a_{p+1} \cdot B_{p+1}(x) + \dots + a_{p+M} \cdot B_{p+M}(x) \quad (6)$$

Las funciones mostradas a continuación constituyen una base de funciones polinómicas del spline:

$$B_0(x) = 1, B_1(x) = x, B_p(x) = x^p \quad (7)$$

$$B_{p+1}(x) = (x - C_1), \dots, B_{p+M} = (x - C_M) \quad (8)$$

En la Figura 1 se muestran las curvas B-spline de base para un determinado grado del polinomio, número de vértices del polígono de control y nodos internos.

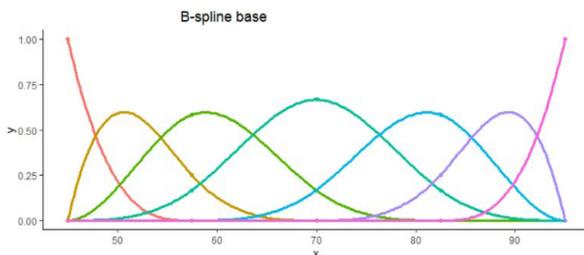


Figura 1. B-spline base para un modelo de regresión.

II. MATERIALES Y MÉTODOS

• Materiales.

Para el desarrollo de la investigación se hace uso de los siguientes materiales:

- Software estadístico R (14).
- Software CAD-CAE.
- Computador y equipos de oficina

• Metodología.

La investigación es de tipo correlacional bivariada, siendo las variables de interés: velocidad lineal del automóvil, fuerza, factor de seguridad (FDS), tiempo de impacto y deformación. El diseño de la investigación es no experimental de corte transversal, porque los datos fueron obtenidos en las condiciones regulares de ocurrencia de los eventos y fueron analizados en una sola oportunidad para obtener modelos de regresión.

De acuerdo al enfoque, la investigación es de tipo cuantitativa, la información corresponde a datos numéricos que permiten establecer relaciones existentes entre las variables. Según la modalidad, la investigación es bibliográfica y de campo. En cuanto a la obtención de los datos de la simulación de impacto vehicular se lo realizó mediante el uso de un software CAD-CAE.

Muestra y variables.

La muestra corresponde a los datos disponibles de las variables involucradas, conforme indica la Tabla 2:

Simulación de impacto vehicular	Velocidad (km/h)	Numérica discreta	Regresora	51
	Fuerza (N)	Numérica continua	Regresora y Respuesta	51
	FDS	Numérica continua	Respuesta	51
	Tiempo de impacto (s)	Numérica continua	Respuesta	51
	Deformación (mm)	Numérica continua	Respuesta	48

Tabla 2. Muestra utilizada en el estudio.

Recolección de información.

Los datos de las variables de la Tabla 2 se generaron en un software CAD-CAE a partir de la simulación del choque de un automóvil contra la parte trasera de la carrocería de un autobús a diferentes velocidades del auto pequeño.

Los valores obtenidos fueron particulares para las condiciones del problema (materiales de automotores y velocidades a las que se producen los impactos). La estructura del autobús fue diseñada con una sección tubular cuadrada de acero galvanizado con dimensiones de 40 mm de largo y 3 mm de espesor.

A continuación, en la Figura 2 se presenta una captura de pantalla de la simulación realizada:

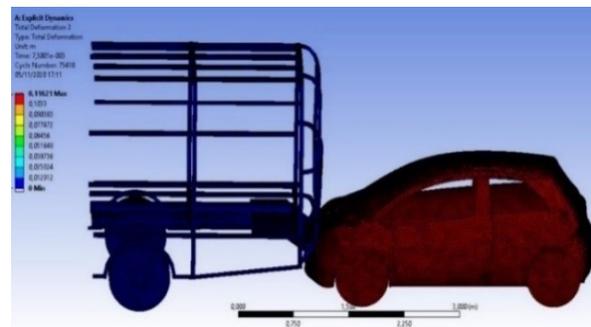


Figura 2. Simulación del impacto del vehículo contra la carrocería del autobús.

Procesamiento de la información y análisis.

Tipos de datos

Los datos son de tipo numérico discreto o numérico continuo. Se aclara que la velocidad de desplazamiento del auto, se considera como una variable numérica discreta y no continua, dado que se planteó una situación en la que se simularon impactos a diferentes velocidades de choque, con incrementos de una unidad dentro del rango 45 a 95 km/h.

Depuración y limpieza de los datos

Se efectuó una clasificación, depuración y limpieza de los datos, mediante utilización de una hoja de cálculo y el software R.

Preparación del código

Se realizó la programación de un script con una codificación que establece los modelos de regresión polinomial y B-splines, tomando en cuenta el lenguaje de código correspondiente y la teoría estadística inherente a los modelos.

