



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



Tecnológico Nacional de México

**Centro Nacional de Investigación
y Desarrollo Tecnológico**

Tesis de Doctorado

**Metodología para el desarrollo de modelos de
comportamiento de sistemas complejos
altamente dinámicos**

Presenta

MTI. Ernesto de la Cruz Nicolás

como requisito para la obtención del grado de

Doctor en Ciencias de la Computación

Director de tesis

Dr. Hugo Estrada Esquivel

Codirector de tesis

Dra. Alicia Martínez Rebollar

Cuernavaca, Morelos, México. Agosto de 2025.

	ACEPTACIÓN DE IMPRESIÓN DEL DOCUMENTO DE TESIS DOCTORAL		Código: CENIDET-AC-006-D20
	Referencia a la Norma ISO 9001:2008 7.1, 7.2.1, 7.5.1, 7.6, 8.1, 8.2.4		Revisión: 0
			Página 1 de 1

Cuernavaca, Mor., a 26 de mayo de 2025

DR. CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO
P R E S E N T E

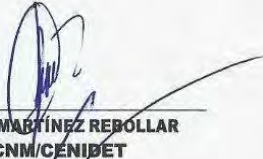
AT'm: DR. JUAN GABRIEL GONZÁLEZ SERNA
PRESIDENTE DEL CLAUSTRO DOCTORAL DEL
DEPARTAMENTO DE CIENCIAS COMPUTACIONALES

Los abajo firmantes, miembros del Comité Tutorial del estudiante **M.C. ERNESTO DE LA CRUZ NICOLAS** manifiestan que después de haber revisado el documento de tesis titulado **"METODOLOGÍA PARA EL DESARROLLO DE MODELOS DE COMPORTAMIENTO DE SISTEMAS COMPLEJOS ALTAMENTE DINÁMICOS"**, realizado bajo la dirección del **Dr. Hugo Estrada Esquivel** y codirección de la **Dra. Alicia Martínez Rebollar**, el trabajo se **ACEPTA** para proceder a su impresión.

ATENTAMENTE

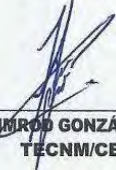
Excelencia en Educación Tecnológica®
"Conocimiento y Tecnología al Servicio de México"


DR. HUGO ESTRADA ESQUIVEL
TECNM/CENIDET


DRA. ALICIA MARTÍNEZ REBOLLAR
TECNM/CENIDET


DR. EDDIE HELBERT CLEMENTE TORRES
TECNM/CENIDET


DR. JAVIER ORTIZ HERNÁNDEZ
TECNM/CENIDET


DR. NIROD GONZÁLEZ FRANCO
TECNM/CENIDET


DR. NOEL GARCÍA DÍAZ
TECNM/ IT. COLIMA

c.c.p: C Verónica Sotelo Boyas / Jefa del Departamento de Servicios Escolares
c.c.p: C Née Alejandro Castro Sánchez / Jefe del Departamento de Ciencias Computacionales
c.c.p: Expediente



Educación
Secretaría de Educación Pública



TECNOLÓGICO
NACIONAL DE MÉXICO



Centro Nacional de Investigación y Desarrollo tecnológico
Subdirección Académica

Cuernavaca Mor, 04/junio/2025

Oficio No. SAC/120/2025

Asunto: Autorización de impresión de tesis

ERNESTO DE LA CRUZ NICOLAS
CANDIDATO AL GRADO DE DOCTOR
EN CIENCIAS DE LA COMPUTACIÓN
P R E S E N T E

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"METODOLOGÍA PARA EL DESARROLLO DE MODELOS DE COMPORTAMIENTO DE SISTEMAS COMPLEJOS ALTAMENTE DINÁMICOS"**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

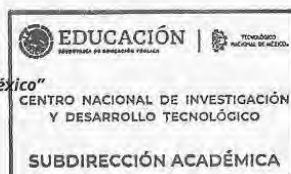
Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

ATENTAMENTE

Excelencia en Educación Tecnológica®

"Conocimiento y Tecnología al Servicio de México"

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO



c.c.p. Departamento de Ciencias Computacionales
Departamento de Servicios Escolares

CMAZ/lmz



2025
Año de
La Mujer
Indígena

Interior Internado Palmira S/N, Col. Palmira,
C. P. 62490, Cuernavaca, Morelos Tel. 01 (777) 3627770, ext. 4104,
e-mail: acad_cenidet@tecnm.mx tecnm.mx | cenidet.tecnm.mx



Dedicatoria

A Dios, por darme salud, marcar el ritmo de este camino y dar sentido a cada paso.

A Alejandra, por ser hogar, calma y fuerza en medio de todo.

A mi familia, por su amor callado, que nunca dejó de sostenerme.

A mis maestros, quienes fueron el punto de partida en cada etapa de mi formación académica. En especial, a la maestra Lorena, al maestro Arturo, a la maestra Juanita y a la doctora Alicia, por dejar huella en mi camino y enseñarme mucho más que contenidos: el valor de aprender con sentido.

A quienes creen, incluso cuando uno duda.

Y al Instituto Tecnológico de Cuautla, por haberme abierto las puertas al inicio de mi carrera profesional y por brindarme la oportunidad de continuar mi formación a nivel doctoral en un entorno que siempre me impulsó a crecer.

Agradecimientos

En primer lugar, expreso mi más profundo agradecimiento a mi Dios por haberme otorgado la salud, la sabiduría y la serenidad necesarias para enfrentar los momentos de frustración, desesperación y desafíos emocionales a lo largo de este camino. Su presencia me acompañó en cada etapa y fue fundamental para alcanzar la culminación de este sueño.

A mi esposa, Odette Alejandra Pliego Martínez, gracias por estar conmigo en cada etapa de este camino. Alejandra, tu paciencia en cada conversación, en los trayectos diarios y en las noches de desvelo, fue un refugio. Nunca te quejaste, siempre estuviste presente con cariño y apoyo. Tu ayuda en matemáticas, tu compañía constante y tu confianza en mí hicieron posible que no abandonara este proyecto. Durante años postergué el doctorado; hoy sé que debía vivirlo contigo. Gracias por comprenderme y por sostenerme. Este logro también te pertenece.

Agradezco de todo corazón a mi mamá, quien fue la primera en insistir, con amor y determinación, en que asistiera a la escuela. Su decisión marcó el inicio de este camino. A mi papá, quien, sin haber tenido la oportunidad de ir a la escuela, fue mi primer maestro. Me enseñó a leer, escribir y hacer operaciones matemáticas, conocimientos con los que ingresé directamente al tercer año de primaria. Gracias a ambos por seguir apoyándonos hasta hoy: por prepararnos el lunch, por estar siempre pendientes, por no mantenerse al margen y por sostener a nuestra familia con consejos y amor. Sin ustedes, este logro no habría sido posible. A mis hermanas, María de Jesús y Refugia, gracias por recibirme siempre con una sonrisa y brindarme su apoyo incondicional, sin condiciones ni cuestionamientos. Su cariño me motiva a ser un mejor hermano cada día. A mis sobrinas Dalia y Ohana, y a mi sobrino Carlos, gracias por contagiarme su alegría y por hacerme sentir parte importante de sus vidas. Gracias por darme el honor de llamarme su “papá Neto”.

A la Dra. Alicia y al Dr. Hugo, mis directores de tesis, les expreso mi más sincero reconocimiento y profundo agradecimiento por su inquebrantable guía, su paciencia generosa y sus observaciones siempre lúcidas y pertinentes. Su acompañamiento constante, la sabiduría compartida a lo largo de este camino, y el respaldo académico que en todo momento me brindaron, han sido pilares fundamentales en la consolidación de este trabajo. Este logro es, en gran parte, testimonio de su vocación formativa y de su calidad humana.

Quiero expresar mi más profundo agradecimiento al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) por el invaluable apoyo brindado a través de la beca otorgada, la cual hizo posible la realización de mis estudios de posgrado y el desarrollo de este trabajo de investigación. De igual manera, agradezco al Tecnológico Nacional de México (TecNM) por proporcionar las instalaciones, recursos tecnológicos y académicos que fueron fundamentales para llevar a cabo este proyecto.

A mis amigos, profesores, compañeros y colegas, por sus aportaciones, discusiones técnicas y colaboración constante, que enriquecieron significativamente esta experiencia académica.

Resumen

El presente trabajo propone una metodología para el desarrollo de modelos de comportamiento en sistemas complejos altamente dinámicos, basada en el análisis de grandes volúmenes de datos. Esta metodología tiene como objetivo proporcionar un marco de trabajo estructurado, que integre fases y métodos orientados al análisis y modelado de sistemas caracterizados por interacciones no lineales, adaptabilidad y fenómenos emergentes. Estos sistemas, están presentes en ámbitos tan diversos como la biología, economía, ingeniería, ecología, sustentabilidad, contaminación y movilidad, entre otros; y requieren un enfoque sistemático que permita su comprensión y representación precisa y eficiente.

Para construir un modelo de comportamiento, es esencial comenzar con la identificación del problema, abordándolo desde una perspectiva de sistema complejo. Esto implica definir claramente los límites del sistema, sus componentes e interacciones. A continuación, se procede a la recolección de datos relevantes que permitan analizar su comportamiento histórico.

El modelado puede abordarse mediante distintos enfoques, tales como la simulación basada en agentes, métodos numéricos, cómputo evolutivo entre otros, que representa a los individuos y sus interacciones; el uso de modelos matemáticos que describen las dinámicas del sistema; o aproximaciones basadas en patrones de comportamiento, entre otras.

Una vez construido un modelo inicial, se realiza un proceso de validación y calibración, comparando las predicciones del modelo con datos reales para asegurar que refleja adecuadamente el comportamiento del sistema. La iteración continua entre modelado y validación resulta relevante para afinar el modelo y mejorar su precisión.

La metodología también contempla la exploración de escenarios, en la cual se analizan diversas condiciones y perturbaciones para observar la respuesta del sistema. Esta etapa es especialmente valiosa para la toma de decisiones y la planificación estratégica, ya que permite anticipar posibles resultados y evaluar riesgos de forma informada.

La comunicación de los resultados constituye un componente importante. Los modelos deben ser presentados con claridad y accesibilidad, de modo que los actores involucrados puedan interpretar los hallazgos y aplicarlos eficazmente.

Palabras clave: Sistemas complejos, Sistemas complejos altamente dinámicos, Datos en tiempo real, Movilidad vehicular, Índice de movilidad, Cómputo evolutivo, Metodología.

Abstract

The methodology for developing behavioral models of highly dynamic complex systems aims to provide a structured framework that integrates phases and methods focused on the analysis and modeling of systems characterized by nonlinear interactions, adaptability, and emergent phenomena. These systems, present in diverse fields such as biology, economics, engineering, ecology, sustainability, pollution, and mobility, among others, require a systematic approach that enables their accurate and efficient understanding and representation.

To construct a behavioral model, it is essential to begin with the identification of the problem, approaching it from a complex systems perspective. This involves clearly defining the system's boundaries, its components, and their interactions. Subsequently, relevant data must be collected to analyze the system's historical behavior.

Modeling can be approached through various methods, such as agent-based simulation, numerical methods, evolutionary computation, among others, which represent individuals and their interactions; the use of mathematical models that describe system dynamics; or behavioral pattern-based approaches.

Once an initial model has been developed, a validation and calibration process are carried out by comparing the model's predictions with real-world data to ensure that it accurately reflects the behavior of the system. Continuous iteration between modeling and validation is key to refining the model and improving its accuracy.

The methodology also includes scenario exploration, in which various conditions and perturbations are analyzed to observe the system's response. This stage is particularly valuable for informed decision-making and strategic planning, as it allows for the anticipation of possible outcomes and risk assessment.

Finally, communicating the results is a fundamental component. Models must be presented in a clear and accessible manner so that stakeholders can interpret the findings and apply them effectively.

Keywords: Complex systems, Highly dynamic complex systems, Real-time data, Vehicular mobility, Mobility index, Evolutionary computation, Methodology.

Contenido

Dedicatoria.....	i
Agradecimientos.....	ii
Resumen	iii
Abstract.....	iv
Lista de figuras	viii
Lista de tablas	xii
Capítulo 1. Introducción.....	1
1.1 Introducción	2
1.2 Motivación	3
1.3 Planteamiento del problema	4
1.4 Propuesta de solución.....	5
1.5 Objetivos de la Investigación.....	6
1.6 Pregunta de investigación.....	7
1.7 Hipótesis	8
1.8 Metodología de solución de la investigación	9
1.9 Contribuciones	12
1.10 Entorno de Investigación.....	12
1.11 Organización de la tesis.....	13
Capítulo 2. Marco Referencial.....	16
2.1 Introducción	17
2.2 Marco teórico	17
2.2.1 Teoría de sistemas.....	17
2.2.2 Teoría de redes.....	18
2.2.3 Redes complejas	19
2.2.4 Teoría de sistemas dinámicos	20
2.2.5 Sistemas complejos	21
2.2.6 Análisis de estabilidad y comportamiento no lineal	23
2.2.7 Sistemas complejos altamente dinámicos	23
2.2.8 Evolución histórica del modelado de sistemas complejos.....	24
2.2.9 Desarrollo de modelos matemáticos y ecuaciones diferenciales	25
2.3 Estado del arte	26

2.3.1 Enfoques y técnicas para el modelado del comportamiento en sistemas complejos	26
2.3.2 Metodologías para la construcción de modelos de sistemas complejos con enfoque en índices o indicadores	31
2.3.3 Metodologías de modelado y simulación para la construcción de modelos de comportamiento en sistemas complejos	36
2.3.4 Metodologías de minería de datos y ciencia de datos con el enfoque de construcción de modelos de comportamiento de sistemas complejos.....	37
2.3.5 Cómputo evolutivo para el modelado matemático con el enfoque de construcción de modelos de comportamiento de sistemas complejos.....	39
2.4 Conclusiones.....	40
Capítulo 3. Metodología propuesta para modelar sistemas complejos altamente dinámicos	42
3.1 Introducción	43
3.2 Vista general de la metodología propuesta para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos	43
3.2.1 Caso de estudio de Movilidad Vehicular	45
3.3 Análisis contextual y de la dinámica del sistema.....	47
3.3.1 Comprensión del sistema complejo dinámico de movilidad vehicular	48
3.3.2 Identificación de entidades y relaciones dinámicas de la movilidad vehicular	57
3.3.3 Definición de variables críticas del sistema complejo de movilidad vehicular	61
3.4 Agregación multinivel y medición de la dinámica compleja	68
3.4.1 Niveles de agregación del sistema complejo dinámico del sistema complejo dinámico de movilidad vehicular	69
3.4.2 Análisis de incertidumbre del sistema complejo dinámico del sistema complejo de movilidad vehicular	73
3.4.3 Análisis de la Identificación de las variables clave de acuerdo con la dinámica del sistema complejo de movilidad vehicular	75
3.5 Recopilación y Monitoreo Continuo de Datos en Tiempo Real	78
3.5.1 Recopilación continua de datos de movilidad vehicular	79
3.5.2 Integración de fuentes dinámicas de datos de movilidad vehicular.....	82
3.5.3 Diseño de sistemas de almacenamiento en tiempo real para los datos de movilidad vehicular	87
3.6 Preprocesamiento y actualización dinámica de datos.....	90
3.6.1 Preprocesamiento en tiempo real de datos de movilidad vehicular	91

3.6.2 Actualización y muestreo dinámico de los datos de movilidad vehicular	99
3.7 Modelado iterativo y desarrollo continuo	102
3.7.1 Desarrollo del modelo dinámico de movilidad vehicular	103
3.7.2 Evaluación y ajuste continuo del modelo de movilidad vehicular	110
3.8 Interpretación, representación e implementación continua	114
3.8.1 Representación dinámica del modelo de movilidad vehicular	115
3.8.2 Implementación adaptativa del modelo de movilidad vehicular	119
3.8.3 Actualización del modelo de movilidad vehicular	124
3.9 Conclusiones	126
Capítulo 4. Validación de la metodología	128
4.1 Introducción	129
4.2 Proceso de validación de la metodología propuesta para el desarrollo de modelos de comportamiento en sistemas complejos altamente dinámicos	130
4.3 Caso de estudio 1: Calidad de aire en la Ciudad de México	132
4.3 Caso de estudio 2: Profundidad de los sismos en México	155
4.5 Conclusiones	167
Capítulo 5. Prototipo de Movilidad vehicular en la Ciudad de México	170
5.1 Introducción	171
5.2 Prototipo de Movilidad Vehicular	171
5.3 Conclusiones	178
Capítulo 6. Resultados	179
6.1 Introducción	180
6.2 Desarrollo de la Metodología para la Modelación de Sistemas Complejos Altamente Dinámicos	180
6.3 Construcción del Índice de Movilidad Vehicular (IMV)	181
6.4 Desarrollo de Índices Complementarios: ICA e IPS	184
6.5 Desarrollo de un Prototipo Web para Evaluación de la Movilidad	185
6.6 Conclusiones	186
Capítulo 7. Conclusiones generales y trabajos futuros	187
7.2 Conclusiones generales	188
7.3 Trabajos futuros	190
7.4 Publicaciones	191
Referencias	193

Lista de figuras

Figura 1.1. Esquema metodológico propuesto para el modelado de sistemas complejos altamente dinámicos.....	6
Figura 1.2. Metodología de solución para el desarrollo de modelos de comportamiento en sistemas complejos altamente dinámicos.....	10
Figura 2.1. Un sistema que interactúa con su entorno	18
Figura 3.1 Metodología propuesta para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos.....	44
Figura 3.2. Fase 1 Análisis contextual y dinámicas del sistema de la metodología propuesta	47
Figura 3.3. Grafo de la movilidad vehicular de una calle a otra de cinco calles de la Ciudad de México.....	49
Figura 3.4. Simulación de la movilidad vehicular de los primeros 100 metros de una calle.....	49
Figura 3.5. Adaptabilidad de la movilidad vehicular	51
Figura 3.6. Representación gráfica de la no linealidad de la Movilidad Vehicular	52
Figura 3.7. Auto organización de la movilidad vehicular mediante el autómata celular	53
Figura 3.8. Valor de entropía de cada segmento de la calle, de acuerdo a las velocidades.....	54
Figura 3.9. Cascada de la congestión de Movilidad Vehicular en algunas zonas geográficas parte 1.....	55
Figura 3.10. Cascada de la congestión de Movilidad Vehicular en algunas zonas geográficas, parte 2.....	55
Figura 3.11. Metodología para la obtención de datos de tráfico.....	58
Figura 3.12. JSON de información de Movilidad vehicular proporcionado por TomTom Traffic.....	59
Figura 3.13. Movilidad vehicular en horario de 7 a 22 horas.....	64
Figura 3.14. Comportamiento de la movilidad vehicular en intervalos de tiempo.....	65
Figura 3.15. Gráfica de densidad de algunas variables de la entidad tráfico.....	66
Figura 3.16. Gráficas Q-Q para visualizar la normalidad de los datos de las variables.....	66
Figura 3.17. Fase 2 Agregación multinivel y medición de la dinámica compleja de la metodología propuesta.....	69
Figura 3.18. División política de la Ciudad de México	70
Figura 3.19. Diferentes niveles de agregación soportados por la aplicación de Here Maps.....	71

Figura 3.20. Conjunto de cuadros delimitadores que cubren a toda la Ciudad de México.....	71
Figura 3.21. Fase 3 Recopilación y monitoreo continuo de datos en tiempo real de la metodología propuesta.....	79
Figura 3.22. Velocidades de la Ciudad de México en los primeros 100 minutos	80
Figura 3.23. Extracción de datos tráfico e incidencias viales en tiempo real mediante un Script en Python	81
Figura 3.24. Esquema del proceso de integración, selección y modelado de datos para la generación de funciones a partir de información vial, de servicios y de tráfico en la Ciudad de México.....	83
Figura 3.25. Obtención de tamaño de muestra para el conjunto de datos de tráfico, incidencias y servicios.....	84
Figura 3.26. Cálculo de distancia entre 2 puntos utilizando ROUTING API de Geoapify	85
Figura 3.27. Distancias entre el tráfico con respecto a las incidencias y a los servicios	85
Figura 3.28. Proceso de Diseño de Sistemas de Almacenamiento en Tiempo Real para Datos de Movilidad Vehicular	89
Figura 3.29. Fase 4 Preprocesamiento y actualización dinámica de datos de la metodología propuesta.....	91
Figura 3.30. Tamaño de la muestra de Tráfico, incidencias y servicios	92
Figura 3.31. Visualización gráfica de la relación entre variables del conjunto de datos de Tráfico, Incidencias y Servicios con la congestión vehicular	93
Figura 3.32. Causalidad entre la variable Number_Segments con respecto a jamFactor	97
Figura 3.33. Tamaño de la muestra de Tráfico, incidencias y servicios con 21 variables	100
Figura 3.34. Fase 5 Modelado iterativo y desarrollo continuo de la metodología propuesta	102
Figura 3.35. Parámetros del programa evolutivo para su ejecución	104
Figura 3.36. Comportamiento de la movilidad vehicular utilizando el IM de una calle	107
Figura 3.37. Un cuadro delimitador de la delegación Benito Juárez de la Ciudad de México.....	108
Figura 3.38. Validación del IM con una red neuronal feedforward	111
Figura 3.39. Mapa cognitivo difuso de las variables del IM	112
Figura 3.40. Sensibilidad de las variables del IM.....	113

Figura 3.41. Fase 6 Interpretación, representación e implementación continua de la metodología propuesta.....	115
Figura 3.42. Fragmento de conjunto de datos de 1, 048, 575 registros, donde cada renglón tiene su valor de IM	116
Figura 3.43. Función de membresía de tipo gaussiano para obtener los valores de pertenencia	117
Figura 3.44. Importancia de las características en el modelo adaptativo.....	120
Figura 3.45. Comportamiento del jamFactor a lo largo de las observaciones	121
Figura 3.46. Comparación del Error Cuadrático Medio (MSE) entre el modelo adaptativo y el modelo estático	122
Figura 4.1 Validación de la metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos.....	130
Figura 4.2. Índice de calidad del aire (AQI).....	133
Figura 4.3. Conjunto de datos de calidad del aire	134
Figura 4.4. Resumen estadístico de las variables de calidad del aire	134
Figura 4.5. Matriz de correlación entre las variables de calidad del aire	135
Figura 4.6. Pairplot de las variables seleccionadas	136
Figura 4.7. Matriz de cargas de los componentes principales que muestra la contribución de cada variable original a los componentes principales	137
Figura 4.8. Resumen de los resultados del MANOVA	137
Figura 4.9. Mapa de calor que representa las medias estandarizadas de las variables del aire por grupo (A, B y C).....	138
Figura 4.10. Diagrama estructural del Modelo de Ecuaciones Estructurales (SEM) que ilustra las relaciones estimadas entre variables meteorológicas y contaminantes	139
Figura 4.11. Distribución de los coeficientes obtenidos mediante la regresión LASSO, destacando a PM2.5 como la variable de mayor importancia, seguida por PM10 y O3.....	140
Figura 4.12. Importancia relativa de las variables determinadas por el modelo Random Forest, en el que PM2.5 tiene la mayor influencia, seguida por PM10 y O3.....	141
Figura 4.13. Diagrama del flujo de procesos para el diseño de sistemas de almacenamiento en tiempo real	143
Figura 4.14. Mapa que muestra las estaciones de monitoreo de calidad del aire en México y otros países.....	144
Figura 4.15. Fragmento de código en Python que utiliza un endpoint de la API de AQICN para solicitar datos de calidad del aire en tiempo real	144

Figura 4.16. Estructura de los datos en formato JSON proporcionados por API de AQICN.....	145
Figura 4.17. Datos de calidad del aire en organizados bajo la estructura de MongoDB	146
Figura 4.18. Conjunto de datos de aire estandarizado entre 0 y 1	146
Figura 4.19. Comparación entre los valores reales del Índice de Calidad del Aire (ICA) y las predicciones generadas por el modelo de programación genética (GP)	150
Figura 4.20. Mapa de calor de correlaciones entre variables	152
Figura 4.21. Modelo adaptativo de calidad del aire	153
Figura 4.22. Evolución del modelo de calidad del aire en tiempo real.....	154
Figura 4.23. Monitoreo del Índice de Calidad del Aire en Plataforma Interactiva.....	154
Figura 4.24. Nivel de agregación por país para el análisis de sismos.....	157
Figura 4.25. Fragmento del conjunto de datos sísmicos utilizados en el estudio, recopilados por el Servicio Sismológico Nacional (SSN)	158
Figura 4.26. Análisis exploratorio de los datos sísmicos en México (1900-2024)	160
Figura 4.27. Correlación de variables del conjunto de datos de sismos.....	161
Figura 4.28. Matriz de Cargas de Componentes en el Análisis de Componentes Principales del conjunto de datos de sismos	162
Figura 4.29. Comportamiento del IP (color azul) con respecto al comportamiento de la profundidad del sismo (color naranja)	165
Figura 5.1. Página principal del prototipo del índice de Movilidad	172
Figura 5.2. Demarcación de las 16 alcaldías de la Ciudad de México.....	173
Figura 5.3. Distribución de semáforos en las vías principales de las alcaldías de la Ciudad de México, en este caso se muestran los de la alcaldía Iztacalco.....	174
Figura 5.4. Conjunto de vías de una alcaldía de interés (en este caso la alcaldía Iztapalapa).....	175
Figura 5.5. Conjunto de puntos de interés(servicios) que rodean a cada calle de interés	176
Figura 5.6. Valor escalar que representa el índice de Movilidad de una calle de interés	177
Figura 5.7. Valor escalar que representa el índice de Red Vial de una calle de interés	177
Figura 6.1. Relación entre el Índice de Movilidad (IM) y diferentes factores.....	182
Figura 6.2. Distribución normal del Índice de Movilidad (IM).....	183
Figura 6.3. Histograma que muestra la distribución de movilidad de acuerdo al Índice de Movilidad (IM)	184

Figura 6.4. Código QR que enlaza al video funcional titulado “Desarrollo de un Prototipo Web para Evaluación de la Movilidad”	185
--	-----

Lista de tablas

Tabla 2.1. Características clave de los sistemas complejos (Mitchell, 2009).....	22
Tabla 3.1. Conjunto de variables de TomTom Traffic	48
Tabla 3.2. Cuadro descriptivo del método de comprensión del sistema complejo dinámico	57
Tabla 3.3. Cuadro comparativo entre aplicaciones de mapas.....	58
Tabla 3.4. Cuadro descriptivo del método identificación de entidades y relaciones dinámicas	61
Tabla 3.5. Descripción de atributos de tráfico, incidencias viales y servicios, parte 1...	62
Tabla 3.6. Descripción de atributos de tráfico, incidencias viales y servicios, parte 2..	63
Tabla 3.7. Cuadro descriptivo del método definición de variables críticas	68
Tabla 3.8 Alcaldías de la Ciudad de México con su correspondiente número de colonias	70
Tabla 3.9. Cuadro descriptivo del método niveles de agregación del sistema complejo dinámico.....	72
Tabla 3.10. Valores ordenados de entropía de cada variable de la movilidad vehicular	74
Tabla 3.11. Cuadro descriptivo del método análisis de incertidumbre del sistema complejo	75
Tabla 3.12. Trabajos relacionados con las variables independientes y dependientes de la movilidad vehicular	77
Tabla 3.13. Cuadro descriptivo del método de identificación de variables relevantes en relación con la dinámica.....	78
Tabla 3.14. Cuadro descriptivo del método recopilación continua	82
Tabla 3.15. Conjunto de variables de distancias entre tráfico, incidencias y servicios, parte 1.....	86
Tabla 3.16. Conjunto de variables de distancias entre tráfico, incidencias y servicios, parte 2.....	86
Tabla 3.17. Cuadro descriptivo del método Integración de fuentes dinámicas de datos	87
Tabla 3.18. Cuadro descriptivo del método diseño de sistemas de almacenamiento en tiempo real.....	90
Tabla 3.19. Valor representativo de medias de cada variable del conjunto de datos de tráfico, incidencias y servicios.....	94

Tabla 3.20. Correlaciones superiores o iguales a 0.7 entre las variables independientes.....	95
Tabla 3.21. <i>Correlación de la variable jamFactor con el conjunto de variables de la Tabla 3.20.</i>	95
Tabla 3.22. Variables más significativas utilizando la matriz de rotación de Análisis Componentes	96
Tabla 3.23. Matriz de cargas factoriales del Modelo de Ecuaciones Estructurales.....	96
Tabla 3.24. Variables con valores de p inferiores al nivel de significancia de 0.05	97
Tabla 3.25. Variables significativas de tráfico, incidencias y servicios.....	98
Tabla 3.26. Cuadro descriptivo del método preprocesamiento en tiempo real	99
Tabla 3.27. Cuadro descriptivo del método actualización y muestreo dinámico	101
Tabla 3.28. Se muestran los 30 mejores modelos de índice de movilidad vehicular.	105
Tabla 3.29. Mejores modelos de acuerdo a la métrica utilizada.....	106
Tabla 3.30. Conjunto de calles del cuadro delimitador marcado en una zona geográfica de la alcaldía Benito Juárez.....	109
Tabla 3.31. Cuadro descriptivo del método modelado dinámico	110
Tabla 3.32. Resultados del modelo evaluado mediante el Error Cuadrático Medio (MSE) y el Coeficiente de Determinación (R^2), que reflejan la precisión y capacidad explicativa del modelo	111
Tabla 3.33. Valores de escenarios para probar la sensibilidad de las variables del IM	112
Tabla 3.34. Cuadro descriptivo del método evaluación y ajuste continuo del modelo	114
Tabla 3.35. Cuadro descriptivo del método representación dinámica del modelo.....	119
Tabla 3.36. Cuadro descriptivo del método implementación adaptativa	124
Tabla 3.37. Cuadro descriptivo del método Actualización del modelo	126
Tabla 4.1. Aplicación de cada fase de la metodología propuesta en los casos de estudio de calidad del aire y profundidad de sismos	131
Tabla 4.2. Nivel de incertidumbre de las variables del conjunto de datos de aire	142
Tabla 4.3. Descripción de variables del conjunto de datos de aire	148
Tabla 4.4. Parámetros utilizados en el programa evolutivo	148
Tabla 4.5. Expresiones matemáticas generadas mediante un algoritmo evolutivo para el comportamiento de la calidad del aire.....	149
Tabla 4.6. Evaluación del desempeño del modelo mediante métricas de error y ajuste	150
Tabla 4.7. Evaluación del desempeño del modelo ajustado.....	150
Tabla 4.8. Descripción de variables del conjunto de datos de sismos.....	163

Tabla 4.9. Parámetros utilizados en el programa evolutivo, para determinar el índice de profundidad de sismos	163
Tabla 4.10. Comparación metodológica entre los casos de estudio de calidad del aire y profundidad de sismos.....	169
Tabla 6.1. Descripción de las variables empleadas en el modelo de movilidad (IM) ...	181

Capítulo 1. Introducción

1.1 Introducción

En las últimas décadas, la necesidad de comprender y modelar sistemas complejos ha cobrado una relevancia creciente en diversas disciplinas científicas y tecnológicas. Estos sistemas, caracterizados por la interacción de múltiples componentes interdependientes, dan lugar a comportamientos no triviales, patrones emergentes y dinámicas no lineales difíciles de predecir mediante enfoques convencionales. La complejidad inherente de estos entornos se intensifica cuando se enfrentan a condiciones altamente dinámicas, es decir, cuando cambian constantemente en el tiempo y presentan una sensibilidad significativa a variaciones en las condiciones iniciales o perturbaciones externas (Mitchell, 2009).

Los sistemas complejos altamente dinámicos se manifiestan en una amplia variedad de fenómenos reales, tales como la movilidad urbana, el cambio climático, los mercados financieros, los ecosistemas naturales, las redes eléctricas inteligentes, entre otros (Ladyman et al., 2013). Su comportamiento es el resultado de interacciones no lineales entre múltiples entidades, lo cual genera desafíos importantes para su análisis, modelado, predicción y gestión (Bar-yam, 1997). La no linealidad, la multiescalaridad, la retroalimentación y la adaptabilidad son algunas de las propiedades que dificultan su representación matemática y su comprensión integral (Nicolis & Prigogine, 1989).

En este contexto, las metodologías tradicionales de modelado, basadas en supuestos estáticos, estructuras rígidas y formulaciones deterministas, resultan insuficientes para capturar la complejidad dinámica, la adaptabilidad y las interacciones no lineales que caracterizan a los sistemas complejos altamente dinámicos (Helbing, 2014). Estas limitaciones impiden una representación fiel de la realidad, particularmente en fenómenos urbanos, ambientales o sociales donde múltiples variables interactúan en tiempo real y de forma interdependiente. Lo anterior evidencia la necesidad de desarrollar enfoques metodológicos que sean flexibles, adaptativos y capaces de integrar información heterogénea, multiescalar y temporalmente variable. Dichos enfoques deben permitir la incorporación de conocimiento estructural, el procesamiento de grandes volúmenes de datos (Hey, 2009), la extracción de patrones significativos, y la modelación de comportamientos emergentes a través de esquemas computacionales robustos. Frente a este panorama, la presente investigación propone una metodología integral para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos, concebida como un marco estructurado y flexible que combina teoría de sistemas, ciencia y minería de datos, construcción de indicadores compuestos y cómputo evolutivo. Esta articulación interdisciplinaria permite abordar la complejidad estructural, la variabilidad temporal y la interacción dinámica de los elementos que conforman estos sistemas, con el propósito de representar matemáticamente sus comportamientos emergentes y contribuir a la toma de decisiones informada y contextualizada.

El diseño de la metodología parte del reconocimiento de que los sistemas complejos no pueden comprenderse mediante el análisis aislado de sus componentes, sino que requieren un enfoque integral que identifique patrones globales a partir de las interacciones locales. Por ello, los modelos deben ser iterativos y adaptativos, capaces de incorporar nuevos datos, ajustarse a cambios estructurales y mantenerse vigentes

en entornos dinámicos. La metodología propuesta fue diseñada y desarrollada tomando como caso de estudio la movilidad vehicular en la Ciudad de México. Su estructura se organiza en seis fases interrelacionadas, cada una con métodos específicos orientados a abordar la complejidad del sistema. Estas fases abarcan desde la contextualización y análisis del entorno urbano, pasando por la recolección y procesamiento de datos, hasta la implementación, representación del modelo y su evaluación continua. El enfoque permite una comprensión integral y dinámica del sistema, considerando su naturaleza cambiante, adaptativa y multiescalar. La validez de esta propuesta se muestra mediante su aplicación en dos casos de estudio: la calidad del aire en la Ciudad de México y la profundidad de los sismos en México.

1.2 Motivación

Los sistemas complejos son un principio básico para entender fenómenos compuestos por múltiples elementos interrelacionados, cuyas interacciones generan propiedades emergentes no explicables desde un análisis aislado. Presentes en disciplinas como la ingeniería, biología, ciencias sociales, economía y ecología, se caracterizan por su no linealidad, autoorganización, adaptabilidad y sensibilidad a condiciones iniciales, lo que dificulta su predicción y control mediante métodos tradicionales. Su complejidad se manifiesta en la interacción simultánea de variables, retroalimentaciones y adaptación continua. Ejemplos como redes de transporte, ecosistemas, sistemas financieros y redes sociales muestran cómo la dinámica global surge de interacciones locales. Entre estos sistemas, destacan los sistemas complejos altamente dinámicos, cuyo comportamiento evoluciona constantemente y presenta alta sensibilidad a perturbaciones. Redes de movilidad urbana, mercados financieros, redes eléctricas inteligentes y ecosistemas sometidos a cambios abruptos ilustran esta categoría.

Modelar estos sistemas a través del análisis de una gran cantidad de datos, exige metodologías capaces de capturar sus interacciones no lineales y evolución en el tiempo. Comprender su dinámica permite anticipar comportamientos, mitigar riesgos y mejorar la toma de decisiones en tiempo real, lo cual es esencial en contextos inciertos y variables. El desafío radica en desarrollar modelos que, además de representar con precisión sus relaciones internas, sean adaptativos y capaces de incorporar información actualizada. Esta capacidad es vital para anticipar eventos como congestiones viales, fallos en redes eléctricas o crisis financieras. Además, contar con modelos dinámicos fortalece la toma de decisiones estratégicas en sectores donde los cambios son rápidos y las consecuencias significativas, como la gestión de emergencias, salud pública, economía y medio ambiente. Frente a ello, se requiere una metodología estructurada que integre técnicas avanzadas de modelado, análisis y simulación. La ausencia de marcos estandarizados limita la utilidad y consistencia de los modelos. Por tanto, una metodología garantiza resultados reproducibles, fomenta la colaboración interdisciplinaria y facilita soluciones eficaces y escalables en diversos contextos.

1.3 Planteamiento del problema

En ingeniería y ciencias aplicadas, los sistemas complejos altamente dinámicos constituyen un campo crítico de estudio debido a su comportamiento no lineal, su estructura interconectada y su evolución temporal impredecible. Estos sistemas se manifiestan en contextos reales como las redes de movilidad urbana, los sistemas financieros, las redes eléctricas inteligentes y los ecosistemas naturales, donde anticipar su comportamiento es fundamental para una toma de decisiones eficaz y sostenible.

Durante las últimas décadas, múltiples enfoques han intentado modelar la dinámica de estos sistemas, utilizando herramientas como redes complejas, teoría del caos, simulaciones basadas en agentes y sistemas dinámicos no lineales. Sin embargo, pese a los avances en modelado y simulación, aún persisten limitaciones significativas para representar con precisión su comportamiento. Estas limitaciones derivan, en gran parte, de la falta de metodologías integradoras que puedan captar simultáneamente la multiescalaridad, la adaptabilidad temporal y la interdependencia entre componentes.

Diversos estudios han abordado aspectos puntuales de estos desafíos. Por ejemplo, algunos han propuesto modelos para analizar la propagación de fallos en redes eléctricas, mientras otros han explorado patrones de congestión en redes de transporte mediante modelos basados en agentes. No obstante, estos enfoques tienden a ser específicos de un dominio y difíciles de generalizar, lo que restringe su aplicabilidad en escenarios complejos y multifactoriales.

El análisis de sistemas complejos presenta, por tanto, varios desafíos interrelacionados: la complejidad y no linealidad de sus interacciones; la integración de datos en distintos niveles espaciales y temporales (multiescalaridad); su evolución dinámica, que requiere modelos adaptativos; y la interdependencia entre subsistemas, que puede generar efectos emergentes e impredecibles.

Ante este panorama, es necesario avanzar hacia una metodología estructurada y transversal que permita representar el comportamiento de sistemas complejos con mayor precisión y aplicabilidad. Esta tesis propone una contribución en ese sentido, desarrollando un marco metodológico que integra herramientas de modelado adaptativo, análisis multiescalar e interconectividad sistémica. A diferencia de trabajos previos, esta propuesta busca no solo mejorar la precisión del modelado, sino también ofrecer una herramienta aplicable a diversos sectores estratégicos como la planificación urbana, la gestión del riesgo ante desastres o la regulación financiera.

En consecuencia, esta investigación no solo fortalece el conocimiento teórico sobre sistemas complejos, sino que responde a una necesidad práctica aún no resuelta en la literatura: contar con una metodología holística, adaptable y replicable que apoye la toma de decisiones en entornos reales altamente dinámicos.

1.4 Propuesta de solución

La metodología propuesta adopta un enfoque iterativo y adaptativo, diseñado específicamente para abordar la complejidad inherente a los sistemas complejos altamente dinámicos. Estos sistemas evolucionan continuamente debido a la interacción de múltiples variables interdependientes, relaciones no lineales y una alta sensibilidad a las condiciones iniciales (Mitchell, 2009; Holland, 1992). Ante estas características, se requiere una metodología capaz de actualizarse conforme cambian los datos y el entorno, manteniendo su validez y aplicabilidad en contextos dinámicos.

El enfoque iterativo se basa en un ciclo continuo de análisis, ajuste y validación, donde cada iteración permite refinar los modelos incorporando nueva información. Este tipo de enfoque ha sido exitosamente aplicado en estudios de simulación basada en agentes y sistemas adaptativos complejos (Axelrod & Tesfatsion, 2006), mejorando la capacidad predictiva y analítica. Por otro lado, el componente adaptativo de la metodología permite integrar datos en tiempo real, responder a cambios abruptos y ajustarse a escenarios imprevistos, como se ha demostrado en aplicaciones en redes de transporte urbano (Zheng et al., 2014), mercados financieros (Farmer & Foley, 2009) y ecosistemas naturales (Levin, 1998).

La propuesta metodológica incorpora técnicas avanzadas como el cómputo evolutivo, el aprendizaje automático y la construcción de indicadores compuestos, todas ellas con respaldo empírico en la literatura. Por ejemplo, el uso de algoritmos evolutivos ha demostrado ser eficaz en la optimización de sistemas no lineales y multiobjetivo (Deb et al., 2002), mientras que el aprendizaje automático se ha utilizado ampliamente para detectar patrones emergentes y anticipar transiciones críticas en sistemas dinámicos (Reichstein et al., 2019).

El marco metodológico se compone de seis fases interrelacionadas:

Fase 1: Análisis contextual y de la dinámica del sistema. Identifica componentes clave, relaciones funcionales y variables críticas, sentando las bases para comprender la estructura y el comportamiento del sistema (Sterman, 2006).

Fase 2: Agregación multinivel y medición de la dinámica. Integra datos desde niveles micro hasta macro, utilizando métricas como entropía, autocorrelación no lineal y sensibilidad al cambio, tal como se propone en estudios de análisis multiescalar (Wilbanks & Kates, 2003).

Fase 3: Recopilación y monitoreo en tiempo real. Implementa mecanismos automatizados de adquisición y actualización continua de datos, una estrategia clave en sistemas de alerta temprana y respuesta rápida (Helbing et al., 2014).

Fase 4: Preprocesamiento y actualización dinámica. Limpia, transforma y estandariza los datos para asegurar consistencia temporal y semántica, siguiendo buenas prácticas de ingeniería de datos en sistemas dinámicos (Kitchin, 2014).

Fase 5: Modelado iterativo y desarrollo continuo. Construye y refina modelos adaptativos mediante técnicas como redes neuronales recurrentes, algoritmos genéticos o simulaciones basadas en agentes, como se ha aplicado en modelos urbanos complejos (Spencer, 2009).

Fase 6: Interpretación, representación e implementación. Traduce los resultados en visualizaciones comprensibles y orientadas a la toma de decisiones, integrando dashboards interactivos o mapas de calor, siguiendo propuestas como las de Griffin (1987) y Nelson (1996).

Estas seis fases conforman una solución metodológica integral, flexible y orientada a la representación precisa y en tiempo real de la dinámica de sistemas complejos. Además de mejorar la capacidad de modelado, la metodología facilita la intervención oportuna y decisiones informadas basadas en datos actualizados y en la evolución constante del sistema (ver Figura 1.1).