Generación del modelo

Se determinaron los modelos de regresión polinomial y B-splines que se ajustaron a los datos disponibles en cada caso mediante el software R (15). Para comparar la bondad de ajuste de los dos modelos obtenidos para cada relación se efectuaron las siguientes actividades:

- Aplicación de la prueba de Wilcoxon para identificar si existen diferencias significativas entre ambos modelos (polinomial y B-splines), con un nivel de significancia de 0.05.
- Generación de un gráfico Q-Q Plot para identificar las diferencias entre los intervalos de confianza de ambos modelos y corroborar los resultados de la prueba de Wilcoxon (16).
- Elaboración de los diagramas de cajas de la variabilidad de la longitud del intervalo de confianza de los dos modelos.
- Construcción de los dos gráficos (modelo polinomial y B-spline) de la dispersión de puntos, curvas e intervalos de confianza en una gráfica.
- Decisión del modelo más idóneo para expresar la relación entre las variables. Para determinar el grado de la regresión polinomial se aplicó la prueba de diferencias significativas ANOVA entre modelos de distinto grado, priorizando por simplicidad el modelo de menor grado entre aquellos que no tienen diferencias significativas. Por su parte, en cuanto a los modelos de regresión no paramétricos B-splines, se evaluó la bondad de ajuste mediante el criterio de información de Deviance (DIC) (17) y el criterio de información de Akaike (AIC) (18).
- Declaración de la aceptación o rechazo de los modelos de regresión obtenidos, con base en las pruebas de hipótesis de validación de los modelos.

Supuestos del modelo de regresión paramétrico polinomial

Los supuestos verificados para los modelos de regresión paramétricos polinomiales fueron:

- Número de observaciones n es mucho mayor que el de parámetros β .
- Homocedasticidad, que implica que la esperanza matemática del error sea 0. $E(\epsilon)=0$ con varianza constante, a través de aplicar la prueba de Breusch-Pagan estudiantilizada (19).

- Residuos del modelo no están correlacionados entre sí, mediante la prueba de Durbin-Watson (20).

Pruebas de hipótesis de validación de los modelos.

Para la determinación de la validez y la bondad de ajuste de los modelos de regresión se aplicaron las siguientes pruebas de hipótesis:

- Prueba de rechazo de la nulidad de los parámetros β 's de cada modelo polinomial y de los coeficientes del modelo B-spline, mediante el estadístico t de Student: $H_0: \beta_1=\beta_2=\dots=\beta_k=0$; $H_1: \beta_j \neq 0$ para al menos un j .
- Prueba de diferencias significativas para la validez de los modelos, mediante la prueba F de Snedecor presentada en la tabla ANOVA: $F > F_{(\alpha, p-1, n-p)}$
- Pruebas de Shapiro-Wilk y Kolmogorov-Smirnov corregida por Lilliefors para comprobar la distribución normal de los residuos de los modelos de regresión $\epsilon \sim N(\mu, \sigma^2)$ (21) (22) (23).
- Pruebas de Durbin-Watson para la independencia de las variables (no autocorrelación) y prueba de homocedasticidad de Breusch-Pagan.
- Se omite un estudio de colinealidad puesto que todos los modelos tienen una sola variable explicativa.

III. RESULTADOS

Los resultados se presentan en tablas que detallan la información de las variables explicativas y regresoras, número de datos procesados, tipo de modelo de regresión, grado del polinomio, número de vértices del polígono de control (modelo B-spline), intercepto y coeficientes del modelo, prueba de hipótesis F, p-valor, RSE, R2 ajustado, MSE, SSE, RSME, SST; pruebas de normalidad de residuos, residuos no correlacionados, homocedasticidad de residuos, diferencias entre modelos, y códigos de significancia.

Todas las tablas se subdividen verticalmente en dos partes, una para el modelo de regresión polinomial y otra para el B-spline, a fin de comparar las características de ambos modelos e identificar la bondad de ajuste respectiva.

Velocidad vs Fuerza

En la Figura 3 se muestra la gráfica de dispersión de puntos de la velocidad versus la fuerza de

impacto, conjuntamente con las curvas de los dos modelos de regresión y las líneas de los intervalos de confianza al 95%:

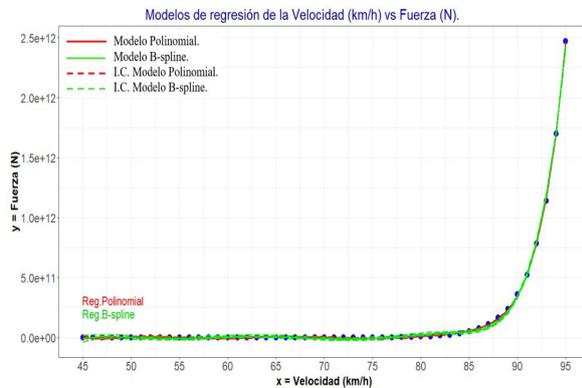


Figura 3. Modelos de regresión polinomial y B-spline de la velocidad vs fuerza.

En la Tabla 3 se presentan los modelos de regresión polinomial y B-spline de la variable velocidad versus la fuerza:

Variables de la regresión			
Tipo de variable		Denominación (Unidad de medida)	
Explicativa o regresor: Respuesta:		Velocidad (km/h)	
		Factor de seguridad FDS (adimensional)	
Número de datos n:		51	
Modelo de regresión Polinomial		Modelo de regresión B-spline	
Grado:	5	Grado:	3 (Cúbico)
		Vértices del polígono de control:	5
Intercepto β_0	-258.6344	Intercepto a_0	33.57894
β_1	21.7104	a_1	-32.38416
β_2	-0.7155	a_2	-32.64817
β_3	0.01167	a_3	-32.84946
β_4	-0.000095	a_4	-28.93164
β_5	3.11×10^{-7}	a_5	-17.16638
F	1537000	F	362900
p-valor	0.000***	p-valor	0.000***
RSE	0.02288	RSE	0.04572
R ² ajustado	0.9999935	R ² ajustado	0.9999724
MSE	0.0004937	MSE	0.002091
SSE	0.0222152	SSE	0.09407996
RSME	0.0222187	RSME	0.0457238
SST	3793.162	SST	3793.162
MAPE	1.00%	-	-
Test de normalidad de los residuos:			
Prueba de Shapiro-Wilk			
W:	0.95485	W:	0.97033
p-valor:	0.05038	p-valor:	0.2282
Prueba de Kolmogorov-Smirnov corregida por Lilliefors			

D:	0.074616	D:	0.095047
p-valor:	0.6809	p-valor:	0.2983
Test de residuos no correlacionados:			
Prueba de Durbin-Watson			
D-W:	0.8168112	D-W:	0.3692189
p-valor:	0.06107	p-valor:	0.02033*
Test de Homocedasticidad de residuos:			
Prueba de Breusch-Pagan estudiantilizada			
BP:	11.256	BP:	5.0543
p-valor:	0.04654*	p-valor:	0.4093
Prueba de Wilcoxon para comprobar diferencias entre las distancias de los IC de los modelos:			
W	82	p-valor	3.58×10^{-16} ***

Tabla 3. Modelos de regresión de la velocidad y fuerza. Códigos de significancia: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1.

La distribución de los puntos sigue una curva en forma de parábola creciente; conforme se incrementa la velocidad del auto, la fuerza del impacto se eleva ostensiblemente, por ello la variable respuesta está en escala logarítmica. Ambos modelos son válidos y presentan intervalos de confianza estrechos, no obstante, el modelo de regresión polinomial es el que ofrece una mejor bondad de ajuste, ya que los residuos son menores que los del modelo B-spline. Es decir que la relación de la variable velocidad y la variable fuerza con las condiciones dadas del problema se expresa mejor mediante la regresión paramétrica polinomial de grado 8.

Velocidad vs FDS

En la Figura 4 se muestra la gráfica de dispersión de puntos de la velocidad versus el FDS, conjuntamente con las curvas de los dos modelos de regresión y las líneas de los intervalos de confianza al 95%:

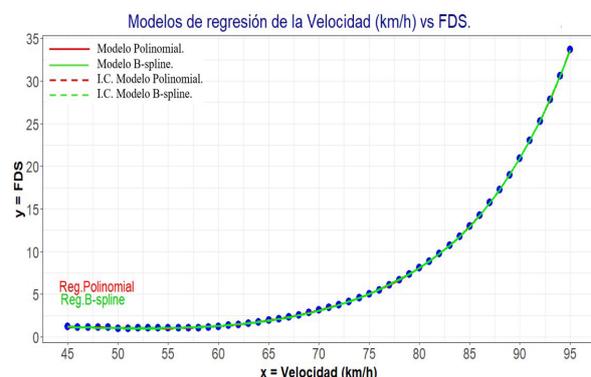


Figura 4. Modelos de regresión polinomial y B-spline de la velocidad vs FDS.