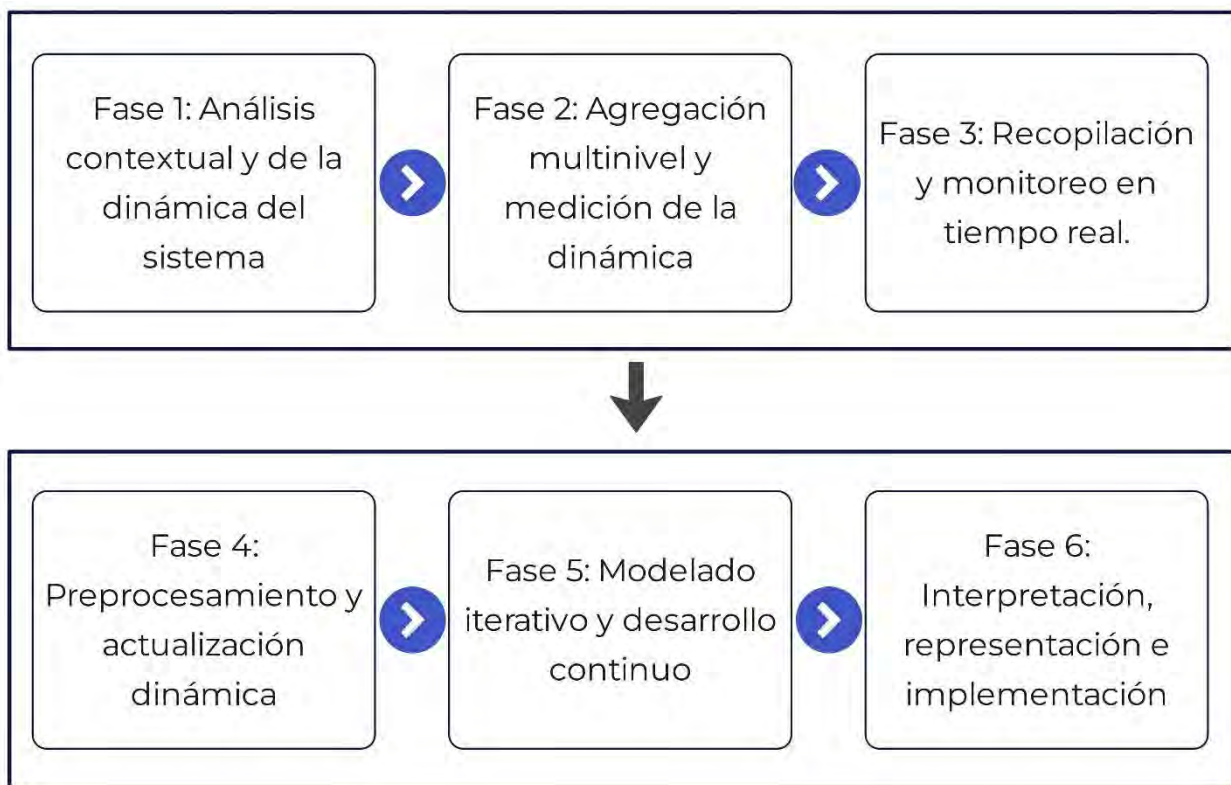


Figura 1.1. Esquema metodológico propuesto para el modelado de sistemas complejos altamente dinámicos

1.5 Objetivos de la Investigación

Este trabajo de investigación profundiza en el papel que desempeña la inferencia basada en modelos en el contexto actual de sistema computacional, un campo donde la interacción entre teoría, datos y simulación es importante para desentrañar los comportamientos de sistemas complejos.

En el marco de esta investigación, cuyo propósito es construir una metodología para desarrollar modelos de comportamiento de sistemas complejos altamente dinámicos, se plantea un objetivo general y una serie de objetivos específicos. Estos objetivos están diseñados para guiar el desarrollo de una metodología robusta y adaptativa para enfrentar las principales dificultades asociadas con la complejidad, la no linealidad, la multiescalaridad y la evolución temporal de estos sistemas. El cumplimiento de estos objetivos permitirá la relevancia y aplicabilidad de los modelos resultantes en situaciones reales. A continuación, se describen el objetivo general y los objetivos específicos.

General: Diseñar y desarrollar una metodología basada en datos, que sea integral, estructurada y utilice en enfoques de ciencia de datos, minería de datos, cómputo evolutivo e indicadores compuestos, orientada a la construcción de modelos de comportamiento para sistemas complejos altamente dinámicos, que permitan representar matemáticamente sus dinámicas e interacciones sistémicas.

Específicos:

- Desarrollar una metodología integral para la construcción de modelos de comportamiento de sistemas complejos altamente dinámicos, basada en enfoques de ciencia de datos, minería de datos, indicadores compuestos y técnicas de desarrollo de software, para lograr una representación precisa y eficiente del comportamiento sistémico.
- Establecer un sistema continuo y automatizado de recolección de datos espaciales y temporales de sistemas complejos altamente dinámicos, con el objetivo de construir un Data Warehouse masivo que sirva como soporte para la generación de modelos de comportamiento.
- Formular y desarrollar modelos de comportamiento precisos para sistemas complejos altamente dinámicos, utilizando la metodología propuesta para capturar con precisión las interacciones y dinámicas de sus variables.
- Comparar la metodología mediante su aplicación en dos casos de estudio representativos, asegurando que los modelos desarrollados reflejen adecuadamente las características y comportamientos de los sistemas analizados.
- Implementar y evaluar los modelos de comportamiento en entornos reales, integrándose en sistemas de soporte a la toma de decisiones para mejorar la comprensión y gestión de los sistemas complejos altamente dinámicos.

1.6 Pregunta de investigación

Los sistemas complejos altamente dinámicos son cada vez más predominantes. Para entender y predecir su comportamiento, se recopilan datos que se caracterizan por la interacción continua y no lineal de múltiples entidades, lo que hace que su comportamiento sea difícil de anticipar mediante métodos tradicionales. La

necesidad de modelos que capturen con precisión la dinámica intrínseca de estos sistemas ha impulsado la búsqueda de nuevas formas para explorar su complejidad inherente. Sin embargo, la falta de un enfoque estandarizado para diseñar y construir estos modelos limita nuestra capacidad para entender y gestionar sus comportamientos.

En este contexto, la presente investigación plantea la pregunta:

¿Es posible que una metodología integral permita la construcción de modelos de comportamiento capaces de representar, con precisión y eficiencia, la dinámica e interacción entre las entidades de sistemas complejos altamente dinámicos, de forma verificable mediante criterios objetivos?

La investigación busca desarrollar un marco metodológico que facilite la comprensión y modelado de sistemas complejos, contribuyendo así al avance en el análisis y gestión de los datos que describen dichos sistemas en diversos campos.

1.7 Hipótesis

En el contexto del modelado de sistemas complejos altamente dinámicos, las hipótesis planteadas tienen como objetivo validar, desde un enfoque estadístico, la eficacia y aplicabilidad de los modelos desarrollados para representar con precisión el comportamiento de estos sistemas. Estas hipótesis se fundamentan en la premisa de que es posible diseñar una metodología integral que combine diversas técnicas de modelado y que, además se adapte a la evolución temporal y a las interacciones no lineales propias de estos sistemas.

A continuación, se presentan las hipótesis formuladas estadísticamente para esta investigación:

- Hipótesis nula (H_{01}): Hipótesis nula (H_{01}): El empleo de metodologías computacionales estructuradas no tiene un impacto significativo en la capacidad para modelar sistemas complejos altamente dinámicos. Hipótesis alternativa (H_{11}): El empleo de metodologías computacionales estructuradas tiene un impacto significativo en la capacidad para modelar sistemas complejos altamente dinámicos.
- Hipótesis nula (H_{02}): No es posible desarrollar modelos de comportamiento de sistemas complejos altamente dinámicos que, a partir de sus variables latentes y observables, permitan caracterizar matemáticamente la interrelación entre las variables y reflejar los diferentes niveles de comportamiento del sistema en función de la variabilidad de los datos. Hipótesis alternativa (H_{12}): Es posible desarrollar modelos de comportamiento de sistemas complejos altamente dinámicos que, a partir de sus variables latentes y observables, permitan caracterizar matemáticamente la interrelación entre las variables y reflejar los diferentes niveles de comportamiento del sistema bajo la perspectiva de la dinamicidad de los datos.

Para la validación de estas hipótesis, se emplearon distintos métodos estadísticos según la naturaleza de los datos y los objetivos de cada modelo, se utilizaron:

- Pruebas de significancia estadística como la ANOVA para evaluar el impacto de las metodologías computacionales sobre la precisión del modelado.
- Modelos de regresión multivariada y análisis de varianza explicada (R^2) para determinar el grado en que las variables observables y latentes explican la variabilidad del comportamiento del sistema.
- Técnicas de validación cruzada, como k-fold cross-validation, para medir la capacidad de generalización de los modelos.
- Análisis de residuos y métricas de ajuste, como el error cuadrático medio (MSE) y el error absoluto medio (MAE), para evaluar el rendimiento predictivo de los modelos.
- Pruebas de normalidad y homocedasticidad, para garantizar los supuestos de los modelos inferenciales aplicados.

1.8 Metodología de solución de la investigación

El diseño de las fases y métodos que conforman la metodología propuesta para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos se elaboró a través de un proceso iterativo, exploratorio y analítico. Este proceso se caracterizó por ajustes y refinamientos continuos, en respuesta tanto al análisis conceptual como a la experimentación empírica, lo que permitió una evolución progresiva del marco metodológico hacia una mayor precisión y adaptabilidad.

La construcción de esta metodología estuvo guiada por una metodología de solución estructurada, que facilitó la identificación y resolución de los desafíos del modelado de sistemas complejos de forma sistemática y organizada, como se muestra en la Figura 1.2. Este enfoque integró principios de modelado, técnicas de análisis de datos y estrategias de validación, garantizando su solidez metodológica y aplicabilidad práctica.

Este enfoque metodológico se alinea con marcos teóricos contemporáneos que defienden el uso de ciclos iterativos y adaptativos en la modelación de fenómenos complejos, tales como los enfoques basados en *Design Science Research* (Hevner et al., 2004; Gregor & Hevner, 2013) y en la construcción incremental de modelos mediante prototipado evolutivo (Wieringa, 2014; Baskerville et al., 2018). Además, se fundamenta en las propuestas de sistemas dinámicos adaptativos aplicados a entornos con alta incertidumbre y múltiples escalas (Mitchell, 2009; Kutz, 2013).

Mediante este proceso, no solo se logró desarrollar una metodología propuesta funcional y adaptable, sino que también se establecieron bases que pueden ser aplicadas en otros contextos de estudio. Esto amplía la relevancia y aplicabilidad del marco metodológico, permitiendo su implementación en la representación y modelado de otros sistemas altamente dinámicos, como la calidad del aire y la profundidad de los sismos.

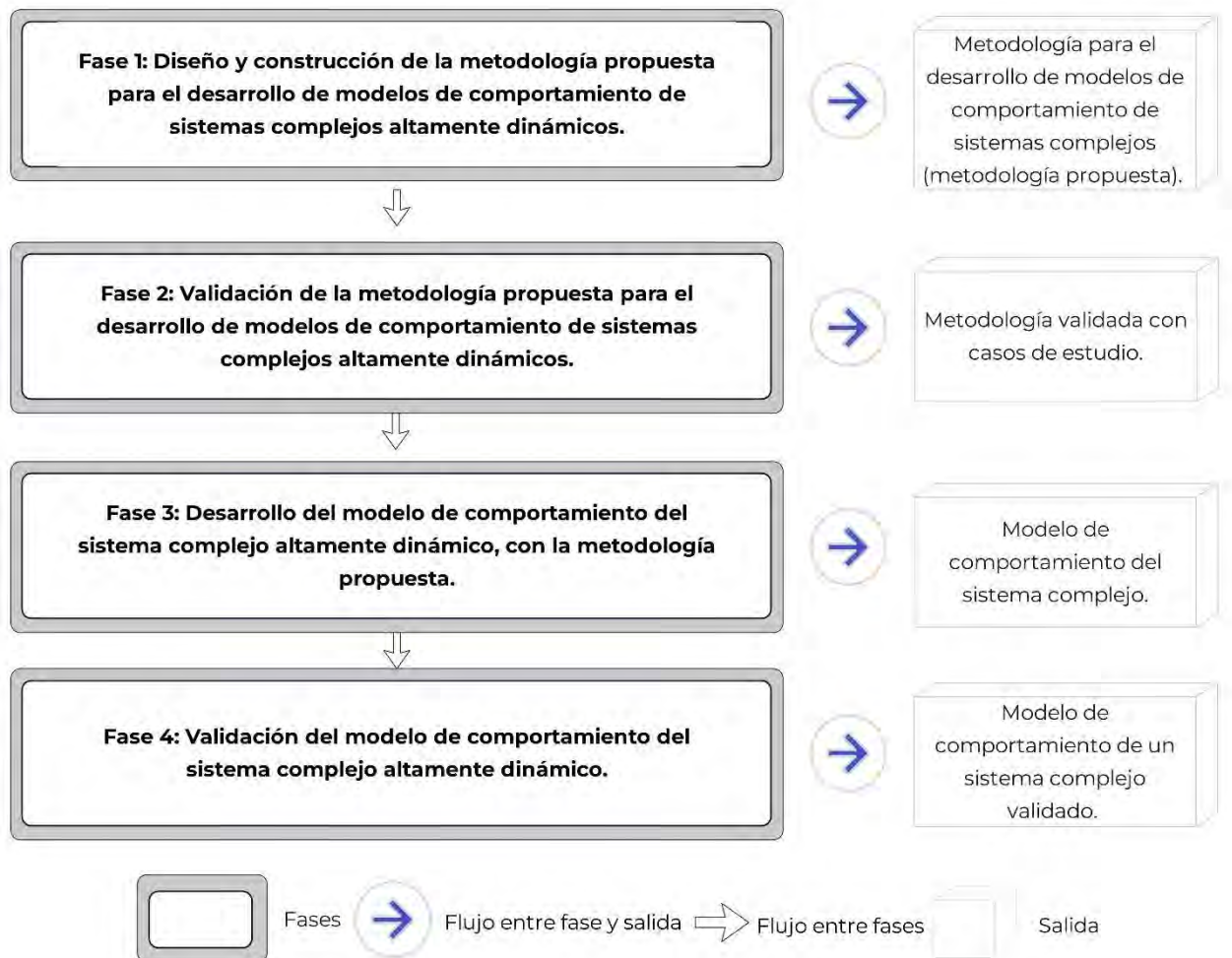


Figura 1.2. Metodología de solución para el desarrollo de modelos de comportamiento en sistemas complejos altamente dinámicos

La metodología de solución para la construcción de una metodología destinada al desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos se organiza en cuatro fases claramente definidas. Estas fases buscan garantizar un enfoque sistemático y robusto que abarque el diseño, la implementación, la validación y el refinamiento de los modelos (ver Figura 1.2). Cada fase cumple un papel esencial en el proceso, facilitando una transición ordenada desde la conceptualización inicial hasta la validación final del modelo. Este enfoque

asegura que el modelo desarrollado represente con precisión las dinámicas intrínsecas y las interacciones características de los sistemas complejos. A continuación, se explica cada fase.

Fase 1. Diseño y construcción de la metodología propuesta: En esta fase se establecen los procesos para el desarrollo de modelos de comportamiento en sistemas complejos altamente dinámicos. Se lleva a cabo un análisis profundo de los principios teóricos y computacionales que permiten la representación de sistemas complejos con múltiples interacciones y variabilidad temporal. Se definen métodos y actividades que compondrán la metodología, asegurando que el marco conceptual sea lo suficientemente flexible para capturar la dinámica del sistema. El resultado de esta fase es una metodología propuesta para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos estructurada para ser evaluada en casos de estudio, estableciendo así las bases teóricas y operativas para su implementación.

Fase 2. Validación de la metodología propuesta: Una vez definida metodología propuesta para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos, es determinante validar su aplicabilidad mediante estudios de caso representativos. En esta fase, se seleccionan sistemas complejos reales (calidad del aire y profundidad de los sismos con perspectiva de sistemas complejos) que permitan evaluar el desempeño del enfoque metodológico en condiciones dinámicas y con variabilidad significativa. Se analiza la efectividad de los métodos propuestos para modelar el comportamiento de estos sistemas, asegurando que la metodología permita capturar patrones, relaciones dinámicas e interacciones emergentes. El resultado de esta fase es una metodología validada con base en evidencia de los modelos desarrollados, lo que garantiza su fiabilidad y aplicabilidad en diferentes dominios.

Fase 3. Desarrollo del modelo de comportamiento del sistema complejo: Con la metodología propuesta para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos validada, se procede a la construcción de un modelo matemático-computacional que represente el comportamiento del sistema complejo altamente dinámico. Este modelo se construye siguiendo los principios metodológicos definidos, asegurando que capture la estructura, interacciones y dinámica inherente del sistema en estudio. En esta etapa, se implementan algoritmos de modelado y análisis de datos, permitiendo la generación de predicciones y escenarios basados en la estructura del sistema. Se integran múltiples fuentes de datos y se aplican técnicas de ajuste para mejorar la precisión y representatividad del modelo. El resultado de esta fase es un modelo funcional del sistema complejo, que refleja su comportamiento dinámico y proporciona herramientas para su análisis y caracterización.

Fase 4. Validación del modelo de comportamiento del sistema complejo: La última fase consiste en evaluar la precisión, robustez y capacidad predictiva del modelo desarrollado. Para ello, se compara el desempeño del modelo con datos reales en diferentes escenarios, asegurando que la representación matemática y computacional del sistema sea consistente con la dinámica observada en la realidad. Se aplican criterios de evaluación cuantitativa y cualitativa, como errores de

aproximación, correlación con datos históricos, estabilidad del modelo y sensibilidad a parámetros significativos.

1.9 Contribuciones

La presente investigación aporta conocimientos teóricos y prácticos que benefician a la comunidad científica y a los profesionales en diversos campos. Las contribuciones de esta investigación abarcan desde el avance en las técnicas de modelado hasta la implementación de herramientas prácticas que pueden ser aplicadas en situaciones reales. Estas contribuciones buscan mejorar la comprensión y gestión de los sistemas complejos altamente dinámicos, proporcionando soluciones efectivas a los desafíos asociados con su complejidad, no linealidad y evolución temporal. A continuación, se detallan las principales contribuciones de esta investigación.

- Construcción metodológica: La tesis propone una metodología estructurada y eficaz para la comprensión y modelado de la dinámica propia de los sistemas complejos altamente dinámicos. Esta propuesta constituye una contribución significativa al ofrecer una guía sistemática que facilita el análisis, representación de comportamientos emergentes en este tipo de sistemas.
- Desarrollo de un modelo matemático que caracteriza el comportamiento de la movilidad vehicular, utilizando la metodología propuesta.
- Desarrollo de un prototipo para la implementación del Índice de Movilidad Vehicular, permitiendo la visualización en tiempo real del desempeño del tráfico vehicular en calles específicas de la Ciudad de México a partir de la interacción de las variables.

1.10 Entorno de Investigación

El entorno de investigación para la construcción de la metodología destinada a modelar sistemas complejos altamente dinámicos desempeña un papel importante, ya que abarca una serie de factores que influyen tanto directa como indirectamente en el desarrollo del estudio. Este entorno tiene como punto de partida el Centro Nacional de Investigación y Desarrollo Tecnológico, el cual proporciona recursos esenciales como infraestructura, el apoyo de directores de tesis, acceso a bases de datos, y plataformas tecnológicas avanzadas. Estos elementos son de suma importancia para el desarrollo y la ejecución de la investigación. Dado que el estudio implica el manejo de grandes volúmenes de datos provenientes de sistemas complejos, el acceso a recursos tecnológicos avanzados es vital. Entre estos recursos destacan el hardware de alto rendimiento, como servidores con múltiples núcleos de procesamiento, gran capacidad de memoria RAM y unidades de almacenamiento de alta velocidad, esenciales para ejecutar simulaciones complejas y procesar grandes volúmenes de datos.

En cuanto a las condiciones de uso, licencias y restricciones, los recursos utilizados están sujetos a los lineamientos institucionales del CENIDET, los cuales contemplan:

- Uso académico y de investigación no comercial, conforme a los términos de uso establecidos en los reglamentos internos del centro.
- Acceso controlado a plataformas y bases de datos, mediante credenciales institucionales y con fines estrictamente científicos.
- Licencias de software académicas o de código abierto (por ejemplo, Python, R, SUMO, Scikit-learn), las cuales permiten su libre uso para investigación bajo sus respectivas licencias (GPL, MIT, Apache, etc.).
- Para los datos utilizados, las bases de datos de terceros (como las provenientes de APIs públicas de tráfico o movilidad urbana) se utilizaron conforme a sus términos y condiciones de uso, sin infringir cláusulas de confidencialidad ni derechos de propiedad intelectual.

El acceso a datos, especialmente en tiempo real, constituye un componente esencial del entorno de investigación, ya que condiciona directamente la calidad, precisión y aplicabilidad de los modelos desarrollados. En el presente estudio, los datos provienen de fuentes diversas, incluyendo instituciones gubernamentales, plataformas privadas y recursos académicos. Entre las principales fuentes gubernamentales se encuentran el Instituto Nacional de Estadística y Geografía (INEGI), que proporciona datos geospaciales y censales; la Secretaría de Movilidad de la Ciudad de México (SEMOVI) y el C5 CDMX, que ofrecen información sobre movilidad, incidentes viales y transporte público; así como la Secretaría del Medio Ambiente (SEDEMA), que aporta registros ambientales que pueden influir en la dinámica del tráfico. Por parte del sector privado, se utilizaron plataformas como HERE Technologies y TomTom, a través de sus APIs, para obtener datos de tráfico en tiempo real, velocidad promedio y congestión, además de herramientas como Geoapify y OpenStreetMap (OSM) para acceder a mapas actualizados, redes viales y puntos de interés. Este enfoque integral garantiza que la información utilizada sea confiable, actualizada y legalmente válida, lo cual fortalece la solidez metodológica de la investigación y su potencial de implementación práctica.

1.11 Organización de la tesis

La organización propuesta para la tesis está diseñada como se describe a continuación:

Capítulo 2. Marco Referencial: Este capítulo proporciona el contexto necesario para comprender la investigación mediante una revisión de literatura relevante y la identificación de estudios previos en el ámbito de los sistemas complejos. Se exploran conceptos y enfoques metodológicos existentes, así como las principales tendencias y desafíos en el modelado del comportamiento de sistemas complejos. Este análisis permite situar la investigación dentro de un marco conceptual amplio y detectar vacíos en el conocimiento que la presente investigación busca tratar.

Capítulo 3. Metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos: Este capítulo expone el diseño detallado de las fases y métodos que conforman la metodología propuesta para el desarrollo de modelos de comportamiento en sistemas complejos altamente dinámicos. Se describe la estructura metodológica, fundamentada en seis fases interdependientes y dieciséis métodos, los cuales han sido diseñados para capturar, analizar y modelar la dinámica de sistemas donde múltiples variables interactúan no linealmente.

La construcción de esta metodología se fundamenta en principios de adaptabilidad, escalabilidad y robustez, permitiendo su aplicación en distintos contextos. Para ilustrar su implementación, se presenta su aplicación en el caso de estudio de movilidad vehicular, el cual ha servido como marco experimental para validar la eficacia de los métodos desarrollados. Cada fase se detalla en términos de sus objetivos, procesos analíticos y técnicas aplicadas, iniciando desde la contextualización del sistema y la identificación de sus componentes críticos hasta el modelado, validación y representación de los resultados obtenidos.

Capítulo 4. Validación de la metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos mediante dos casos de estudio: Este capítulo evalúa la metodología propuesta mediante su aplicación en dos casos de estudio representativos, con el fin de mostrar su capacidad para modelar sistemas complejos altamente dinámicos en distintos contextos. El primer caso examina la calidad del aire en la Ciudad de México, un sistema influido por factores ambientales, sociales y económicos. El segundo caso se enfoca en la profundidad de los sismos en México, considerando variables geológicas, tectónicas y sísmicas desde un enfoque geoespacial. A través del análisis de datos históricos y en tiempo real, se identifican patrones de profundidad y su impacto en la propagación sísmica.

Capítulo 5. Prototipo de movilidad vehicular en la Ciudad de México: Este capítulo detalla el desarrollo de un prototipo en ambiente web diseñado para la interacción dinámica del usuario con el modelo integrado del Índice de Movilidad Vehicular (IMV). Este prototipo permite la evaluación cuantitativa y en tiempo real del desempeño de la movilidad vehicular en calles y zonas geográficas específicas, proporcionando un análisis detallado basado en datos cuantitativos reales de desempeño. La implementación del IMV dentro de la plataforma web no solo optimiza la accesibilidad y visualización de la información, sino que también facilita la toma de decisiones fundamentadas en patrones de movilidad, permitiendo la identificación de áreas críticas y la formulación de estrategias para la gestión eficiente del tráfico.

Capítulo 6. Resultados: Este capítulo analiza los cuatro principales resultados de la investigación, resaltando su aporte al estudio de la movilidad vehicular y su aplicabilidad en diversos contextos. El primer resultado es la metodología propuesta, estructurada en seis fases y conformada por dieciséis métodos interdependientes, diseñada para modelar eficazmente el comportamiento de sistemas complejos altamente dinámicos. El segundo resultado es el Índice de Movilidad Vehicular (IMV), construido a partir de dicha metodología. El tercer resultado es un modelo matemático que relaciona emisiones contaminantes, condiciones meteorológicas y movilidad vehicular para analizar la calidad del aire en la Ciudad de México. El cuarto

resultado es el modelo del Índice de Profundidad de Sismos, que utiliza datos geoespaciales, históricos y en tiempo real para identificar patrones sísmicos en México. Incorpora variables geológicas y tectónicas para estimar cómo la profundidad de los sismos afecta la propagación de ondas y su impacto en la superficie, apoyando la gestión del riesgo sísmico.

Capítulo 7. Conclusiones y trabajos futuros: Se sintetizan los hallazgos y reflexiones relevantes de la investigación, destacando su impacto en el modelado de sistemas complejos altamente dinámicos. Se discuten las limitaciones del estudio, se sugieren direcciones para futuras investigaciones y se resaltan las contribuciones teóricas y prácticas de la metodología desarrollada. Esta sección proporciona una visión integral de los resultados obtenidos y su aplicabilidad en diferentes dominios.

Capítulo 2. Marco Referencial

2.1 Introducción

El estudio de los sistemas complejos altamente dinámicos ha sido un desafío interdisciplinario en diversas áreas del conocimiento, desde la física y la biología hasta la ingeniería y las ciencias sociales. La necesidad de comprender y modelar estos sistemas radica en su capacidad para exhibir comportamientos emergentes, no lineales y adaptativos, que los hacen difíciles de analizar mediante enfoques tradicionales. En este capítulo, se establecen las bases teóricas que sustentan la metodología propuesta para el desarrollo de modelos de comportamiento en estos sistemas complejos altamente dinámicos, integrando conceptos de la teoría de sistemas, teoría de redes, redes complejas y sistemas dinámicos. A través de una revisión de teorías relacionadas con el modelado de sistemas complejos, se busca proporcionar un contexto estructurado que permita no solo comprender las propiedades críticas de los sistemas complejos altamente dinámicos, sino también desarrollar enfoques metodológicos adecuados para su modelado y análisis. La interconexión de estos conceptos con herramientas matemáticas y computacionales permite establecer un marco de trabajo para estudiar la complejidad inherente a estos sistemas y proponer estrategias que faciliten su estudio y aplicación en contextos del mundo real.

2.2 Marco teórico

La comprensión y modelado de sistemas complejos altamente dinámicos han sido objeto de creciente interés en diversas disciplinas a lo largo del tiempo. La necesidad de ilustrar estos sistemas, caracterizados por su complejidad y variabilidad, ha llevado al desarrollo y perfeccionamiento de diversas teorías, metodologías y herramientas. Esta sección presenta los antecedentes que forman la base de la investigación en metodologías para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos.

2.2.1 Teoría de sistemas

La teoría de sistemas es un término sin una definición formalmente acordada. Es un término que se ha utilizado en una variedad de disciplinas y se encuentra frecuentemente en la literatura sobre sistemas. Gracias a que el término se ha usado en diversas disciplinas y tiene múltiples significados (Adams, 2012).

Ludwig Von Bertalanffy (Hofkirchner & Schafranek, 2011) fue el creador de la teoría general de sistemas. Su trabajo se centró en la teoría de sistemas orgánicos, donde analizó el equilibrio termodinámico en estado estacionario en organismos vivos como sistemas abiertos. En la investigación afirma: “La correspondencia formal de principios generales, independientemente del tipo de relaciones o fuerzas entre los componentes, conduce a la concepción de una “teoría general de sistemas”.

James Grier Miller (Miller, 1975) desarrolló la teoría de los sistemas vivos, que se enfoca en cómo operan todos los sistemas vivos, cómo se mantienen y cómo evolucionan. Como afirma Miller, los sistemas vivos son sistemas abiertos y autoorganizados que poseen características particulares de la vida y su interacción con el entorno, a través

de intercambios de información, materiales y energía. La teoría de Miller organiza estos sistemas en una jerarquía de ocho niveles: Células, Órganos, Organismos, Grupos, Organizaciones, Comunidades, Sociedades y Sistemas Supranacionales. Los niveles superiores son más complejos que los inferiores, ya que los niveles superiores no solo heredan las características de los niveles inferiores, sino que también incluyen propiedades emergentes que no están presentes en los niveles inferiores.

Un sistema es una porción de la realidad que percibimos como una entidad independiente dentro de esa misma realidad. Lo que se encuentra fuera de este sistema se denomina entorno. La conexión entre el sistema y su entorno se establece a través de magnitudes, frecuentemente expresadas como funciones del tiempo, conocidas como entradas y salidas. El sistema recibe influencias a través de las funciones de entrada y, a su vez, ejerce una influencia sobre el entorno mediante las funciones de salida, como se muestra en la Figura 2.1.

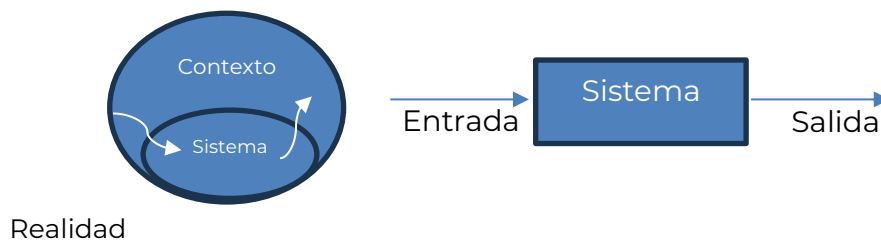


Figura 2.1. *Un sistema que interactúa con su entorno*

2.2.2 Teoría de redes

La teoría de redes es una rama de la teoría de grafos que se desarrolló para analizar cuestiones relacionadas con la conectividad y la optimización en sistemas que pueden representarse mediante nodos y enlaces. Este enfoque teórico ha sido empleado para investigar la estabilidad y el funcionamiento de redes como las eléctricas, las redes sociales humanas y las dinámicas de redes tróficas. A medida que creció el interés en la estructura de las redes, se identificaron ciertas generalidades, entre ellas, la capacidad de describir muchas redes con una estructura de "mundo pequeño" o "libre de escala".

La estructura y dinámica de las redes es variada como las redes genéticas o la World Wide Web se entienden mejor como redes con una topología compleja. Una característica común en muchas redes grandes es que las conexiones entre los vértices siguen una distribución de ley de potencia sin escala. Este fenómeno resulta de dos mecanismos generales: (i) las redes crecen continuamente mediante la incorporación de nuevos vértices, y (ii) los nuevos vértices se conectan preferentemente a nodos que ya tienen muchas conexiones. Un modelo que incorpora estos dos factores reproduce las distribuciones estacionarias sin escala observadas, sugiriendo que la evolución de grandes redes está guiada por fenómenos robustos de autoorganización que trascienden las particularidades de cada sistema individual (Albert-László Barabási, 1999).

La teoría de redes se ha aplicado a sistemas dinámicos acoplados que se han empleado para modelar diversos fenómenos como osciladores biológicos, arreglos de uniones Josephson, medios excitables, redes neuronales, juegos espaciales, redes de control genético y otros sistemas autoorganizados. Generalmente, se considera que la topología de conexión es totalmente regular o completamente aleatoria (Duncan J. Watts, 2018).

2.2.3 Redes complejas

Las redes complejas son estructuras compuestas por nodos interconectados por enlaces (relaciones o interacciones) que exhiben patrones de organización no triviales. Este tipo de redes se encuentran en una amplia variedad de sistemas naturales y artificiales, desde redes biológicas hasta infraestructuras tecnológicas y sociales. El estudio de redes complejas ha ganado relevancia en virtud de su capacidad para describir y analizar el comportamiento colectivo de sistemas altamente dinámicos y con múltiples interacciones entre sus componentes (Newman, 2003).

Uno de los aspectos de las redes complejas es su capacidad para modelar interacciones no lineales entre entidades, lo cual es esencial para entender sistemas donde las relaciones entre componentes no son predecibles de forma directa a partir de las propiedades individuales de los nodos. Estas redes se caracterizan por la existencia de propiedades emergentes, es decir, características globales que no pueden ser explicadas solo observando los elementos por separado (Boccaletti et al., 2006).

Algunas propiedades de las redes complejas que han sido identificadas en diversos sistemas son: Una conectividad heterogénea, es decir no siguen un patrón de conectividad uniforme, sino que presentan una distribución de grado (número de conexiones por nodo) heterogénea (Albert-Lázló Barabási, 1999). Este tipo de redes suelen seguir una distribución libre de escala (Dorogovtsev & Mendes, 2002). Las redes complejas tienden a organizarse en módulos o comunidades, donde los nodos dentro de un módulo están densamente conectados entre sí que con nodos de otros módulos. Esta estructura permite a las redes escalar en tamaño manteniendo una organización interna, lo cual es vital en sistemas biológicos y sociales (Fortunato, 2010). Muchas redes complejas muestran la propiedad de pequeño mundo, lo que implica que cualquier par de nodos puede estar conectado a través de un número relativamente pequeño de pasos o enlaces. Esta característica facilita la rápida propagación de información o influencias a través de la red, tal como ocurre en la difusión de noticias en redes sociales o la transmisión de enfermedades en redes epidemiológicas (Duncan J. Watts*, 2018). Son altamente robustas ante fallos aleatorios en los nodos o enlaces, ya que la estructura general tiende a mantenerse conectada. Sin embargo, son vulnerables a ataques dirigidos a los nodos conectados o críticos, lo que puede causar una fragmentación significativa de la red (Albert et al., 2000). Este tipo de análisis es relevante en infraestructuras tecnológicas, como las redes eléctricas o de telecomunicaciones (Crucitti et al., 2003).

El estudio de redes complejas se fundamenta en teorías de grafos, que permiten representar matemáticamente las relaciones entre nodos y analizar propiedades

como la centralidad, la transitividad o el clustering (Peters, 2017). Estas métricas facilitan la comprensión de la estructura y dinámica del sistema y pueden ser aplicadas para modelar el comportamiento de sistemas complejos en campos tan diversos como la biología, la economía y la ingeniería (Newman, 2003).

En el contexto de la modelización de sistemas complejos altamente dinámicos, las redes complejas proporcionan un marco teórico poderoso para capturar las interacciones y dependencias entre los distintos componentes del sistema. Al representar los sistemas como redes de interacciones, es posible analizar su comportamiento en diferentes escalas, desde lo local (interacciones entre nodos vecinos) hasta lo global (patrones emergentes en toda la red) (Boccaletti et al., 2006).

2.2.4 Teoría de sistemas dinámicos

La Teoría de Sistemas Dinámicos se ocupa del análisis de sistemas que evolucionan con el tiempo y poseen fundamento en las matemáticas. Este enfoque ofrece un marco conceptual para comprender cómo los sistemas complejos se desarrollan y responden a factores tanto internos como externos, así como a diferentes condiciones iniciales. Se aplica en una amplia gama de disciplinas, como la física, biología, economía y ciencias sociales, donde las interacciones entre los componentes de un sistema suelen generar comportamientos no lineales e impredecibles (Strogatz, 2015).

En términos generales, un sistema dinámico se describe a través de ecuaciones diferenciales que capturan cómo cambia el estado del sistema con el tiempo. Estas ecuaciones pueden ser lineales o no lineales, y son de particular interés en sistemas complejos, ya que los sistemas no lineales pueden exhibir fenómenos como el caos o una alta sensibilidad a las condiciones iniciales (Lorenz, 1963). Este enfoque resulta necesario para analizar las interacciones dentro de los sistemas complejos, ya que incluso pequeñas variaciones pueden tener efectos amplificados en su evolución a largo plazo.

Un concepto en esta teoría es el de los atractores, que representan los estados o conjuntos de estados a los cuales tiende un sistema tras una larga evolución. Estos atractores pueden ser puntos fijos, ciclos o configuraciones complejas, como los atractores caóticos, que son comunes en sistemas no lineales y dinámicos (Thompson & Stewart, 2002). La identificación de estos atractores resulta esencial para comprender los patrones estables dentro del comportamiento de un sistema, incluso en presencia de alta complejidad e imprevisibilidad (Kaplan et al., 1996).

En sistemas complejos, las bifurcaciones pueden conducir al surgimiento de dinámicas nuevas, como la transición de un estado estable a uno caótico o el establecimiento de ciclos periódicos (John Guckenheimer & Holmes, 1983). Este concepto es vital para comprender cómo los sistemas complejos pueden cambiar abruptamente por razón de influencias externas, modificando radicalmente su estructura y dinámica (May, 1976).

En la modelización de sistemas complejos altamente dinámicos, la Teoría de Sistemas Dinámicos aporta herramientas valiosas para representar y analizar cómo los

componentes de un sistema interactúan a lo largo del tiempo, generando fenómenos emergentes o transiciones críticas. Esto es especialmente relevante cuando se buscan modelar sistemas con múltiples escalas temporales y espaciales, como en el caso de la movilidad urbana, donde la congestión y los patrones de tráfico pueden evolucionar rápidamente y no lineal como resultado de las variaciones en la infraestructura o cambios ambientales (Helbing, 2001).

La capacidad de esta teoría para capturar tanto la dinámica local (interacciones entre componentes) como la dinámica global (patrones emergentes a nivel sistémico) es esencial para diseñar modelos de comportamiento que capturen con exactitud la complejidad inherente a los sistemas altamente dinámicos (Ott, 2002). En campos como la planificación urbana, el análisis de redes de transporte y la modelización ambiental, la Teoría de Sistemas Dinámicos permite prever el comportamiento a largo plazo y anticipar posibles bifurcaciones, ayudando así en la toma de decisiones (Sterman, 2006).

2.2.5 Sistemas complejos

Los sistemas complejos son conjuntos formados por múltiples elementos interrelacionados cuya dinámica colectiva no puede explicarse únicamente a partir del comportamiento aislado de cada componente. Estos sistemas se caracterizan por presentar interacciones no lineales entre sus partes, lo que da lugar a fenómenos emergentes que no pueden predecirse de manera directa a partir de las propiedades individuales de sus elementos (ver Tabla 2.1). Esta complejidad inherente exige enfoques de modelado y análisis capaces de capturar la dinámica global resultante de dichas interacciones. En un sistema complejo, las interacciones entre sus partes crean dinámicas que no pueden descomponerse ni analizarse linealmente, lo que implica que el todo es mayor que la suma de sus partes (Ceïich, 2015). Estos sistemas abarcan diversas áreas como biología, economía, ciencias sociales e ingeniería, y su análisis es esencial para comprender fenómenos como la evolución de redes, ecosistemas y la distribución del tráfico vehicular.

Uno de los rasgos principales de los sistemas complejos es la emergencia, donde patrones globales surgen de interacciones locales entre sus componentes. Los sistemas complejos, además, son sensibles a las condiciones iniciales, de modo que pequeñas diferencias pueden conducir a resultados completamente distintos, como lo ilustra el concepto de caos (Lorenz, 1963). Otro aspecto esencialmente necesario de la complejidad es la retroalimentación. Los sistemas complejos suelen tener bucles de retroalimentación positiva o negativa que amplifican o estabilizan ciertos comportamientos. En el contexto de la movilidad urbana, por ejemplo, un aumento en el uso de una carretera puede provocar mayor congestión, lo que a su vez impulsa a los conductores a utilizar rutas alternativas, afectando el comportamiento del sistema globalmente (Helbing, 2001).

Por otra parte, muchos sistemas complejos son también adaptativos, es decir, tienen la capacidad de cambiar y ajustar su estructura interna en respuesta a condiciones externas. Este tipo de sistemas, conocidos como Sistemas Complejos Adaptativos (CAS por sus siglas en inglés), se exploran en áreas como la evolución biológica, los

mercados financieros y las redes sociales (Holland & H, 1992). En este contexto, los CAS poseen la capacidad de autoorganizarse, lo que significa que pueden configurarse sin necesidad de una guía central.

La jerarquía de complejidad es otra característica común en estos sistemas, lo que implica que están organizados en niveles o capas de interacciones. Cada nivel tiene su propia dinámica, pero también interactúa con otros niveles, influyendo en la dinámica global del sistema (Simon, 1991). En los sistemas urbanos, por ejemplo, las decisiones individuales de movilidad (como elegir tomar el autobús en lugar de conducir) pueden tener efectos a nivel de ciudad, impactando la congestión y la planificación del transporte.

Tabla 2.1. Características clave de los sistemas complejos (Mitchell, 2009)

<i>Característica</i>	<i>Descripción breve</i>
Emergencia	El comportamiento global del sistema no puede predecirse únicamente a partir de las propiedades de sus partes.
Retroalimentación	Las acciones dentro del sistema generan efectos que retroalimentan el comportamiento de los elementos.
No linealidad	Pequeños cambios pueden provocar efectos desproporcionados; las relaciones causa-efecto no son directas.
Adaptabilidad	El sistema puede cambiar su comportamiento en función de estímulos internos o del entorno.
Jerarquía	El sistema está organizado en múltiples niveles o escalas que interactúan entre sí.
Autoorganización	El sistema tiende a generar orden sin intervención externa centralizada.
Sensibilidad a condiciones iniciales	Pequeñas diferencias al inicio pueden conducir a resultados muy distintos (efecto mariposa).
Apertura	Intercambia energía, materia o información con su entorno; no es un sistema cerrado.
Redundancia	Existen componentes que pueden suplir a otros ante fallos, fortaleciendo la resiliencia del sistema.
Modularidad	El sistema está formado por subcomponentes relativamente independientes pero interconectados.
Multiplicidad de agentes	Compuesto por múltiples entidades autónomas que interactúan entre sí.
Evolución temporal	El sistema cambia constantemente a lo largo del tiempo, mostrando dinámica continua.
Heterogeneidad	Los elementos del sistema no son uniformes; tienen distintas funciones, propiedades o comportamientos.
Acoplamiento dinámico	Las relaciones entre componentes varían en el tiempo; no son fijas ni permanentes.
Memoria	El comportamiento del sistema puede depender de estados pasados (historia o trayectoria).