En la Tabla 4 se presentan los modelos de regresión polinomial y B-spline de la variable velocidad en km/h versus la variable FDS:

Variables de la regresión			
Tipo de variable		Denominación (Unidad de medida)	
Explicativa o regresor: Respuesta:		Velocidad (km/h)	
		Factor de seguridad FDS (adimensional)	
Número de datos n:		51	
Modelo de regresión Polinomial		Modelo de regresión B-spline	
Grado:	5	Grado:	3 (Cúbico)
		Vértices del polígono de control:	5
Intercepto β_0	-258.6344	Intercepto a_0	33.57894
β_1	21.7104	a_1	-32.38416
β_2	-0.7155	a_2	-32.64817
β_3	0.01167	a_3	-32.84946
β_4	-0.000095	a_4	-28.93164
β_5	3.11×10^{-7}	a_5	-17.16638
F	1537000	F	362900
p-valor	0.000***	p-valor	0.000***
RSE	0.02288	RSE	0.04572
R ² ajustado	0.9999935	R ² ajustado	0.9999724
MSE	0.0004937	MSE	0.002091
SSE	0.0222152	SSE	0.09407996
RSME	0.0222187	RSME	0.0457238
SST	3793.162	SST	3793.162
MAPE	1.00%	-	-
Test de normalidad de los residuos:			
Prueba de Shapiro-Wilk			
W:	0.95485	W:	0.97033
p-valor:	0.05038	p-valor:	0.2282
Prueba de Kolmogorov-Smirnov corregida por Lilliefors			
D:	0.074616	D:	0.095047
p-valor:	0.6809	p-valor:	0.2983
Test de residuos no correlacionados:			
Prueba de Durbin-Watson			
D-W:	0.8168112	D-W:	0.3692189
p-valor:	0.06107	p-valor:	0.02033*
Test de Homocedasticidad de residuos:			
Prueba de Breusch-Pagan estudiantilizada			
BP:	11.256	BP:	5.0543
p-valor:	0.04654*	p-valor:	0.4093
Prueba de Wilcoxon para comprobar diferencias entre las distancias de los IC de los modelos:			
W	82	p-valor	3.58×10^{-16} ***

Tabla 4. Modelos de regresión de la velocidad y FDS. Códigos de significancia: 0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 '.' 1.

La distribución de los puntos sigue una curva en forma de parábola creciente, conforme se incrementa la velocidad de movimiento del auto, el FDS de diseño requerido también crece. Comparativamente cuando la velocidad de desplazamiento del vehículo que impacta a la carrocería del autobús no excede de 60 km/h, el FDS de diseño aproximadamente no supera el valor de 2, pero a medida que se eleva la

velocidad el FDS requerido para que la estructura no se deforme está en torno a 35. Ambos modelos son válidos y tienen intervalos de confianza estrechos, no obstante, el modelo de regresión polinomial es el que ofrece una mejor bondad de ajuste, ya que los residuos son menores que los del modelo B-spline.

Es decir que la relación de la variable velocidad y la variable FDS con las condiciones dadas del problema se expresa de mejor manera con el uso de la regresión paramétrica polinomial de grado 5.

Velocidad vs Tiempo de impacto

En la Figura 5 se presenta la gráfica de dispersión de puntos de la velocidad versus el tiempo de impacto, conjuntamente con las curvas de los dos modelos de regresión y las líneas de los intervalos de confianza al 95%:

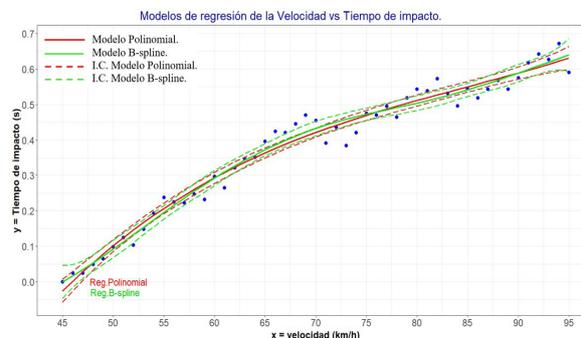


Figura 5. Modelos de regresión polinomial y B-spline de la velocidad vs tiempo de impacto.

En la Tabla 5 se presentan los modelos de regresión polinomial y B-spline de la variable velocidad en km/h versus el tiempo de impacto en segundos:

Variables de la regresión			
Tipo de variable		Denominación (Unidad de medida)	
Explicativa o regresor:		Fuerza (N)	
Respuesta:		Deformación (mm)	
Número de datos n:		48	
Modelo de regresión Polinomial		Modelo de regresión B-spline	
Grado:	8	Grado:	4
		Vértices del polígono de control:	6
Intercepto β_0	3.986846	Intercepto a_0	12.3073
β_1	8.577×10^{-10}	a_1	-11.5149
β_2	-3.331×10^{-20}	a_2	-7.5879
β_3	5.500×10^{-31}	a_3	-3.6727
β_4	-4.52×10^{-42}	a_4	4.4501
β_5	1.98×10^{-53}	a_5	-7.4979
β_6	-4.641×10^{-65}	a_6	0
β_7	5.421×10^{-77}	a_7	0
β_8	-2.45×10^{-89}	a_8	0
F	11.87	F	449.7
p-valor	2.23×10^{-8} ***	p-valor	0.000***
RSE	2.185	RSE	0.4832
R ² ajustado	0.649063	R ² ajustado	0.9828409
MSE	4.774958	MSE	0.2334722