Robustez	Capacidad del sistema para mantener su funcionalidad ante perturbaciones o fallos.
Resiliencia	Capacidad del sistema para recuperarse tras perturbaciones significativas.
Capacidad de aprendizaje	Algunos sistemas complejos (biológicos o artificiales) pueden modificar su comportamiento basado en experiencias previas.
Propagación de efectos	Una perturbación local puede propagarse y afectar al sistema en su conjunto.
Incertidumbre estructural	La estructura del sistema no es completamente conocida o está en evolución constante.

2.2.6 Análisis de estabilidad y comportamiento no lineal

El análisis de estabilidad y comportamiento no lineal ha sido un pilar en el estudio de sistemas complejos altamente dinámicos, con sus orígenes en las matemáticas y la física. A partir de los primeros trabajos de Henri Poincaré, se sentaron las bases que permiten comprender cómo pequeñas variaciones en las condiciones iniciales pueden dar lugar a dinámicas impredecibles en sistemas no lineales (Poincaré, 1892). Estos sistemas suelen mostrar una gran sensibilidad a las perturbaciones, lo que llevó al desarrollo de la teoría del caos. En esta línea, los estudios de Edward Lorenz demostraron que, incluso en modelos deterministas, bajo ciertas condiciones, puede surgir un comportamiento inherentemente impredecible (Lorenz, 1963). Este enfoque ha sido esencial para analizar fenómenos dominados por la no linealidad, como el tráfico vehicular, los ecosistemas o la economía. El análisis de estabilidad, particularmente mediante el método de Lyapunov, ha sido importante para determinar en qué condiciones un sistema permite regresar a un estado de equilibrio tras una perturbación, lo cual es vital en el modelado de sistemas dinámicos propensos a la inestabilidad (Lyapunov, 1992).

Hoy en día, los modelos no lineales han adquirido gran relevancia en la ingeniería de tráfico, donde se emplean para describir la congestión y las interacciones entre vehículos (Kerner, 2004). Del mismo modo, en biología, estos modelos han sido esenciales para explicar la dinámica de poblaciones y los equilibrios en ecosistemas (May, 1976). En este contexto, el análisis de estabilidad y comportamiento no lineal es indispensable para prever la respuesta de los sistemas ante perturbaciones, utilizando herramientas como el análisis de bifurcación y las ecuaciones diferenciales parciales no lineales, que permiten describir las transiciones súbitas entre distintos estados de equilibrio (Strogatz, 2015). Estas referencias históricas y metodológicas resaltan la importancia de los métodos no lineales y del análisis de estabilidad en la comprensión de los sistemas complejos y altamente dinámicos.

2.2.7 Sistemas complejos altamente dinámicos

Los sistemas complejos altamente dinámicos se caracterizan por la interacción de múltiples elementos en un entorno en constante evolución, lo que genera patrones emergentes y cambios impredecibles en el tiempo. A diferencia de los sistemas

complejos tradicionales, estos presentan una alta sensibilidad a perturbaciones externas, dinámicas no lineales e interacciones adaptativas, lo que provoca una evolución continua y, en muchos casos, irreversible (Boccaletti et al., 2006). Entre sus propiedades destacan el dinamismo y la evolución constante, donde la estructura y el comportamiento del sistema cambian por motivo de la influencia de factores internos y externos (Prokopenko, 2014); la retroalimentación y las interacciones adaptativas, que permiten ajustes en función de nuevas condiciones, pudiendo generar efectos acumulativos o bifurcaciones en la evolución del sistema (Sayama, 2015); la no linealidad y sensibilidad a condiciones iniciales, lo que implica que pequeñas variaciones en los parámetros pueden inducir cambios significativos en su evolución, dificultando su predicción exacta (Ott, 2002); la emergencia y la autoorganización dinámica, donde el comportamiento del sistema surge de la interacción entre sus componentes sin la necesidad de un control centralizado (Dirk Helbing, 2012); y la multiescalaridad, dado que los procesos dentro del sistema ocurren a múltiples escalas temporales y espaciales, permitiendo que los efectos locales se amplifiquen hasta generar impactos globales (Thurner et al., 2018). Ejemplos representativos incluyen las redes de tráfico y movilidad urbana, donde las decisiones individuales de los conductores y la infraestructura vial generan patrones de congestión en constante cambio; los sistemas climáticos, donde múltiples variables como la temperatura, la humedad y los vientos interactúan de forma dinámica y no lineal; los sistemas económicos y financieros, en los que la interacción entre agentes, políticas económicas y eventos externos pueden desencadenar fluctuaciones abruptas; y las redes neuronales biológicas, que se reconfiguran continuamente en función de estímulos externos y procesos internos de aprendizaje.

2.2.8 Evolución histórica del modelado de sistemas complejos

La evolución del modelado de sistemas complejos ha sido un proceso dinámico y multidisciplinario, que ha incorporado ideas de diversas áreas, como matemáticas, física, biología y teoría de sistemas. A finales del siglo XIX y principios del XX, el estudio de estos sistemas comenzó a consolidarse gracias a los aportes de Ludwig von Bertalanffy, quien presentó la teoría general de sistemas, subrayando la interconexión y las relaciones entre los distintos componentes de un sistema (Ludwig von Bertalanffy, 1974). En los años 60, el surgimiento de la cibernética, impulsada por Norbert Wiener, estableció las bases para comprender la retroalimentación y el control en sistemas dinámicos, lo que facilitó el avance en el modelado de comportamientos complejos (Wiener, 1985).

Durante la década de 1970, el desarrollo de simulaciones por computadora permitió a los investigadores explorar modelos que anteriormente resultaban difíciles de analizar. Este período se fusionaron conceptos de teoría de redes, teoría del caos y dinámica de sistemas. La investigación de Edward Lorenz sobre el caos determinista demostró que sistemas aparentemente simples pueden manifestar comportamientos extremadamente complejos, presentando nuevos retos y oportunidades en el modelado de fenómenos naturales (Lorenz, 1963).

Con el avance de la tecnología y el aumento de la capacidad computacional, los modelos de sistemas complejos se volvieron más elaborados. En campos como la

biología de sistemas, la ecología y la economía, se empezaron a aplicar enfoques de modelado como la simulación basada en agentes y los modelos de redes complejas, que permiten captar la complejidad inherente y las interacciones entre los componentes de un sistema. En el siglo XXI, la creciente interconexión entre disciplinas ha fomentado el desarrollo de enfoques integrados para el modelado de sistemas complejos, promoviendo la investigación interdisciplinaria y la solución de problemas complejos en áreas como el cambio climático, la salud pública y la dinámica.

2.2.9 Desarrollo de modelos matemáticos y ecuaciones diferenciales

El desarrollo de modelos matemáticos y el empleo de ecuaciones diferenciales en el análisis de sistemas complejos altamente dinámicos ha sido un enfoque importante en múltiples campos científicos y tecnológicos. Estos modelos permiten describir la evolución temporal de dichos sistemas, especialmente cuando están sometidos a cambios continuos en sus estados. Las ecuaciones diferenciales ordinarias y parciales son ejemplos prominentes que proporcionan herramientas para representar interacciones dinámicas y no lineales entre las entidades de un sistema. La investigación de (Strogatz, 2015), los modelos fundamentados en ecuaciones diferenciales resultan esenciales para comprender el comportamiento emergente en sistemas no lineales como redes de osciladores acoplados y sistemas caóticos. (Abarbanel, 2013) añade que estos modelos también facilitan la predicción de comportamientos futuros en sistemas biológicos, económicos y sociales, al combinar información empírica con marcos teóricos.

En el contexto de los sistemas complejos, las ecuaciones diferenciales son base para describir fenómenos como la transmisión de información, el comportamiento colectivo de multitudes y la dinámica de redes de transporte. (Murray, 2002) resalta la importancia de las ecuaciones diferenciales para modelar sistemas biológicos, donde la complejidad surge de interacciones a múltiples escalas temporales y espaciales. Este enfoque ha sido elegido para representar la dinámica de poblaciones, los flujos de tráfico y otros sistemas físicos. Además, los avances en la solución numérica de estas ecuaciones han mejorado la precisión de los modelos, particularmente en situaciones donde las soluciones analíticas no son viables.

Por otro lado, (Barabási, 2012) introduce el concepto de redes complejas, destacando el uso de ecuaciones diferenciales para modelar la dinámica de redes interconectadas, como las neuronales o sociales. Este enfoque ha permitido avanzar en la comprensión de cómo pequeñas variaciones en las condiciones iniciales pueden producir cambios significativos en el comportamiento del sistema, un fenómeno conocido como la sensibilidad a las condiciones iniciales o efecto mariposa. Igualmente, (Ruey S. Tsay, 2019) muestran, a través de su estudio de sistemas caóticos, cómo las ecuaciones diferenciales permiten identificar patrones en sistemas que inicialmente parecen caóticos o desordenados.

No obstante, aunque estos modelos ofrecen una base sólida para el análisis y la predicción, también presentan limitaciones que deben ser consideradas. Una de las principales es la dependencia de datos precisos y completos: en la práctica, los

sistemas reales pueden estar sujetos a ruido, incertidumbre o registros parciales, lo que afecta la calidad de los resultados (Abarbanel, 2013; Kalnay, 2005). Además, los modelos pueden incurrir en sobreajuste si se calibran excesivamente a un conjunto de datos específico, perdiendo capacidad de generalización ante nuevas condiciones (Hastie, Tibshirani & Friedman, 2017). Otro desafío importante es que, en sistemas altamente no lineales, errores pequeños en las condiciones iniciales pueden amplificarse drásticamente, afectando la estabilidad y confiabilidad de las predicciones (Doss, 2021; Strogatz, 2008). Por ello, es necesario acompañar el uso de ecuaciones diferenciales con técnicas de validación cruzada, análisis de sensibilidad y evaluación de errores, que permitan garantizar que los modelos desarrollados sean tanto interpretables como útiles en contextos reales (Saltelli et al., 2008; Oberkampff & Roy, 2025).

2.3 Estado del arte

El modelado de sistemas complejos altamente dinámicos ha experimentado una evolución significativa a partir de enfoques interdisciplinarios que integran técnicas matemáticas, computacionales y orientadas a datos para comprender y representar sus dinámicas emergentes. El acceso creciente a herramientas analíticas avanzadas ha facilitado la construcción de modelos cada vez más precisos y adaptativos. No obstante, persisten desafíos relevantes relacionados con la integración coherente de múltiples metodologías, la escalabilidad de los modelos y la validación efectiva en entornos reales.

En este contexto, la presente sección analiza el estado del arte entorno a las principales metodologías utilizadas para el modelado de sistemas complejos, abarcando enfoques de simulación, modelado de comportamiento, minería de datos, ciencia de datos y cómputo evolutivo. La revisión se llevó a cabo mediante un análisis sistemático de literatura científica indexada en bases de datos reconocidas como Scopus, IEEE Xplore, SpringerLink y ScienceDirect, utilizando combinaciones de palabras clave relacionadas con “modelado de sistemas complejos”, “dinámica de sistemas”, “modelado de comportamientos”, “modelos basados en datos” entre otros.

Se priorizaron publicaciones de alto impacto de los últimos diez años, considerando criterios de relevancia, rigor metodológico y aplicabilidad práctica. Asimismo, se categorizaron los enfoques revisados en función de su naturaleza teórica, empírica o híbrida, identificando sus fortalezas, limitaciones y áreas de oportunidad. Esta revisión tiene como propósito contextualizar la investigación en relación con los avances representativos del campo, identificar vacíos metodológicos y establecer los fundamentos conceptuales que sustentan la propuesta desarrollada en el presente trabajo.

2.3.1 Enfoques y técnicas para el modelado del comportamiento en sistemas complejos

El modelado del comportamiento de sistemas complejos constituye un desafío multidisciplinario que requiere la integración de diversas técnicas analíticas y computacionales para capturar la dinámica emergente de múltiples entidades

interconectadas. En este contexto, las técnicas de modelado han evolucionado para enfrentar problemas donde las interacciones locales generan comportamientos globales no triviales, como en el tráfico vehicular, la propagación de enfermedades, los mercados financieros y las redes neuronales.

Entre las técnicas relevantes se encuentra la minería de datos para el modelado del comportamiento de sistemas complejos, ya que permite identificar patrones, tendencias y relaciones ocultas en grandes volúmenes de datos. Su aplicación facilita la extracción de parámetros de relevancia para la construcción de modelos representativos de dinámicas emergentes en estos sistemas. A través de métodos como el análisis de correlaciones, la segmentación, la clasificación y las técnicas de aprendizaje automático, la minería de datos proporciona información estructurada para la calibración y validación de modelos. Esta técnica es particularmente útil cuando las interacciones entre los componentes del sistema generan comportamientos no lineales o cuando la estructura del sistema es incierta y requiere una aproximación basada en evidencia empírica. Al integrar la minería de datos con enfoques de modelado como los modelos basados en agentes, redes neuronales o sistemas dinámicos, es posible mejorar la precisión predictiva y adaptabilidad de los modelos, permitiendo una representación realista de la evolución y las interacciones dentro de sistemas complejos. (Zhang et al., 2024).

En este contexto, las redes complejas han emergido como una herramienta para modelar y analizar la dinámica de sistemas complejos, convirtiéndose en el enfoque predominante en diversas disciplinas. Estas redes permiten representar las interacciones de orden superior entre los componentes de un sistema, proporcionando una visión estructurada de las relaciones subyacentes. Se ha encontrado evidencia reciente que demuestra cómo las interacciones de orden superior pueden inducir comportamientos colectivos únicos que no se explican fácilmente mediante interacciones simples entre pares. Estos comportamientos colectivos resaltan la importancia de considerar estructuras complejas para comprender la dinámica de los sistemas en su totalidad (Battiston et al., 2021).

Sin embargo, para capturar no solo la estructura de las interacciones, sino también su evolución temporal y la influencia de múltiples variables interdependientes, los sistemas dinámicos representan un marco teórico y práctico indispensable para modelar, analizar y comprender los sistemas complejos del mundo real, a causa de su capacidad para capturar la evolución temporal de un sistema, así como las interacciones entre múltiples variables interdependientes. Estos modelos son esenciales para representar fenómenos que cambian a lo largo del tiempo, reflejando dinámicas no lineales, interdependencias entre componentes y propiedades emergentes que no pueden observarse al analizar los elementos de forma aislada. La versatilidad de los sistemas dinámicos permite su aplicación en diversos campos interdisciplinarios, en los que se analizan problemáticas propias de la biología, economía, ingeniería, ciencias sociales, entre otros.

En biología, por ejemplo, los sistemas dinámicos se utilizan para modelar el crecimiento poblacional, las interacciones ecológicas entre especies y la propagación de enfermedades, ofreciendo herramientas para prever escenarios críticos y

desarrollar estrategias de mitigación. En economía, estos modelos permiten analizar ciclos económicos, dinámicas de mercados financieros y patrones de consumo, ayudando a comprender las fluctuaciones globales y sus impactos. En ingeniería, los sistemas dinámicos son importantes para diseñar sistemas de control industrial, optimizar recursos energéticos y prever el comportamiento de estructuras bajo diversas condiciones operativas. Por su parte, en las ciencias sociales, se utilizan para modelar la difusión de información en redes sociales, la dinámica de grupos humanos y los procesos de toma de decisiones colectivas.

La integración de tecnologías avanzadas, como la inteligencia artificial, el aprendizaje automático y el Big Data, ha ampliado enormemente las capacidades de los sistemas dinámicos. Estas tecnologías permiten realizar simulaciones complejas en tiempo real, manejar incertidumbres en sistemas con información incompleta y desarrollar modelos adaptativos que evolucionan con los datos disponibles (Bagdasaryan, 2012).

A la par de estos avances tecnológicos, los enfoques teóricos han evolucionado para proporcionar nuevas perspectivas en el análisis de sistemas complejos. El matroide invariante y sus bases (M, BM), combinados con la teoría de grafos de Ramsey, constituyen un enfoque teórico versátil para el análisis de fenómenos emergentes en sistemas complejos. Este enfoque se basa en la identificación de dos magnitudes centrales: la potencia del fenómeno emergente y la complejidad estructural asociada, las cuales permiten describir, cuantificar y analizar cómo las interacciones locales y las estructuras globales de un sistema generan comportamientos emergentes. La potencia del fenómeno emergente mide la intensidad o el impacto de dicho fenómeno sobre el sistema, mientras que la complejidad estructural evalúa las interdependencias, relaciones y configuraciones que subyacen al fenómeno.

A diferencia de los enfoques físicos tradicionales, el matroide simplifica los cálculos al trabajar con configuraciones geométricas complejas, como la pregeometría. Esta simplicidad no solo reduce la necesidad de grandes cantidades de datos, sino que también permite realizar análisis ágiles y adaptativos. La integración de la teoría de grafos de Ramsey añade un nivel adicional de profundidad analítica al enfoque, al permitir explorar patrones específicos y condiciones necesarias para la aparición de ciertas configuraciones en sistemas con múltiples componentes. Esto cobra especial importancia en contextos en los que los elementos del sistema presentan interacciones no lineales y dan lugar a propiedades emergentes que dependen de la complejidad estructural de dichas interacciones. La teoría de grafos de Ramsey proporciona herramientas matemáticas robustas para identificar estas condiciones y establecer relaciones entre las configuraciones locales y los patrones globales (Bila et al., 2021).

Desde esta perspectiva, la representación matemática de sistemas complejos se ha extendido más allá del ámbito teórico hacia aplicaciones en disciplinas como la economía, la gestión de recursos y el estudio de impactos urbanos. Una estrategia eficaz para atender esta diversidad consiste en traducir las interacciones y dependencias dentro del sistema a un formato análogo al de las reacciones químicas. Este enfoque de mapeo estructural facilita la representación del sistema y permite la aplicación de modelos matemáticos ya establecidos para evaluar su dinámica y

estabilidad. Al estructurar formalmente las relaciones entre los elementos del sistema, es posible identificar patrones recurrentes, evaluar posibles transiciones críticas y desarrollar estrategias de intervención precisas para optimizar su comportamiento.

Lo que hace único al enfoque de mapeo de sistemas complejos es su capacidad para ser lo suficientemente general como para adaptarse y escalar a diferentes disciplinas, y a la vez ser lo suficientemente detallado como para proporcionar información práctica y valiosa en escenarios reales. Por ejemplo, en mercados económicos, este método podría identificar las relaciones entre los agentes del mercado y predecir cómo un cambio en un componente crítico podría afectar al sistema en su conjunto. En la gestión de recursos, podría modelar el impacto de la explotación o conservación de recursos naturales, evaluando la sostenibilidad de distintas estrategias. En el caso de la suburbanización, podría analizar los efectos de cambios demográficos y estructurales en comunidades específicas, ayudando a prever impactos en infraestructura, economía local y medio ambiente (Pardi & Burina, 2024).

Para fortalecer estos modelos y mejorar su capacidad predictiva, es esencial recurrir a técnicas estadísticas que permitan analizar y comprender fenómenos en una amplia variedad de disciplinas. Estas herramientas, aplicables tanto a modelos de respuesta única como a modelos de múltiples respuestas, representan un avance significativo en el campo del modelado al proporcionar el tratamiento de sistemas con alta dimensionalidad, interacciones no lineales e incertidumbre inherente. La combinación del mapeo estructural con técnicas estadísticas robustas no solo optimiza la representación de los sistemas complejos, sino que también proporciona una base cuantitativa sólida para la toma de decisiones basada en datos, mejorando la capacidad de predicción y el desarrollo de soluciones estratégicas.

Entre las técnicas representativas, se encuentran aquellas destinadas a la discriminación de modelos, un proceso importante para identificar el modelo que mejor se ajusta a los datos disponibles y que ofrece la mayor capacidad predictiva. Estas estrategias incluyen el uso de parámetros intrínsecos y no intrínsecos, métodos bayesianos y técnicas de discriminación basadas en la probabilidad. Los parámetros intrínsecos se centran en las propiedades del modelo, mientras que los no intrínsecos pueden variar en función del contexto o los datos específicos. Por otro lado, los métodos bayesianos aportan un enfoque probabilístico que integra conocimiento previo y gestiona la incertidumbre con alta solidez. Por ello, las técnicas probabilísticas resultan útiles para comparar modelos y valorar su rendimiento en términos numéricos (Iyengar & Rao, 1983).

Por otro lado, la teoría del caos ofrece un marco unificador para las principales métricas de complejidad desarrolladas, permitiendo capturar propiedades esenciales como la sensibilidad a las condiciones iniciales y patrones dinámicos aparentemente aleatorios. Esta combinación metodológica no solo amplía las capacidades del aprendizaje automático, sino que también ofrece herramientas robustas y adaptativas para tratar problemas complejos en diversos campos como la ingeniería, la biología y la economía (Gao & Xu, 2021).

En este proceso de discriminación y optimización de modelos, la incorporación de la técnica de inteligencia artificial aplicadas al modelado matemático de sistemas complejos busca no solo comprender sus interacciones y efectos no lineales, sino también proporcionar herramientas innovadoras para afrontar su dinámica intrínseca. Estas técnicas permiten analizar el comportamiento de los sistemas a partir de grandes volúmenes de datos, extrayendo patrones y regularidades que de otro modo serían difíciles de detectar. Sus aplicaciones son amplias y destacan en campos como el reconocimiento de imágenes y voz, el análisis de series temporales, la predicción de eventos y la optimización de procesos (Zhao, 2024).

En paralelo, la representación matemática de estos sistemas requiere herramientas que integren tanto los componentes deterministas como los estocásticos, permitiendo capturar la incertidumbre inherente en su evolución. Bajo esta perspectiva, las ecuaciones diferenciales estocásticas (SDEs) y las ecuaciones diferenciales parciales estocásticas (SPDEs) representan herramientas matemáticas esenciales para modelar y analizar sistemas complejos sometidos a influencias aleatorias. Estas ecuaciones permiten capturar dinámicas que dependen tanto de componentes deterministas como de factores estocásticos, siendo particularmente útiles en contextos donde la incertidumbre juega un papel importante, como en procesos biológicos, fenómenos físicos o sistemas financieros. Comprender la dinámica de los sistemas modelados mediante SDEs y SPDEs es indispensable para obtener predicciones precisas y tomar decisiones informadas basadas en dichos modelos.

Sin embargo, las técnicas numéricas tradicionales a menudo enfrentan serias limitaciones al evaluar las complejidades y propiedades inherentes a estas ecuaciones, como la no linealidad, las dependencias temporales y las fluctuaciones estocásticas. Esta investigación profundiza en una selección de métodos numéricos diseñados específicamente para superar estos desafíos, destacando su eficacia y fiabilidad en la resolución de SDEs y SPDEs. Además, se abordan los retos técnicos que surgen durante la implementación de estos métodos, incluyendo cuestiones de estabilidad, precisión y eficiencia computacional, y se presentan estrategias innovadoras para afrontarlos. Este análisis no solo busca optimizar el uso de herramientas numéricas en el estudio de sistemas estocásticos, sino también fomentar su aplicación en una amplia variedad de disciplinas científicas y técnicas (Patients et al., 2012).

El desarrollo de modelos matemáticos efectivos no solo depende de la formulación de ecuaciones y su resolución numérica, sino también de la capacidad de extraer conocimiento significativo a partir de los datos. A partir de este enfoque las inferencias basadas en modelos son para comprender la complejidad de los sistemas y para desarrollar modelos que no solo sean confiables, sino también altamente informativos. Este trabajo de investigación profundiza en el papel que desempeña la inferencia basada en modelos en el contexto actual de sistema computacional, un campo donde la interacción entre teoría, datos y simulación es importante para desentrañar los comportamientos de sistemas complejos.

Las inferencias en la práctica y examinan las bases teóricas y materiales que sustentan su eficacia, destacando los principios que las hacen valiosas para el análisis y la

predicción. A través de su estudio, subrayan la importancia del razonamiento basado en modelos en las actividades diarias de resolución de problemas científicos. También enfatizan el papel crítico de los procesos de depuración en modelos parciales o incompletos, los cuales no solo mejoran la calidad de los modelos, sino que también facilitan avances en la inferencia científica y el descubrimiento de nuevos conocimientos, especialmente en sistemas complejos donde la incertidumbre es una constante (MacLeod, 2020).

2.3.2 Metodologías para la construcción de modelos de sistemas complejos con enfoque en índices o indicadores

En la literatura, se han desarrollado y adaptado diversas metodologías para la construcción de modelos de sistemas complejos con un enfoque basado en índices, lo que ha representado una contribución significativa al campo. Estas metodologías suelen organizarse en etapas claramente definidas, orientadas a la representación, análisis e interpretación de la complejidad inherente al sistema. A diferencia de los indicadores, que reflejan una única variable o dimensión del sistema, los índices integran múltiples indicadores en una medida compuesta, permitiendo una visión holística y sintetizada del fenómeno analizado. La implementación de estas metodologías posibilita modelar de manera rigurosa la dinámica y las interacciones entre los componentes del sistema, proporcionando un marco analítico robusto que facilita tanto la evaluación integral como la toma de decisiones fundamentadas.

El esquema de las fases de la metodología para la construcción de índices, a juicio de (Bonaño, 2002), especifica las condiciones técnicas que, a priori, deben exigirse a un índice:

- Existencia y determinación: La función matemática que define el indicador debe existir y tener una solución perfectamente determinada.
- Exhaustividad: El indicador compuesto debe ser tal que aproveche al máximo, sin redundancia y en forma útil la información suministrada por los indicadores y variables que lo componen.
- Monotonía: El indicador ha de responder positivamente al cambio positivo de las componentes y viceversa. Ello obliga, en algunos casos, a cambiar el signo de las variables que lo componen cuyas correlaciones pudieran estar invertidas.
- Unicidad: El indicador compuesto ha de ser único para una situación dada.
- Invariancia: El indicador debe ser invariante frente a cambios de origen o de escala de sus componentes.
- Homogeneidad: La función matemática que define al indicador compuesto: debería ser homogénea de grado 1.

De acuerdo a la Organización para la Cooperación Económica y Desarrollo (OECD, 2008) las fases que debe contener la metodología de desarrollo de índices son las siguientes:

Marco teórico: Proporciona la base para la selección y combinación de variables. En un indicador compuesto significativo bajo una aptitud para el propósito.

Selección de datos: Se debe analizar la solidez analítica, mensurabilidad, cobertura de países y pertinencia de los indicadores para el fenómeno que se mide y la relación entre sí.

Imputación de datos faltantes: Es necesario para proporcionar un conjunto de datos completo (por ejemplo, por medio de imputación única o múltiple).

- *Análisis multivariable:* Debe usarse para estudiar la estructura general del conjunto de datos, evaluar su idoneidad y guiar las elecciones metodológicas posteriores.
- *Normalización:* Debe llevarse a cabo para que las variables sean comparables.
- *Ponderación y agregación:* Debe hacerse a lo largo de las líneas de la teoría subyacente estructura.
- *Análisis de incertidumbre y sensibilidad:* Debe llevarse a cabo para evaluar la robustez del material compuesto indicador en términos del mecanismo para incluir o excluir un indicador, el esquema de normalización, la imputación de faltantes datos, la elección de los pesos y el método de agregación.
- *Volver a los datos:* Es necesario revelar los principales impulsores de un bien o mal en general. actuación. La transparencia es primordial para un buen análisis y elaboración de políticas.
- *Enlaces a otros indicadores:* Debe hacerse para correlacionar el indicador compuesto (o sus dimensiones) con indicadores existentes (simples o compuestos), así como para identificar vínculos a través de regresiones.
- *Visualización de los resultados:* Debe recibir la debida atención, dado que la visualización puede influir (o ayudar a mejorar) la interpretabilidad del indicador compuesto.

A criterio de (Schuschny & Soto, 2009), la construcción de un indicador compuesto con un enfoque metodológico riguroso implica el análisis de varios aspectos importantes, entre ellos:

- *Desarrollo de un marco conceptual:* En esta etapa se deben seleccionar las variables y establecer la manera en que éstas se combinarán entre sí a nivel conceptual.

- *Selección de los indicadores:* La selección de cada uno de los indicadores debe realizarse sobre la base de su relevancia, su calidad, la frecuencia, de muestreo y su disponibilidad al dominio público.
- *Análisis multivariado:* La relación que debe existir entre las variables deben ser cuidadosamente analizadas antes de proceder a integrar toda la información en un solo indicador. Para ello se puede recurrir a diversos métodos de síntesis estadística. El análisis combinado de la información puede realizarse sobre la base de dos criterios, a saber:
- *Análisis de relaciones entre variables:* agrupando las variables en sub-indicadores, definiendo una estructura anidada y determinando si la selección de éstas, resulta apropiada para describir el fenómeno en cuestión.
- *Estudio de relaciones entre unidades de análisis:* estableciendo grupos de unidades de análisis vinculadas por su similitud.
- *Imputación de datos perdidos:* En esta etapa se debe analizar el tipo de información perdida para implementar técnicas de imputación de datos adecuadas.
- *Normalización de los datos:* La existencia de varias fuentes de información ocasiona que las variables que se seleccionan, se encuentren expresadas nominalmente en unidades económicas o físicas (dólares PPP, \$, m, km, km², ha, yardas, pulgadas, cm, Tn, kg, por ejemplo), en fracciones de una determinada unidad (cantidades per cápita, por hectárea, por km², en kg/m³, etc.), en forma de tasas de variación o como porcentajes de variación, unidades con relación a un año base, etc. Por lo anterior es necesario normalizar la información para evitar la congregación de variables de unidades de medida distintas y la aparición de fenómenos dependientes de la escala.
- *Ponderación de la información:* En esta etapa se establecen los diferentes factores de peso que den cuenta de la importancia relativa de cada indicador en el agregado. La manera en que se pese la información disponible definirá en forma determinante el valor final del indicador.
- *Agregación de la información:* En esta etapa se agregan todas las variables o sub-indicadores en un indicador sintético.
- *Análisis de robustez y sensibilidad:* Para incrementar la transparencia del proceso de diseño del indicador. Es necesario juzgar lo realizado críticamente y evaluar la sensibilidad frente a cambios en la elección de las variables pues podría ocurrir que pequeños cambios de la arquitectura del indicador compuesto den lugar a grandes alteraciones de los valores o rankings obtenidos.

La construcción de un indicador compuesto con rigor metodológico, conforme a (Blasco-Blasco et al., 2020), requiere el análisis de diversas cuestiones, entre ellas:

- *Selección de indicadores parciales:* Estudiar y analizar detenidamente qué variables incluir a partir de su relevancia.
- *Normalización:* Las unidades y el rango en las que están medidos los indicadores deben ser comparables.
- *Elección de las ponderaciones:* Se estudia la importancia de cada variable en el indicador teniendo en cuenta el marco teórico, y en función de estudios previos o juicio experto se puede determinar el peso de cada una.
- *Método de agregación:* Determinar el objetivo de maximizar o minimizar el valor de cada indicador parcial.

En el estado del arte se proponen diferentes etapas o procesos que deben seguir las metodologías para la realización de minería de datos, se analizaron las etapas de las siguientes metodologías.

KDD (Knowledge Discovery from Databases): La metodología de KDD consta de cinco fases diferenciadas (Orallo José Hernández et al., 2004).

- *Fase de integración y recopilación de datos:* En esta fase se obtienen los datos de distintas fuentes de información y se integran bajo un formato de estandarización.
- *Fase de selección, limpieza y transformación:* En esta fase se corrigen los datos, se eliminan los datos, y se decide qué hacer con aquellos datos que estén incompletos. También se realiza una selección de aquellos datos que son relevantes para el proceso de extracción de conocimiento que se desea realizar.
- *Fase de minería de datos:* En esta fase se decide cuál es la tarea (agrupar, clasificar, etc.) que se va a realizar y se elige el método que se va a emplear para ello.
- *Fase de evaluación e interpretación:* En esta fase se evalúa e interpreta los patrones observados, con el fin de poder entender el resultado obtenido.
- *Fase de difusión y uso:* La última fase del proceso de KDD tiene como objetivo utilizar el nuevo conocimiento adquirido y hacer que dicho conocimiento sea empleado por todos los usuarios posibles.

CRISP-DM (Cross Industry Standard Process for Data Mining), esta metodología consta de seis fases que se describen a continuación (Espinosa Zúñiga, 2020).

- *Comprensión del problema o negocio:* En esta etapa se identifica, se entiende y se delimita la problemática, así como la identificación de los requisitos, supuestos, restricciones y beneficios del proyecto.
- *Comprensión de datos:* En esta etapa se realiza la recolección de datos a partir de fuentes identificadas. Se realiza la descripción de datos en base al tipo, formato, volumetría y significado de cada dato. Así mismo, se realiza la exploración de datos

mediante pruebas estadísticas básicas que permitan conocer las propiedades de los datos a fin de entenderlos lo mejor posible.

- *Preparación de datos:* En esta etapa se realiza la limpieza de los datos que consiste en normalizar, discretizar y tratar los valores ausentes. Se realiza el tratamiento de valores duplicados e imputación de datos. Así mismo se realiza la creación de indicadores que potencien la capacidad predictiva. Durante esta fase, se transforma el formato o la estructura de ciertos datos sin modificar su significado, con el objetivo de prepararlos para técnicas particulares en el proceso de modelado.
- *Modelado:* En esta etapa se realiza la selección de la técnica de modelado en base al problema a resolver, los datos disponibles, las herramientas de minería de datos disponibles, así como el dominio de la técnica elegida. Se realiza la selección de datos de prueba, entrenamiento y de validación. La obtención del modelo se realiza mediante un proceso iterativo enfocado en la modificación de sus parámetros.
- *Evaluación del modelo:* En esta etapa se determina la calidad del modelo con base en el análisis de ciertas métricas estadísticas del mismo, comparando los resultados con resultados previos, o bien, analizando los resultados con apoyo de expertos en el dominio del problema. Con base en los resultados de esta etapa se determina seguir con la última fase de la metodología, regresar a alguna de las etapas anteriores o incluso partir de cero con un nuevo proyecto.
- *Implementación del modelo:* Esta etapa explota, mediante acciones concretas, el conocimiento adquirido mediante el modelo. Aquí también es importante documentar los resultados con claridad para el usuario final y asegurarse de que todas las etapas de la metodología se documenten debidamente para hacer una revisión del proyecto a fin de obtener lecciones aprendidas durante el proceso. Asimismo, monitorear las acciones para detectar áreas de oportunidad o incluso nuevos problemas.

SEMMA (Sample, Explore, Modify, Model, Assess), esta metodología consta de cinco fases que se describen a continuación (Omari Firas, 2023):

- *Muestreo (Sample):* Identificación y selección del tamaño de las muestras con el objetivo de que el conjunto de datos sea suficientemente grande para que resulte representativo o para permitir el procesamiento.
- *Exploración (Explore):* Se realiza la identificación de las relaciones, tendencias que existe entre las variables y cualquier otra información para comprender el fenómeno del mundo real.
- *Modificación (Modify):* En esta etapa se realiza la creación, selección y transformación de las variables para dirigir el proceso de modelado. Se utiliza para encontrar valores atípicos, reducir el número de variables y seleccionar las más significativas.

- *Modelado (Model)*: En esta etapa se determina el modelo adecuado para predecir la o las variables de salida a partir de la o las variables de entrada, seleccionando algoritmos como redes neuronales, árboles de decisiones, regresión logística, series temporales, y análisis en componentes principales.
- *Evaluación (Assess)*: En esta etapa se evalúa el funcionamiento del modelo en su conjunto en cuanto a fiabilidad y utilidad.

2.3.3 Metodologías de modelado y simulación para la construcción de modelos de comportamiento en sistemas complejos

Las metodologías de modelado y simulación para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos han progresado significativamente en años recientes, reflejando la creciente complejidad de estos sistemas en múltiples disciplinas, como la ingeniería, la economía y las ciencias sociales. Estas metodologías están diseñadas para captar la naturaleza no lineal y multivariada de los sistemas complejos, donde el comportamiento emergente es determinante para entender el funcionamiento general del sistema.

Uno de los enfoques destacados en este contexto es la simulación basada en agentes (SBA), que permite modelar las interacciones entre individuos (o agentes) y su entorno. Esta técnica ha demostrado ser especialmente útil para investigar cómo las decisiones de los agentes influyen en el comportamiento colectivo del sistema (Cruz, 2019). La SBA se integra eficazmente con otras metodologías, como las metodologías híbridas, que combinan simulaciones discretas con enfoques continuos. Estas últimas buscan capturar la complejidad de los sistemas mediante la integración de diferentes tipos de modelado y simulación, facilitando así una representación precisa de las dinámicas interactivas (Nguyen et al., 2021).

El uso de redes neuronales y técnicas de inteligencia artificial (IA) en el modelado de sistemas complejos ha emergido como una tendencia significativa que complementa estas metodologías. Estas herramientas permiten representar comportamientos emergentes y no lineales, lo que puede llevar a predicciones exactas sobre el comportamiento del sistema (Dahia & Szabo, 2024). Además, el análisis de sensibilidad se vuelve importante en este contexto, ya que permite a los investigadores comprender cómo las variaciones en los parámetros del modelo pueden influir en los resultados, facilitando así la validación y calibración de modelos complejos (Grznár et al., 2024). A su vez, la incorporación de técnicas de análisis de datos masivos (big data) transforma la forma en que se recopilan y analizan los datos, enriqueciendo el proceso de modelado y mejorando la robustez de los modelos desarrollados.

Con el avance en plataformas de simulación y herramientas de software ha hecho que estas metodologías sean accesibles y aplicables en diversas áreas. Herramientas como AnyLogic, NetLogo y MATLAB han sido esenciales para la implementación de modelos complejos, permitiendo a investigadores y profesionales crear simulaciones sofisticadas sin necesidad de una programación intensiva (George-Williams et al., 2022). Esta evolución en las herramientas de simulación y en las metodologías empleadas refleja un movimiento hacia una mayor integración y colaboración en el

estudio de sistemas complejos, lo que abre nuevas posibilidades para entender y gestionar dinámicas altamente cambiantes en el mundo real.

2.3.4 Metodologías de minería de datos y ciencia de datos con el enfoque de construcción de modelos de comportamiento de sistemas complejos

Las metodologías de minería de datos han experimentado un desarrollo significativo en los últimos años, adaptándose a las crecientes necesidades de análisis y comprensión de sistemas complejos altamente dinámicos. Este progreso se ha visto impulsado por el aumento en la disponibilidad de grandes volúmenes de datos y la sofisticación de las técnicas de análisis, que permiten extraer patrones y conocimientos útiles para la toma de decisiones en diversas áreas como la ingeniería, la biología, la economía y las ciencias sociales. La minería de datos proporciona herramientas poderosas para modelar el comportamiento de sistemas complejos al identificar relaciones no evidentes entre variables, permitiendo a los investigadores y profesionales comprender mejor la dinámica de estos sistemas.

Uno de los enfoques destacados en minería de datos es el aprendizaje automático, que se basa en algoritmos que pueden aprender de los datos y hacer predicciones o clasificaciones sin intervención humana directa. Tal como exponen (Ramírez et al., 2020), las técnicas de aprendizaje supervisado, como los árboles de decisión y las máquinas de soporte vectorial, han demostrado ser efectivas para predecir comportamientos emergentes en sistemas complejos, permitiendo una mayor precisión en las simulaciones y modelos. Además, el aprendizaje no supervisado, que incluye técnicas como el clustering y la reducción de dimensionalidad, para descubrir patrones ocultos y agrupar datos similares, lo que resulta valioso para la comprensión de la estructura interna de los sistemas (Andres & Sandoval, 2024).

La integración de técnicas de minería de datos con el análisis de series temporales ha permitido a los investigadores modelar y predecir el comportamiento dinámico de los sistemas a lo largo del tiempo. La modelación de series temporales combina datos históricos con algoritmos de minería para identificar tendencias, ciclos y patrones estacionales, proporcionando así una visión integral de cómo un sistema evoluciona y se comporta en diferentes condiciones (Gallardo-Pérez et al., 2020). Esta combinación es particularmente relevante en áreas como la economía y la ecología, donde los sistemas suelen presentar fluctuaciones significativas por como consecuencia de diversos factores.

Otra tendencia relevante es el uso de algoritmos de redes neuronales profundas, que han revolucionado el campo de la minería de datos al permitir el procesamiento de grandes volúmenes de datos no estructurados, como imágenes y texto. Estas técnicas han demostrado ser efectivas para modelar comportamientos complejos y no lineales en sistemas dinámicos, lo que permite realizar predicciones precisas. La capacidad de las redes neuronales profundas para extraer características relevantes automáticamente las convierte en una herramienta valiosa para el análisis de datos complejos. Además, la minería de datos está cada vez más integrada con técnicas de big data, que permiten el procesamiento y análisis de grandes volúmenes de datos en tiempo real. La combinación de estas metodologías con plataformas como Hadoop y

Spark ha facilitado el desarrollo de sistemas capaces de manejar flujos de datos masivos, lo que es importante para sistemas que operan en entornos dinámicos y cambiantes. La capacidad de analizar datos en tiempo real permite a las organizaciones responder rápidamente a las dinámicas del sistema, mejorando así la toma de decisiones.

En años recientes, las metodologías en ciencia de datos han experimentado una evolución notable, centrada en el desarrollo de modelos que pueden describir el comportamiento de sistemas complejos altamente dinámicos. Esta evolución ha sido impulsada por la creciente disponibilidad de grandes volúmenes de datos y la necesidad de extraer conocimiento útil a partir de estos datos para la toma de decisiones en tiempo real. Las metodologías actuales integran técnicas de análisis estadístico, aprendizaje automático e inteligencia artificial, permitiendo a los investigadores y profesionales modelar fenómenos complejos y dinámicos en áreas como la salud, la economía, la ingeniería y el medio ambiente.

Una de las tendencias predominantes en la ciencia de datos es el uso de enfoques basados en el aprendizaje automático para analizar y predecir comportamientos emergentes en sistemas dinámicos. Conforme a (Méndez Gurrola et al., 2020), estos enfoques no solo facilitan la identificación de patrones en grandes conjuntos de datos, sino que también permiten la construcción de modelos predictivos que pueden ajustarse y adaptarse a cambios en el entorno. Además, la implementación de algoritmos de aprendizaje profundo ha demostrado ser eficaz en la modelización de relaciones no lineales y complejas, lo que permite comprender el comportamiento de sistemas dinámicos.

El uso de la minería de datos es otra metodología significativa en la ciencia de datos. Esta técnica permite a los investigadores descubrir patrones ocultos y relaciones significativas en los datos, lo que es de gran relevancia para el desarrollo de modelos precisos en sistemas complejos. (Urbina-Nájera et al., 2020) enfatizan que la minería de datos, combinada con técnicas de visualización, ayuda a interpretar los resultados y a comunicar hallazgos a las partes interesadas, facilitando así una mejor toma de decisiones. Además, las metodologías de ciencia de datos han incorporado cada vez más el uso de análisis de series temporales, que es esencial para modelar comportamientos en sistemas que evolucionan a lo largo del tiempo. Este enfoque permite a los investigadores realizar pronósticos y detectar tendencias, lo que es importante para la gestión y control de sistemas complejos (Gallardo-Pérez et al., 2020). Las técnicas de análisis de series temporales han sido aplicadas exitosamente en áreas como la gestión de recursos hídricos y la planificación urbana, donde la dinámica temporal de los sistemas es un factor decisivo (Box et al., 2015).