SSE	186.2233	SSE	9.572361
RSME	2.185168	RSME	0.4831896
SST	639.4965	SST	639.4965
Test de normalidad de los residuos:			
Prueba de Shapiro-Wilk			
W:	0.96138	W:	0.95802
p-valor:	0.1148	p-valor:	0.08395
Prueba de Kolmogorov-Smirnov corregida por Lilliefors			
D:	0.066969	D:	0.10452
p-valor:	0.8524	p-valor:	0.2109
Test de residuos no correlacionados:			
Prueba de Durbin-Watson			
D-W:	0.09927003	D-W:	0.2574849
p-valor:	0.072152	p-valor:	0.049731*
Test de Homocedasticidad de residuos:			
Prueba de Breusch-Pagan estudiantilizada			
BP:	9.7254	BP:	9.5885
p-valor:	0.2848	p-valor:	0.1431
Prueba de Wilcoxon para comprobar diferencias entre las distancias de los IC de los modelos:			
W	2233	p-valor	0.000***

Tabla 5. Modelos de regresión de la velocidad y tiempo de impacto. Códigos de significancia: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1.

La distribución de los puntos sigue una curva con forma parecida a una función logarítmica.

Conforme se incrementa la velocidad de movimiento del auto también aumenta el tiempo de duración del impacto, aunque el ritmo de crecimiento va disminuyendo a medida que la velocidad se acerca a la máxima del ensayo de 95 km/h. Ambos modelos ofrecen una aceptable bondad de ajuste, aunque con intervalos de confianza que dejan por fuera a algunos puntos.

Los dos modelos son válidos, pero el modelo de regresión B-spline es el que ofrece una mejor bondad de ajuste, ya que los residuos son menores que los del modelo polinomial. Es decir que la relación de la velocidad y tiempo de duración del impacto con las condiciones dadas se expresa mejor a través de una regresión no paramétrica B-spline de grado 3.

Fuerza vs Deformación

En la Figura 6 se presenta la gráfica de dispersión de puntos de la fuerza versus la deformación se indica, conjuntamente con las curvas de los dos modelos de regresión y las líneas de los intervalos de confianza al 95%:

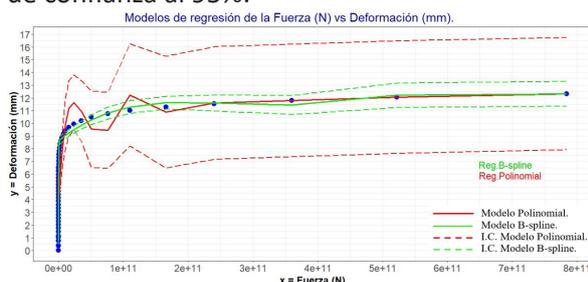


Figura 6. Modelos de regresión polinomial y B-spline de fuerza vs deformación.

En la Tabla 6 se presentan los modelos de regresión polinomial y B-spline de la variable fuerza en Newton versus la variable deformación en milímetros:

Variables de la regresión			
Tipo de variable		Denominación	
		(Unidad de medida)	
Explicativa o regresor:		Fuerza (N)	
Respuesta:		Deformación (mm)	
Número de datos n:		48	
Modelo de regresión Polinomial		Modelo de regresión B-spline	
Grado:	8	Grado:	4
		Vértices del polígono de control:	6
Intercepto	3.986846	Intercepto a_0	12.3073
β_0			
β_1	8.577×10^{-10}	a_1	-11.5149
β_2	-3.331×10^{-20}	a_2	-7.5879
β_3	5.500×10^{-31}	a_3	-3.6727
β_4	-4.52×10^{-42}	a_4	4.4501
β_5	1.98×10^{-53}	a_5	-7.4979
β_6	-4.641×10^{-65}	a_6	0
β_7	5.421×10^{-77}	a_7	0
β_8	-2.45×10^{-89}	a_8	0
F	11.87	F	449.7
p-valor	$2.23 \times 10^{-8***}$	p-valor	0.000***
RSE	2.185	RSE	0.4832
R ² ajustado	0.649063	R ² ajustado	0.9828409
MSE	4.774958	MSE	0.2334722
SSE	186.2233	SSE	9.572361
RSME	2.185168	RSME	0.4831896
SST	639.4965	SST	639.4965
Test de normalidad de los residuos:			
Prueba de Shapiro-Wilk			
W:	0.96138	W:	0.95802
p-valor:	0.1148	p-valor:	0.08395
Prueba de Kolmogorov-Smirnov corregida por Lilliefors			
D:	0.066969	D:	0.10452
p-valor:	0.8524	p-valor:	0.2109
Test de residuos no correlacionados:			
Prueba de Durbin-Watson			
D-W:	0.09927003	D-W:	0.2574849
p-valor:	0.072152	p-valor:	0.049731*
Test de Homocedasticidad de residuos:			
Prueba de Breusch-Pagan estudiantilizada			
BP:	9.7254	BP:	9.5885
p-valor:	0.2848	p-valor:	0.1431
Prueba de Wilcoxon para comprobar diferencias entre las distancias de los IC de los modelos:			
W	2233	p-valor	0.000***

Tabla 6. Modelos de regresión de la velocidad y tiempo de impacto. Códigos de significancia: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1.

La distribución de los puntos sigue una curva con forma parecida a una función logarítmica.

Conforme se incrementa la fuerza de impacto del auto sobre la carrocería del autobús también aumenta la deformación de la estructura, aunque el ritmo de crecimiento va disminuyendo a medida que la fuerza se hace más grande.

La curva del modelo polinomial no ofrece un ajuste aceptable, por lo que el modelo es rechazado, también se observa que el intervalo de confianza al 95% es demasiado ancho al mismo tiempo que la curva tiene aristas en forma de dientes de sierra. Por el contrario, la curva del modelo de regresión B-spline brinda un ajuste aceptable a pesar de que existen algunos puntos que quedan fuera de intervalo de confianza.

El modelo de regresión B-spline tiene una menor longitud del intervalo de confianza, lo que significa que tiene una mejor capacidad de minimizar los residuos. Por consiguiente, la relación de la variable explicativa fuerza de impacto y la variable respuesta deformación de la estructura tubular de acero de la carrocería del autobús, se expresa mejor a través de la regresión no paramétrica B-spline de grado 4.

IV DISCUSIÓN

Los resultados obtenidos se muestran en la Tabla 7. Se establecieron cuatro modelos de regresión simples polinomiales e igual número de modelos de regresión B-splines para el problema de la simulación de impacto de un vehículo en la parte posterior de la carrocería de un autobús.

En cuanto a la bondad de ajuste, el coeficiente de determinación R² ajustado en todos los casos analizados presentó un valor cercano a 1 y el test F de bondad de ajuste de la tabla ANOVA arrojó que los modelos son idóneos sin mayor complicación.

Problema	Variable	Aspecto	Modelo de regresión Polinomial	Modelo de regresión B-spline
Simulación de impacto vehicular entre un auto y un bus.	Velocidad vs Fuerza	Validez del modelo	✓	✓
		Grado	8	5, 7 v. p. c.
	Velocidad vs FDS	Validez del modelo	✓	✓
		Grado	5	3, 5 v. p. c.
	Velocidad vs Tiempo de impacto	Validez del modelo	✓	✓
		Grado	3	3, 6 v. p. c.
	Fuerza vs Deformación	Validez del modelo	x	✓
		Grado	8	4, 6 v. p. c.

Tabla 7. Resumen general para comparación de los dos modelos de regresión.

Tabla 7.

✓ Modelo válido.

x Modelo no válido.

v. p. c. vértices del polígono de control.

Sombreado en color azul el modelo más apropiado en cada caso.

ramétricos polinomiales se ajustan bien cuando las curvas tienen forma parabólica o siguen un patrón sin cambios abruptos de curvatura. Mientras que los modelos de regresión no paramétricos B-spline brindan un mejor ajuste cuando las curvas tienen forma de campana con cambios de curvatura más abruptos. Atendiendo al principio de simplicidad de los modelos y que éstos no afecten el suavizamiento del modelo, se consideraron modelos polinomiales de hasta grado 8 y B-splines de hasta grado 5 con 7 vértices de control.

En la práctica no se suelen establecer modelos de regresión con polinomios de grado elevado, debido a la complejidad para la obtención de los valores de la variable respuesta a partir de los valores de la variable regresora.

Una desventaja de obtener modelos complejos es que no resulta factible determinar una relación de crecimiento directa o inversamente proporcional entre las variables.

Por este motivo los modelos de regresión estudiados pueden resultar útiles para la predicción de los valores de las variables respuesta, pero no para resolver problemas de clasificación que son de interés en el ámbito de aprendizaje automático, que suelen establecerse con regresiones lineales (24).

La regresión polinomial tiene una estructura más sencilla, dado que consiste en una expresión algebraica con la variable elevada a distintos exponentes en forma progresiva, ajustada con el uso de coeficientes (25).