El uso de herramientas y plataformas de ciencia de datos ha facilitado la implementación de estas metodologías. Herramientas como Python, R y plataformas de análisis en la nube han proporcionado a los científicos de datos los recursos necesarios para procesar y llevar a cabo un análisis eficiente de grandes volúmenes de datos (Malberti et al., 2021). Estas herramientas no solo mejoran la productividad, sino que también permiten una mayor colaboración entre equipos multidisciplinarios en el desarrollo de modelos para sistemas complejos.

2.3.5 Cómputo evolutivo para el modelado matemático con el enfoque de construcción de modelos de comportamiento de sistemas complejos

El cómputo evolutivo es una disciplina dentro de la inteligencia artificial y la computación que se inspira en los mecanismos de la evolución biológica para resolver problemas complejos mediante técnicas de optimización y búsqueda. Su fundamento radica en la aplicación de procesos análogos a la selección natural, como la reproducción, la mutación y la supervivencia de los más aptos, con el propósito de encontrar soluciones óptimas en entornos de alta dimensionalidad y no linealidad. A diferencia de los métodos deterministas convencionales, que dependen de formulaciones analíticas o técnicas de búsqueda exhaustiva, el cómputo evolutivo facilita la exploración eficiente de grandes espacios de soluciones, adaptándose a contextos donde la estructura matemática del problema puede ser incierta o dinámica (Goldberg, 1989).

El modelado matemático en este contexto se basa en la formulación de una función de aptitud que cuantifica la calidad de cada posible solución dentro de una población. Esta función actúa como criterio de evaluación dentro del proceso evolutivo, permitiendo medir el desempeño de cada individuo en la resolución del problema planteado. De esta manera, la evolución de la población se orienta hacia regiones del espacio de búsqueda que presentan mejores valores en términos de optimización. En modelos matemáticos altamente complejos, donde las ecuaciones diferenciales o los sistemas de ecuaciones tradicionales resultan insuficientes para describir el comportamiento del sistema, los algoritmos evolutivos proporcionan un enfoque alternativo basado en la heurística adaptativa.

El proceso inicia con una población inicial de soluciones candidatas, que son sometidas a evaluación a través de la función de aptitud. Posteriormente, se aplican operadores evolutivos como la cruce (crossover), que intercambia información entre soluciones para generar nuevas combinaciones, y la mutación, que introduce pequeñas variaciones para incrementar la diversidad y evitar la convergencia prematura a óptimos locales. A lo largo de múltiples generaciones, las soluciones más prometedoras son seleccionadas para reproducirse, permitiendo que la población evolucione progresivamente hacia una solución óptima o casi óptima. Este enfoque ha demostrado ser particularmente útil en problemas donde el espacio de búsqueda es demasiado amplio o no estructurado, como en la optimización de sistemas dinámicos, la predicción de series temporales y la identificación de patrones en grandes volúmenes de datos (Goldberg, 1989).

En el ámbito del cómputo evolutivo, la programación genética representa una extensión del modelado matemático, ya que permite la generación automática de modelos en forma de programas o ecuaciones simbólicas. Inspirada en la evolución biológica, esta técnica opera mediante la representación de soluciones en estructuras manipulables, como árboles sintácticos o expresiones algebraicas, que evolucionan a lo largo de múltiples generaciones. Cada modelo generado es evaluado en base a su capacidad predictiva o explicativa respecto al problema en cuestión. Gracias a la aplicación de operadores evolutivos, la programación genética ha sido empleada en la construcción de ecuaciones diferenciales, modelos de simulación y en la

identificación de relaciones funcionales dentro de datos experimentales, lo que la convierte en una herramienta importante para la inferencia de modelos matemáticos en dominios donde las relaciones subyacentes son desconocidas o demasiado complejas para ser expresadas mediante métodos tradicionales (Koza, 1994).

El concepto de aptitud juega un papel importante en la evolución de estos modelos. La función de aptitud asigna un valor cuantitativo a cada individuo en función de su desempeño en la resolución del problema planteado. Cuanto mayor sea la puntuación de aptitud, más efectiva será la solución dentro del contexto del modelo matemático. Durante el proceso evolutivo, estos valores guían la selección de soluciones, favoreciendo aquellas con mayor ajuste a los datos o mayor capacidad de generalización. La correcta definición de la función de aptitud es esencial para garantizar la eficacia del algoritmo evolutivo y la calidad de los modelos obtenidos, ya que determina la dirección del proceso de optimización y evita sesgos en la exploración del espacio de soluciones (Goldberg, 1989).

Dentro del proceso evolutivo, el mejor individuo es aquel cuya puntuación de aptitud supera a la de los demás dentro de la población. Este individuo representa la mejor solución identificada hasta ese momento y es utilizado como referencia para la generación de nuevas soluciones en iteraciones posteriores. La evolución del mejor individuo a lo largo del tiempo refleja el aprendizaje adaptativo del algoritmo y su capacidad para mejorar progresivamente la calidad de los modelos generados. A raíz de que la estructura del problema puede cambiar o refinarse durante el proceso evolutivo, la definición de la "mejor solución" es dinámica y puede ajustarse a medida que el modelo matemático se perfecciona (Goldberg, 1989).

2.4 Conclusiones

Un sistema complejo altamente dinámico, por su naturaleza, tiene múltiples capas de interacción entre elementos y variables, lo que significa que cualquier perturbación o cambio en una parte del sistema puede afectar al sistema completo con efectos impredecibles. Las metodologías existentes para el desarrollo de índices y la minería de datos están más enfocadas en extraer patrones a partir de datos estáticos o en situaciones de cambio controlado, pero carecen de mecanismos para modelar adecuadamente la dinámica temporal y la retroalimentación entre componentes del sistema, aspectos primordiales en sistemas complejos. Asimismo, las metodologías de desarrollo de software, aunque ofrecen valiosas herramientas para la organización de proyectos y la entrega de software eficiente, no están diseñadas para capturar las dinámicas cambiantes y las interacciones no lineales que caracterizan a los sistemas complejos.

Este análisis de la literatura identifica un área de oportunidad al no encontrar una metodología específica o marco de trabajo ampliamente aceptado para el desarrollo de modelos de comportamiento en sistemas complejos altamente dinámicos. Esto genera una importante área de oportunidad, ya que la capacidad de modelar estos sistemas es cada vez más relevante en una amplia gama de campos, desde la ingeniería hasta las ciencias sociales, pasando por la biología, la economía y el análisis de datos a gran escala. Estos sistemas complejos pueden incluir ejemplos como redes

de tráfico urbano, ecosistemas biológicos, mercados financieros, redes de comunicación, entre otros, donde las dinámicas no son estáticas y requieren una constante reevaluación y ajuste.

La falta de metodologías diseñadas específicamente para estos sistemas no solo limita la capacidad para entender y predecir el comportamiento de dichos sistemas, sino que también restringe el desarrollo de soluciones más eficientes y adaptativas en contextos prácticos. Una metodología orientada a modelar el comportamiento en estos sistemas integra componentes en tiempo real, la retroalimentación continua, y la adaptación a las perturbaciones del entorno, permitiendo así la creación de modelos que no solo representan el estado actual del sistema, sino que también sean capaces de evolucionar con él.

Por tanto, se hace evidente la necesidad de diseñar una metodología que aborde específicamente el modelado de comportamiento en sistemas complejos altamente dinámicos. Esta metodología debe permitir integrar tanto el análisis estático como dinámico de las interacciones entre componentes. Además, deberían ser lo suficientemente flexibles como para adaptarse a diferentes dominios, ya que los sistemas complejos se encuentran en una variedad de áreas de estudio. Esta flexibilidad también sería determinante para poder implementar dicha metodología en contextos aplicados, en los que los modelos permitirían tanto entender el sistema como intervenir eficazmente en su funcionamiento.

Capítulo 3. Metodología propuesta para modelar sistemas complejos altamente dinámicos

3.1 Introducción

El modelado de sistemas complejos altamente dinámicos requiere una metodología estructurada que integre principios teóricos y herramientas computacionales capaces de capturar su naturaleza emergente, no lineal y multiescalar. En este capítulo, se presenta el desarrollo de una metodología sistemática para la construcción de modelos de comportamiento, detallando cada una de sus fases, desde la conceptualización y definición del sistema hasta su validación e interpretación. La propuesta metodológica se fundamenta en la interacción de enfoques provenientes de la teoría de sistemas, redes complejas y dinámica de sistemas, incorporando técnicas de modelado matemático y análisis de datos masivos. A través de una arquitectura modular y flexible, se busca establecer un marco analítico robusto que permita representar con precisión la evolución y las interacciones de los componentes del sistema en distintos escenarios.

3.2 Vista general de la metodología propuesta para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos

La metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos constituye un marco metodológico estructurado y adaptativo (ver Figura 3.1), cuyo objetivo principal es desarrollar modelos de comportamiento capaces de representar las interacciones dinámicas y emergentes entre los elementos que conforman el sistema complejo. A través de una integración de técnicas analíticas, modelado computacional y procesamiento de datos.

El enfoque propuesto facilita la identificación de patrones subyacentes, la evaluación de la influencia de cada componente en la dinámica global del sistema y la formulación de estrategias de intervención basadas en modelos de comportamiento. Además, permite entender la complejidad inherente de estos sistemas mediante una estructura metodológica flexible, para adaptarse a distintas fuentes de datos, condiciones cambiantes y entornos altamente inciertos.

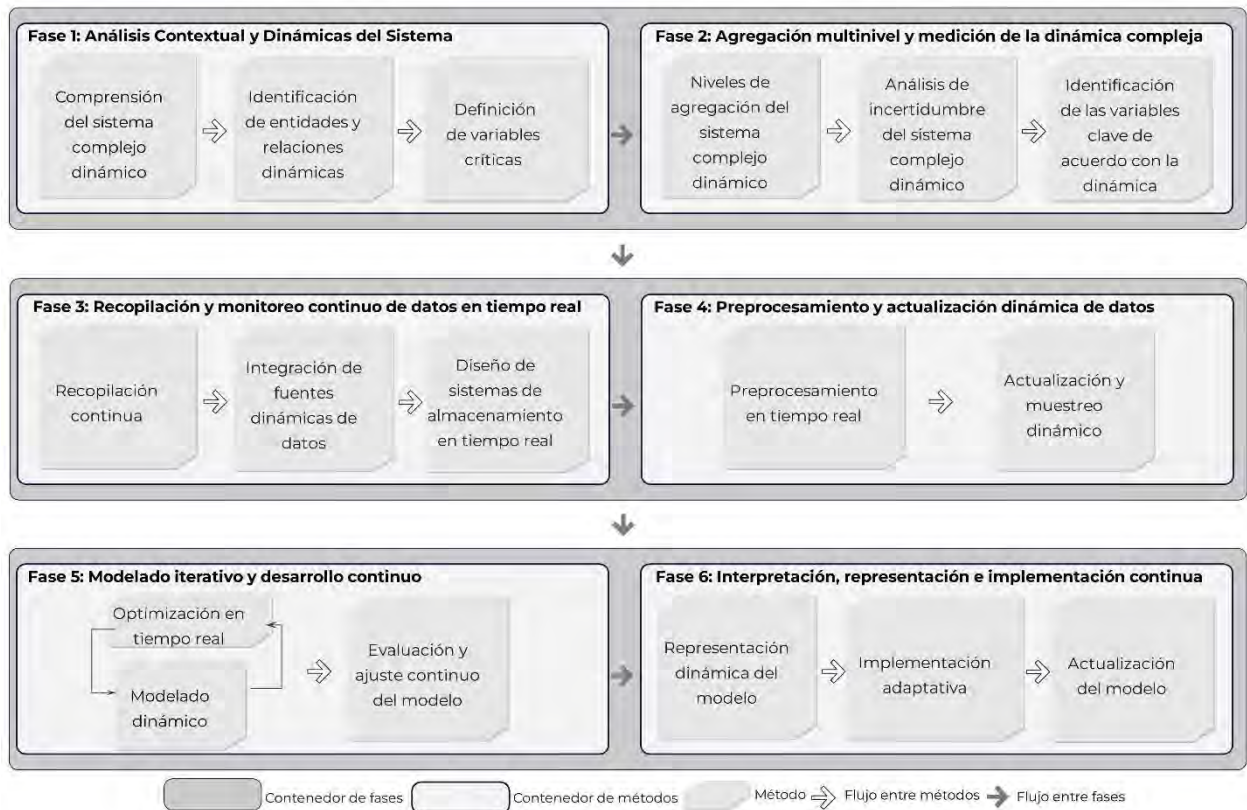


Figura 3.1 Metodología propuesta para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos

La metodología propuesta inicia con la Fase 1: Análisis Contextual y Dinámicas del Sistema, donde se identifican los elementos esenciales del sistema, sus componentes principales y las interacciones internas y externas que lo definen. En esta fase, se establecen los fundamentos conceptuales y metodológicos para su análisis, considerando factores como la naturaleza de los datos, la escala de observación y las posibles restricciones del sistema. La entrada de esta fase consiste en información estructural y conceptual del sistema bajo estudio, mientras que la salida es un marco conceptual que permite comprender las interdependencias y dinámicas del sistema. Posteriormente, la Fase 2: Agregación Multinivel y Medición de la Dinámica Compleja se centra en evaluar la incertidumbre y variabilidad en los datos con el propósito de identificar patrones emergentes y comprender las relaciones subyacentes entre las distintas variables.

A partir de esta caracterización, la Fase 3: Recopilación y Monitoreo Continuo de Datos en tiempo real realiza la recolección de grandes volúmenes de información de forma continua y estructurada, asegurando su actualización, relevancia y cobertura integral. En esta fase, se implementan mecanismos de adquisición de datos que pueden incluir sensores, registros históricos, flujos de datos en tiempo real o bases de datos interconectadas, aplicaciones de mapas, dependiendo de la naturaleza del sistema. La entrada está conformada por la definición de métricas y variables establecidas en la fase anterior, la Fase 4: Preprocesamiento y Actualización Dinámica de Datos se

encarga de transformar y adecuar los datos mediante procesos de limpieza, normalización y selección de subconjuntos representativos.

Con los datos preparados, se procede a la Fase 5: Modelado Iterativo y Desarrollo Continuo, donde se implementan técnicas de modelado matemático y computacional para representar el comportamiento del sistema. En esta etapa, se construyen modelos dinámicos que pueden incluir enfoques basados en ecuaciones diferenciales, redes complejas, cómputo evolutivo, inteligencia artificial o simulaciones computacionales, dependiendo de la naturaleza del sistema y los objetivos de la investigación. La Fase 6: Interpretación, Representación e Implementación Continua se centra en el análisis de los modelos obtenidos, su validación con datos reales y su implementación en entornos experimentales o de simulación. En esta fase, se interpretan los resultados, se generan representaciones visuales que faciliten su comprensión y se establecen estrategias de implementación para su aplicación en escenarios reales.

En conjunto, estas seis fases permiten enfrentar integralmente los desafíos asociados. al modelado de sistemas complejos altamente dinámicos, garantizando que los modelos generados sean representativos, adaptativos y aplicables en diferentes contextos. La estructura metodológica no solo proporciona un enfoque sistemático para la construcción de modelos, sino que también permite su evolución y mejora continua, garantizando que el conocimiento producido sea susceptible de refinamiento y aplicación eficiente en el análisis de sistemas reales.

3.2.1 Caso de estudio de Movilidad Vehicular

El caso de estudio en movilidad vehicular, un problema caracterizado por su complejidad inherente y la interacción no lineal de múltiples factores que generan dinámicas emergentes y adaptativas, sirve como ejemplo para ilustrar la metodología presentada en la Figura 3.1. El análisis de los métodos en sus distintas fases sigue un desarrollo progresivo y estructurado, permitiendo mostrar su implementación de forma clara y sistemática.

El caso de estudio de la movilidad vehicular jugó un papel esencial en este proceso, sirviendo como una base para explorar y comprender las interacciones y comportamientos característicos de los sistemas altamente dinámicos para implementar las fases y los métodos de la metodología para modelar sistemas complejos altamente dinámicos. Este caso no solo representó un ejemplo tangible de las dinámicas complejas, sino que también proporcionó un marco para validar las hipótesis iniciales, evaluar las fases y métodos propuestos y ajustar los enfoques empleados.

La elección de la movilidad vehicular en la Ciudad de México como objeto de estudio no fue arbitraria, sino resultado de una decisión estratégica, ya que este fenómeno encapsula características de gran trascendencia que la metodología buscaba modelar. El problema de la movilidad vehicular constituye un sistema complejo altamente dinámico influido por múltiples factores interdependientes. Entre ellos, destacan la densidad vehicular, que afecta la fluidez del tránsito; la infraestructura vial,

cuya calidad y diseño determinan la capacidad de la red para atender el tráfico; y los horarios pico, que reflejan patrones de actividad humana en momentos fundamentales del día. Asimismo, las condiciones climáticas, como la lluvia o la neblina, pueden alterar significativamente los patrones de conducción, mientras que eventos imprevistos, como accidentes o manifestaciones, generan interrupciones localizadas. Otros factores como la regulación del tránsito mediante semáforos y programas como el “Hoy No Circula,” el acceso al transporte público, los patrones de desarrollo urbano, la implementación de tecnologías de control de tráfico y el comportamiento de los conductores, también juegan un papel importante. La interacción de estos factores es dinámica tanto temporal como espacialmente, presentando desafíos analíticos que impulsaron el uso de técnicas robustas y adaptables para comprender y modelar su complejidad.

Otro aspecto que respaldó la elección de este caso de estudio fue la disponibilidad de conjuntos de datos representativos y accesibles, esenciales para el análisis de la movilidad vehicular. Entre las fuentes más relevantes se encuentran las bases de datos proporcionadas por organismos gubernamentales, como el Instituto Nacional de Estadística y Geografía (INEGI) y la Secretaría de Movilidad de la Ciudad de México, que recopilan información detallada sobre el flujo vehicular, infraestructura y patrones de uso. También sobresalen los datos generados por sistemas de monitoreo en tiempo real, como cámaras de tránsito, sensores instalados en calles y avenidas, y aplicaciones de mapas como Waze, Google Maps, HERE Maps, TomTom Traffic, Mapbox, entre otras. Estas herramientas proporcionan información sobre la congestión vehicular, velocidades actuales, presencia de incidentes, así como datos relacionados con el horario, el día y las condiciones específicas del tráfico, lo que permite un análisis más detallado y dinámico de la movilidad. Además, los datos históricos de incidentes de tráfico registrados por las aseguradoras y servicios de emergencia permiten analizar la relación entre eventos específicos y su impacto en la red vial. En las siguientes secciones, se detalla el desarrollo de cada fase implementada en la construcción de modelos de comportamiento para sistemas complejos altamente dinámicos.

Durante el proceso de implementación de los métodos en cada fase, se identificaron las actividades asociadas a cada uno, estableciendo así un conjunto estructurado de acciones recomendadas. Estas actividades se presentan en tablas organizadas que incorporan los elementos esenciales para proporcionar un marco de trabajo claro y estructurado. Las tablas que muestran las actividades sugeridas para cada método tienen cuatro columnas, descripción, actividades, roles y salidas, a continuación, se describe cada columna.

Descripción: Expone de forma clara el propósito del método y su importancia dentro del proceso general de modelado de sistemas complejos altamente dinámicos.

Actividades: Detalla las acciones recomendadas para implementar el método, asegurando que los pasos a seguir sean claros y ejecutables.

Roles: Especifica los perfiles profesionales, expertos o conocimientos necesarios para ejecutar las actividades, fomentando la colaboración interdisciplinaria y el aprovechamiento de conocimientos especializados.

Salidas: Enumera los productos esperados de la implementación del método, como diagramas, reportes, modelos o datos procesados, que son esenciales para el avance hacia las fases posteriores.

A continuación, se presentan las actividades correspondientes a cada fase de la metodología en el contexto del caso de estudio sobre movilidad vehicular, a partir del cual se han derivado las actividades genéricas que conforman los métodos de cada fase. Las actividades se organizan en tablas que resaltan, a un nivel abstracto, las acciones vinculadas a cada método, permitiendo su aplicación en diferentes contextos al seguir la metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos. La definición de estas actividades es el resultado directo del análisis del caso de estudio, proporcionando un marco estructurado y replicable que facilita la implementación de la metodología en otros sistemas complejos altamente dinámicos.

3.3 Análisis contextual y de la dinámica del sistema

Esta fase se orienta a la comprensión del fenómeno en su contexto específico, e incluye la identificación de variables, sus interrelaciones y los factores externos que influyen en su comportamiento. Para ello, se emplean análisis exploratorios y diagramas de causalidad, los cuales permiten mapear las relaciones entre variables y delimitar el alcance del sistema. Con este propósito, se diseñan y construyen tres métodos para la fase 1 que proporcionan una base conceptual para el desarrollo de los modelos, asegurando que estos representen con precisión la dinámica real del sistema analizado (ver Figura 3.2).



Figura 3.2. Fase 1 Análisis contextual y dinámicas del sistema de la metodología propuesta

Para comprender las acciones necesarias en la Fase 1: Análisis Contextual y de la Dinámica del Sistema, se necesita delinear el procedimiento aplicado en el caso de estudio sobre movilidad vehicular. Este análisis presenta una descripción estructurada de la ejecución de cada actividad dentro de los métodos empleados, permitiendo una visión clara de su implementación en el sistema complejo de movilidad vehicular.

3.3.1 Comprensión del sistema complejo dinámico de movilidad vehicular

Para analizar la movilidad vehicular desde la perspectiva de un sistema complejo, es importante identificar y comprender una serie de características inherentes a este tipo de sistemas, en esta etapa se aplicaron diversas técnicas desde la teoría de sistemas complejos, se emplearon grafos ponderados para representar interconexión, autómatas celulares (Nagel-Schreckenberg) para simular emergencia y autoorganización, la ecuación de Burgers para analizar no linealidad, y el modelo LWR para evaluar adaptabilidad. En ciencia de datos, se trabajó con un conjunto real de datos de movilidad, aplicando procesamiento, simulación y visualización en Python. Desde la minería de datos, se identificaron patrones y anomalías a través de modelos simplificados y simulaciones computacionales. Finalmente, se construyeron indicadores compuestos como la entropía de Shannon (para incertidumbre), un índice de congestión emergente y un indicador de conectividad dinámica, permitiendo sintetizar fenómenos clave del tráfico urbano.

Para caracterizar la movilidad vehicular como un sistema complejo, se llevó a cabo un proceso estructurado que permitió analizar sus dinámicas e interacciones no lineales. Este proceso permitió identificar los factores principales que influyen en su comportamiento, estableciendo un marco analítico sólido para su modelado y representación matemática. Este proceso se describe a continuación.

Paso 1: Se empleó un conjunto de datos de movilidad vehicular compuesto por 2,000 registros y 8 variables, cuya descripción se presenta en la Tabla 3.1. Un fragmento de estos datos está disponible en el siguiente enlace: <https://goo.su/iyQ4zo>

Tabla 3.1. Conjunto de variables de TomTom Traffic

Variable	Descripción
Frc	Tipo de carretera.
Currentspeed	La velocidad actual del tráfico en la carretera en metros por segundo (m/s) o kilómetros por hora (km/h).
Freeflowspeed	La velocidad libre de flujo en la carretera en metros por segundo (m/s) o kilómetros por hora (km/h).
Currenttraveltime	El tiempo de viaje actual en la carretera en segundos.
Freeflowtraveltime	El tiempo de viaje libre de flujo en la carretera en segundos.
Confidence	La confianza en los datos de tráfico para la carretera.
Roadclosure	Indicador de cierre de carretera.
Coordinates	Las coordenadas geográficas de la ubicación de la carretera.

Paso 2: Para identificar la característica de interconexión, se verifica la conectividad en la movilidad vehicular mediante la representación de un grafo (ver Figura 3.3), en el cual los vértices representan las calles y los arcos modelan las conexiones entre ellas, posibilitando el desplazamiento de una vía a otra. El peso de cada arco corresponde al tiempo de desplazamiento entre dos calles, el cual puede variar en función de la saturación vial, la hora del día, el día de la semana y otras variables contextuales. Esta representación gráfica evidencia la interconexión dinámica dentro del sistema de

movilidad vehicular, permitiendo la identificación de patrones emergentes y comportamientos no lineales que van más allá de la simple agregación de sus componentes individuales.

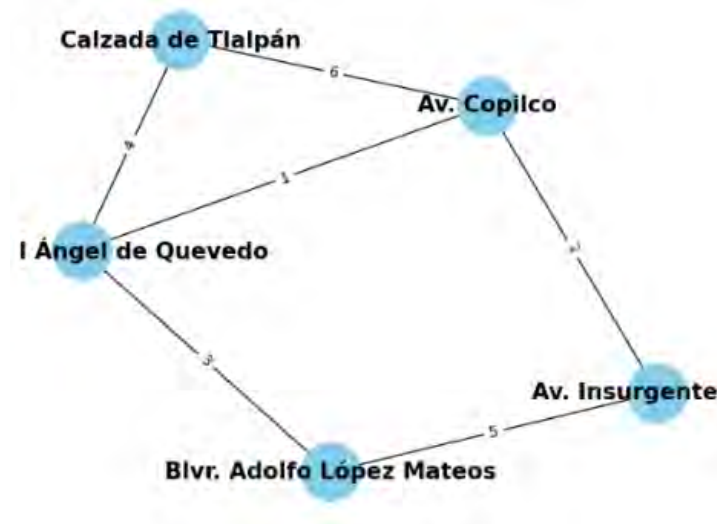


Figura 3.3. Grafo de la movilidad vehicular de una calle a otra de cinco calles de la Ciudad de México

Paso 3: Para comprobar la característica de emergencia en la movilidad vehicular, se implementó el modelo Nagel-Schreckenberg (Na-Sch) mediante un script en Python. Este modelo de autómatas celulares permite simular la evolución del tráfico a partir de reglas de transición sencillas que consideran factores como la velocidad máxima de cada vehículo, la distancia al vehículo precedente y la probabilidad de frenado aleatorio. El modelo Na-Sch se basa en una estructura matemática unidimensional L de celdas con condiciones de frontera periódicas, donde N vehículos se desplazan siguiendo reglas de actualización de posición $x(i, t)$ y velocidad $v(i, t)$ en el tiempo t . La simulación realizada con el modelo Na-Sch muestra la formación de patrones de congestión a partir de reglas locales de interacción entre vehículos, como se ilustra en la Figura 3.4. Este tipo de modelización proporciona una herramienta para comprender y predecir la dinámica emergente de la movilidad vehicular desde una perspectiva de sistemas complejos.

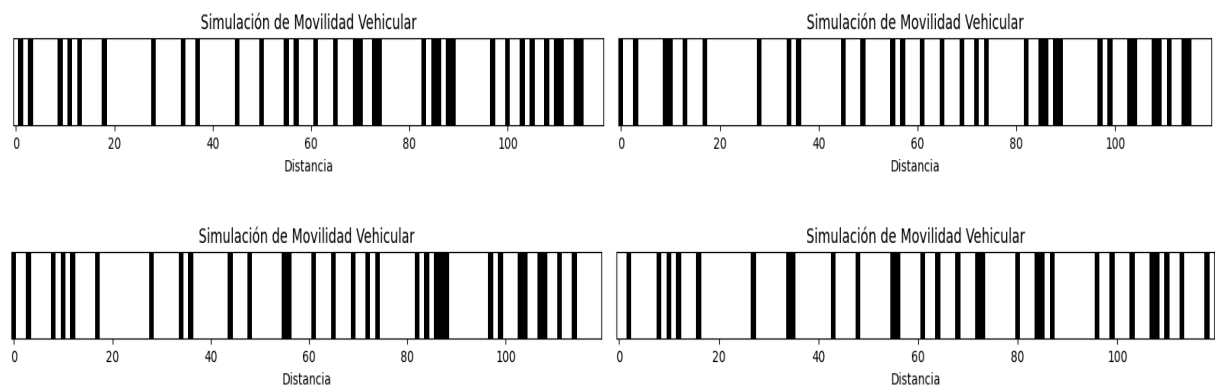


Figura 3.4. Simulación de la movilidad vehicular de los primeros 100 metros de una calle

Paso 4: Para comprobar la característica de adaptabilidad en la movilidad vehicular, se utilizó el modelo de Lighthill-Whitham-Richards (LWR), una ecuación de estado que describe la dinámica del tráfico en una red de carreteras. Este modelo representa cómo varía la densidad vehicular en función del flujo de entrada y salida de vehículos, mediante una ecuación de transporte conservativa que relaciona la densidad de tráfico (ρ) con el flujo de tráfico (q):

$$\frac{\partial \rho}{\partial t} + \frac{\partial q}{\partial x} = 0 \quad (3.1)$$

Donde:

ρ es la densidad de tráfico (vehículos por unidad de longitud)

q es el flujo de tráfico (vehículos por unidad de tiempo)

t representa el tiempo

x la posición a lo largo de la carretera.

La ecuación 3.1 expresa la conservación del número de vehículos en un tramo de carretera, donde el primer término refleja los cambios en la densidad en un punto fijo de la vía, mientras que el segundo describe la variación del flujo a lo largo de la misma. Para analizar la adaptabilidad del tráfico, se implementó el modelo LWR en Python, permitiendo visualizar la relación entre la densidad vehicular y su capacidad de ajuste dinámico, como se muestra en la Figura 3.5. En este gráfico, el eje x representa la posición en la carretera y el eje y la densidad vehicular en cada punto. La densidad se mide como la fracción de ocupación de cada celda, reflejando qué porcentaje de la vía está ocupado en un instante dado. Cada punto en la línea muestra la evolución de la densidad en una celda específica durante la simulación, donde un aumento en la densidad se traduce en un incremento en la altura del punto, indicando congestión, mientras que una disminución señala menor ocupación, así mismo un valor negativo es una variación en la densidad o una tasa de cambio, no la densidad directa. Este comportamiento evidencia la capacidad del sistema de ajustarse a las condiciones del tráfico en tiempo real, resaltando la naturaleza adaptativa de la movilidad vehicular.

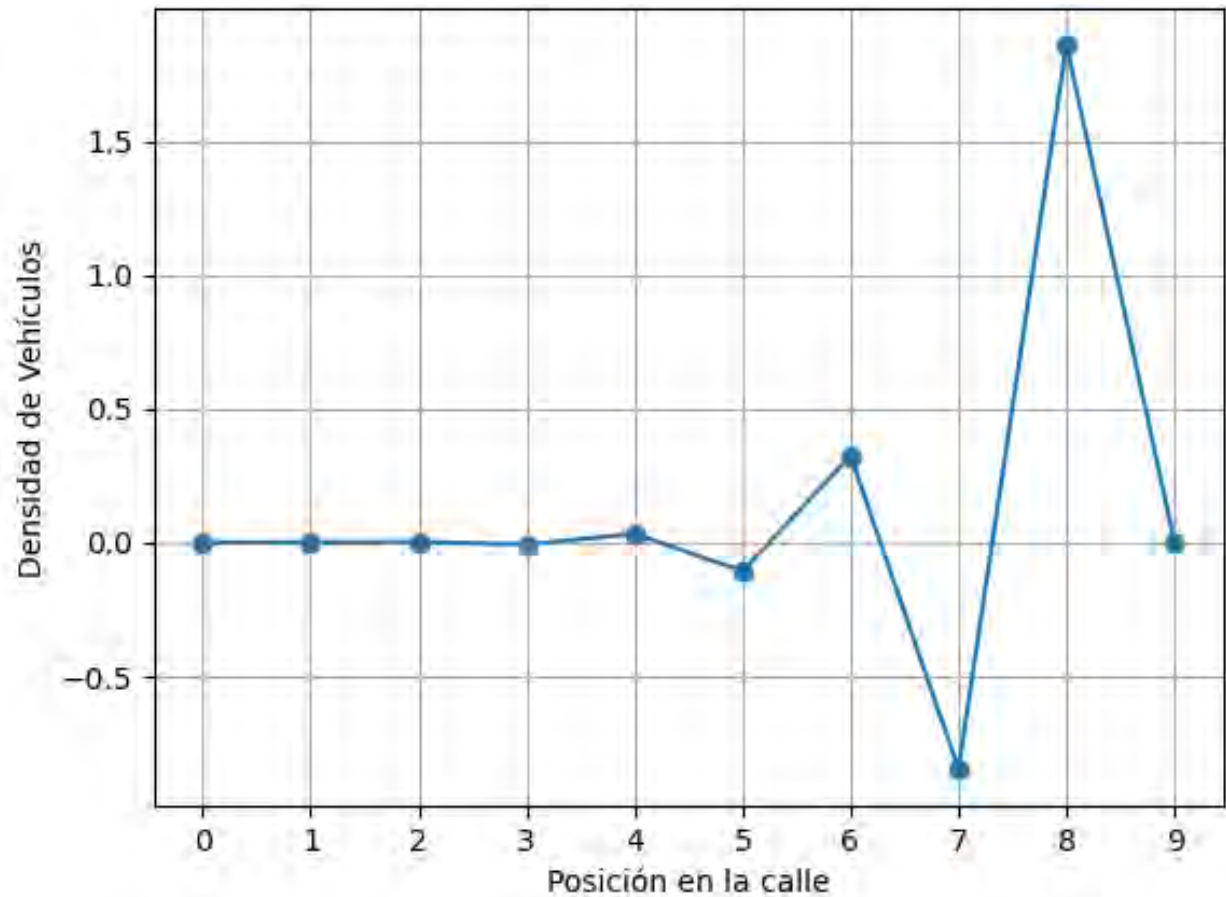


Figura 3.5. Adaptabilidad de la movilidad vehicular

Paso 5: Para verificar la característica de no linealidad en la movilidad vehicular, se emplea el modelo de flujo de tráfico de Burgers, el cual permite describir el comportamiento del tráfico en una carretera. Este modelo se expresa mediante la siguiente ecuación:

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u)}{\partial x} = 0 \quad (3.2)$$

Donde:

ρ Es la densidad de tráfico (vehículos por unidad de longitud).

t es el tiempo.

x es la posición a lo largo de la calle.

u es la velocidad del flujo de tráfico.

$$u = V_{\max} \left(1 - \frac{\rho}{\rho_{\max}} \right) \quad (3.3)$$

Donde:

$V_{\max}$ Es la velocidad máxima.

$\rho_{\max}$ Es la densidad máxima.

El modelo de Burgers se implementó en lenguaje Python. El gráfico resultante se muestra en la Figura 3.6 describe la evolución de la densidad de la movilidad a lo largo del tiempo para diferentes posiciones a lo largo de la calle.

La densidad de tráfico varía tanto en el espacio (calle) como en el tiempo. Esto se muestra mediante las diferentes líneas en el gráfico, donde cada línea representa la densidad de la movilidad en una posición específica de la carretera en función del tiempo. Los cambios en la forma de las curvas a lo largo del tiempo, se puede inferir cómo la congestión se propaga a lo largo de la calle. A medida que la densidad de tráfico cambia, también lo hace el flujo de tráfico, ya que están relacionados por la ecuación de Burgers. Se puede observar si el sistema de movilidad tiende a estabilizarse o si experimenta oscilaciones o inestabilidades a lo largo del tiempo. Esto se refleja en la forma y la evolución de las curvas de densidad de la movilidad.

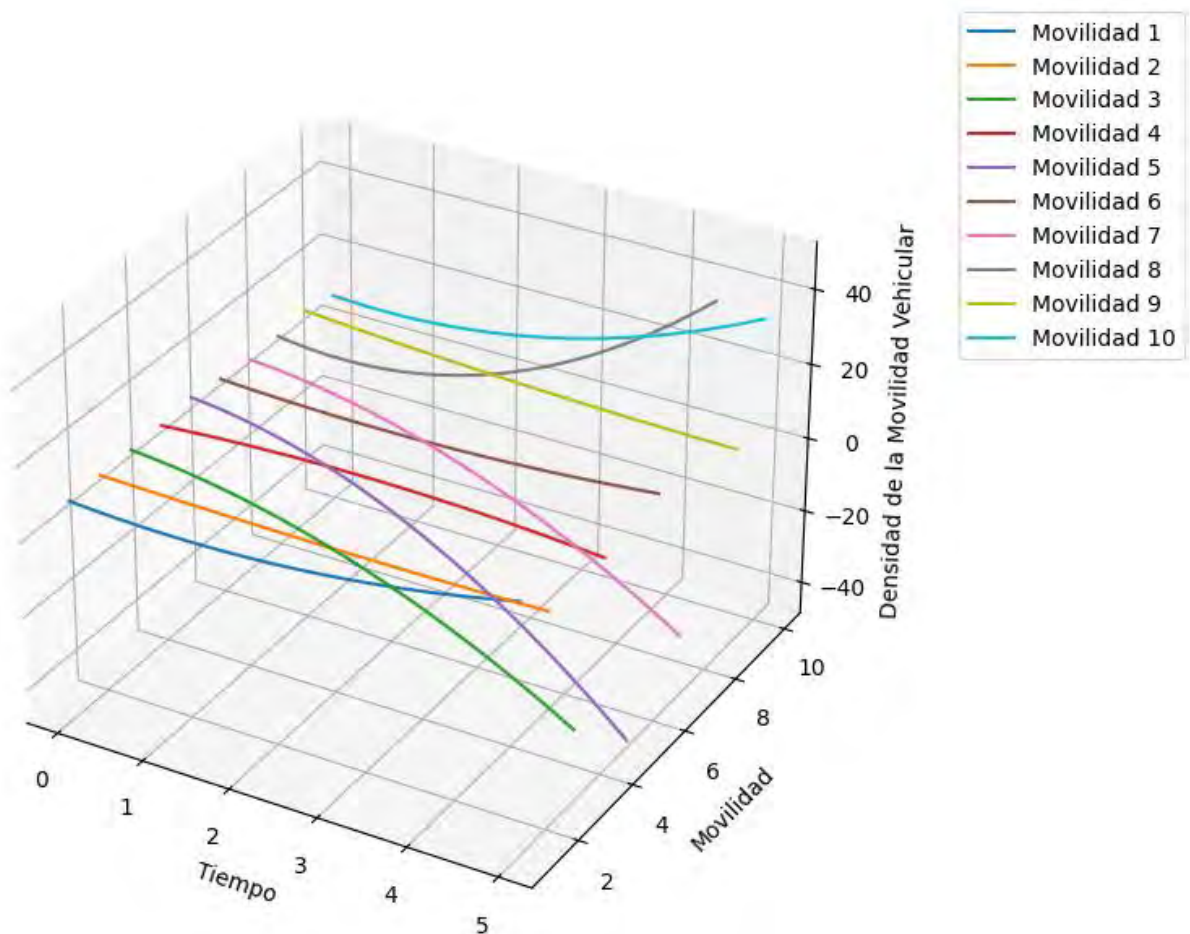


Figura 3.6. Representación gráfica de la no linealidad de la Movilidad Vehicular

Paso 6: Para verificar la autoorganización de la movilidad vehicular se implementó el modelo Na-Sch (Nagel-Schreckenberg) utilizando el lenguaje Python. El comportamiento asociado a la autoorganización en la movilidad vehicular es

representado gráficamente. Tomando en cuenta la movilidad vehicular bajo diversas circunstancias, como se muestra en la Figura 3.7, se observa serie de mapas de calor que ilustran el fenómeno de autoorganización de la movilidad vehicular en una calle. En cada subgráfico, el eje horizontal representa la longitud de la calle, mientras que el eje vertical indica el ancho de la vía, es decir, las diferentes posiciones transversales o carriles. Cada celda del mapa está coloreada según un valor que representa la velocidad de los vehículos en esa posición: los tonos claros (amarillos) indican baja velocidad o ausencia de vehículos, mientras que los tonos oscuros (azul profundo) reflejan alta velocidad o concentración vehicular. Al observar los diferentes mapas en secuencia, se aprecia cómo el sistema vehicular evoluciona desde una distribución aparentemente aleatoria de velocidades hacia una disposición más estructurada. Este comportamiento emergente refleja un proceso de autoorganización, donde los vehículos, en ausencia de una coordinación centralizada, tienden a formar patrones coherentes y repetitivos a través de interacciones locales. Se pueden distinguir zonas densas alternadas con vacíos, lo cual sugiere una optimización adaptativa del espacio vial.

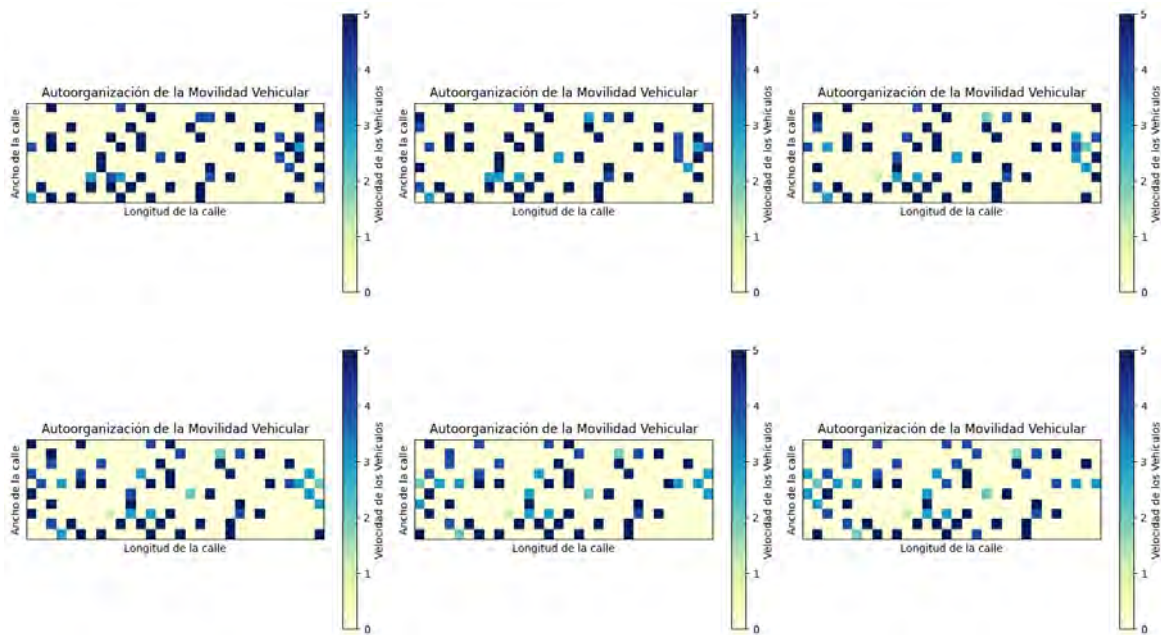


Figura 3.7. Auto organización de la movilidad vehicular mediante el autómata celular

Paso 7: Para comprobar la característica de incertidumbre en la movilidad vehicular, se empleó el modelo de entropía de Shannon. La ecuación asociada a este modelo es la siguiente:

$$H(X) = - \sum_{x \in X} P(x) \log_2 P(x) \quad (3.4)$$

Donde:

$H(X)$ representa la entropía del conjunto de datos X .