De su parte la regresión B-spline está estructurada por una base formada por curvas splines que también tienen polinomios y en los que es necesario establecer un determinado número de vértices del polígono de control (26).

Se debe destacar la utilidad práctica de la aplicación de los modelos de regresión en la simulación de impacto como parte del diseño de la estructura de la carrocería de autobuses, se detectó que la fuerza del impacto crece ostensiblemente a medida que se va incrementando la velocidad del vehículo que está en movimiento, el FDS de diseño también aumenta con la velocidad al igual que el tiempo de impacto, pero de forma más lenta.

El FDS cuanto mayor es, implica que se necesitará sobredimensionar el diseño, lo que implica utilizar un espesor mayor para el diseño de la estructura del autobús (4). En cuanto a la relación de la fuerza de impacto con la deformación de la estructura del autobús se observó que a partir de la magnitud de fuerza de 1011 N la deformación tendió a ser constante.

En cuanto a investigaciones similares desarrolladas acerca de la relación entre las magnitudes velocidad y tiempo de impacto en las colisiones vehiculares, se puede destacar la llevada a cabo por parte de Mastandrea y Vangi. Los investigadores hallaron una relación directamente proporcional entre la velocidad y el tiempo de duración del impacto en un vehículo golpeado por otro, aunque los rangos de velocidades fueron menores a los del presente estudio, sin embargo, la curva muestra una tendencia creciente similar a la de la Figura 5, lo que corrobora la relación existente entre ambas magnitudes (27).

V. CONCLUSIONES

Los modelos estadísticos de regresión paramétrica polinomial y no paramétrica B-splines tienen sus propias particularidades. En el primer caso, demanda el cumplimiento de supuestos para los residuos: distribución normal, no autocorrelación y homocedasticidad.

La estructura de un modelo de regresión polinomial consiste en una expresión algebraica con la variable elevada a exponentes consecutivos de entre los cuales el mayor determina el grado del polinomio y para el ajuste se emplean diferentes coeficientes para cada término.

La regresión B-spline está estructurada por una base formada por curvas splines y en los que es necesario establecer un determinado número de vértices del polígono de control.

En el fenómeno simulado se estableció como variable regresora a la velocidad del auto pequeño en el momento de la colisión, cuyos valores oscilaron entre 45 y 95 km/h.

Según la distribución de los puntos de la velocidad

versus la fuerza, son directamente proporcionales, la elevación de la velocidad provocó que la fuerza tienda a incrementarse ostensiblemente, con un rango de valores fluctuantes entre 11935 y 2.47×10^{12} N.

La distribución de los puntos de la velocidad versus el FDS y de la velocidad vs el tiempo de impacto también expresaron relaciones directamente proporcionales, siendo que los rangos de datos fluctuaron entre 1 y 33.66 para el FDS, y entre 0 y 0.67 segundos para el tiempo.

La distribución de puntos de la fuerza de impacto versus la deformación en la estructura es directamente proporcional, aunque la deformación tendió a presentar incrementos menores conforme se eleva la fuerza de impacto. El rango de datos de deformación osciló entre 0 y 13.11 mm.

La relación de la velocidad y la fuerza se explica de manera óptima mediante un modelo de regresión polinomial de grado 8. La relación de la velocidad y el FDS mediante un modelo de regresión polinomial de grado 5. La relación de la velocidad y el tiempo de impacto se explica con un modelo de regresión polinomial cúbico. La fuerza y deformación se relacionan con un modelo de regresión B-spline de grado 4, con 6 vértices del polígono de control.

Los modelos de regresión paramétricos polinomiales se ajustan bien cuando las curvas tienen forma parabólica o siguen un patrón sin cambios abruptos de curvatura, adaptándose mejor a las relaciones de la simulación de impacto vehicular. Por su parte, los modelos de regresión no paramétricos B-splines brindan un mejor ajuste cuando las curvas tienen forma de campana con cambios de curvatura más abruptos.

En el diseño de las carrocerías de autobuses se recomienda hacer ensayos experimentales en diferentes escenarios, tanto para los vehículos que como para varios materiales constructivos. Adicionalmente se sugiere desarrollar investigaciones destinadas a comparar los dos modelos aquí estudiados, pero destinados a otras aplicaciones, para corroborar los resultados obtenidos en este trabajo.

VI. REFERENCIAS

1. Jurewicz C, Sobhani A, Woolley J, Dutschke J, Corben B. Exploration of Vehicle Impact Speed – Injury Severity Relationships for Application in Safer Road Design. *Transportation Research Procedia*. 2016;14:4247-56.