$P(x)$ es la probabilidad de que ocurra el evento x en el conjunto de datos.

El símbolo $\sum_{x \in X}$ indica la suma sobre todos los eventos x en el conjunto de datos.

El término $-P(x)\log_2 P(x)$ calcula la contribución de cada evento x a la entropía total, donde $\log_2$ es el logaritmo en base 2.

Se implementó en el lenguaje Python el modelo de Shannon para cuantificar la incertidumbre de las calles donde se lleva a cabo la movilidad vehicular, como se muestra en la Figura 3.8. En este contexto, a mayor entropía, mayor es la incertidumbre asociada al resultado $H(X)$. Cuando todos los posibles resultados tienen la misma probabilidad de ocurrir, la entropía alcanza su valor máximo, lo que indica una alta incertidumbre. En cambio, si un resultado es casi seguro (es decir, su probabilidad se aproxima a 1), la entropía es baja, reflejando una menor incertidumbre en el sistema.

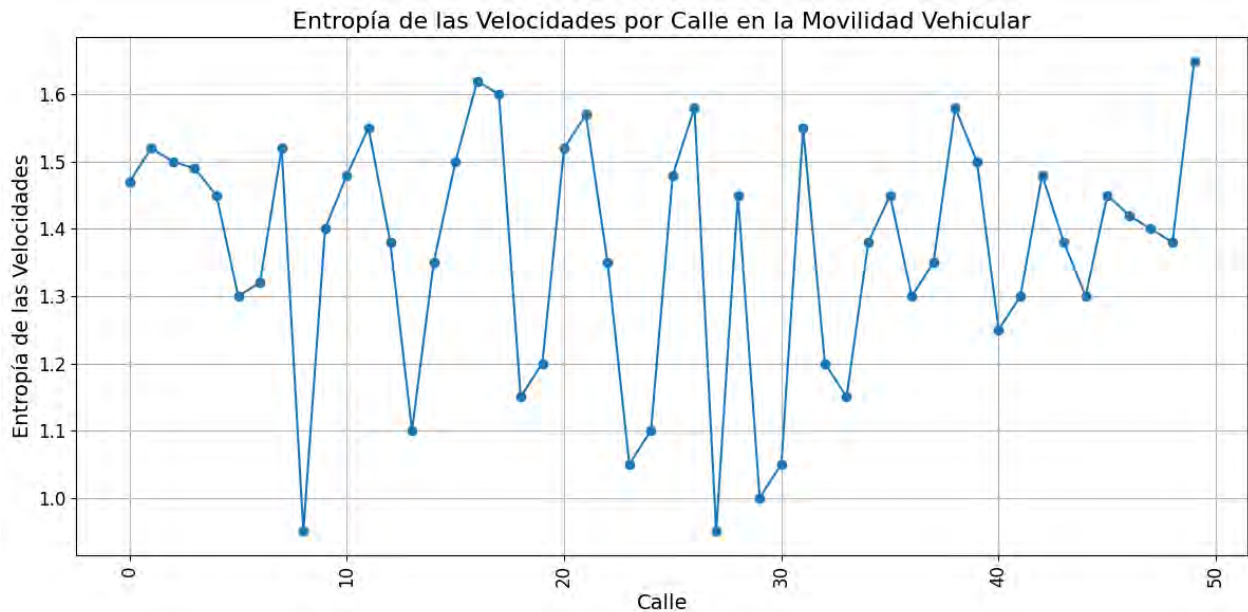


Figura 3.8. Valor de entropía de cada segmento de la calle, de acuerdo a las velocidades

Paso 8: Para comprobar la característica de cascadas de efectos de la movilidad vehicular se realizó la red compleja de la movilidad vehicular utilizando el modelo de red aleatoria de Erdős-Rényi que describe la probabilidad de que dos nodos están conectados por un borde en una red con un cierto número de nodos y una probabilidad de conexión. La función matemática que describe la probabilidad de conexión p en el modelo de Erdős-Rényi.

$$p = \frac{m}{\binom{n(n-1)}{2}} \quad (3.5)$$

Donde:

p es la probabilidad de conexión entre dos nodos

m es el número real de bordes en la red

n Es el número de nodos en la red.

Se implementó el modelo de Erdős-Rényi en el lenguaje de Python para identificar la cascada de efecto de la movilidad vehicular donde se evidencia el fenómeno de

cascada de congestión en distintas zonas geográficas. En la Figura 3.9, se observa un estado inicial de la red con pocos nodos afectados (representados en rojo), lo que indica que aún no se ha producido una propagación significativa de la congestión. La mayoría de los nodos se mantienen en un estado normal (azul), lo que sugiere que la red conserva su funcionalidad y no presenta una cascada activa. En contraste, la Figura 3.10. Revela cómo, con el paso del tiempo o ante un cambio en las condiciones del sistema de movilidad, la congestión se extiende rápidamente a través de los enlaces de la red, afectando a un número considerable de nodos. Esta propagación, conocida como cascada de efectos, es característica de sistemas complejos interconectados, donde una perturbación local puede desencadenar impactos sistémicos. Visualmente, la red con cascada activa se distingue por un aumento evidente en la proporción de nodos rojos y la densidad de conexiones que los rodean. Cuantitativamente, este fenómeno puede medirse mediante indicadores como el porcentaje de nodos afectados, la profundidad de propagación, el tiempo de difusión o métricas de centralidad. Cualitativamente, la cascada se identifica por el cambio abrupto en el patrón de color y la pérdida de uniformidad en la distribución de estados, lo que evidencia una pérdida de control local y la emergencia de un comportamiento colectivo no deseado.

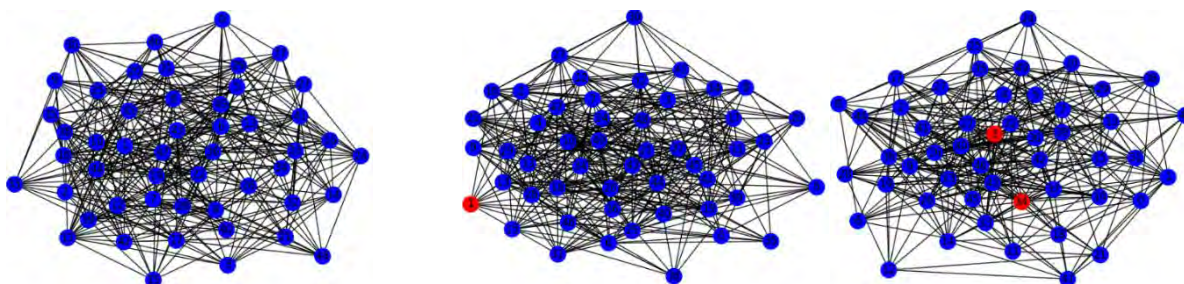


Figura 3.9. *Cascada de la congestión de Movilidad Vehicular en algunas zonas geográficas parte 1*

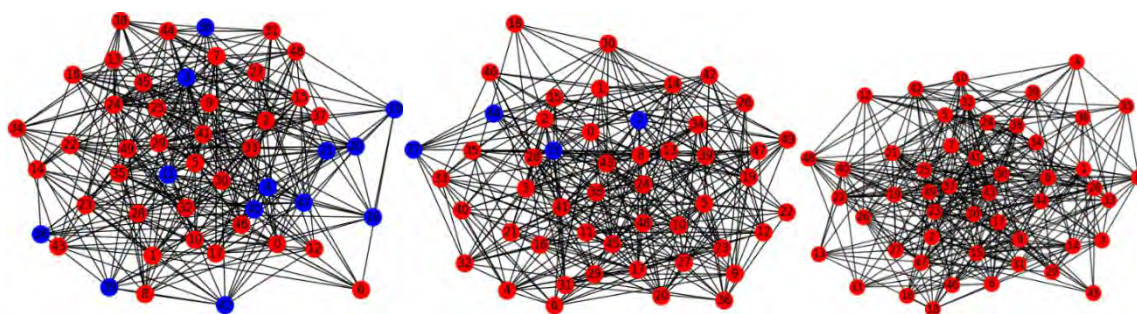


Figura 3.10. *Cascada de la congestión de Movilidad Vehicular en algunas zonas geográficas, parte 2*

Como resultado del análisis, se determinó que el problema de la movilidad vehicular presenta las características propias de un sistema complejo. Esta conclusión se fundamenta en la evaluación de múltiples modelos que permitieron identificar propiedades emergentes, comportamiento no lineal, interacciones locales e incertidumbre, entre otros factores. La validación de estas características refuerza la necesidad de enfoques metodológicos que capturen la dinámica subyacente del sistema y permitan una representación más precisa de su evolución.

Desde este punto de vista, el estudio del marco conceptual de la movilidad vehicular como un sistema complejo ha requerido una exploración profunda de sus interacciones dinámicas y no lineales. La aplicación de herramientas analíticas y computacionales, como autómatas celulares, modelos de entropía de Shannon y ecuaciones de flujo vehicular, ha permitido analizar fenómenos como la incertidumbre y la autoorganización en la dinámica del tráfico. Estos enfoques han evidenciado cómo pequeñas variaciones en las condiciones iniciales pueden generar efectos significativos en la movilidad, resaltando la importancia de considerar este fenómeno desde una perspectiva sistémica e interconectada.

A partir de las actividades desarrolladas en el marco del método de comprensión del sistema complejo dinámico de movilidad vehicular, se ha definido un conjunto de actividades sugeridas (ver

Tabla **3.2**) que facilitan la caracterización y análisis del sistema. Estas actividades proporcionan una guía estructurada, permitiendo su aplicación parcial o total, así como la incorporación de otras acciones complementarias en función de los requerimientos específicos del estudio y la naturaleza del sistema analizado.

El proceso inicia con la comprensión de los conceptos y del contexto teórico del problema, entendido como un sistema complejo donde se identifican las relaciones, dinámicas y características inherentes. Para ello, se plantean diversas actividades, incluyendo el análisis del contexto, la definición del alcance del problema y la investigación del entorno y antecedentes. Asimismo, se enfatiza la identificación de variables y sus interrelaciones, la evaluación de intereses y expectativas, y la recopilación de requisitos funcionales y no funcionales, asegurando una visión integral del sistema. A partir de esta información, se diseñan modelos conceptuales y diagramas de flujo, los cuales permiten estructurar gráficamente las interacciones y procesos dentro del sistema., facilitando la generación de propuestas de solución basadas en evidencia y en la dinámica observada.

La ejecución de estas actividades se recomienda la participación de diversos roles especializados del dominio. Cada uno de estos roles aporta una perspectiva única y complementaria, asegurando que el análisis y modelado del sistema refleje su complejidad y funcionalidad con efectividad. Como salidas del proceso, se generan productos para la estructuración del estudio, tales como el documento de identificación del problema, la lista de requisitos identificados, el informe de contexto y análisis del sistema, el mapa de interrelaciones del sistema, los documentos de requisitos funcionales y no funcionales, diagramas conceptuales y de flujo, así como propuestas de solución y modelos preliminares.

Tabla 3.2. Cuadro descriptivo del método de comprensión del sistema complejo dinámico

Fase de análisis o contextualización	
Descripción	Comprender conceptos y el contexto teórico del problema como un sistema complejo, identificando relaciones, dinámicas y características inherentes al sistema.
Actividades	- Realizar análisis de contexto. - Definir el alcance del problema. - Investigar el entorno y antecedentes. - Identificar variables relevantes. - Analizar interrelaciones del sistema. - Evaluar intereses y expectativas. - Recopilar requisitos funcionales y no funcionales. - Diseñar modelos conceptuales y diagramas de flujo. - Proponer alternativas y generar ideas de solución.
Roles	- Expertos del dominio. - Usuarios. - Investigadores. - Analistas de requisitos. - Modeladores de sistemas. - Equipo de desarrollo. - Equipo de toma de decisiones.
Salidas	- Documento de identificación del problema. - Lista de requisitos identificados. - Informe de contexto y análisis del sistema. - Mapa de interrelaciones del sistema. - Documentos de requisitos funcionales y no funcionales. - Diagramas conceptuales y de flujo. - Propuestas de solución y modelos preliminares.

3.3.2 Identificación de entidades y relaciones dinámicas de la movilidad vehicular

En el proceso de identificación de entidades y relaciones dinámicas en la movilidad vehicular, se aplicaron técnicas de sistemas complejos, como el análisis de interconectividad entre entidades, la representación estructural mediante diagramas de relaciones dinámicas y la caracterización del sistema a través de propiedades como emergencia y no linealidad. Desde la ciencia de datos, se emplearon estrategias de recolección automatizada de datos vía APIs (Here Maps y TomTom), procesamiento de datos JSON, y análisis temporal y espacial de atributos de tráfico. En cuanto a minería de datos, se realizó la extracción de variables significativas, categorización funcional de entidades y detección de relaciones relevantes mediante análisis estructurado. Para la construcción de indicadores, se identificaron atributos clave como velocidad, flujo y tiempo de viaje, utilizados para sintetizar el estado del sistema en métricas cuantificables.

En este caso para identificar las entidades y variables que intervienen en la movilidad vehicular y cómo interactúan entre sí, se realizó un proceso de observación, recopilación de datos y análisis. Los pasos se describen a continuación.

Paso 1, Selección de la muestra. Se optó por utilizar un método de muestreo por conveniencia, el cual implica la selección de las avenidas de la Ciudad de México para la muestra basándose en la experiencia y accesibilidad del grupo de investigación. Se eligieron dos de las avenidas importantes y representativas de la movilidad vehicular

en la Ciudad de México, siendo la Avenida Insurgentes y la Avenida Tlalpan las seleccionadas por su relevancia y acceso.

Paso 2, Definición de la herramienta para la obtención de datos. En la literatura, se encontraron diversas herramientas para recopilar datos sobre movilidad vehicular, que incluyen entrevistas, encuestas, consultas a expertos, sensores, cámaras de vídeo y aplicaciones de mapas. Para esta investigación, se ha elegido utilizar las aplicaciones de mapas como principal fuente de información. Con el fin de elegir la aplicación de mapas adecuada para la obtención de datos de tráfico, se llevó a cabo un análisis comparativo, cuyos resultados se presentan en la Tabla 3.3. El análisis comparativo se basó en una serie de criterios mostrados en la tabla, entre los cuales se destaca los datos que proporciona, si ofrece una API, costo y los formatos de datos. Tras una cuidadosa consideración, se determinó que Here Maps y TomTom Traffic son las opciones sobresalientes para cumplir con los objetivos de la investigación.

Tabla 3.3. Cuadro comparativo entre aplicaciones de mapas

Aplicación de mapas	Tráfico	Incidencias	Api	Libre	Formato JSON	Formato imagen
Here maps	Si	Si	Si	Si	Si	Si
Google	Si	Si	Si	No	Si	Si
Tom tom	Si	Si	Si	Si	Si	Si
Amap	No	No	No	No	No	No
Mapbox	Si	No	Si	Si	Si	Si
Baidu	Si	No	Si	Si	Si	Si
Zenrin	Si	No	No	No	Si	Si
Nvr info	Si	Si	No	No	No	No
Esri	Si	No	No	No	Si	Si
Azuremaps	Si	Si	Si	No	Si	Si

Los datos de movilidad vehicular fueron recopilados mediante una metodología propia que incluye: la selección de una aplicación de mapas, la obtención de su API, el diseño del Endpoint para la solicitud de datos, la ejecución de dicho Endpoint y el almacenamiento de la información obtenida, como se muestra en la Figura 3.11. La recolección se automatizó para ejecutarse cada 10 minutos entre las 7:00 a.m. y las 10:00 p.m. durante 30 días. Esto generó 45,912 registros con 14 atributos. Un fragmento en formato JSON está disponible en el siguiente enlace <https://goo.su/1vazi>.



Figura 3.11. Metodología para la obtención de datos de tráfico

Paso 4, Identificación de entidades de la movilidad vehicular. Para la identificación de entidades, se analiza la información de movilidad proporcionada por TomTom Traffic. Por cada solicitud, TomTom Traffic devuelve un JSON (JavaScript Object Notation) que

incluye información como el tipo de calle (frc), velocidad actual (currentSpeed), velocidad de flujo libre (freeFlowSpeed), tiempo de viaje actual (currentTravelTime), tiempo de viaje en flujo libre (freeFlowTravelTime), si la calle está cerrada (roadClosure), así como el anexo a estas variables el día, mes, año, hora, minuto y segundo correspondientes de captura de datos, como se muestra en la Figura 3.12.

```
{'flowSegmentData': {'frc': 'FRC0',
  'currentSpeed': 95,
  'freeFlowSpeed': 95,
  'currentTravelTime': 1311,
  'freeFlowTravelTime': 1311,
  'confidence': 1,
  'roadClosure': False,
  'coordinates': {'coordinate': [{ 'latitude': 19.05644603109388,
    'longitude': -99.24117571656619},
    { 'latitude': 19.056628408161256, 'longitude': -99.24108318035516},
    { 'latitude': 19.056993161694418, 'longitude': -99.2408672625294},
    { 'latitude': 19.057781768133474, 'longitude': -99.24029192869558},
    { 'latitude': 19.057978879544947, 'longitude': -99.24014172499072},
    { 'latitude': 19.058074107723318, 'longitude': -99.24006259982475},
    { 'latitude': 19.05828999503007, 'longitude': -99.23990300838832},
    { 'latitude': 19.05852204386213, 'longitude': -99.23972598259331},
    { 'latitude': 19.05887879156995, 'longitude': -99.23944435064668},
    { 'latitude': 19.05964584011634, 'longitude': -99.23887169902186},
    { 'latitude': 19.05972767848862, 'longitude': -99.23881939594607},
    { 'latitude': 19.059908705141506, 'longitude': -99.23871881310798},
    { 'latitude': 19.06004821832649, 'longitude': -99.23865980450964},
    { 'latitude': 19.060219896230095, 'longitude': -99.2386007959113},
    { 'latitude': 19.060344594328267, 'longitude': -99.2385659271941},
    { 'latitude': 19.060481413544146, 'longitude': -99.23854178731297},
    { 'latitude': 19.060677174947354, 'longitude': -99.23852435295436},
    { 'latitude': 19.060771054777476, 'longitude': -99.23852167074534},
    { 'latitude': 19.06145237515887, 'longitude': -99.23860750143383},
    { 'latitude': 19.0621819394263, 'longitude': -99.23872820083953},
    { 'latitude': 19.06229190062072, 'longitude': -99.2387402707801},
```

Figura 3.12. JSON de información de Movilidad vehicular proporcionado por TomTom Traffic

Como resultado del proceso de identificación de entidades que conforman el sistema complejo de movilidad vehicular, se han definido los principales elementos que intervienen en su dinámica. Las entidades identificadas incluyen: tipo de calle (frc), velocidad actual (currentSpeed), velocidad en flujo libre (freeFlowSpeed), tiempo de viaje actual (currentTravelTime), tiempo de viaje en flujo libre (freeFlowTravelTime), presencia de cierre de calle (roadClosure), día, mes, año, hora y minuto. Esta caracterización permite establecer un marco estructurado para representar el sistema desde una perspectiva basada en entidades observables y medibles, esenciales para comprender el comportamiento colectivo del flujo vehicular.

La identificación de estas entidades se realizó a partir de un análisis de los componentes que integran el entorno dinámico del tráfico urbano. Estas variables mantienen interacciones dinámicas, no lineales y mutuamente influyentes, lo que da

lugar a comportamientos emergentes propios de los sistemas complejos. Por ejemplo, se ha detectado una relación inversa entre la densidad vehicular y la velocidad actual, donde un aumento en el número de vehículos en circulación tiende a disminuir la velocidad promedio, especialmente en horarios de alta demanda y en vías con menor jerarquía funcional. Asimismo, cuando se presentan cierres viales, se observa un incremento considerable en el tiempo de viaje actual, superando notablemente el tiempo estimado bajo condiciones de flujo libre, lo que evidencia el impacto de perturbaciones locales en la eficiencia global del sistema.

También se han identificado patrones horarios y temporales relevantes. Durante las horas pico, la velocidad actual disminuye sistemáticamente, mientras que el tiempo de viaje se prolonga, lo cual confirma una alta sensibilidad del sistema a variaciones temporales. Además, el tipo de calle influye de forma directa sobre la velocidad en flujo libre: vías primarias, como arterias o ejes viales, permiten velocidades significativamente más altas que calles secundarias o terciarias. Estas observaciones reflejan cómo la estructura vial, la dinámica temporal y las perturbaciones externas convergen para producir comportamientos complejos y adaptativos.

A partir de las actividades desarrolladas en el marco del método de identificación de entidades y relaciones dinámicas en la movilidad vehicular, se ha establecido un conjunto de actividades sugeridas (ver Tabla 3.4) que permiten definir la composición del sistema complejo, identificando tanto sus componentes como sus interacciones dinámicas. Este enfoque proporciona una estructura, permitiendo su aplicación parcial o total y facilitando la incorporación de nuevas acciones en función de los requerimientos específicos del estudio y la complejidad inherente al sistema analizado.

El método se basa en la identificación de las entidades que conforman el sistema complejo, el análisis de sus relaciones dinámicas y la comprensión de cómo estas interacciones influyen en el comportamiento global del sistema. Para ello, se llevan a cabo diversas actividades, tales como la identificación de los elementos esenciales, el análisis de sus interacciones, la categorización de las entidades a partir de su función o rol, y la definición de sus atributos importantes. Además, se diseñan diagramas representativos de las relaciones dinámicas, se establecen las características esenciales y se elaboran descripciones detalladas de cada entidad y sus interacciones, asegurando un modelo claro y estructurado del sistema.

La implementación de este método requiere la participación de diversos roles especializados. Cada uno de estos roles aporta conocimientos específicos que permiten una representación integral y precisa del sistema. Como salidas del proceso, se genera una lista de entidades identificadas, acompañada de diagramas que representan sus relaciones, descripciones detalladas de los atributos relevantes, una matriz jerárquica de interacciones y especificaciones tanto funcionales como dinámicas de las entidades.

Tabla 3.4. Cuadro descriptivo del método identificación de entidades y relaciones dinámicas

Interacción de las entidades del sistema	
Descripción	Identificar las entidades que conforman un sistema complejo, analizar sus relaciones dinámicas y comprender cómo estas interacciones influyen en el comportamiento global del sistema.
Actividades	- Identificar elementos importantes dentro del sistema. - Analizar componentes y sus interacciones. - Agrupar entidades atendiendo a su función o rol. - Definir atributos centrales de las entidades. - Crear diagramas que representen las relaciones dinámicas. - Identificar características esenciales y roles de las entidades. - Crear descripciones detalladas de cada entidad y sus interacciones.
Roles	- Analistas de sistemas. - Expertos del dominio. - Especialistas en clasificación y modelado. - Diseñadores de estructuras conceptuales. - Modeladores de relaciones dinámicas.
Salidas	- Lista de entidades identificadas. - Diagramas de relaciones entre entidades. - Descripciones detalladas de atributos significativos. - Matriz jerárquica de interacciones. - Especificaciones funcionales y dinámicas de las entidades.

3.3.3 Definición de variables críticas del sistema complejo de movilidad vehicular

En la definición de variables críticas del sistema complejo de movilidad vehicular se emplearon técnicas de sistemas complejos, como la identificación de interdependencias no lineales, análisis de retroalimentación entre variables y clasificación en variables de entrada, estado y salida para construir modelos dinámicos. Desde la ciencia de datos, se aplicaron métodos de recolección masiva y procesamiento de grandes volúmenes de datos (más de 11 millones de registros), análisis exploratorio, graficación radial y pruebas de distribución para evaluar la densidad y normalidad de variables clave como la velocidad. En minería de datos, se utilizó segmentación temporal, detección de patrones recurrentes y evaluación de correlaciones entre variables temporales y de flujo, lo que permitió identificar momentos críticos de congestión. En cuanto a la construcción de indicadores, se derivaron métricas compuestas a partir de la combinación de atributos como velocidad, longitud y tiempo para representar fenómenos como la fluidez y la criticidad del tráfico.

Para llevar a cabo el análisis de interacción entre las variables que componen el sistema complejo de movilidad vehicular, se empleó un proceso de representación gráfica con el fin de visualizar claramente el comportamiento de la movilidad vehicular en función de la interacción entre sus variables. A continuación, se detallan los pasos seguidos en este proceso.

Paso 1, Selección de la muestra. Se utilizó un conjunto de datos de 11, 000, 000 registros de las 307 calles de la Alcaldía Tlalpan (número de calles detectadas por la infraestructura de la aplicación de mapas) con 41 atributos, recolectadas con Here Maps (en el período de agosto-diciembre 2022) y Geoapify para obtener los datos de

servicios, los atributos se describen en la Tabla 3.5 y Tabla 3.6, el tipo de muestreo utilizado fue un **muestreo aleatorio simple**.

Tabla 3.5. Descripción de atributos de tráfico, incidencias viales y servicios, parte 1

Variable del modelo	Variable en el conjunto de datos	Descripción
X1	Bbox_Incident	Cuadro delimitador de donde se origina la incidencia.
X2	Bbox_Traffic	Cuadro delimitador de donde origina el tráfico.
X3	confidence	El valor del campo de confianza indica la proporción de datos en tiempo real incluidos en el cálculo de velocidad. Es un valor
X4	Free_Flow	La velocidad de referencia (en metros por segundo) a lo largo de la carretera cuando no hay tráfico.
X5	Length	Longitud del segmento.
X6	Number_Segments	Número de segmentos que contiene la vialidad monitoreada.
X7	Speed	La velocidad esperada (en metros por segundo) a lo largo de la carretera; no excederá el límite de velocidad legal.
X8	Speed_Uncapped	La velocidad esperada (en metros por segundo) a lo largo de la carretera; puede exceder el límite de velocidad legal.
X9	Daily_Traffic	Día del monitoreo de tráfico.
X10	Day_Off_Traffic	Valor de 1 cuando el día es festivo y 0 cuando no lo es.
X11	Traffic_Hour	Hora del monitoreo de tráfico.
X12	Traffic_Month	Mes del monitoreo de tráfico.
X13	Traffic_Minute	Minuto del monitoreo de tráfico.
X14	Traffic_Day_Number	Número de día del monitoreo de tráfico (1= lunes, 2=martes...7=domingo).
X15	Work_Day_Traffic	Monitoreo del tráfico en semana laboral y fin de semana (semana laboral=1, fin de semana=0).
X16	Traffic_Block_Type	Monitoreo de tráfico por bloques de horario (bloque 1=0-3 horas, bloque 2=4-7 horas, bloque 3=8-11 horas.....bloque 6: 20-23 horas.
X17	Road_Closed	Indica si la calle está cerrada por la incidencia (0 abierto y 1 cerrado).
X18	Criticality	Representa la gravedad del incidente. Tiene los siguientes valores: bajo - menos severo, menor, importante y crítico - más severo.
X19	Type	Describe el tipo de incidente: accidente, construcción, congestión, avería de vehículo, clima entre otros.

Tabla 3.6. Descripción de atributos de tráfico, incidencias viales y servicios, parte 2

Variable del modelo	Variable en el conjunto de datos	Descripción
X20	i14	Proporciona una descripción textual más larga, posiblemente con información de ubicación de la incidencia.
X21	i2	Id de la forma de la incidencia.
X22	i7	Hora de inicio de la incidencia.
X23	i8	Hora de fin de la incidencia.
X24	Incident_Day	Día del monitoreo de la incidencia.
X25	Day_Off_Incident	Valor de 1 cuando el día es festivo y 0 cuando no lo es del monitoreo de la incidencia.
X26	Incident_Hour	Hora del monitoreo de la incidencia.
X27	Incident_Month	Mes del monitoreo de la incidencia.
X28	Incident_Minute	Minuto del monitoreo de la incidencia.
X29	Incident_Day_Number	Número de día del monitoreo de incidencia (1= lunes, 2=martes...7=domingo).
X30	Work_Day_Incident	Monitoreo de la incidencia en semana laboral y fin de semana (semana laboral=1, fin de semana=0).
X31	Category_Number	Categoría de la incidencia.
X32	Time	Formato de 24 horas a minutos.
X33	name	Descripción del servicio.
X34	lon	Coordenadas de longitud de la ubicación del servicio.
X35	lat	Coordenadas de latitud de la ubicación del servicio.
X36	numcategoria	Clasificación del servicio.
X37	colorcategoria	Descripción en color del servicio.
X38	bboxservicio	Cuadro delimitador de donde se ubica el servicio.
X39	Start_Time	Hora de inicio de la incidencia.
X40	End_Time	Hora de fin de la incidencia.
X41	incidencia_tipobloque	Monitoreo de incidencia por bloques de horario (bloque 1=0-3 horas, bloque 2=4-7 horas, bloque 3=8-11 horas.....bloque 6: 20-23 horas.

Paso 2, Representación gráfica de la Movilidad vehicular. Se diseñaron gráficas radiales utilizando datos de diferentes días para ofrecer una representación visual del comportamiento de la movilidad vehicular a lo largo del tiempo en la Avenida

Insurgentes y la Avenida Tlalpan de la Ciudad de México. Estas gráficas proporcionan una perspectiva única al mostrar cómo ciertos aspectos de la movilidad, como la velocidad, el flujo de tráfico, la densidad, varían en diferentes momentos del día y en días específicos. Al utilizar gráficas radiales, podemos identificar fácilmente patrones recurrentes o anomalías en el flujo de tráfico, lo que nos permite comprender mejor las tendencias y los factores que influyen en la movilidad vehicular en distintas condiciones y momentos del día. Este enfoque visual es útil para visualizar cómo interactúan las variables de la movilidad vehicular como se muestran en la Figura 3.13 y Figura 3.14.

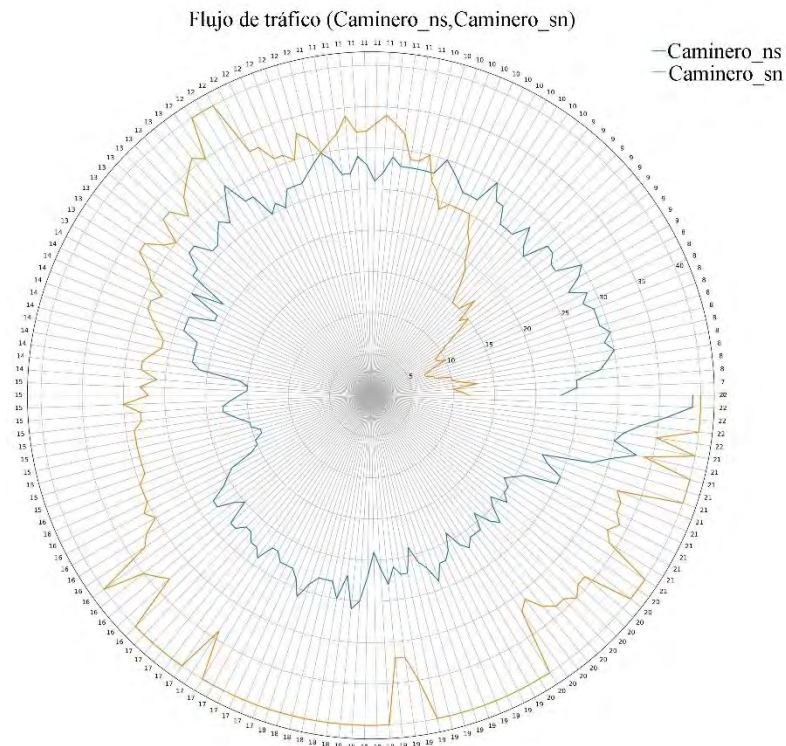


Figura 3.13. Movilidad vehicular en horario de 7 a 22 horas

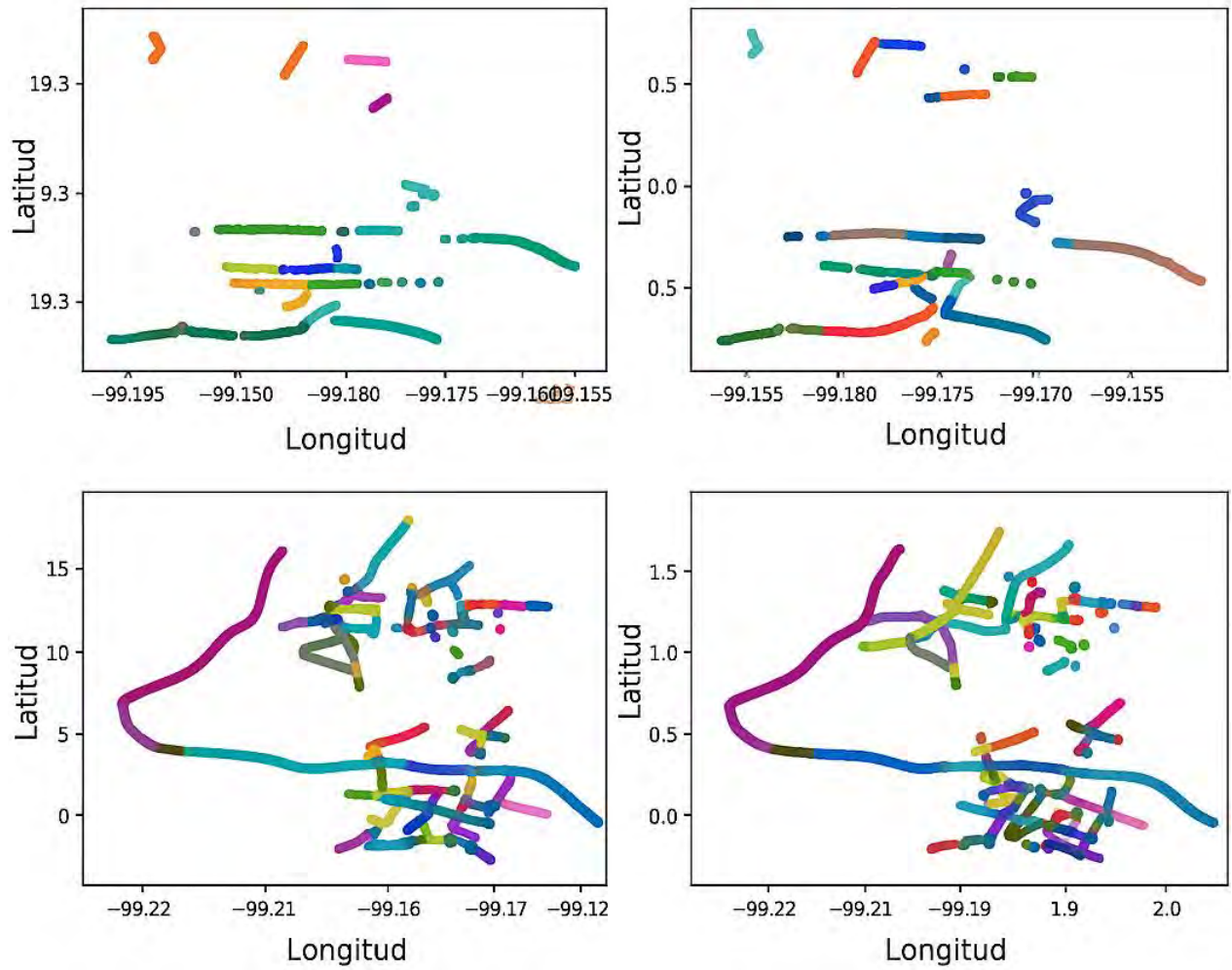


Figura 3.14. Comportamiento de la movilidad vehicular en intervalos de tiempo

La densidad y la normalidad de algunas variables de la entidad tráfico se muestran en la Figura 3.15 y Figura 3.16.

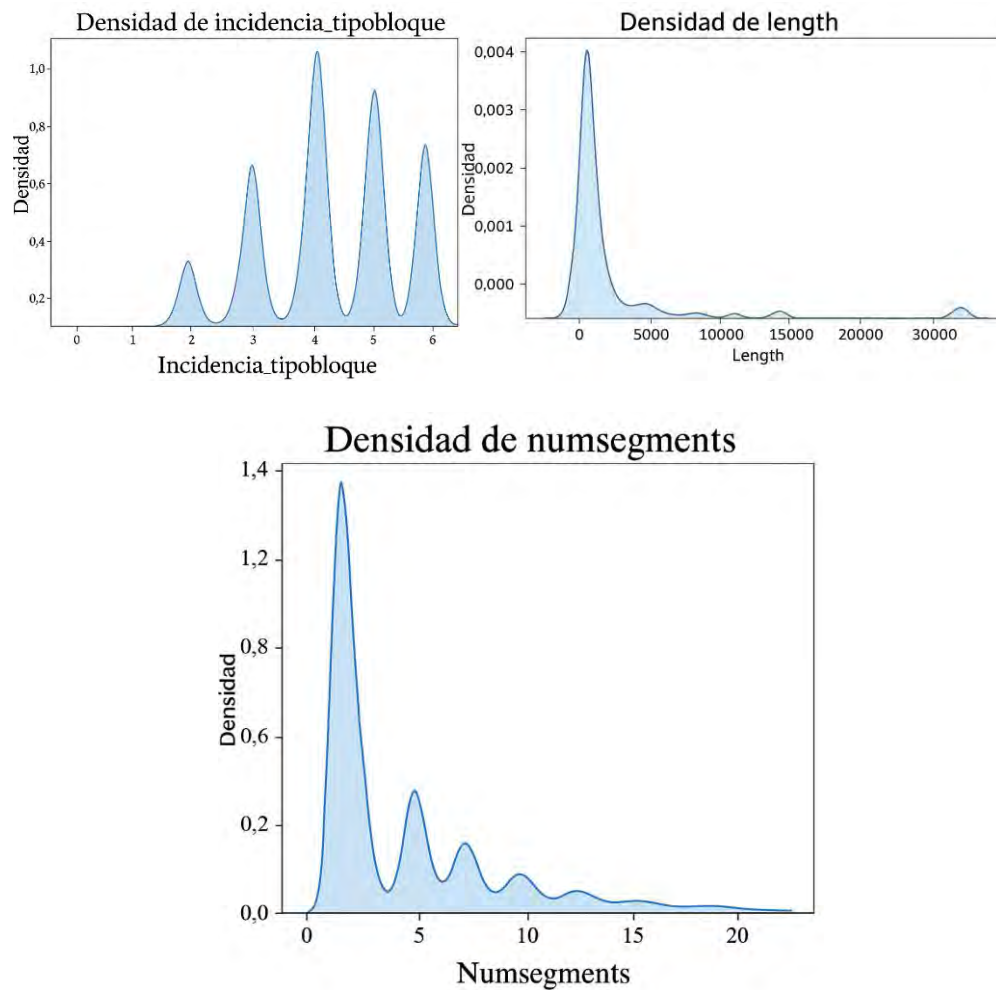


Figura 3.15. Gráfica de densidad de algunas variables de la entidad tráfico

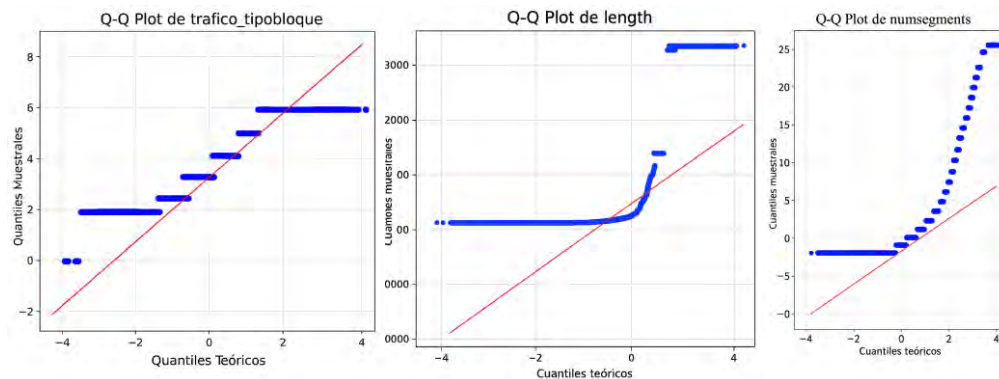


Figura 3.16. Gráficas Q-Q para visualizar la normalidad de los datos de las variables.

Como resultado, se encontró la interacción entre la variable de **velocidad** y otras variables temporales, como **día, mes, hora y minuto**. Al visualizar estos datos, se

identificó una clara relación entre la velocidad y el tiempo del día. Por ejemplo, se observó que la velocidad tiende a variar significativamente dependiendo la hora del día, con fluctuaciones notables durante ciertos períodos. En particular, se observó una variabilidad significativa en la velocidad durante las horas pico de tráfico. Durante estos períodos, que suelen coincidir con las horas de mayor actividad en las vías, se observa una mayor dispersión en los valores de velocidad registrados. Esto sugiere que el flujo de tráfico y las **condiciones de la carretera** pueden ser variables importantes que influyen en la velocidad de los vehículos durante estos momentos críticos. Además, al analizar la velocidad vehicular en función de los minutos específicos dentro de cada hora, se identificó una variabilidad notable en los patrones de tráfico. Para llevar a cabo este análisis, se desagregaron los registros temporales en intervalos de un minuto, permitiendo observar microfluctuaciones en el comportamiento del flujo vehicular. Se aplicaron medidas estadísticas descriptivas como la media, mediana, desviación estándar y coeficiente de variación para evaluar la dispersión y estabilidad de la velocidad a lo largo del tiempo. Se identificó que, incluso dentro de una misma hora, los valores de velocidad pueden fluctuar significativamente, lo que indica que factores como la **congestión del tráfico**, los **semáforos** y las **condiciones climáticas** pueden tener un impacto inmediato en la velocidad de los vehículos en diferentes momentos.

El análisis de las variables críticas del sistema de movilidad vehicular ha permitido una comprensión de cómo estas relaciones afectan la dinámica del tráfico. La interdependencia entre estas variables no solo influye en la eficiencia del flujo vehicular, sino que también impacta la sostenibilidad del sistema y la calidad de vida en entornos urbanos. Identificar y caracterizar estas conexiones para definir las variables que determinan el comportamiento del sistema, proporcionando una base para el desarrollo de modelos que reflejen con precisión su complejidad.

A partir de las actividades desarrolladas en el marco del método de definición de variables críticas del sistema complejo de movilidad vehicular, se ha establecido un conjunto de actividades sugeridas (ver Tabla 3.7) que permiten identificar y seleccionar las variables que integran la dinámica del sistema. Estas actividades proporcionan una estructura para la definición de las variables críticas, permitiendo su implementación parcial o total y facilitando la incorporación de nuevas actividades al contexto específico del sistema complejo analizado.

El proceso inicia con la identificación de las variables, aquellas que tienen un impacto significativo en la evolución del sistema. Para ello, se llevan a cabo actividades como el análisis de datos históricos, la revisión de literatura especializada y la consulta con expertos del dominio. Posteriormente, se procede a la clasificación y categorización de las variables, diferenciándose en variables de entrada, de estado y de salida. Además, se evalúan las relaciones funcionales entre variables, determinando su grado de influencia y estableciendo modelos de interdependencia.

La correcta identificación de estas variables requiere la intervención de diversos roles especializados. Estos roles contribuyen con su conocimiento técnico para garantizar que las variables seleccionadas sean representativas y relevantes para la comprensión y predicción del comportamiento del sistema. Como salidas del proceso, se obtiene

una lista de variables críticas identificadas, acompañada de matrices de relaciones e interdependencias, modelos conceptuales y matemáticos, descripciones detalladas de las funciones y efectos de cada variable en la dinámica del sistema, y especificaciones sobre su recolección, procesamiento y análisis.

Tabla 3.7. Cuadro descriptivo del método definición de variables críticas

Variables relevantes	
Descripción	Analizar las relaciones entre las variables del sistema para identificar aquellas que tienen un impacto crítico en el comportamiento del sistema, considerando tanto interacciones lineales como no lineales.
Actividades	- Enumerar todas las variables relevantes asociadas a las entidades. - Clasificar las variables en independientes y dependientes - Identificar relaciones y dependencias entre las variables. - Analizar la naturaleza de las interacciones (lineales, no lineales). - Realizar simulaciones y modelado matemático para evaluar el impacto de las variables. - Generar diagramas y matrices de dependencia. - Determinar las variables más influyentes en el sistema.
Roles	- Analistas de datos. - Matemáticos y estadísticos. - Expertos del dominio. - Modeladores de sistemas. - Diseñadores de simulaciones. - Especialistas en análisis de interacciones.
Salidas	- Lista de variables críticas identificadas. - Diagramas de relaciones entre variables. - Matriz de dependencia entre variables. - Descripción detallada de las interacciones. - Documentación que prioriza variables influyentes. - Informe final sobre interacciones y su impacto.

3.4 Agregación multinivel y medición de la dinámica compleja

Esta fase se centra en analizar y estructurar las interacciones dentro del sistema complejo desde diferentes niveles de agregación, permitiendo una comprensión profunda de las relaciones dinámicas que emergen. Esta fase contempla tanto la consolidación de datos en niveles jerárquicos como la identificación de patrones importantes que influyen en el comportamiento global del sistema.

La Fase 2 se compone de tres métodos principales: los niveles de agregación del sistema complejo dinámico, el análisis de incertidumbre del sistema y la identificación de variables determinantes en función de su comportamiento dinámico, como se muestra en la Figura 3.17. Cada uno de estos métodos considera aspectos esenciales para evaluar tanto la dinámica como la estructura del sistema en distintos niveles de agregación. Para su comprensión y aplicación, las actividades correspondientes a cada método se detallan en tablas específicas.



Figura 3.17. Fase 2 Agregación multinivel y medición de la dinámica compleja de la metodología propuesta

Para comprender las acciones necesarias en la Fase 2 Agregación multinivel y medición de la dinámica compleja, se requiere delinear el procedimiento aplicado en el caso de estudio sobre movilidad vehicular. Este análisis presenta una descripción estructurada de la ejecución de cada actividad dentro de los métodos empleados, permitiendo una visión clara de su implementación en el sistema complejo de movilidad vehicular.

3.4.1 Niveles de agregación del sistema complejo dinámico del sistema complejo dinámico de movilidad vehicular

El nivel de agregación se refiere al grado de detalle con el que se representa un sistema, y en el caso de los sistemas complejos dinámicos, como la movilidad vehicular, determina la escala espacial, temporal y funcional con la que se agrupan las entidades individuales (por ejemplo, calles, intersecciones, manzanas, colonias o cuadros delimitadores) para su análisis e interpretación. En el presente estudio, el análisis multiescalar del sistema de movilidad vehicular se llevó a cabo mediante herramientas propias de la teoría de sistemas complejos. Se aplicaron enfoques como la clasificación jerárquica de componentes, que permite identificar subestructuras funcionales dentro del sistema, y el análisis multiescalar, que facilita la observación de patrones de comportamiento a través de distintos niveles de agregación. Un ejemplo concreto de esta jerarquía se observa en la transición del nivel micro (calle o segmento vial), al nivel meso (colonia o cuadrante urbano), y posteriormente al nivel macro (zona metropolitana). Este paso gradual de escala permitió analizar, cómo los patrones de congestión en calles específicas influyen en la movilidad de toda una colonia, y cómo la suma de estas afectaciones locales puede repercutir en la eficiencia del sistema vial metropolitano.

Para delimitar espacialmente las zonas de análisis y capturar el comportamiento emergente en cada escala, se empleó una segmentación estructurada mediante cuadros delimitadores (bounding boxes) georreferenciados. Desde la ciencia de datos, se aplicaron métodos de segmentación espacial, procesamiento masivo de datos geográficos y análisis comparativo de fuentes heterogéneas como HERE Maps con el objetivo de validar la granularidad y la consistencia espacial de los datos utilizados.

Para llevar a cabo el análisis de la literatura sobre los niveles de agregación del sistema complejo de movilidad vehicular, se realizó un estudio específico sobre los niveles de agregación que permite Here Maps para la obtención de datos de tráfico. Este análisis permitió identificar la granularidad con la que la plataforma representa la movilidad vehicular, desde niveles individuales hasta niveles agregados más amplios.

Paso 1, Selección de la muestra. Se decidió realizar un **muestreo por conveniencia**, seleccionando **una ciudad**, en este caso la Ciudad de México de entre las ciudades de la República Mexicana. Esta elección se fundamentó en el hecho de que la Ciudad de México es reconocida como una de las urbes más congestionadas a nivel mundial, ocupando el puesto 22 tal como lo indica el informe de INRIX Traffic 2023.

Paso 2, Identificación de los niveles de agregación para la Ciudad de México. De acuerdo a la división política administrativa de la Ciudad de México (ver Figura 3.18), esta comprende las **16 alcaldías** que conforman la Ciudad de México, cada una con su propio gobierno y autoridades locales, como se describe en la Tabla 3.8.



Figura 3.18. División política de la Ciudad de México

Tabla 3.8 Alcaldías de la Ciudad de México con su correspondiente número de colonias

Alcaldía	Número de colonias	Alcaldía	Número de colonias
Tlalpan	207	Magdalena Contreras	40
Milpa alta	39	Coyoacán	96
Xochimilco	107	Miguel Hidalgo	83
Iztapalapa	319	Azcapotzalco	84
Álvaro obregón	224	Venustiano Carranza	68
Gustavo a. Madero	178	Cuauhtémoc	41
Tláhuac	74	Benito Juárez	53
Cuajimalpa de Morelos	45	Iztacalco	37

Paso 3, Identificación de los niveles de agregación que soporta la aplicación de mapas de Here Maps: Se revisó la documentación de Here Maps para identificar los niveles de agregación disponibles para cubrir zonas geográficas y extraer datos de

movilidad vehicular. Estos niveles incluyen áreas circulares, cuadros delimitados y segmentos por longitud (ver Figura 3.19).

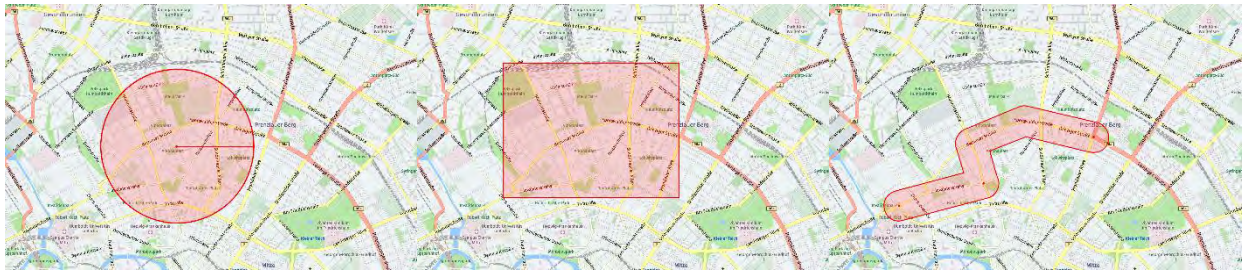


Figura 3.19. *Diferentes niveles de agregación soportados por la aplicación de Here Maps*

Usando cuadros delimitadores, es posible obtener un nivel de agregación por calle a partir de las aplicaciones de mapas. Al combinar varios cuadros, se cubre toda la Ciudad de México (ver Figura 3.20).



Figura 3.20. *Conjunto de cuadros delimitadores que cubren a toda la Ciudad de México*

Como resultado, el nivel de agregación para obtener datos de movilidad vehicular se centra en las **calles** dentro de cada alcaldía de la Ciudad de México. Para lograr esto, se adopta un enfoque que consiste en agrupar un conjunto de calles utilizando **cuadros delimitadores o bounding boxes**. Este método implica dividir el área de cada alcaldía en secciones más pequeñas y manejables, delimitadas por coordenadas geográficas específicas.

Se concluye que la obtención de datos de movilidad vehicular puede realizarse a distintos niveles de agregación, utilizando los métodos proporcionados por HERE Maps u otras fuentes de adquisición de datos. Esta flexibilidad permite adaptar el análisis a diferentes escalas y necesidades, permitiendo una representación precisa del comportamiento de la movilidad vehicular.

A partir de las actividades desarrolladas en el marco del método de niveles de agregación del sistema complejo dinámico, se ha establecido un conjunto de actividades sugeridas (ver Tabla 3.9) que permiten estructurar la segmentación del

sistema en distintos niveles jerárquicos, facilitando su análisis en múltiples escalas espaciales y temporales. Estas actividades proporcionan una estructura propuesta flexible, permitiendo su aplicación parcial o total en base a los requerimientos del estudio y la naturaleza del sistema complejo analizado.

El método se enfoca en el análisis de los diferentes niveles de agregación dentro del sistema complejo dinámico, estableciendo grados de granularidad y clasificando sus componentes para comprender su comportamiento en diversas escalas. Para ello, se desarrollan una serie de actividades, como la investigación de estudios previos sobre niveles de agregación, la recopilación de información sobre patrones y tendencias en la literatura, y la evaluación de enfoques existentes para definir estos niveles. Además, se lleva a cabo la agrupación y categorización de los niveles de agregación conforme a los criterios comunes, la identificación de variables asociadas a cada nivel y la generación de representaciones visuales que faciliten la interpretación de la estructura del sistema.

Para la correcta implementación de este método, es esencial la participación de diversos roles especializados. Cada uno de estos actores contribuye con su experiencia en la identificación de niveles de agregación adecuados, asegurando una representación estructurada y funcional del sistema complejo. Como salidas del proceso, se obtiene un resumen de hallazgos sobre niveles de agregación, acompañado de una taxonomía estructurada y una matriz de clasificación de niveles. Asimismo, se genera una lista de variables relevantes por nivel, representaciones gráficas que ilustran las distintas escalas de agregación y un documento final que sintetiza patrones, tendencias y conclusiones del análisis.

Tabla 3.9. Cuadro descriptivo del método niveles de agregación del sistema complejo dinámico

Nivel de análisis para el sistema complejo	
Descripción	Analizar los diferentes niveles de agregación del sistema dinámico complejo, definiendo grados de granularidad y clasificando componentes para entender su comportamiento a diversas escalas.
Actividades	- Investigar estudios previos sobre niveles de agregación en sistemas. - Recopilar información relevante sobre patrones y tendencias en la literatura. - Evaluar enfoques existentes para definir niveles de agregación. - Agrupar y categorizar los niveles de agregación a partir de los criterios comunes. - Diseñar una taxonomía para clasificar niveles de agregación. - Identificar las variables asociadas a cada nivel de agregación. - Generar representaciones visuales de los niveles definidos.
Roles	- Investigadores. - Analistas de tendencias y patrones. - Especialistas en clasificación. - Analistas de datos. - Expertos en impacto de niveles de agregación. - Diseñadores de sistemas y modelos estructurales.
Salidas	- Resumen de hallazgos sobre niveles de agregación. - Taxonomía de niveles de agregación. - Matriz de clasificación de niveles. - Lista de variables relevantes por nivel. - Representaciones gráficas de los niveles definidos. - Documento final que describe patrones, tendencias y conclusiones del análisis.

3.4.2 Análisis de incertidumbre del sistema complejo dinámico del sistema complejo de movilidad vehicular

En el análisis de incertidumbre del sistema complejo de movilidad vehicular se aplicaron técnicas de sistemas complejos, como la evaluación de entropía para identificar comportamientos emergentes e impredecibles en variables interdependientes, y el análisis de propagación de incertidumbre en sistemas dinámicos. Desde la ciencia de datos, se utilizaron estrategias de muestreo aleatorio, procesamiento masivo de datos multivariable y visualización de densidad y normalidad, apoyadas en representaciones gráficas como Q-Q plots y análisis de dispersión. En minería de datos, se aplicó la selección de características basada en entropía (feature selection) y se identificaron variables críticas mediante umbrales de información. En la construcción de indicadores, se utilizó la entropía de Shannon como métrica formal de incertidumbre, permitiendo priorizar variables con mayor impacto en la dinámica del sistema. Además, desde la física estadística, se aplicaron conceptos de desorden y entropía informacional para medir la aleatoriedad estructural del sistema; desde las matemáticas aplicadas, se utilizaron funciones de probabilidad, transformaciones logarítmicas y métricas de dispersión para el análisis cuantitativo; y desde la computación, se desarrollaron algoritmos en Python para el cálculo eficiente de la entropía y la automatización del filtrado de variables.

Para realizar el análisis de entropía de las variables del sistema complejo de movilidad vehicular, se llevó a cabo un proceso de análisis de entropía de un conjunto de variables a partir de un conjunto de datos que incluye datos de tráfico, incidencias viales y servicios. A continuación, se detallan los pasos seguidos en este proceso.

Paso 1, Selección de la muestra. Se utilizó un conjunto de datos de 20, 000, 000 registros con 41 variables descritas en la que corresponden a los datos de las 16 alcaldías de la Ciudad de México (datos recolectados en el período de agosto 2022-junio 2023). Se utilizó un muestreo aleatorio simple para seleccionar el conjunto de datos.

Paso 2, Selección del modelo de entropía. La selección del modelo para el cálculo de entropía de las variables de movilidad vehicular se optó por el modelo de Shannon que se muestra con la siguiente ecuación.

$$H(X) = - \sum_{x \in X} P(x) \log_2 P(x) \quad (3.6)$$

Donde:
X es la variable aleatoria.

Paso 3, Obtención de entropía de las variables con el modelo de Shannon, se realizó el cálculo de entropía de las variables de movilidad vehicular, descritas en la Tabla 3.5.

Los valores de entropía de las variables fueron calculados utilizando la ecuación 3.6 basada en el modelo de Shannon, el cual permite cuantificar el grado de incertidumbre asociado a una variable dentro de un sistema. Posteriormente, los resultados fueron ordenados de menor a mayor, dado que las variables con menor

entropía presentan un nivel reducido de incertidumbre y, por tanto, una mayor capacidad explicativa dentro del modelo. Con el objetivo de filtrar únicamente aquellas variables con mayor valor informativo, se estableció un umbral de entropía igual a 10. Este valor se determinó empíricamente tras analizar la distribución de entropías en el conjunto de datos, observando que las variables con valores superiores a dicho umbral tendían a comportarse de manera más aleatoria, aportando menor consistencia al modelo. De esta manera, el umbral de 10 permitió conservar aquellas variables con mayor relevancia estructural y reducir el ruido informacional en el proceso de modelado; las variables con entropía mayor a 10 no se tomaron en cuenta, ver la Tabla 3.10.

Tabla 3.10. Valores ordenados de entropía de cada variable de la movilidad vehicular

Número	Variable	Valor	Número	Variable	Valor
1	Road_Closed	0	14	Incident_Day	3.16992026
2	Day_Off_Traffic	0.50188384	15	Traffic_Minute	3.58377537
3	Day_Off_Incident	0.50188384	16	Incident_Minute	3.58383797
4	Criticality	0.72930232	17	Incident_Hour	4.20317612
5	Work_Day_Traffic	0.91758918	18	Traffic_Hour	4.42943189
6	Work_Day_Incident	0.91781269	19	Free_Flow	5.46871293
7	Incident_Month	0.91825872	20	Bbox_Traffic	5.76946977
8	Category_Number	1.16102762	21	Bbox_Incident	5.76946977
9	Number_Segments	1.49113744	22	Speed	5.87462317
10	Type	2.18833606	23	Speed_Uncapped	6.03088299
11	Traffic_Day_Number	2.72574448	24	Time	7.98038035
12	Incident_Day_Number	2.72574834	25	Length	9.61173521
13	Daily_Traffic	3.16991705			

Las variables que tienen un umbral de valor de entropía menor a 10 establecido en este trabajo de investigación, se consideran de mayor impacto en el desempeño de movilidad vehicular son: **Road_Closed, Day_Off_Traffic, Day_Off_Incident, Criticality, Work_Day_Traffic, Work_Day_Incident, Incident_Month, Category_Number, Number_Segments, Type, Traffic_Day_Number e Incident_Day_Number** por mencionar algunos. Estas variables se destacan como las variables de mayor impacto que contribuyen al desempeño de la movilidad vehicular.

La utilización de la entropía como medida de incertidumbre ha resaltado la importancia de considerar la variabilidad y la aleatoriedad de las variables relacionadas con el desempeño de movilidad vehicular, proporcionando una comprensión profunda de las variables que contribuyen en el comportamiento de la movilidad vehicular.

A partir de las actividades desarrolladas en el marco del método de análisis de incertidumbre del sistema complejo dinámico de la movilidad vehicular, se ha definido un conjunto de actividades sugeridas (ver Tabla 3.11) que permiten cuantificar y gestionar la variabilidad inherente al sistema. Estas actividades facilitan la identificación de fuentes de incertidumbre, la evaluación de su impacto en la dinámica del sistema y la caracterización de su propagación a lo largo del tiempo.

El método se basa en el análisis de la incertidumbre inherente en sistemas dinámicos complejos, e incluye la evaluación de distribuciones de probabilidad, entropía y variabilidad de las variables, con el fin de comprender su impacto en la evolución del sistema. La implementación de este método requiere la participación de diversos roles especializados. La interacción entre estos roles permite un análisis riguroso de la incertidumbre y su efecto sobre la movilidad vehicular, asegurando la aplicación de técnicas adecuadas para su reducción o control. Como salidas del proceso, se obtiene una lista de variables con niveles de incertidumbre identificados, junto con sus distribuciones de probabilidad y gráficos de densidad. También se generan cálculos de entropía y métricas de variabilidad, visualizaciones que ilustran la propagación de la incertidumbre, y un informe detallado sobre el análisis de incertidumbre y sus implicaciones en el sistema.

Tabla 3.11. Cuadro descriptivo del método análisis de incertidumbre del sistema complejo

Entropía de las variables del sistema complejo	
Descripción	Analizar la incertidumbre inherente en los sistemas dinámicos complejos, evaluando las distribuciones de probabilidad, entropía y variabilidad de las variables primordiales para entender su impacto en el comportamiento del sistema.
Actividades	- Identificar las variables importantes del sistema. - Evaluar la incertidumbre con la entropía de Shannon entre otros. - Analizar distribuciones de probabilidad de las variables - Determinar las variables con mayor impacto en la incertidumbre del sistema. - Generar visualizaciones de la variabilidad y la propagación de incertidumbre. - Documentar los hallazgos y proponer estrategias para mitigar la incertidumbre.
Roles	- Analistas de datos. - Matemáticos y estadísticos. - Expertos del dominio. - Modeladores de sistemas dinámicos. - Especialistas en gestión de incertidumbre.
Salidas	- Lista de variables con niveles de incertidumbre identificados. - Distribuciones de probabilidad y gráficos de densidad para las variables relevantes. - Cálculos de entropía y métricas de variabilidad. - Visualizaciones de propagación de incertidumbre. - Informe detallado sobre el análisis de incertidumbre. - Recomendaciones para mitigar incertidumbres críticas.

3.4.3 Análisis de la Identificación de las variables clave de acuerdo con la dinámica del sistema complejo de movilidad vehicular

En la identificación de variables clave relacionadas con la dinámica del sistema complejo de movilidad vehicular se aplicaron técnicas de sistemas complejos, como la descomposición estructural del sistema en variables dependientes e independientes y el análisis de relaciones causales multivariadas, permitiendo comprender la influencia de múltiples factores sobre la congestión. Desde la ciencia de datos, se utilizó un protocolo de revisión sistemática, análisis exploratorio de literatura y minería bibliográfica en bases de datos académicas, junto con el uso de técnicas de visualización de relaciones para priorizar variables con mayor influencia. En minería de datos, se aplicaron métodos de correlación, análisis de impacto y selección de características mediante filtrado estadístico y métricas de información. En la construcción de indicadores, se definió la congestión como variable síntesis del desempeño urbano, integrando múltiples factores explicativos como velocidad,

densidad, volumen y condiciones viales. Se incorporó también cómputo evolutivo como marco conceptual para refinar iterativamente la selección de variables relevantes a partir de su contribución al ajuste del modelo. Adicionalmente, se utilizaron técnicas de matemáticas aplicadas, como regresión multivariable, análisis de varianza y matrices de influencia; desde la física de redes, se adoptó el enfoque de centralidad e influencia nodal en redes de movilidad; y desde la computación, se desarrollaron herramientas automatizadas para extracción de patrones y clasificación jerárquica de variables.

Para realizar el análisis de la literatura para identificar las variables dependientes e independientes del sistema complejo de movilidad vehicular, se llevó a cabo un proceso de análisis de la literatura. A continuación, se detallan los pasos seguidos en este proceso.

Paso 1, Protocolo de búsqueda para la revisión sistemática sobre movilidad vehicular: Este protocolo se llevó a cabo realizando lo siguiente:

1. Pregunta de investigación:

¿Cuáles son las variables dependientes e independientes más estudiadas en la investigación sobre movilidad vehicular y sus efectos?

2. Bases de datos a utilizar:

PubMed, Scopus, Web of Science y Google Scholar

3. Términos de búsqueda:

Movilidad vehicular, Tráfico vehicular, Transporte urbano, Congestión del tráfico, Seguridad vial, Impacto ambiental del transporte, Infraestructura vial, Políticas de transporte y Factores de comportamiento del conductor.

4. Estrategia de búsqueda:

Se utilizaron operadores booleanos (AND, OR), se utilizaron comillas para buscar términos exactos y el uso de* para buscar variaciones de una palabra.

5. Criterios de inclusión:

Estudios publicados en inglés o español, trabajos que involucren variables relacionadas con la movilidad vehicular y trabajos que incluyan una clara identificación de variables dependientes e independientes.

6. Criterios de exclusión: Investigaciones no relacionadas con la movilidad vehicular, trabajos que no identifiquen claramente variables dependientes e independientes y estudios que no estén disponibles en texto completo.

Resultados, se obtuvieron conjunto de trabajos de la literatura que tratan del tráfico, transporte y movilidad vehicular, en las que se identifican diversas variables independientes y dependientes, como se muestra en la Tabla 3.12.

Tabla 3.12. Trabajos relacionados con las variables independientes y dependientes de la movilidad vehicular

Autor	Título del artículo	Variable dependiente	Variables independientes
(Duranton & turner, 2011)	The Fundamental Law of Road Congestion: Evidence from US Cities	Congestión de tráfico	Dimensiones de las calles, presencia de autopistas, cultura de conducción y transporte público.
(Jin et al., 2023)	Spatio-temporal graph neural point process for traffic congestion event prediction	Congestión de tráfico	Estructura de la red de carreteras, niveles de velocidad, tiempo.
(Perveen et al., 2017)	Evaluating transport externalities of urban growth: a critical review of scenario-based planning methods.	Congestión de tráfico	Crecimiento de la población, patrones de uso de la tierra, infraestructura de transporte, impactos ambientales.
(Yang et al., 2020)	The Marginal Cost of Traffic Congestion and Road Pricing: Evidence from a Natural Experiment in Beijing	Densidad del tráfico	Velocidad, restricciones de conducción, tarifa vial.
(Duan et al., 2023)	Spatiotemporal dynamics of traffic bottlenecks yields an early signal of heavy congestions	Atasco de tráfico	Propagación del tráfico, tamaño del atasco, velocidad.
(Li et al., 2024)	Analysis of Urban Congestion Traceability: The Role of the Built Environment	Congestión de tráfico	Señalización móvil, ubicación geográfica, densidad de población, densidad de carreteras.
(Arti et al., 2022)	Urban traffic congestion: its causes-consequences-mitigation	Congestión de tráfico	Infraestructura de carreteras, la gestión del flujo de tráfico, la composición del tráfico, el control de semáforos y las condiciones locales.
(Mohammed almatar, 2023)	Traffic congestion patterns in the urban road network: (Dammam metropolitan area)	Congestión de tráfico	Horas pico de la tarde, horas pico de la mañana.
(Huang & Loo, 2023)	Urban traffic congestion in twelve large metropolitan cities: A thematic analysis of local news contents, 2009–2018	Congestión de tráfico	Ciudades de alta densidad, política de transporte.
(X. Zhao et al., 2022)	Study on Identification and Prevention of Traffic Congestion Zones Considering Resilience-Vulnerability of Urban Transportation Systems	Congestión de tráfico	Volumen de tráfico, millas de congestión severa, tiempo de retraso, velocidad de operación promedio, condición de la carretera y número de carriles.
(Ramana et al., 2023)	A Vision Transformer Approach for Traffic Congestion Prediction in Urban Areas	Flujo de tráfico	Población, opciones de transporte e infraestructura de transporte.

El análisis de diversos estudios sobre movilidad vehicular revela que la congestión de tráfico emerge como la variable dependiente principal, dado su impacto determinante en la eficiencia y el desempeño del sistema de transporte. Su comportamiento está condicionado por múltiples factores, lo que la convierte en un indicador para evaluar la dinámica y el rendimiento de la movilidad urbana.

A partir de las actividades desarrolladas en el marco del método de identificación de las variables con la dinámica del sistema complejo de movilidad vehicular, se ha definido un conjunto de actividades sugeridas (ver Tabla 3.13) que permite la selección de los parámetros relevantes en la evolución del sistema. Este proceso facilita la

diferenciación entre variables de alto impacto y aquellas con menor influencia, asegurando que el modelo capture con precisión las relaciones esenciales que determinan su comportamiento y evolución.

Se aplican técnicas como análisis de regresión, correlación y métodos de entropía para cuantificar la influencia de las variables en la evolución del sistema. Se generan visualizaciones que facilitan la interpretación y priorización de variables significativas. Este proceso requiere colaboración interdisciplinaria para lograr un enfoque integral. Como resultado, se obtiene una lista priorizada de variables, diagramas de relaciones, análisis de impacto y recomendaciones para construir y optimizar el modelo de movilidad vehicular.

Tabla 3.13. Cuadro descriptivo del método de identificación de variables relevantes en relación con la dinámica

Variables relevantes del sistema complejo	
Descripción	Identificar las variables de importancia que influyen en la dinámica del sistema, analizando su interacción, importancia relativa y efecto sobre el comportamiento global.
Actividades	- Realizar una revisión de la literatura para identificar variables relevantes. - Analizar la relación entre variables independientes y dependientes. - Evaluar el impacto de cada variable en la dinámica del sistema. - Aplicar técnicas como análisis de regresión, correlaciones y métodos de entropía. - Generar visualizaciones para representar las interacciones dinámicas. - Priorizar las variables de mayor impacto en función de su relevancia en el sistema.
Roles	- Investigadores en sistemas dinámicos. - Analistas de datos. - Especialistas en modelado matemático. - Estadísticos y matemáticos. - Expertos del dominio del sistema.
Salidas	- Lista priorizada de variables base. - Diagramas de relaciones entre variables. - Resultados del análisis de impacto de las variables. - Visualizaciones de interacciones dinámicas. - Documento con recomendaciones para el modelado del sistema.

3.5 Recopilación y Monitoreo Continuo de Datos en Tiempo Real

Esta fase se centra en establecer un proceso eficiente para la captura, almacenamiento y procesamiento de datos provenientes de diversas fuentes en tiempo real. Esto garantiza que la información sea actualizada y refleje las condiciones dinámicas del sistema complejo en todo momento.

En esta fase, se implementan mecanismos para integrar datos de aplicaciones de mapas (como TomTom, HERE Maps y Geoapify) y otras fuentes relevantes. Esto incluye la configuración de pipelines de datos que automatizan la extracción, transformación y carga (ETL) de la información hacia bases de datos optimizadas para consultas en tiempo real, como MongoDB.

La Fase 3 está conformada por tres métodos principales: recopilación continua, integración de fuentes dinámicas de datos y diseño de sistemas de almacenamiento en tiempo real, como se muestra en la Figura 3.21. Esta fase tiene como objetivo asegurar la captura, integración y monitoreo constante de datos en tiempo real, aspectos para realizar el análisis y modelado de sistemas complejos altamente dinámicos.



Figura 3.21. Fase 3 Recopilación y monitoreo continuo de datos en tiempo real de la metodología propuesta

Para comprender las acciones necesarias en la Fase 3 Recopilación y Monitoreo Continuo de Datos en Tiempo Real., es primordial delinear el procedimiento aplicado en el caso de estudio sobre movilidad vehicular. Este análisis presenta una estructura clara y detallada de la ejecución de cada actividad dentro de los métodos empleados, permitiendo una visión clara de su implementación en el sistema complejo de movilidad vehicular.

3.5.1 Recopilación continua de datos de movilidad vehicular

En la recopilación continua de datos del sistema complejo de movilidad vehicular se aplicaron técnicas de sistemas complejos, como el seguimiento dinámico de variables en sistemas no estacionarios, y la captura de estados del sistema en alta frecuencia para representar su evolución temporal. Desde la ciencia de datos, se integraron procesos de automatización mediante scripts en Python para el consumo de APIs (Here Maps y Geoapify), programación de ventanas temporales de muestreo optimizadas mediante análisis de estabilidad y variabilidad de la velocidad, y configuración de flujos de captura, validación y almacenamiento de datos. En minería de datos, se aplicó el filtrado de registros redundantes y el control de calidad de datos en tiempo real, así como el análisis de series temporales para ajustar el intervalo óptimo de recolección. En la construcción de indicadores, los datos capturados se estructuraron en variables clave como velocidad, flujo, tipo de incidente y ubicación geográfica, habilitando el monitoreo continuo del sistema. Complementariamente, se emplearon técnicas de teoría de muestreo estadístico (como muestreo aleatorio y por conveniencia), estructuras de datos eficientes en informática aplicada (almacenamiento escalable y serialización JSON), y conceptos de física computacional, como el análisis de fluctuaciones en sistemas abiertos. Para llevar a cabo la recopilación continua de datos del sistema complejo de movilidad vehicular, se implementó un proceso de recolección de datos. A continuación, se detallan los pasos seguidos en este proceso.

Paso 1, Seleccionar la muestra. Se decidió realizar un muestreo por conveniencia, eligiendo la **Ciudad de México**.

Paso 2, Seleccionar el tamaño de la muestra. El tamaño de la muestra fueron las 6708 calles de las **16 alcaldías de la Ciudad de México** de acuerdo a la cobertura de Here Maps.

Paso 3, Identificar los datos a recolectar. Los datos a recolectar relacionados con la movilidad vehicular incluyen información sobre el **tráfico, incidencias viales y servicios**.

Paso 4, Periodicidad de captura de datos. Para definir los intervalos de tiempo de captura para la obtención de datos de tráfico e incidencias viales, se realizó un análisis de las velocidades promedio en ventanas de tiempo de 1, 2, 3, 4, 5, 10, 15 y 20 minutos. Se observó que no había una variabilidad significativa en la velocidad de la movilidad vehicular en intervalos de un minuto, mientras que en intervalos de 15 o 20 minutos se presentaban saltos bruscos en la variabilidad de la velocidad. Como resultado, se decidió seleccionar un intervalo de captura de datos de 5 minutos. Se muestra una adecuada variabilidad de la velocidad en 5 minutos como se muestra en la Figura 3.22.



Figura 3.22. Velocidades de la Ciudad de México en los primeros 100 minutos

Paso 5, Automatización de la recolección de datos: Se desarrolló un script en Python para automatizar la extracción de datos de tráfico e incidentes viales mediante la API de Here Maps. Además, para obtener información sobre servicios, se utilizó la API de Geopify, se muestra un fragmento de extracción de datos en la Figura 3.23.

```

C:\WINDOWS\system32\cmd.exe
Python 3.7.4 Shell
> python script.py
[123 rows x 13 columns]
  description  length  ... jamTendency  colortrafico
0  Eje 1 N      1866.0  ...      NaN      green
1  Calz San Simón  59.0  ...      NaN      green
2  Prol Guerrero  723.0  ...      1.0      green
3  Av del Trabajo/C Boleo  963.0  ...      NaN      green
4  Lázaro Cárdenas/Eje Central  85.0  ...      1.0      green
...
118 Paseo de la Jacarandas  1053.0  ...      NaN      green
119 Eje 1/Rayón/Calle Mosqueta  634.0  ...      NaN      green
120 Eje 1 Pte/Bucarell  570.0  ...      NaN      red
121 Eje 1 Pte/Bucarell  477.0  ...      NaN      green
122 Saturno  684.0  ...      NaN      green

[184 rows x 13 columns]
  description  length  ... jamTendency  colortrafico
0  Prolongación División del Norte  827.0  ...      NaN      green
1  Avenida 16 de Septiembre  112.0  ...      NaN      green
2  Calle Heliotropo  500.0  ...      NaN      green
3  Calz de las Bombas  819.0  ...      1.0      green
4  Calz del Hueso/Eje 1 Oriente  624.0  ...      NaN      green
...
179 Calz Mexico Xochimilco  1390.0  ...      NaN      green
180 C Vía Tlalpan  850.0  ...      NaN      green
181 Calz Redención  1079.0  ...      NaN      green
182 Avenida Guadalupe I Ramirez  265.0  ...      NaN      green
183 C Padre Selva/Camino a Nativitas  30.0  ...      NaN      green

```

Figura 3.23. Extracción de datos tráfico e incidencias viales en tiempo real mediante un Script en Python

Se recopilaban tres conjuntos de datos principales: tráfico vehicular, incidencias viales y servicios urbanos. Este último incluye puntos de atracción como centros de salud, espacios de recreación, zonas laborales, entre otros. Se puede consultar una muestra de los conjuntos de datos mencionados en el siguiente repositorio: <https://goo.su/TFtyzgh>

La movilidad vehicular en las calles de la Ciudad de México se actualiza cada 2 minutos de acuerdo a Here Maps. Sin embargo, el análisis gráfico de los datos históricos muestra que los cambios en la movilidad son mínimos en ese intervalo. Al aumentar el rango de actualización a 10 minutos, se observan variaciones significativas. Por esta razón, se decidió establecer el intervalo de actualización en 5 minutos, logrando un equilibrio entre precisión y variabilidad de los datos.

A partir de las actividades desarrolladas en el marco del método de recopilación continua de datos de movilidad vehicular, se ha definido un conjunto de actividades sugeridas (ver Tabla 3.14) que permite estructurar la adquisición de datos en tiempo real, permitiendo la captura sistemática de información relevante para el análisis del sistema complejo. Este enfoque facilita la integración de múltiples fuentes dinámicas de datos, asegurando la disponibilidad de registros actualizados que reflejen la evolución del sistema.

El método se fundamenta en la implementación de un sistema de recolección continua, diseñado para obtener información con eficiencia y precisión, permitiendo alimentar modelos dinámicos de movilidad vehicular. Para ello, se llevan a cabo diversas actividades, tales como la identificación de fuentes de datos relevantes, el diseño de un sistema automatizado para la captura continua de datos y la configuración de una periodicidad de recolección óptima, que garantice una representación detallada de la dinámica vehicular.

Para la implementación exitosa de este método, es requisito la participación de distintos roles. Cada uno de estos roles contribuye a garantizar que el proceso de recolección sea eficiente, robusto y adaptable a la evolución del sistema de movilidad vehicular. Como salidas del proceso, se obtiene un sistema automatizado de recolección de datos, que proporciona información con alta frecuencia y consistencia, asegurando la captura de datos en tiempo real. Además, se implementa un

almacenamiento escalable y seguro, acompañado de informes sobre la calidad y periodicidad de los datos capturados. Se genera una documentación técnica del sistema, que detalla su diseño, configuración y operación, asegurando su mantenimiento y actualización a lo largo del tiempo.

Tabla 3.14. Cuadro descriptivo del método recopilación continua

Obtención continua de datos	
Descripción	Desarrollar un sistema de captura continua de datos en tiempo real, eficiente y preciso, que sirva como base para análisis dinámicos y decisiones fundamentadas.
Actividades	- Identificar las fuentes de datos relevantes para la recolección. - Diseñar un sistema automatizado orientado a la captura continua de datos. - Configurar una periodicidad de recolección adecuada. - Garantizar la calidad y consistencia de los datos recopilados mediante validaciones. - Monitorear el sistema para detectar fallos o interrupciones en la captura de datos. - Almacenar los datos en un sistema diseñado para soportar acceso y análisis en tiempo real.
Roles	- Ingenieros de datos. - Especialistas en automatización. - Administradores de bases de datos. - Analistas de calidad de datos. - Desarrolladores de sistemas en tiempo real.
Salidas	- Sistema automatizado de recolección de datos. - Datos recopilados con alta frecuencia y consistencia. - Almacenamiento confiable y escalable para datos continuos. - Informes sobre la calidad y periodicidad de los datos capturados. - Documentación técnica del sistema implementado.

3.5.2 Integración de fuentes dinámicas de datos de movilidad vehicular

En la integración de fuentes dinámicas del sistema complejo de movilidad vehicular, se aplicaron técnicas propias de los sistemas complejos, articulando flujos de datos interdependientes para representar interacciones espacio-temporales entre tráfico, incidencias y servicios. Este enfoque favorece una comprensión holística del sistema de movilidad. Desde la perspectiva de ciencia de datos, se implementaron estrategias de recolección multifuente mediante APIs (Here Maps, Geoapify), seguidas de procesos de normalización, estructuración y enriquecimiento semántico de los datos mediante el cálculo de variables derivadas, como la distancia entre eventos o el tiempo de respuesta ante una incidencia. En minería de datos, se aplicaron técnicas de limpieza, fusión y emparejamiento condicional, utilizando uniones y productos cruzados combinados con filtros basados en correspondencia espacial (por coordenadas geográficas) y semántica (por tipo de evento o servicio). Además, se diseñaron indicadores que cuantifican relaciones funcionales, espaciales y temporales entre las entidades involucradas. Se incorporaron principios de álgebra relacional y lógica difusa para la formulación de reglas de combinación de datos heterogéneos. Desde la física computacional, se adoptó el uso de distancias efectivas basadas en rutas reales, en lugar de métricas euclidianas, para una representación precisa de la conectividad urbana. En paralelo, desde la ingeniería de software, se desarrollaron scripts automatizados en Python para ejecutar tareas de recolección, integración y cálculo de relaciones causales de manera recurrente (ver Figura 3.24).

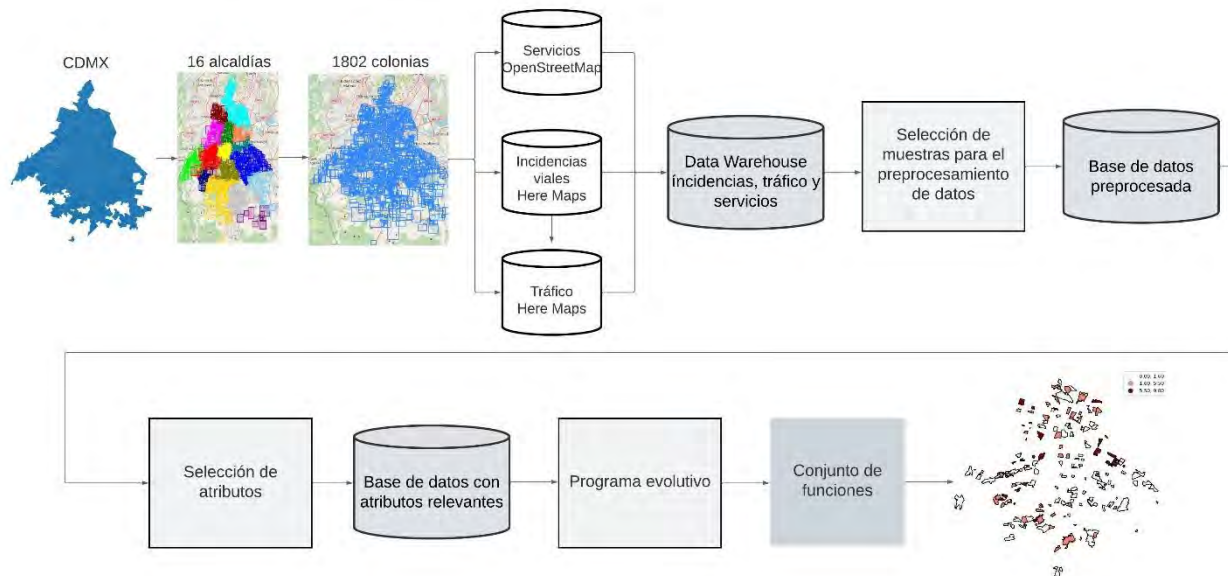


Figura 3.24. Esquema del proceso de integración, selección y modelado de datos para la generación de funciones a partir de información vial, de servicios y de tráfico en la Ciudad de México

Para llevar a cabo Integración de fuentes dinámicas de datos del sistema complejo de movilidad vehicular, se implementó un proceso de integración y almacenamiento. A continuación, se detallan los pasos seguidos en este proceso.

Paso 1, Seleccionar la muestra. Se decidió realizar un muestreo aleatorio simple.

Paso 2, Determinar tamaño de la muestra: Se utilizó el conjunto de datos de tráfico, incidencias y servicios, que incluye 100, 000, 000 registros de tráfico, 10, 000, 000 registros de incidencias viales (recolectados durante el período agosto 2022-diciembre 2023) y 32,000 registros de servicios, los datos de tráfico e incidencias se eligieron con muestreo aleatorio simple. Con el fin de determinar el tamaño adecuado de la muestra, se empleó la fórmula para el cálculo de proporciones, como se muestra en la siguiente ecuación (Cochran, 1977).

$$n = \frac{Z^2 \cdot p \cdot (1-p)}{E^2} \quad (3.7)$$

Donde:

n es el tamaño de la muestra.

Z es el valor Z correspondiente al nivel de confianza deseado (1.96 para el 95% de confianza).

p es la proporción esperada (en este caso 0.5).

E es el margen de error (0.05 para un 5% de margen de error).

Se utilizó la calculadora de <https://goo.su/Fzzxlo> que implementa la ecuación 3.7, para determinar el tamaño de las muestras, los resultados se muestran en la Figura 3.25.

a) b) c)

Figura 3.25. Obtención de tamaño de muestra para el conjunto de datos de tráfico, incidencias y servicios

Obteniendo como resultado de tamaño de la muestra para tráfico de 3, 699, 482 registros, incidencias de 2, 775, 403 registros y servicios de 31,737.