2. Moreno-Zulca PA, Llanes-Cedeño EA, Guaña-Fernández WV, Jima-Matailo JC. Análisis estructural de un bus por el método de elementos finitos. *Polo del Conocimiento*. 2020;5(1):799-837.
3. Portillo M, Chacón R, Moreno M, Bongiorno F. Simulación y análisis de una prueba de choque de un automóvil tipo deportivo, utilizando un software basado en el método de los elementos finitos. *Revista Ciencia e Ingeniería*. 2011;32(1):69-77.
4. Cárdenas D, Escudero J, Quizhpi S, Amaya M. Propuesta de diseño estructural para buses de carrocería interprovincial. *Ingenius Revista de Ciencia y Tecnología*. 2014;11(1):42-52.
5. Pikūnas A, Pumputis V, Sadauskas V. The influence of vehicles speed on accident rates and their consequences. *Transport*. 2004;19(1):15-25.
6. Stimson JA, Carmines EG, Zeller RA. Interpreting Polynomial Regression. *Sociological Methods & Research*. mayo de 1978;6(4):515-24.
7. Meilán-Vila A, Opsomer J, Francisco-Fernández M, Crujeiras R. A goodness-of-fit test for regression models with spatially correlated errors. *TEST*. 2020;29(3):728-49.
8. Ostertagová E. Modelling using Polynomial Regression. *Procedia Engineering*. 2012;48:500-6.
9. Montgomery DC, Runger GC. *Applied statistics and probability for engineers*. 3.a ed. New York, USA: John Wiley & Sons, Inc.; 2003. 976 p.
10. Aparicio J, Martínez M, Morales J. Modelos lineales aplicados en R [Internet]. 1.a ed. Elche, España; 2006 [citado 29 de abril de 2021]. 215 p. Disponible en: <https://umh3067.edu.umh.es/wp-content/uploads/sites/240/2013/02/Modelos-Lineales-Aplicados-en-R.pdf>
11. Racine JS. The crs Package [Internet]. 2021 [citado 3 de abril de 2021]. Disponible en: <https://cran.r-project.org/web/packages/crs/vignettes/crs.pdf>
12. Paluszny M, Prautzsch H, Böhm W. *Métodos de Bézier y B-splines*. Karlsruhe: Univ.-Verl; 2005. 303 p.
13. Racine JS. *A primer on Regression Splines*. 2019.
14. Yli-Heikkilä M. *Data Science Languages*. 2019;2.
15. Garibaldi LA, Oddi FJ, Aristimuño FJ, Behnisch AN. Modelos estadísticos en lenguaje R [Internet]. 1.a ed. Buenos Aires, Argentina: UNRN; 2019. 261 p. Disponible en: https://rid.unrn.edu.ar/bitstream/20.500.12049/5789/2/garibaldi_lenguajeR_eunrn.pdf
16. Woolson RF. Wilcoxon signed-Rank test. En: *Wiley encyclopedia of clinical trials* [Internet]. Wiley; 2007. p. 1-3. Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780471462422.eoct979>
17. Meyer R. Deviance Information Criterion (DIC). En: *Wiley StatsRef: Statistics Reference Online* [Internet]. 1.a ed. Wiley; 2016 [citado 7 de mayo de 2022]. p. 1-6. Disponible en: <https://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat07878>
18. Cavanaugh JE, Neath AA. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Comp Stat*. 2019;11(3):1-11.
19. Koenker R. A note on studentizing a test for heteroscedasticity. *Journal of Econometrics*. 1981;17(1):107-12.
20. Tillman JA. The Power of the Durbin-Watson Test. *Econometrica*. 1975;43(5/6):959-74.
21. Yap BW, Sim CH. Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*. 2011;81(12):2141-55.
22. Pedrosa I, Juarros-Basterretxea J, Robles-Fernández A, Basteiro J, García-Cueto E. Pruebas de bondad de ajuste en distribuciones simétricas, ¿qué estadístico utilizar? *Univ Psychol*. 2014;14(1):245-54.
23. Saculinggan M, Balase EA. Empirical Power Comparison of Goodness of Fit Tests for Normality in the Presence of Outliers. *J Phys: Conf Ser*. 2013;435:1-11.
24. Pečkov A. A machine learning approach to polynomial regression [Internet] [Tesis Doctoral]. [Ljubljana, Slovenia]: Jožef Stefan International Postgraduate School; 2012. Disponible en: http://kt.ijs.si/theses/phd_aleksandar_peckov.pdf
25. Rencher AC, Schaalje GB. *Linear models in Statistics*. 2nd ed. Hoboken, N.J: Wiley-Interscience; 2008. 672 p.
26. Kaňka M. *Segmented Regression Based on B-Splines with Solved Examples*. 2015;20.
27. Mastandrea M, Vangi D. Influence of braking force in low-speed vehicle collisions. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*. 2005;219(2):151-64.