Paso 3, Integrar los conjuntos de tráfico, incidencias viales y servicios: Para la integración de datos se utilizó el producto cruz utilizando la siguiente ecuación:

$$T_{(IS)} = |T| \times |I| \times |S| \quad (3.8)$$

Donde:

$T_{(IS)}$ es la integración de los conjuntos de tráfico, incidencias viales y servicios.

I es el conjunto de datos de incidencias viales.

T es el conjunto de datos de tráfico.

S es el conjunto de datos de servicios.

Para tener un solo conjunto de datos de Tráfico_Incidencias_Servicios (T_{I_S}), se implementaron los valores del paso 2 en la ecuación descrita en el paso 3, y la cantidad de registros es la que se muestra a continuación, con 41 variables.

$$T_{(IS)} = |3699482| \times |2775403| \times |31737| = 50,000,000$$

Nota: El resultado del producto cruz solo se incluyeron los registros que tienen una relación entre la calle donde se origina el tráfico con respecto a donde ocurre la incidencia y el servicio, las que no tenían relación no se incluyeron.

Paso 4, Cálculo de nuevas variables. Para establecer una relación de causalidad entre las incidencias viales y los servicios al tráfico, se calcularon las distancias entre tres pares de elementos: la calle y las incidencias viales, la calle y la ubicación de los servicios. Para este cálculo se utilizó la API de enrutamiento (Routing API) de Geoapify, la cual permite obtener la distancia entre dos puntos mediante rutas reales, como se muestra Figura 3.26.

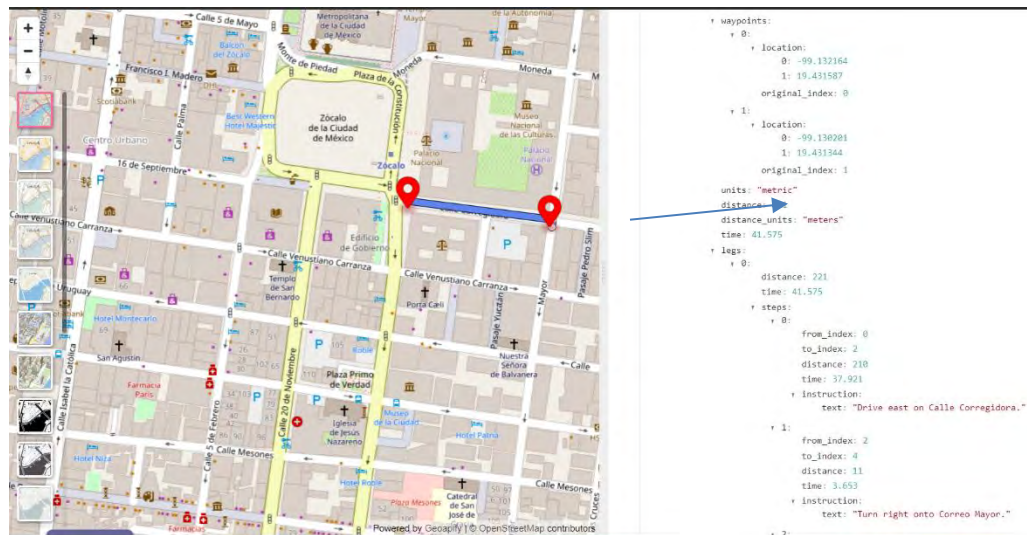


Figura 3.26. Cálculo de distancia entre 2 puntos utilizando ROUTING API de Geoapify

A partir del conjunto de datos Tráfico_Incidencias_Servicios, se realizó el cálculo de distancias espaciales entre las calles asociadas al flujo vehicular y los puntos correspondientes a incidencias y servicios (como salud, trabajo y recreación). Este procedimiento se implementó mediante la API de Geoapify, utilizando la técnica conocida como cálculo de distancias geoespaciales punto a línea (point-to-line distance calculation), como se muestra en la Figura 3.27, así como las nuevas variables que se presentan en la Tabla 3.15 y Tabla 3.16.

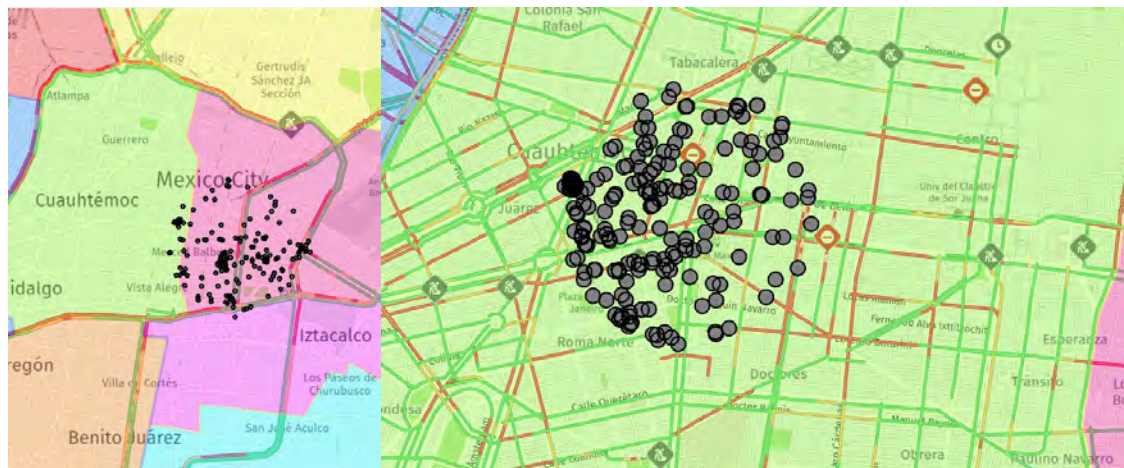


Figura 3.27. Distancias entre el tráfico con respecto a las incidencias y a los servicios

Tabla 3.15. *Conjunto de variables de distancias entre tráfico, incidencias y servicios, parte 1*

No.	Variable	Descripción de la variable
1	d1_d	Distancia en metros entre el inicio de incidencia con la ubicación de servicios.
2	d1_t	Distancia en segundos entre el inicio de incidencia con la ubicación de servicios.
3	d1_nc	Distancia en número de calles entre el inicio de incidencia con la ubicación de servicios.
4	d2_d	Distancia en metros entre la mitad de la incidencia con la ubicación de servicios.
5	d2_t	Distancia en segundos entre la mitad de la incidencia con la ubicación de servicios.
6	d2_nc	Distancia en número de calles entre la mitad de la incidencia con la ubicación de
7	d3_d	Distancia en metros entre el fin de incidencia con la ubicación de servicios.
8	d3_t	Distancia en segundos entre el fin de incidencia con la ubicación de servicios.
9	d3_nc	Distancia en número de calles entre el fin de incidencia con la ubicación de servicios.
10	d4_d	Distancia en metros entre el inicio de tráfico con la ubicación de servicios.
11	d4_t	Distancia en segundos entre el inicio de tráfico con la ubicación de servicios.
12	d4_nc	Distancia en número de calles entre el inicio de tráfico con la ubicación de servicios.
13	d5_d	Distancia en metros entre la mitad de tráfico con la ubicación de servicios.
14	d5_t	Distancia en segundos entre la mitad de tráfico con la ubicación de servicios.
15	d5_nc	Distancia en número de calles entre la mitad de tráfico con la ubicación de servicios.
16	d6_d	Distancia en metros entre el fin de tráfico con la ubicación de servicios.
17	d6_t	Distancia en segundos entre el fin de tráfico con la ubicación de servicios.
18	d6_nc	Distancia en número de calles entre el fin de tráfico con la ubicación de servicios.
19	d7_d	Distancia en metros entre el inicio de la incidencia con el inicio de tráfico.
20	d7_t	Distancia en segundos entre el inicio de la incidencia con el inicio de tráfico.
21	d7_nc	Distancia en número de calles entre el inicio de la incidencia con el inicio de tráfico.
22	d8_d	Distancia en metros entre el inicio de la incidencia con la mitad de tráfico.
23	d8_t	Distancia en segundos entre el inicio de la incidencia con la mitad de tráfico.
24	d8_nc	Distancia en número de calles entre el inicio de la incidencia con la mitad de tráfico.
25	d9_d	Distancia en metros entre el inicio de la incidencia con el fin de tráfico.
26	d9_t	Distancia en segundos entre el inicio de la incidencia con el fin de tráfico.
27	d9_nc	Distancia en número de calles entre el inicio de la incidencia con el fin de tráfico.
28	d10_d	Distancia en metros entre la mitad de la incidencia con el inicio de tráfico.
29	d10_t	Distancia en segundos entre la mitad de la incidencia con el inicio de tráfico.
30	d10_nc	Distancia en número de calles entre la mitad de la incidencia con el inicio de tráfico.
31	d11_d	Distancia en metros entre la mitad de la incidencia con la mitad de tráfico.
32	d11_t	Distancia en segundos entre la mitad de la incidencia con la mitad de tráfico.

Tabla 3.16. *Conjunto de variables de distancias entre tráfico, incidencias y servicios, parte 2*

No.	Variable	Descripción de la variable
33	d11_nc	Distancia en número de calles entre la mitad de la incidencia con la mitad de tráfico.
34	d12_d	Distancia en metros entre la mitad de la incidencia con el fin de tráfico.
35	d12_t	Distancia en segundos entre la mitad de la incidencia con el fin de tráfico.
36	d12_nc	Distancia en número de calles entre la mitad de la incidencia con el fin de tráfico.
37	d13_d	Distancia en metros entre el fin de la incidencia con el inicio de tráfico.
38	d13_t	Distancia en segundos entre el fin de la incidencia con el inicio de tráfico.
39	d13_nc	Distancia en número de calles entre el fin de la incidencia con el inicio de tráfico.
40	d14_d	Distancia en metros entre el fin de la incidencia con la mitad de tráfico.
41	d14_t	Distancia en segundos entre el fin de la incidencia con la mitad de tráfico.
42	d14_nc	Distancia en número de calles entre el fin de la incidencia con la mitad de tráfico.
43	d15_d	Distancia en metros entre el fin de la incidencia con el fin de tráfico.
44	d15_t	Distancia en segundos entre el fin de la incidencia con el fin de tráfico.
45	d15_nc	Distancia en número de calles entre el fin de la incidencia con el fin de tráfico.

A partir del método de integración de fuentes dinámicas de datos, se definieron actividades sugeridas (ver Tabla 3.17) para unificar información de diversas plataformas en un sistema de análisis. Este proceso asegura la incorporación de datos heterogéneos, mejorando la representación y precisión del sistema de movilidad vehicular. La integración fortalece el modelo al ofrecer una visión holística y dinámica del sistema complejo.

El método contempla la recopilación, procesamiento y consolidación de datos desde APIs de tráfico, sensores IoT y bases externas. Su implementación requiere roles especializados para garantizar una integración eficiente, escalable y confiable, optimizando el modelado y análisis del sistema.

Como salidas del proceso, se obtiene un sistema con datos dinámicos integrados, acompañado de modelos de datos unificados y optimizados que mejoran la estructura y organización de la información. Asimismo, se implementan herramientas de monitoreo para garantizar la continuidad y estabilidad de la integración, junto con informes de calidad y validación de datos que permiten evaluar la precisión y consistencia de la información recopilada.

Tabla 3.17. Cuadro descriptivo del método Integración de fuentes dinámicas de datos

Integración de datos	
Descripción	Integrar múltiples fuentes de datos dinámicos para construir un sistema consolidado que permita el análisis y la toma de decisiones en tiempo real, garantizando la coherencia y accesibilidad de los datos.
Actividades	- Identificar y seleccionar las fuentes de datos dinámicas relevantes (APIs, sensores IoT, bases de datos externas). - Establecer conexiones para la recopilación continua de datos en tiempo real. - Diseñar procesos de integración para combinar datos de diferentes formatos y estructuras. - Implementar mecanismos de validación y limpieza de datos para garantizar su calidad. - Generar un modelo unificado de datos consolidados. - Automatizar la integración mediante herramientas y scripts programáticos.
Roles	- Ingenieros de datos. - Especialistas en integración de sistemas. - Desarrolladores de software. - Analistas de calidad de datos. - Administradores de bases de datos.
Salidas	- Sistema consolidado con datos dinámicos integrados. - Modelos de datos unificados y optimizados. - Herramientas de monitoreo para garantizar la continuidad de la integración. - Informes de calidad y validación de datos. - Documentación técnica sobre procesos y fuentes de datos integradas.

3.5.3 Diseño de sistemas de almacenamiento en tiempo real para los datos de movilidad vehicular

En el diseño de sistemas de almacenamiento en tiempo real para el sistema complejo de movilidad vehicular se aplicaron técnicas de sistemas complejos, orientadas a la gestión de flujos de datos multivariables, interdependientes y de alta frecuencia, facilitando la representación continua del estado del sistema en entornos urbanos dinámicos. Desde la ciencia de datos, se implementó un pipeline automatizado de recolección, transformación y carga (ETL) a través de scripts en Python y consumo de APIs, estructurando datos heterogéneos en formatos JSON/XML para su

almacenamiento eficiente. Se utilizaron bases de datos NoSQL como MongoDB, optimizadas para consultas simultáneas y alta escalabilidad, y se aplicaron técnicas de diseño de modelos clave-valor para minimizar latencia y maximizar acceso concurrente. En minería de datos, se desarrollaron procesos de validación estructural, consistencia temporal y filtrado de registros redundantes, y se incorporaron reglas para la detección de patrones de actualización de datos en tiempo real. En la construcción de indicadores, se generaron variables derivadas (distancias, tiempos, relaciones espaciales) que capturan la interacción entre tráfico, incidencias y servicios, enriqueciendo semánticamente el modelo de almacenamiento. Además, se integraron conceptos de teoría de la información (para evaluar redundancia y carga informativa de atributos), técnicas de sistemas distribuidos (para tolerancia a fallos y consistencia eventual), y fundamentos de física computacional, como el modelado de estados del sistema a partir de flujos entrantes.

Para llevar a cabo el diseño de sistemas de almacenamiento en tiempo real para los datos de movilidad vehicular. A continuación, explican los pasos:

Paso 1. Petición de datos: En este primer paso, el sistema envía solicitudes de datos a aplicaciones de mapas como Here Maps y TomTom Traffic mediante un script desarrollado en Python. Este script utiliza protocolos de comunicación estándar, como HTTP/2, para obtener datos actualizados en tiempo real relacionados con la movilidad vehicular. Las solicitudes fueron diseñadas específicamente para recopilar información sobre velocidad promedio, densidad de tráfico, incidencias viales y la disponibilidad de servicios urbanos, asegurando así una base de datos completa y relevante para el análisis.

Paso 2. Respuesta de datos: Una vez realizada la solicitud, las fuentes externas responden enviando los datos en formatos estructurados, como JSON o XML. Estos datos se almacenan en directorios para mantener una organización de los datos.

Paso 3. Estructura de almacenamiento de datos: En esta etapa, los datos recopilados son procesados y organizados en un formato clave-valor para su almacenamiento eficiente en la base de datos. Este proceso permite que la información sea fácilmente accesible y utilizable en tiempo real.

Paso 4. Almacenamiento de datos: Los datos se almacenan en un sistema de base de datos diseñado para manejar grandes volúmenes de información en tiempo real. En este caso, se utiliza un sistema NoSQL, como MongoDB, que permite consultas rápidas y escalabilidad. Esta etapa asegura que los datos estén organizados, sean fácilmente accesibles y estén disponibles para análisis dinámicos, modelado y toma de decisiones en tiempo real.

Con los pasos anteriores se tiene el diseño de sistemas de almacenamiento en tiempo real para los datos de movilidad vehicular, como se muestra en la Figura 3.28.

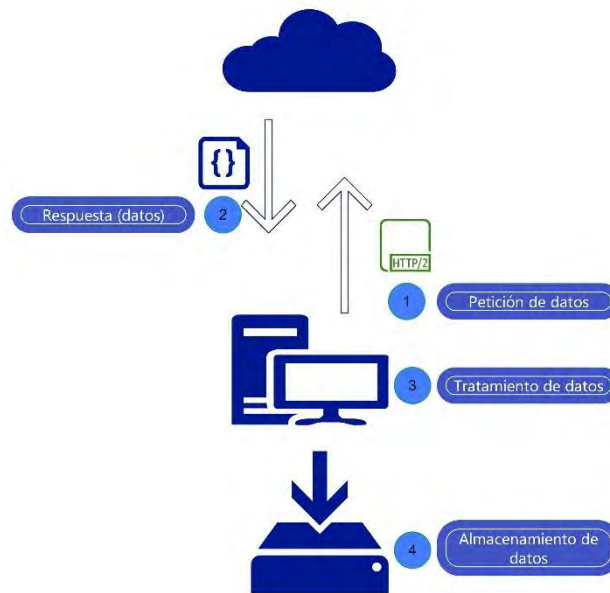


Figura 3.28. Proceso de Diseño de Sistemas de Almacenamiento en Tiempo Real para Datos de Movilidad Vehicular

Se obtiene un conjunto de datos de tráfico, incidencias y servicios con 86 variables (es la suma de las variables de la Tabla 3.5 y Tabla 3.15). Se puede consultar una muestra de los conjuntos de datos mencionados en el siguiente repositorio: <https://goo.su/TFtyzgh>.

La integración de datos entre las variables de tráfico, incidencias y servicios se vuelve compleja por la influencia combinada de estos factores en la movilidad vehicular. El desempeño de la movilidad es afectado por el lugar donde ocurre el tráfico, la calle donde se origina la incidencia y la ubicación del servicio. Aunque esta interrelación produce numerosas combinaciones, resulta necesario para un análisis preciso de la movilidad vehicular. Por ello, se han generado nuevas variables que muestran las distancias entre las variables de tráfico, incidencias y servicios, facilitando un análisis más detallado de su interacción y su impacto en la movilidad.

A partir de las actividades desarrolladas en el marco del método de diseño de sistemas de almacenamiento en tiempo real para los datos de movilidad vehicular, se ha definido un conjunto de actividades sugeridas (ver Tabla 3.18) que permiten la estructuración y optimización de infraestructuras tecnológicas para la gestión eficiente de datos en sistemas complejos.

El método se centra en el diseño de una infraestructura de almacenamiento de alto rendimiento, idóneo de gestionar datos en tiempo real con alta frecuencia y gran volumen, optimizada para sistemas dinámicos complejos. Para ello, se llevan a cabo diversas actividades, tales como la identificación de los requisitos de almacenamiento de tecnologías adecuadas como bases de datos NoSQL y sistemas distribuidos, y el diseño de una estructura de datos optimizada para consultas rápidas y eficientes.

La implementación de este método requiere la participación de distintos roles. La colaboración entre estos roles permite el diseño de una infraestructura robusta, adaptable y eficiente, asegurando que el sistema de almacenamiento responda a las necesidades del análisis de movilidad vehicular en entornos urbanos complejos. Como salidas del proceso, se obtiene una infraestructura de almacenamiento implementada y validada, con un modelo de datos optimizado para el procesamiento en tiempo real. Asimismo, se logra la integración exitosa con fuentes de datos dinámicas, lo que permite la disponibilidad y actualización continua de la información. Se generan además documentos técnicos que describen el diseño del sistema, su configuración y mejores prácticas para su mantenimiento y escalabilidad.

Tabla 3.18. Cuadro descriptivo del método diseño de sistemas de almacenamiento en tiempo real

Almacenamiento de datos en tiempo real	
Descripción	Diseñar una infraestructura de almacenamiento adecuada para gestionar datos en tiempo real, con alta frecuencia y gran volumen, optimizada para sistemas dinámicos complejos
Actividades	- Identificar los requisitos de almacenamiento y procesamiento en tiempo real. - Seleccionar tecnologías adecuadas (bases de datos NoSQL, sistemas distribuidos). - Diseñar la estructura del modelo de datos para consultas rápidas. - Integrar fuentes de datos en tiempo real (sensores, APIs, IoT). - Implementar sistemas de monitoreo y alerta para garantizar la integridad de los datos. - Optimizar el sistema para manejar picos de datos y escalabilidad futura.
Roles	- Ingenieros de datos. - Especialistas en bases de datos. - Expertos en tecnologías de almacenamiento distribuido. - Desarrolladores de sistemas en tiempo real. - Analistas de requisitos tecnológicos.
Salidas	- Infraestructura de almacenamiento implementada y validada. - Modelo de datos optimizado para tiempo real. - Integración exitosa con fuentes de datos dinámicas. - Documentación técnica del diseño del sistema. - Informes de rendimiento y pruebas de escalabilidad.

3.6 Preprocesamiento y actualización dinámica de datos

Esta fase está centrada en garantizar la calidad, consistencia y relevancia de la información necesaria para el análisis y modelado de sistemas complejos altamente dinámicos. Esta etapa incluye actividades como la limpieza de datos, donde se identifican y eliminan valores atípicos, inconsistentes o duplicados, además del tratamiento de datos faltantes mediante técnicas de imputación. Asimismo, se realiza la normalización y transformación de los datos para uniformar unidades, ajustar rangos y corregir distribuciones sesgadas o no lineales. Otro aspecto primordial es la actualización dinámica, que permite la integración de datos en tiempo real provenientes de múltiples fuentes, asegurando que los modelos se adapten a los cambios del sistema y mantengan su pertinencia. La Fase 4 está compuesta por dos métodos principales: preprocesamiento en tiempo real y actualización y muestreo dinámico, como se muestra en la Figura 3.29. Esta fase se enfoca en garantizar la calidad y relevancia de los datos mediante su actualización continua y preprocesamiento dinámico, procesos esenciales para el análisis efectivo de sistemas complejos altamente dinámicos.



Figura 3.29. Fase 4 Preprocesamiento y actualización dinámica de datos de la metodología propuesta

Para comprender las acciones necesarias en la Fase 4 Preprocesamiento y actualización dinámica de datos, es esencial delinear el procedimiento aplicado en el caso de estudio sobre movilidad vehicular. Este análisis presenta una estructura clara y ordenada, la ejecución de cada actividad dentro de los métodos empleados, permitiendo una visión clara de su implementación en el sistema complejo de movilidad vehicular.

3.6.1 Preprocesamiento en tiempo real de datos de movilidad vehicular

En el preprocesamiento en tiempo real del sistema complejo de movilidad vehicular se aplicaron técnicas de sistemas complejos para depurar, transformar y estructurar datos provenientes de múltiples entidades interactivas, como tráfico, incidencias y servicios, en entornos dinámicos y no lineales. Desde la ciencia de datos, se implementaron procedimientos de transformación (raíz cuadrada, z-score, cúbica), normalización y estandarización por min-max scaling, aplicados de manera automatizada con scripts en Python sobre grandes volúmenes de datos, permitiendo su uniformidad y comparabilidad. En minería de datos, se aplicaron técnicas de selección de características basadas en análisis de medias, correlación de Spearman, análisis de componentes principales (PCA), modelos de ecuaciones estructurales y pruebas de causalidad de Granger, permitiendo identificar variables clave con impacto directo sobre el factor de congestión (jamFactor). En la construcción de indicadores, se analizaron relaciones espacio-temporales y comportamientos emergentes del sistema. Además, se incorporaron técnicas de estadística para el manejo de valores atípicos, conceptos de análisis funcional de datos (FDA) para evaluar variabilidad multitemporal, y principios de computación distribuida para asegurar eficiencia y escalabilidad del preprocesamiento en sistemas intensivos de datos.

Para llevar a cabo el preprocesamiento de los datos del sistema complejo de movilidad vehicular, se realizaron los siguientes pasos.

Paso 1, Seleccionar la población: Se seleccionó una población de 50, 000, 000 de registros con 86 variables, obtenidos del producto cruz.

Paso 2, Determinar el tamaño de la muestra: Para seleccionar el tamaño de la muestra se utilizó la ecuación 3.7, con la calculadora que lo implementa de <https://goo.su/Fzzxlo> se obtuvo un resultado de tamaño de muestra de 3, 567, 503 con 86 variables, así como se muestra en la Figura 3.30.

Figura 3.30. *Tamaño de la muestra de Tráfico, incidencias y servicios*

Paso 3, transformación de los datos: La transformación llevada a cabo abarcó la utilización de la raíz cuadrada, el z-score y la transformación cúbica, aplicadas mediante las ecuaciones 3.9, 3.10 y 3.11, respectivamente, en el conjunto de datos de tráfico, incidencias y servicios. Este proceso tuvo como objetivo normalizar los datos, ajustándose para que los datos siguieran una distribución más uniforme y facilitando su análisis posterior.

$$\text{Variable normalizada} = \sqrt{(\text{Variable original})} \quad (3.9)$$

$$z = \frac{x - \mu}{\sigma} \quad (3.10)$$

Donde:

x es el valor individual de la variable.

μ es la media de la distribución de la variable.

σ es la desviación estándar de la distribución de la variable.

$$\text{Variable normalizada} = \sqrt[3]{(\text{Variable original})} \quad (3.11)$$

Las ecuaciones 3.9, 3.10 y 3.11 fueron implementados en un script de Python y aplicados a la muestra de 3, 567, 503 registros con 86 variables. Los hallazgos derivados de la normalización del conjunto de datos de tráfico, incidencias y servicios señalan que la aplicación de la transformación cúbica resulta en una notable mejora en la normalidad de este conjunto de datos.

Paso 4, estandarización de los datos: Para el conjunto de datos de tráfico, incidencias y servicios de 3, 567, 503 registros con 86 variables, que incluye variables con escalas diferentes, se realizó un proceso de estandarización que estandarizó los valores dentro de un rango de 0 a 1. Esta estandarización se efectuó utilizando la ecuación número 3.12.

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3.12)$$

Donde:

x es el valor original que se desea normalizar.

x_min es el valor mínimo en el conjunto de datos original.

x_max es el valor máximo en el conjunto de datos original.

La estandarización produjo que todas las variables del conjunto de datos de tráfico, incidencias y servicios fueran ajustadas a un intervalo de valores que oscila entre 0 y 1. Este proceso garantiza que todas las variables estuvieran en una escala uniforme y comparable entre sí.

Paso 5, Selección de variables: Para la selección de variables se utilizaron las siguientes técnicas:

- Identificación de variables que impactan en la movilidad vehicular mediante el cálculo de las medias de cada variable independiente con respecto a la variable dependiente (jamFactor) que es el factor de atasco, como se muestra en la Figura 3.37.

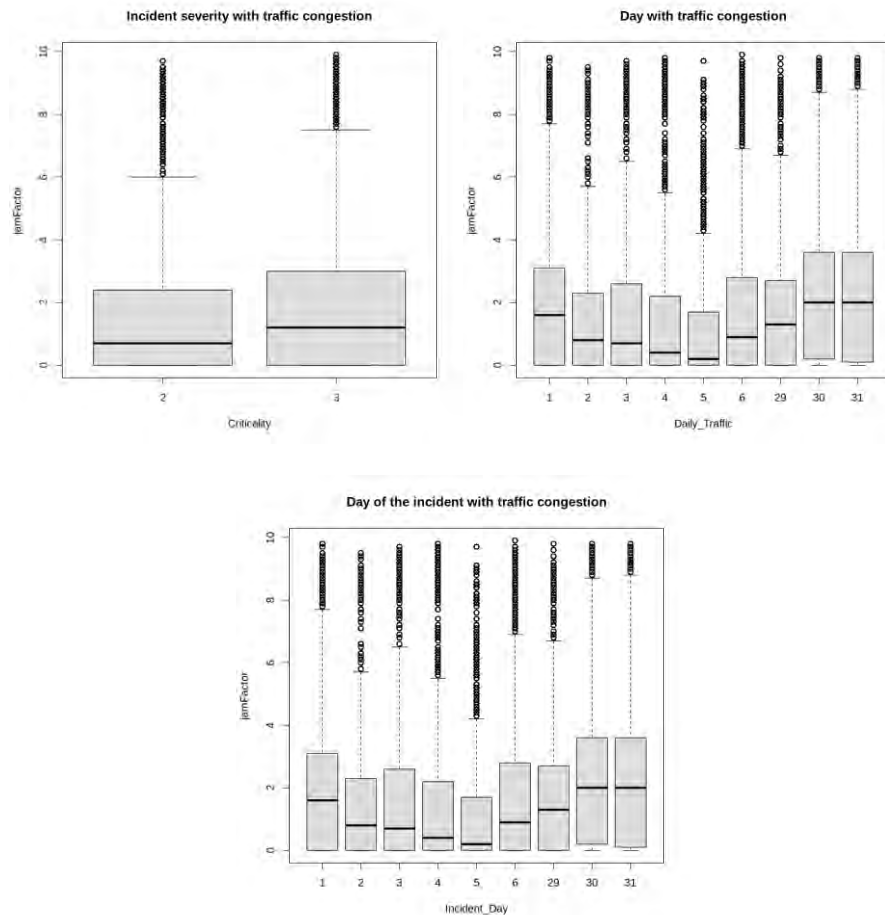


Figura 3.31. Visualización gráfica de la relación entre variables del conjunto de datos de Tráfico, Incidencias y Servicios con la congestión vehicular

Las variables de mayor impacto en el desempeño de la movilidad vehicular con el cálculo de las medias, se muestran en la Tabla 3.19.

Tabla 3.19. Valor representativo de medias de cada variable del conjunto de datos de tráfico, incidencias y servicios

No.	Variable	Media	No.	Variable	Media
1	Bbox_Traffic	32.319297	17	Criticality	2.802658
2	Free_Flow	10.815257	18	Type	220.314207
3	jamFactor	1.909658	19	Description	6.667524
4	Length	2301.60852	20	Olr	2.711761
5	Number_Segments	0.927815	21	Start_Time	608.070831
6	Speed	9.210867	22	End_Time	191.787018
7	Speed_Uncapped	9.421041	23	Incident_Day	11.173299
8	Daily_Traffic	11.171208	24	Incident_Dayinhabil	0.099745
9	Day_Off_Traffic	0.099745	25	Incident_Hour	14.300548
10	Traffic_Hour	13.45928	26	Incident_Month	10.700803
11	Traffic_Minute	27.226993	27	Incident_Minute	27.300381
12	Traffic_Day_Number	3.909158	28	Incident_Day_Number	3.909862
13	Work_Day_Traffic	0.699212	29	Work_Day_Incident	0.699082
14	Traffic_Block_Type	3.987062	30	incidencia_tipobloque	4.180406
15	Bbox_Incidents	32.319297	31	Category_Number	12.255365
16	Road_Closed	1	32	Bbox_Servicio	32.319297

- Identificación de variables que impactan en la movilidad vehicular mediante la correlación de Spearman. Para identificar las variables con mayor impacto en la movilidad vehicular, se empleó el coeficiente de correlación de Spearman. Esta técnica fue seleccionada debido a que permite medir la fuerza y dirección de relaciones monótonas entre variables, sin requerir que los datos sigan una distribución normal ni que exista una relación estrictamente lineal entre ellos. Así mismo la elección de Spearman se justifica por la naturaleza de los datos de tráfico utilizados en el estudio. Durante el análisis exploratorio inicial, se observó que muchas de las variables relevantes como velocidad, densidad vehicular, tiempo de viaje, congestión o incidentes no presentan un comportamiento normal, sino que tienden a estar sesgadas, contener valores atípicos o mostrar distribuciones multimodales, especialmente en horarios pico o en zonas de alta variabilidad. Se estableció un umbral mínimo de 0.7 como criterio de correlación aceptable. Los resultados obtenidos se presentan en la Tabla 3.20.

Tabla 3.20. Correlaciones superiores o iguales a 0.7 entre las variables independientes

Variable 1	Variable 2	Valor de correlación
Speed	Free_Flow	0.76523058
Speed_Uncapped	Free_Flow	0.75527344
Free_Flow	Speed	0.76523058
Speed_Uncapped	Speed	0.99645253
Free_Flow	Speed_Uncapped	0.75527344
Speed	Speed_Uncapped	0.99645253
Incident_Day	Daily_Traffic	0.99754481
Traffic_Block_Type	Traffic_Hour	0.98022436
Incident_Day_Number	Traffic_Day_Number	0.99940959
Work_Day_Incident	Work_Day_Traffic	0.99969206
Traffic_Hour	Traffic_Block_Type	0.98022436
Daily_Traffic	Incident_Day	0.99754481
Incidencia_Tipobloque	Incident_Hour	0.97056214
Traffic_Day_Number	Incident_Day_Number	0.99940959
Work_Day_Traffic	Work_Day_Incident	0.99969206

Una vez identificadas las variables con mayor correlación entre las variables independientes, se procedió a calcular la correlación de Spearman entre las variables de la Tabla 3.20 con la variable dependiente factor de atasco (jamFactor) y el resultado se muestra en la Tabla 3.21.

Tabla 3.21. Correlación de la variable jamFactor con el conjunto de variables de la Tabla 3.20.

Congestión de tráfico	Variables	Valor de correlación
Jamfactor	Number_Segments	0.2687642
Jamfactor	Speed_Uncapped	-0.7378836
Jamfactor	Work_Day_Traffic	0.1270429
Jamfactor	Description	-0.1250564
Jamfactor	Work_Day_Incident	0.1271877
Jamfactor	Free_Flow	-0.2272753
Jamfactor	Length	-0.1526781
Jamfactor	Speed	-0.7261867
Jamfactor	Traffic_Hour	0.1530089
Jamfactor	Traffic_Day_Number	-0.1548273
Jamfactor	Traffic_Block_Type	0.1534324
Jamfactor	Type	-0.1271393
Jamfactor	End_Time	0.1222923
Jamfactor	Incident_Day_Number	-0.1549883

- La identificación de variables que afectan la movilidad vehicular se realizó usando la matriz de rotación del Análisis de Componentes Principales (PCA). Se seleccionaron las variables con cargas superiores a 0.7, destacando las de mayor impacto (ver Tabla 3.22).

Tabla 3.22. Variables más significativas utilizando la matriz de rotación de Análisis Componentes

No.	Variable	Valor	No.	Variable	Valor	No.	Variable	Valor
1	Length	0.99	9	Olrr	0.91	17	Speed	0.75
2	Start_Time	0.98	10	Category_Number	0.90	18	Incident_Minute	0.73
3	Incident_Month	0.97	11	Description	0.89	19	Traffic_Minute	0.73
4	Criticality	0.96	12	Traffic_Block_Type	0.88	20	jamFactor	0.73
5	Number_Segments	0.95	13	Type	0.82	21	Bbox_Servicio	0.71
6	Road_Closed	0.94	14	Bbox_Traffic	0.81	22	Daily_Traffic	0.70
7	Incident_Hour	0.94	15	End_Time	0.81	23	Incident_Day	0.70
8	Traffic_Hour	0.93	16	incidencia_tipobloque	0.80			

- Identificación de variables que impactan en la movilidad vehicular mediante la matriz de cargas factoriales del Modelo de Ecuaciones Estructurales. Una vez establecidas las ecuaciones de tráfico, incidencias y servicios con sus correlaciones correspondientes, se obtuvieron los valores que se muestran en la Tabla 3.23.

Tabla 3.23. Matriz de cargas factoriales del Modelo de Ecuaciones Estructurales

Variable latente	Operador	Variable	Valor
Tráfico	=~	Free_Flow	0.5002061
Tráfico	=~	Jamfactor	0
Tráfico	=~	Length	1.00004338
Tráfico	=~	Speed	0.34515264
Tráfico	=~	Speed_Uncapped	0.33619433
Tráfico	=~	Traffic_Hour	0.03514468
Tráfico	=~	Traffic_Minute	0.01307886
Tráfico	=~	Traffic_Day_Number	0.00194125
Tráfico	=~	Traffic_Block_Type	0.03676568
Incidencias	=~	Bbox_Incidents	0.06720955
Incidencias	=~	Type	-0.23748091
Incidencias	=~	Start_Time	1.00296711
Incidencias	=~	End_Time	0.56983903
Incidencias	=~	Incident_Day	0.01422786
Incidencias	=~	Incident_Hour	-0.01648256
Incidencias	=~	Incident_Minute	-0.01805015
Incidencias	=~	Incident_Day_Number	-0.00035456
Incidencias	=~	Incidencia_Tipobloque	-0.02422802
Incidencias	=~	Category_Number	-0.03245275
Servicio	=~	Bbox_Servicio	1

- Identificación de variables que impactan en la movilidad vehicular mediante la prueba de Granger. Para identificar la causalidad entre dos variables mediante la prueba de Granger se establecen las siguientes hipótesis: Hipótesis nula (H0): No hay causalidad entre las dos variables. En este caso la serie temporal de una variable no puede predecir la otra variable. Hipótesis alternativa (H1): Existe causalidad entre las dos variables. La serie temporal de una variable puede predecir la otra variable. Si el valor p es menor que el nivel de significancia, en este caso de 0.05, entonces se rechaza la hipótesis nula y concluimos que hay evidencia suficiente para sugerir que la serie temporal de una variable influye en la otra variable. Si el valor p es mayor que el nivel de significancia, no se rechaza la hipótesis nula y no hay suficiente evidencia para afirmar que una

variable causa la otra. En la Tabla 3.24 se muestran las variables que su valor P es menor a 0.05.

Tabla 3.24. Variables con valores de p inferiores al nivel de significancia de 0.05

X	Y	Valor f	Valor p	Grados de libertad
Type	jamFactor	15.8806	0.0011	3
Speed_Uncapped	jamFactor	12.6906	0.0053	3
Speed	jamFactor	12.0518	0.0072	3
Number_Segments	jamFactor	70.3036	0.000	3
Free_Flow	jamFactor	35.8523	0.0024	3

Se muestra la causalidad entre la variable Number_Segments con jamFactor y se muestra cierta causalidad, como se visualiza en la Figura 3.32.

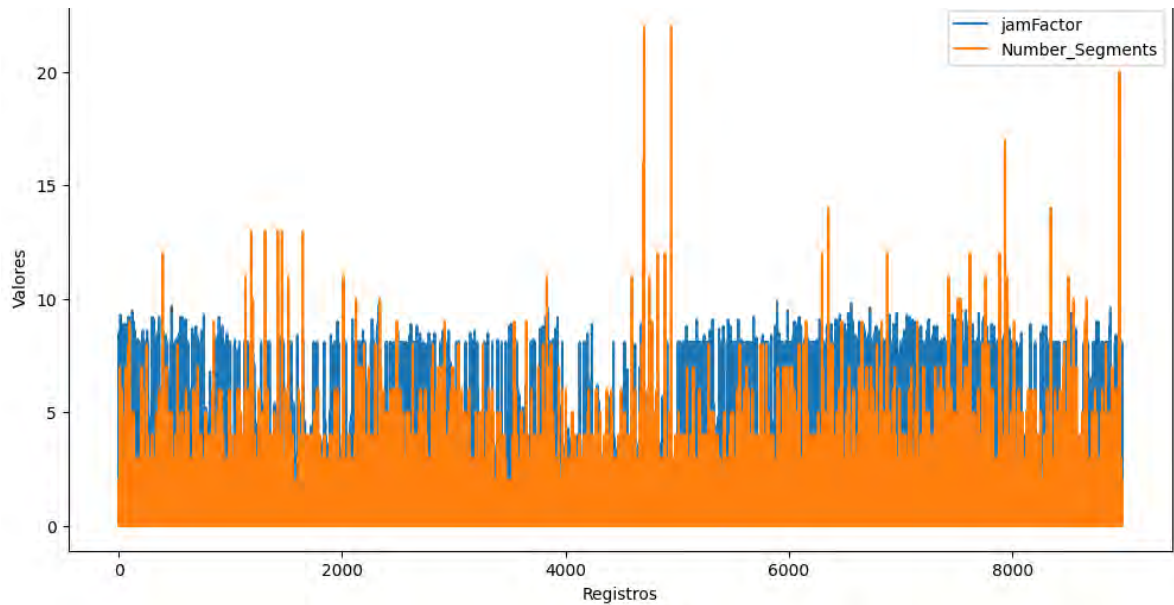


Figura 3.32. Causalidad entre la variable Number_Segments con respecto a jamFactor

A partir de la aplicación de diversas técnicas para la identificación de las variables relevantes, se obtuvo un conjunto de datos compuesto por 3,567,503 registros y 21 variables, proporcionando una base robusta para el análisis del sistema. La distribución y características de estas variables se presentan en la

Tabla 3.25, lo que permite una evaluación detallada de su impacto en la dinámica del sistema estudiado.

Tabla 3.25. Variables significativas de tráfico, incidencias y servicios

Id	Variable	Descripción
X0	Bbox_Traffic	Cuadro delimitador de donde origina el tráfico.
X1	Free_Flow	La velocidad de referencia (en metros por segundo) a lo largo de la carretera cuando no hay tráfico.
X2	Length	Longitud del segmento.
X3	Number_Segments	Número de segmentos que contiene la vialidad monitoreada.
X4	Speed	La velocidad esperada (en metros por segundo) a lo largo de la carretera; no excederá el límite de velocidad legal.
X5	Speed_Uncapped	La velocidad esperada (en metros por segundo) a lo largo de la carretera; puede exceder el límite de velocidad legal.
X6	Daily_Traffic	Día del monitoreo de tráfico.
X7	Traffic_Hour	Hora del monitoreo de tráfico.
X8	Traffic_Minute	Minuto del monitoreo de tráfico.
X9	Traffic_Day_Number	Número de día del monitoreo de tráfico (1= lunes, 2=martes...7=domingo).
X10	Traffic_Block_Type	Monitoreo de tráfico por bloques de horario (bloque 1=0-3 horas, bloque 2=4-7 horas, bloque 3=8-11 horas.....bloque 6: 20-23 horas.
X11	Type	Describe el tipo de incidente: accidente, construcción, avería de vehículo, clima entre otros.
X12	Start_Time	Hora de inicio de la incidencia
X13	End_Time	Hora de fin de la incidencia
X14	Incident_Day	Día del monitoreo de la incidencia
X15	Incident_Hour	Hora del monitoreo de la incidencia.
X16	Incident_Minute	Minuto del monitoreo de la incidencia.
X17	Incident_Day_Number	Número de día del monitoreo de incidencia (1= lunes, 2=martes...7=domingo).
X18	incidencia_tipobloque	Monitoreo de incidencia por bloques de horario (bloque 1=0-3 horas, bloque 2=4-7 horas, bloque 3=8-11 horas.....bloque 6: 20-23 horas.
X19	Category_Number	Categoría de la incidencia.
X20	Bbox_Servicio	Cuadro delimitador de donde se ubica el servicio.

El conjunto de datos fue mejorado mediante la aplicación de una transformación cúbica, seguida de un proceso de escalado que normaliza los valores en un rango entre 0 y 1. Estas operaciones no solo mejoraron la distribución de los datos, sino que también permitieron la aplicación de técnicas de selección de características, permitiendo una identificación precisa de las variables con mayor impacto dentro del conjunto de datos.

A partir de las actividades desarrolladas en el marco del método de preprocesamiento en tiempo real de datos de movilidad vehicular, se ha definido un conjunto de actividades sugeridas (ver Tabla 3.26) que permiten la depuración, transformación y estandarización de la información del sistema complejo antes de su análisis. Este proceso permite que los datos recopilados en entornos dinámicos sean coherentes, precisos y utilizables, eliminando redundancias, corrigiendo errores y gestionando valores atípicos o incompletos. La implementación de este método es importante para asegurar que la información utilizada en modelos sea representativa y confiable, optimizando su integración en sistemas de análisis en tiempo real.

Para la implementación eficiente de este método, es esencial la participación de roles especializados. Cada uno de estos roles contribuye a garantizar que el preprocesamiento sea robusto, escalable y permita gestionar grandes volúmenes de datos dinámicos. Como salidas del proceso, se obtiene un conjunto de datos limpios, normalizados y estandarizados en tiempo real, asegurando que la información procesada esté lista para su análisis inmediato.

Tabla 3.26. Cuadro descriptivo del método preprocesamiento en tiempo real

Tratamiento de los datos en tiempo real	
Descripción	Realizar el preprocesamiento de datos en tiempo real, asegurando la calidad, consistencia y relevancia de los datos para el análisis y la toma de decisiones inmediatas.
Actividades	- Identificar los formatos y estructuras de datos a procesar en tiempo real. - Aplicar técnicas de limpieza de datos, como eliminación de valores atípicos y corrección de errores. - Normalizar y estandarizar los datos utilizando métodos como z-score o escalado a un rango fijo. - Transformar los datos para ajustarlos a distribuciones normales si es necesario. - Implementar procesos automatizados de preprocesamiento para manejar datos en flujo continuo. - Evaluar la calidad de los datos en tiempo real y registrar métricas de validación.
Roles	- Ingenieros de datos. - Analistas de calidad de datos. - Desarrolladores de sistemas en tiempo real. - Especialistas en limpieza y transformación de datos. - Administradores de sistemas de monitoreo.
Salidas	- Datos limpios, normalizados y estandarizados en tiempo real. - Modelos automatizados de preprocesamiento implementados. - Informes de calidad y consistencia de los datos procesados. - Registro de errores y correcciones aplicadas. - Documentación de los procesos y herramientas de preprocesamiento.

3.6.2 Actualización y muestreo dinámico de los datos de movilidad vehicular

En el proceso de actualización y muestreo dinámico del sistema complejo de movilidad vehicular se aplicaron técnicas propias de sistemas complejos, enfocadas en representar adecuadamente la evolución temporal y espacial de un sistema multivariable sujeto a cambios constantes. Desde la ciencia de datos, se utilizó un enfoque adaptativo de muestreo aleatorio simple, guiado por el cálculo estadístico del tamaño de muestra óptimo (ecuación de proporciones) y criterios de significancia, con la finalidad de reducir la carga computacional manteniendo representatividad y robustez inferencial. En minería de datos, se emplearon mecanismos de selección estratificada y reducción de dimensionalidad que preservan la variabilidad estructural y los patrones de interés, asegurando que los subconjuntos reflejen adecuadamente el comportamiento del sistema original. En la construcción de indicadores, se evaluaron métricas de densidad, congestión e impacto de servicios, para preservar atributos clave en la muestra seleccionada. Además, se incorporaron conceptos de estadística para ajustar la probabilidad de selección de muestras ante nueva evidencia, técnicas de series de tiempo para evaluar la estabilidad temporal de los registros, y fundamentos de teoría de redes para garantizar la conservación de relaciones críticas entre entidades del sistema (tráfico, incidencia y servicio).

Para llevar a cabo el muestreo de los datos del sistema complejo de movilidad vehicular, se implementó un proceso de muestreo. A continuación, se detallan los pasos seguidos en este proceso.

Paso 1, Identificación de la población de datos: Se seleccionó un conjunto de datos de tráfico, incidencias y servicios con 3, 567, 503 registros y 21 variables.

Paso 2, Definición del tamaño de la muestra: Para seleccionar el tamaño de la muestra se utilizó la ecuación 3.7, con la calculadora que lo implementa de <https://goo.su/Fzzxlo>

se obtuvo un resultado de tamaño de muestra de 1, 849, 742 con 21 variables, así como se muestra en la Figura 3.33:

Nivel de confianza: ☒ 95% ☐ 99%

Margen de Error:

Población:

Tamaño de Muestra:

Figura 3.33. *Tamaño de la muestra de Tráfico, incidencias y servicios con 21 variables*

Se obtuvo un conjunto de datos integrado de tráfico, incidencias y servicios, compuesto por 1,849,742 registros y 21 variables. Este conjunto de datos fue sometido a un proceso de transformación y estandarización, permitiendo optimizar su estructura y mejorar su interpretabilidad. Además, mediante técnicas de selección de características, se identificaron y preservaron las variables de mayor relevancia, garantizando su utilidad para el análisis y modelado del sistema.

Se realizó un proceso de muestreo sobre los datos de movilidad vehicular con el objetivo de obtener una muestra representativa del conjunto de datos original. Esta selección garantiza la preservación de las características esenciales del sistema, permitiendo un análisis eficiente y estadísticamente válido para el experimento.

A partir de las actividades desarrolladas en el marco del método de actualización y muestreo dinámico de los datos de movilidad vehicular, se ha definido un conjunto de procedimientos sugeridos (ver Tabla 3.27) que permiten la incorporación continua de datos recientes y la selección eficiente de subconjuntos representativos para su análisis posterior. Este enfoque garantiza que el sistema conserve una base de datos actualizada, relevante y no redundante, optimizando la detección de patrones de movilidad en tiempo real, al mismo tiempo que se evita el almacenamiento innecesario de información obsoleta o repetitiva. La definición de las ventanas de actualización se realiza en función del comportamiento dinámico del sistema y de la disponibilidad de datos. En este estudio, se optó por ventanas móviles de actualización cada 60 minutos, lo que permite capturar cambios significativos en las condiciones del tráfico a lo largo del día sin incurrir en una sobrecarga computacional. Este intervalo fue determinado a partir de un análisis exploratorio de la variabilidad horaria del tráfico y permite un equilibrio entre granularidad temporal y eficiencia operativa. Cada nueva ventana activa un proceso de muestreo adaptativo, en el cual se priorizan segmentos viales con alta variabilidad en indicadores clave (como velocidad, densidad o presencia de incidentes), mientras que aquellos segmentos con comportamiento estable pueden ser muestreados con menor frecuencia. Este mecanismo asegura una cobertura representativa del sistema sin necesidad de almacenar o procesar la totalidad de los datos continuamente, lo que resulta especialmente útil para aplicaciones en tiempo real o entornos con restricciones de almacenamiento.

El método se basa en la implementación de técnicas de muestreo dinámico y actualización continua para seleccionar datos relevantes dentro del sistema complejo de movilidad. Para ello, se desarrollan diversas actividades importantes, tales como la definición de los objetivos específicos del muestreo, la identificación de preguntas relevantes de investigación, y la delimitación de la población objetivo. Posteriormente, se selecciona el marco muestral más adecuado, eligiendo entre enfoques como muestreo aleatorio, estratificado, sistemático o por conglomerados, dependiendo de las características de los datos y del sistema analizado. Además, se diseña un proceso de actualización continua del muestreo, garantizando que la representatividad de los datos se mantenga a medida que el sistema evoluciona. Para ello, se calcula el tamaño de muestra óptimo, considerando niveles de confianza y márgenes de error, permitiendo que los resultados sean estadísticamente robustos. Para la correcta implementación de este método, se requiere la participación de roles especializados. La colaboración de estos roles permite estructurar un proceso de muestreo flexible y adaptativo, asegurando la precisión y pertinencia de los datos recopilados en entornos de movilidad complejos.

Como salidas del proceso, se obtiene una lista de objetivos de muestreo dinámico bien definidos, un marco muestral seleccionado, y un plan detallado con los criterios de inclusión y exclusión. Además, se generan datos representativos y actualizados, acompañados de instrumentos para la recolección continua de datos y un informe de análisis de muestreo, en el cual se extrapolan los resultados al sistema complejo.

Tabla 3.27. Cuadro descriptivo del método actualización y muestreo dinámico

Muestreo de datos	
Descripción	Implementar técnicas de muestreo dinámico y actualización continua para seleccionar datos representativos en sistemas complejos, ajustándose a los cambios en el tiempo y garantizando la relevancia de los datos.
Actividades	- Establecer los objetivos específicos del muestreo dinámico. - Identificar preguntas determinantes de investigación y definir la población objetivo. - Seleccionar el marco muestral más adecuado (aleatorio, estratificado, sistemático, por conglomerados, etc.). - Diseñar un proceso de actualización continua del muestreo para ajustar la representatividad. - Calcular el tamaño de muestra necesario dependiendo de los niveles de confianza y el margen de error. - Implementar herramientas para la recolección dinámica de datos. - Realizar análisis estadísticos descriptivos y extrapolar los resultados al sistema completo. - Evaluar y ajustar el diseño del muestreo conforme cambian las condiciones del sistema.
Roles	- Investigadores en sistemas dinámicos. - Analistas de datos. - Estadísticos. - Diseñadores de metodologías de muestreo. - Administradores de bases de datos dinámicas.
Salidas	- Lista de objetivos de muestreo dinámico. - Marco muestral seleccionado. - Plan detallado de muestreo y criterios de inclusión/exclusión. - Datos obtenidos con representatividad y actualización constante - Instrumentos para la recolección continua de datos. - Informe de análisis de muestreo y resultados extrapolados. - Recomendaciones para ajustes futuros en el muestreo.

3.7 Modelado iterativo y desarrollo continuo

Esta fase está centrada en la construcción, ajuste y mejora progresiva de los modelos que representan el comportamiento de sistemas complejos altamente dinámicos. Esta etapa adopta un enfoque iterativo, permitiendo la incorporación de nuevos datos y aprendizajes obtenidos durante el proceso, con el objetivo de optimizar la precisión, eficiencia y capacidad de caracterización de los modelos.

En esta fase, se diseñan y prueban diferentes enfoques de modelado, evaluando su desempeño mediante métricas como el Error Cuadrático Medio (MSE), la Raíz del Error Cuadrático Medio (RMSE), el Error Medio Absoluto (MAE) y el coeficiente de determinación r^2 , entre otras. Los modelos generados son sometidos a pruebas utilizando conjuntos de datos no vistos previamente, lo que permite identificar patrones y comportamientos emergentes. A partir de los resultados obtenidos, los modelos se refinan mediante un proceso iterativo, incorporando los cambios en las condiciones del sistema y cumpliendo con los requerimientos específicos del análisis. Este enfoque iterativo garantiza una mejora continua en la precisión y robustez de los modelos desarrollados.

La Fase 5 comprende tres métodos principales: modelado dinámico, optimización en tiempo real y evaluación y ajuste continuo del modelo, como se muestra en la Figura 3.34. Esta fase está orientada a la construcción, mejora y adaptación constante de los modelos, en respuesta a la naturaleza dinámica y evolutiva de los sistemas complejos.



Figura 3.34. Fase 5 Modelado iterativo y desarrollo continuo de la metodología propuesta

Para comprender las acciones necesarias en la Fase 5 Modelado iterativo y desarrollo continuo, es necesario delinear el procedimiento aplicado en el caso de estudio sobre movilidad vehicular. Este análisis presenta una estructura clara y ordenada de la ejecución de cada actividad dentro de los métodos empleados, permitiendo una visión clara de su implementación en el sistema complejo de movilidad vehicular.

3.7.1 Desarrollo del modelo dinámico de movilidad vehicular

En el desarrollo del modelo dinámico de movilidad vehicular mediante programación genética se emplearon diversas técnicas avanzadas de sistemas complejos, inteligencia artificial y modelado computacional. Desde la perspectiva de sistemas complejos, se aplicó la modelación basada en agentes y la simulación de comportamiento emergente para capturar la naturaleza dinámica, no lineal y adaptativa del tráfico urbano. En ciencia de datos, se utilizaron estrategias de validación cruzada, partición de conjuntos de entrenamiento/prueba y análisis de sensibilidad para asegurar la robustez del modelo. En la minería de datos, se recurrió a selección de características, reducción de dimensionalidad y detección de patrones en grandes volúmenes de datos multivariantes. El cómputo evolutivo, mediante programación genética, permitió explorar automáticamente estructuras funcionales complejas optimizando ecuaciones matemáticas simbólicas sin imponer estructuras previas, facilitando la generación de modelos interpretables.

Sin embargo, dependiendo del contexto del sistema complejo que se pretende modelar, pueden emplearse diversas técnicas de aproximación, como métodos numéricos (interpolación, extrapolación, diferencias finitas e integración numérica), que permiten estimar la evolución del sistema en el tiempo. También pueden utilizarse enfoques basados en regresión simbólica, que derivan expresiones matemáticas que describen las relaciones entre variables, así como optimización metaheurística, a través de algoritmos genéticos, búsqueda de enjambre de partículas (PSO) y otros algoritmos evolutivos, para encontrar estructuras óptimas de modelos. Asimismo, la neuro evolución, que combina programación genética con redes neuronales, permite mejorar la capacidad predictiva del modelo, mientras que los sistemas de inferencia difusa y lógica borrosa facilitan la representación de incertidumbres y comportamientos no lineales. Otras herramientas, como las transformaciones matemáticas (wavelets y análisis de Fourier), permiten procesar señales y detectar patrones ocultos en los datos, y los modelos de sistemas dinámicos, basados en ecuaciones diferenciales ordinarias (EDOs) y ecuaciones diferenciales parciales (EDPs).

El objetivo de este método es construir un modelo que capture las complejas interacciones y dinámicas de la movilidad vehicular. Para ello, se define un conjunto inicial de posibles soluciones (programas) que se evalúan atendiendo su capacidad para caracterizar correctamente el comportamiento de la movilidad vehicular basado en datos históricos y en tiempo real. A través de iteraciones sucesivas, los mejores programas se seleccionan y combinan para producir nuevas generaciones de soluciones mejoradas. Este proceso continúa hasta que se encuentra un modelo que ofrece un comportamiento de ajuste preciso y confiable de la movilidad vehicular.

Para llevar a cabo el desarrollo del modelo de movilidad vehicular del sistema complejo de movilidad vehicular, se realizaron los siguientes pasos:

Paso 1, Establecer las dimensiones del conjunto de datos: Se seleccionó una muestra de 1, 479, 794 registros con 21 variables.

Paso 2, Descripción de las variables utilizadas del conjunto de datos: Las variables utilizadas para la generación de modelos se enlistan a continuación, **Bbox_Traffic**, **Free_Flow**, **Length**, **Number_Segments**, **Speed**, **Speed_Uncapped**, **Daily_Traffic**, **Traffic_Hour**, **Traffic_Minute**, **Traffic_Day_Number**, **Traffic_Block_Type**, **Type**, **Start_Time**, **End_Time**, **Incident_Day**, **Incident_Hour**, **Incident_Minute**, **Incident_Day_Number**, **incidencia_tipobloque**, **Category_Number**, **Bbox_Servicio**, son las variables de mayor impacto de acuerdo al experimento de selección de características y se describen en la

Tabla **3.25**.

Paso 3, establecer los parámetros del evolutivo utilizado: Los parámetros del evolutivo fueron una población de 300, 500 generaciones, parámetros de validación cruzada, así como se muestran en la Figura 3.35, los parámetros se establecieron en base a varios experimentos realizados (sintonización).

```

=====
*****Terminales*****
['Bbox_Traffic', 'Free_Flow', 'Length', 'Number_Segments', 'Speed', 'Speed_Uncapped', 'Daily_Traffic', 'Traffic_Hour', 'Traffic_Minute',
'Traffic_Day_Number', 'Traffic_Block_Type', 'Type', 'Start_Time', 'End_Time', 'Incident_Day', 'Incident_Hour', 'Incident_Minute',
'Incident_Day_Number', 'incidencia_tipobloque', 'Category_Number', 'Bbox_Servicio']
=====
*****Función de referencia*****
funcion_referencia: ['jamFactor']
=====
*****Funciones*****
function_set: ['add', 'sub', 'mul', 'div', 'sqrt', 'log', 'abs', 'neg', 'inv', 'sin', 'cos', 'tan']
=====
*****Métrica de selección de fitness*****
metric:'mse', 'mae', 'maxerror', 'r2'
=====
*****Parámetros del algoritmo evolutivo*****
population: 300
generations: 500
tournament_size: 20
stopping_criteria: 0.0
const_range: (-1, 1)
init_depth: (2, 6)
init_method: half and half
p_crossover: 0.9
p_subtree_mutation: 0.01
p_hoist_mutation: 0.01
p_point_mutation: 0.01
p_point_replace: 0.05
max_samples: 0.9
warm_start: False
parsimony_coefficient: 0.001
random_state: None
low_memory: False
n_jobs: 1
verbose: 0
=====
*****Parámetros de la validación cruzada*****
KFold(n_splits=5, random_state=42, shuffle=True)

```

Figura 3.35. Parámetros del programa evolutivo para su ejecución

Paso 4, Establecer el número de ejecuciones: Se ejecutó el programa evolutivo 30 veces.

Tras la ejecución del algoritmo evolutivo en 30 ejecuciones, se generaron 30 modelos óptimos del Índice de Movilidad Vehicular. Estos modelos representan las mejores soluciones obtenidas a lo largo del proceso y se presentan en la Tabla 3.28, permitiendo un análisis comparativo de su desempeño y características.

Tabla 3.28. Se muestran los 30 mejores modelos de índice de movilidad vehicular

[illegible]

En la Tabla 3.29 se muestran los valores obtenidos de las métricas: MSE (Error Cuadrático Medio), MAE (Error Medio Absoluto), MAXERROR (Error Máximo), r^2 (Coeficiente de determinación) utilizadas en las 30 ejecuciones del experimento y se muestra en color amarillo el mejor modelo de acuerdo a la métrica utilizada.

Tabla 3.29. Mejores modelos de acuerdo a la métrica utilizada

Número de Ecuación	MSE	Número de Ecuación	MAE	Número de Ecuación	MaxError	Número de Ecuación	R ²
9	0.5997479	9	0.539065	19	14.6357324	20	-0.626244
29	1.3896642	29	0.7511468	29	10.7146896	2	-0.608308
19	1.59813025	19	0.855809	1	10.28889	4	-0.608308
7	1.93369993	7	0.895497	10	10.28889	5	-0.608308
17	2.0888517	17	1.1630231	11	10.28889	8	-0.608308
28	2.6585328	22	1.1782227	12	10.28889	16	-0.608308
1	3.2298834	28	1.2122734	14	10.28889	21	-0.608308
10	3.2298834	1	1.2174888	15	10.28889	23	-0.608308
11	3.2298834	10	1.2174888	24	10.28889	26	-0.608308
12	3.2298834	11	1.2174888	25	10.28889	18	-0.1856590
14	3.2298834	12	1.2174888	2	9.699	3	0.0252462
15	3.2298834	14	1.2174888	4	9.699	27	0.0942317
24	3.2298834	15	1.2174888	5	9.699	22	0.3093476
25	3.2298834	24	1.2174888	8	9.699	13	0.3899251
30	3.3801857	25	1.2174888	16	9.699	6	0.3998289
6	3.46711876	6	1.2586625	18	9.699	30	0.4148773
13	3.5243322	30	1.303430	20	9.699	1	0.4408952
22	3.9898190	13	1.3222123	21	9.699	10	0.4408952
27	5.2325184	27	1.394260	23	9.699	11	0.4408952
3	5.63103981	18	1.590469	26	9.699	12	0.4408952
18	6.84941541	3	1.7658481	17	9.02273133	14	0.4408952
2	9.2910098	2	1.875264	22	8.999	15	0.4408952
4	9.2910098	4	1.875264	28	8.96666571	24	0.4408952
5	9.2910098	5	1.875264	27	8.899	25	0.4408952
8	9.2910098	8	1.875264	3	7.76794919	28	0.53979817
16	9.2910098	16	1.875264	13	7.71428573	17	0.6384120
21	9.2910098	21	1.875264	30	6.48033251	7	0.6652694
23	9.2910098	23	1.875264	6	6.31052629	19	0.7233577
26	9.2910098	26	1.875264	7	5.78890598	29	0.7594440
20	9.3946264	20	2.000866	9	4.36409917	9	0.89618142

Se realizó al selección del mejor modelo basado en el Error Cuadrático Medio, que corresponde a la ecuación 9, mostrado en la Tabla 3.29. La ecuación es la siguiente:

$$IM = 2 \cdot X_1 + 3 \cdot \sqrt{X_0} + \sqrt{\log(X_6)} - 5 \cdot \log(X_{11}) + \sin(\log(X_5)) + \sin(\cos(\sin(X_1 + \sqrt{X_2} - \log(X_{20}) + \sin(\log(X_5)) - 0.729))) - 1.458 \quad (3.13)$$

Donde:

IM es el índice de Movilidad.

X0 es el cuadro delimitador de donde se origina el tráfico.

X1 es la velocidad de referencia.

X2 es la longitud del segmento (calle).

X5 es la velocidad esperada.

X6 es el día del monitoreo de tráfico.

X11 describe el tipo de incidente: accidente, construcción, congestión, avería de vehículo, clima entre otros.

X20 es el cuadro delimitador de donde se ubica el servicio.

La interpretación del modelo encontrado del Índice de Movilidad (IM) es una medida compleja y multifacética que involucra diversas variables del entorno de tráfico, condiciones incidentales y servicios. Las variables de velocidad de referencia y esperada, el servicio, el tipo de incidente y el día del tráfico, todos contribuyen de diferentes maneras, lineales y no lineales, al valor final del IM. Esto sugiere que el

modelo trata de capturar la naturaleza dinámica y a veces caótica de la movilidad vehicular.

Representación gráfica de la caracterización de la movilidad vehicular utilizando el Índice de Movilidad $IM = 2 \cdot X_1 + 3 \cdot \sqrt{X_0} + \sqrt{\log(X_6)} - 5 \cdot \log(X_{11}) + \sin(\log(X_5)) + \sin(\cos(\sin(X_1 + \sqrt{X_2} - \log(X_{20}) + \sin(\log(X_5)) - 0.729))) - 1.458$ con respecto al valor de referencia (jamFactor). El IM en color azul se ajusta al comportamiento de jamFactor en color naranja, como se puede observar en la Figura 3.36.

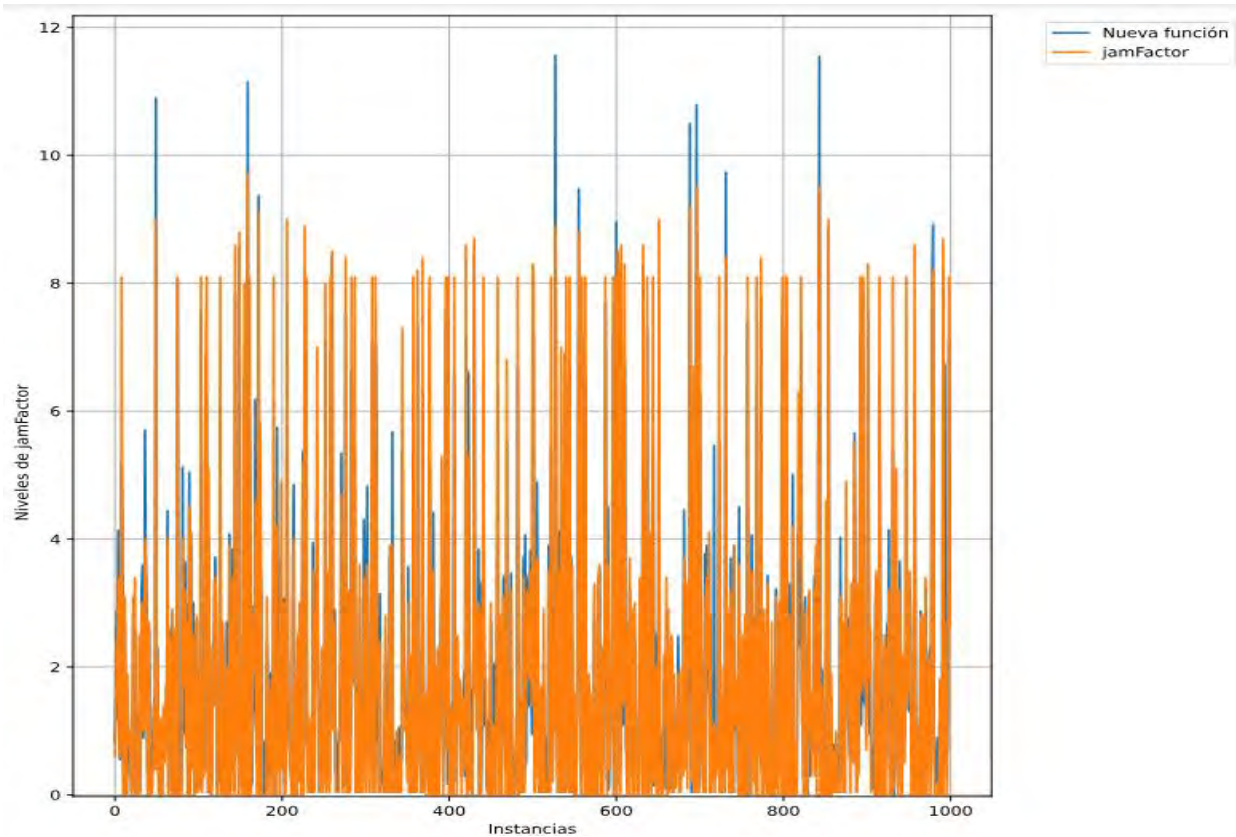


Figura 3.36. Comportamiento de la movilidad vehicular utilizando el IM de una calle

El modelo para calcular el Índice de Movilidad de la Red Vial (IMRV) se define como el promedio de los Índices de Movilidad (IM) individuales de cada calle o segmento dentro de la red vial. La fórmula es la siguiente:

$$IMRV = \frac{1}{n} \sum_{i=1}^n IM_i \quad (3.14)$$

Donde:

IMRV (Índice de Movilidad de la Red Vial) es el valor promedio que representa la movilidad general de toda la red vial considerada (cuadro delimitador). Un valor más alto de IMRV indica una mala movilidad en la red vial, mientras que un valor más bajo una movilidad vehicular buena.

IM (Índice de Movilidad de la calle o segmento): Cada IM individual es una medida de la movilidad para una calle o segmento específico. Los valores de IM reflejan las condiciones locales del tráfico, incluyendo

velocidad de referencia, tipo de incidentes, velocidad esperada, y otras variables que afectan la movilidad en esa área particular.

n (Cantidad de calles o segmentos): Este es el número total de calles o segmentos considerados en la red vial.

Para calcular el IMRV, se suma el índice de movilidad (IM) de cada una de las n calles o segmentos y luego se divide esta suma por el número total de calles o segmentos. Este promedio proporciona una visión general de cómo está la movilidad vehicular de toda la red vial, integrando las variaciones y particularidades de cada segmento individual.

El significado práctico del IMRV es una herramienta útil para autoridades de tráfico y urbanistas, ya que permite evaluar la eficiencia y el rendimiento global de la red vial. Al identificar las áreas con IM altos, se pueden dirigir los esfuerzos de mejora hacia esos segmentos específicos para optimizar la movilidad general. Además, al monitorear el IMRV a lo largo del tiempo, se pueden detectar tendencias y evaluar el impacto de políticas o intervenciones específicas en la red vial.

Para determinar el índice de movilidad de la red vial utilizando $\frac{1}{n} \sum_{i=1}^n IM_i$ con el $IM = 2 \cdot X_1 + 3 \cdot \sqrt{X_0} + \sqrt{\log(X_6)} - 5 \cdot \log(X_{11}) + \sin(\log(X_5)) + \sin(\cos(\sin(X_1 + \sqrt{X_2} - \log(X_{20}) + \sin(\log(X_5)) - 0.729))) - 1.458$ un cuadro delimitador de la ciudad de México que corresponde las calles de Cumbres de Maltrata, Av División del Norte/Av Cuauhtémoc, Av Eugenia/Eje 5 S, Eje Central, Avenida Doctor Jose Maria Vertiz, Eje 1 Poniente / Avenida Universidad, como se muestra en la Figura 3.37.

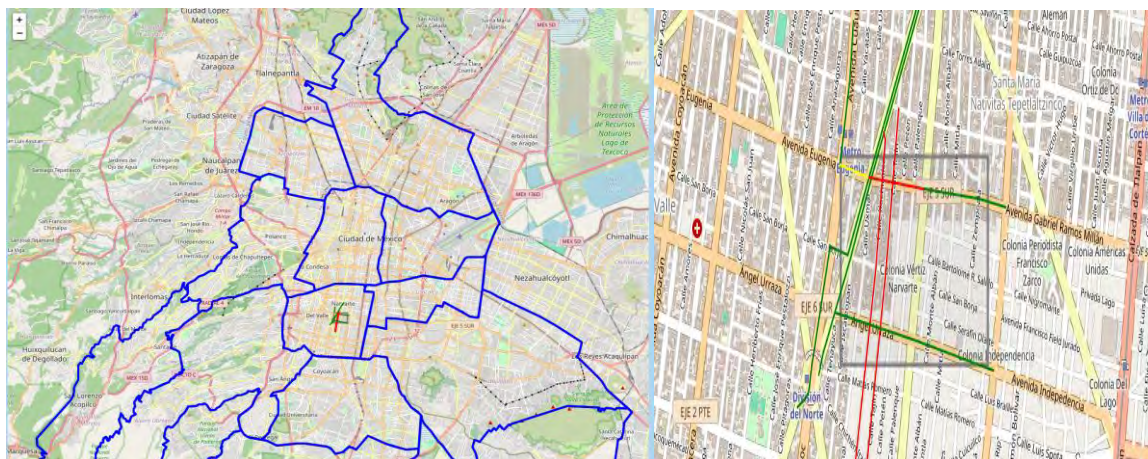


Figura 3.37. Un cuadro delimitador de la delegación Benito Juárez de la Ciudad de México

El cuadro delimitador proporciona un conjunto de calles, tal como se mostró en la Figura 3.37. Las características individuales de cada calle, junto con su respectivo Índice de Movilidad (IM) se muestran en la Tabla 3.30. Posteriormente, al considerar todas las calles que componen el cuadro delimitador, se calcula el Índice de Movilidad de la Red Vial (IMRV).

Tabla 3.30. Conjunto de calles del cuadro delimitador marcado en una zona geográfica de la alcaldía Benito Juárez

Vía	Bbox_traffic	Free_flow	Length	Speed_uncapped	Daily_traffic	Type	Bbox_servicio	IM
Cumbres de maltrata	49	11.1111	172	6.6666	4	108	49	33.8690
Av división del norte/av cuauhtémoc	49	9.1666	567	5.0	4	101	49	30.1028
Av eugenia/eje 5 s s	49	9.4444	721	4.1666	4	101	49	30.6048
Eje central	49	16.1111	836	17.7777	4	101	49	44.2601
Avenida doctor jose maria vertiz	49	10.0	470	3.3333	4	101	49	31.1272
Eje 1 poniente / avenida universidad	49	7.2222	842	6.1111	4	101	49	26.289
IMRV								32.72

Un Índice de Movilidad de la Red Vial (IMRV) de 32.72 indica una evaluación general del estado de movilidad dentro del área delimitada por el cuadro. Esto significa que, en general, el tráfico puede estar fluyendo a un ritmo razonable, pero puede haber ciertos puntos de congestión o áreas donde el tráfico se ve afectado.

A partir de las actividades desarrolladas en el marco del método de desarrollo del modelo dinámico de movilidad vehicular, se ha definido un conjunto de actividades sugeridas (ver Tabla 3.31) que permiten la construcción de una representación matemática y computacional del sistema, integrando sus principales componentes y dinámicas subyacentes. Este enfoque posibilita la formalización de las relaciones entre variables, facilitando el análisis de distintos escenarios del sistema complejo y permite la exploración de estrategias para mejorar su desempeño. El modelado dinámico fortalece la capacidad de ajuste del modelo, asegurando una aproximación precisa al comportamiento emergente del sistema complejo, adaptándose a nuevas condiciones y garantizando su aplicabilidad en entornos cambiantes.

El desarrollo del modelo dinámico se enfoca en la generación de representaciones matemáticas de la movilidad, combinando técnicas avanzadas de optimización en tiempo real. Para ello, se implementan diversas actividades, en la programación genética para mejorar la precisión de las caracterizaciones de comportamiento. Además, se realizan validaciones mediante la comparación de los resultados del modelo con datos reales para garantizar su fidelidad y capacidad de caracterización. A medida que se recopilan nuevos datos, los modelos dinámicos son ajustados y refinados, optimizando su desempeño y permitiendo una mejora continua.

Para la correcta ejecución de este método, es indispensable la participación de roles especializados. La colaboración interdisciplinaria de estos roles garantiza la implementación de un modelo preciso, eficiente y adaptable a distintas condiciones

de movilidad vehicular. Como salidas del proceso, se obtiene un modelo dinámico validado y ajustado en tiempo real, se generan informes detallados de simulaciones y validaciones, documentando la precisión del modelo y su capacidad para representar escenarios complejos. Asimismo, se registran métricas de mejora del desempeño del sistema, permitiendo evaluar su efectividad en la predicción y gestión de la movilidad vehicular.

Tabla 3.31. Cuadro descriptivo del método modelado dinámico

Desarrollo del modelo	
Descripción	Desarrollar modelos dinámicos que representen el comportamiento del sistema y aplicar técnicas de optimización en tiempo real para mejorar el desempeño y la toma de decisiones en sistemas complejos.
Actividades	- Identificar las variables y parámetros vitales para el modelado dinámico. - Diseñar modelos matemáticos que representen las dinámicas del sistema. - Implementar algoritmos de optimización en tiempo real. - Validar los modelos mediante simulaciones y comparación con datos reales. - Ajustar los modelos dinámicos conforme se recopilan nuevos datos. - Evaluar el desempeño del sistema optimizado y registrar métricas de mejora.
Roles	- Modeladores de sistemas dinámicos. - Ingenieros en optimización. - Analistas de datos. - Matemáticos y estadísticos. - Desarrolladores de software en tiempo real.
Salidas	- Modelos dinámicos validados y ajustados en tiempo real. - Algoritmos de optimización implementados. - Informes de simulaciones y validaciones de modelos. - Métricas de mejora del desempeño del sistema. - Documentación técnica de los modelos y procesos de optimización.

3.7.2 Evaluación y ajuste continuo del modelo de movilidad vehicular

En esta fase se lleva a cabo la evaluación y ajuste del modelo del índice de movilidad obtenido durante el modelado dinámico. Durante la evaluación y ajuste continuo del modelo de movilidad vehicular, se integraron técnicas avanzadas provenientes de sistemas complejos, ciencia de datos y optimización evolutiva. Desde la perspectiva de los sistemas complejos, se aplicó análisis de sensibilidad multivariable y mapas cognitivos difusos para modelar las interacciones no lineales y las retroalimentaciones entre variables del sistema. En ciencia de datos, se utilizaron redes neuronales artificiales (ANN) para la validación cruzada del modelo, junto con técnicas de aprendizaje continuo (online learning) para ajustar el modelo en función de nuevos datos. En la construcción de indicadores, se aplicaron técnicas de regresión simbólica evolutiva para optimizar fórmulas del índice de movilidad, mientras que en minería de datos se recurrió a técnicas de selección de características, feature ranking y validación con datos históricos y reales. Desde la física y las matemáticas aplicadas, se integraron conceptos de teoría del control adaptativo, análisis de estabilidad dinámica, y modelado estocástico para capturar la variabilidad inherente al tráfico urbano. Además, el uso de mapas de calor y simulaciones en espacio de fase permitió visualizar la evolución del sistema bajo distintos escenarios.

Para realizar la evaluación y el ajuste continuo del modelo de movilidad vehicular, se implementó un proceso estructurado que comprende los siguientes pasos detallados a continuación.

Paso 1, Evaluación del modelo: Para evaluar el modelo de $IM = 2X_1 + 3\sqrt{X_0} + \sqrt{\log(X_6)} - 5\log(X_{11}) + \sin(\log(X_5)) + \sin(\cos(\sin(X_1 + \sqrt{X_2} - \log(X_{20}) + \sin(\log(X_5)) - 0.729))) - 1.458$ se empleó un conjunto de datos compuesto por siete variables ($X_0, X_1, X_2, X_5, X_6, X_{11}$ y X_{20}) descritos en la

Tabla 3.25) y los 369,948 registros de prueba. Se optó por utilizar una red neuronal feedforward básica con tres capas densamente conectadas para validar el IM. La primera capa densa consta de 32 neuronas, la segunda capa densa de 16 neuronas y la tercera capa densa de una sola neurona, que es la capa de salida. El modelo se optimizó utilizando el algoritmo Adam, con un total de cien épocas de entrenamiento.

La evaluación arrojó un valor de Error Cuadrático Medio (MSE): 0.012700295166609412 y de Coeficiente de determinación (R^2): 0.9998179834436258. Un valor de Error Cuadrático Medio (MSE) de 0.0127 indica que, en promedio, las predicciones del modelo están desviadas del valor real por un pequeño margen cuadrático. En cuanto al coeficiente de determinación con valor de 0.9998 indica una adecuada proporción de la variabilidad en la variable dependiente que puede explicarse por las variables independientes del modelo (ver Tabla 3.32).

Tabla 3.32. Resultados del modelo evaluado mediante el Error Cuadrático Medio (MSE) y el Coeficiente de Determinación (R^2), que reflejan la precisión y capacidad explicativa del modelo

MÉTRICA	VALOR	INTERPRETACIÓN
Error Cuadrático Medio (MSE)	0.0127002952	Las predicciones se desvían del valor real por un margen cuadrático muy pequeño en promedio.
Coeficiente de Determinación (R^2)	0.9998179834	El modelo explica el 99.98% de la variabilidad en la variable dependiente.

El modelo presenta un ajuste adecuado a los valores reales de la movilidad vehicular, como se observa en la Figura 3.38.

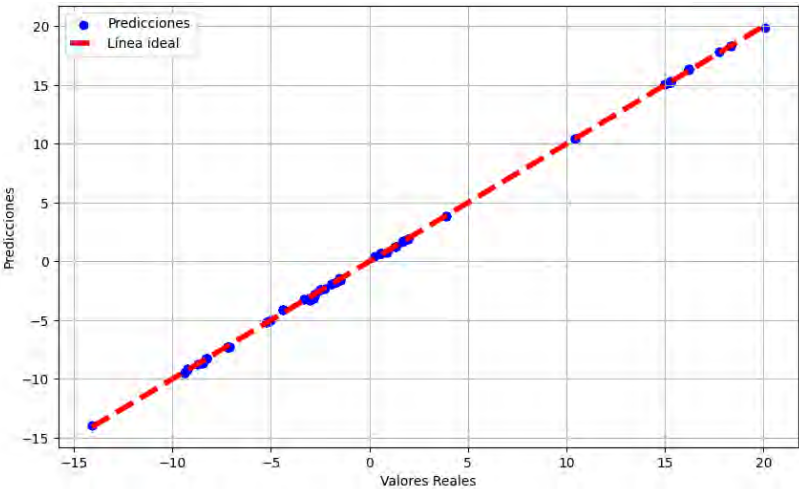


Figura 3.38. Validación del IM con una red neuronal feedforward

Paso 2, ajuste y pruebas al modelo: Se realizaron pruebas de sensibilidad para evaluar cómo los cambios en los valores de las variables críticas afectan los resultados del modelo. Se utilizó Mental Modeler (Gray et al., 2013) para crear un mapa cognitivo difuso que permite visualizar cómo cada variable impacta al resto de las variables relacionadas con la movilidad vehicular, el mapa con las variables del IM, se muestran en la Figura 3.39.

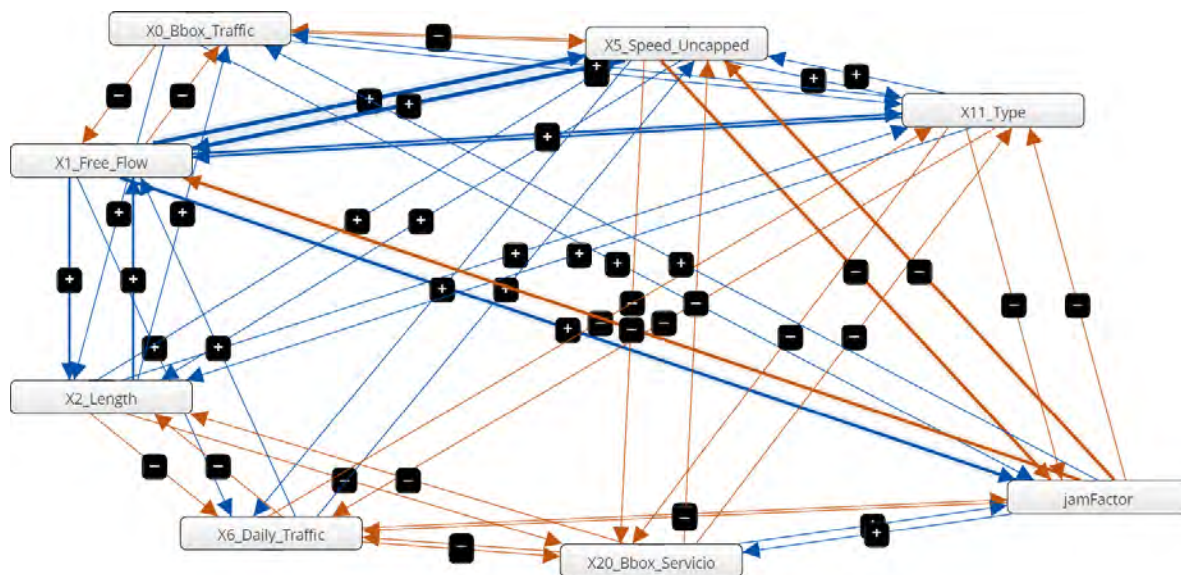


Figura 3.39. Mapa cognitivo difuso de las variables del IM

Para visualizar la sensibilidad de las variables del IM, se realizaron cinco escenarios que muestran cómo la relación de las variables del IM impacta en el comportamiento del jamFactor, los valores de cada escenario se muestran en la Tabla 3.33 .

Tabla 3.33. Valores de escenarios para probar la sensibilidad de las variables del IM

Escenario	Bbox_Traffic	Free_Flow	Length	Speed_Uncapped	Daily_Traffic	Type	Bbox_Servicio
1	0.00	0.76	0.82	0.00	0.26	0.10	0.00
2	0.10	0.74	0.12	0.35	0.83	0.67	0.33
3	0.77	0.01	0.88	0.44	0.99	0.61	0.22
4	0.89	0.56	0.09	0.40	0.95	0.17	0.17
5	0.64	0.55	0.04	0.16	0.92	0.04	0.50

Los diferentes valores que pueden tener las variables del IM impactan el comportamiento de la movilidad vehicular, como se muestra en la Figura 3.40.

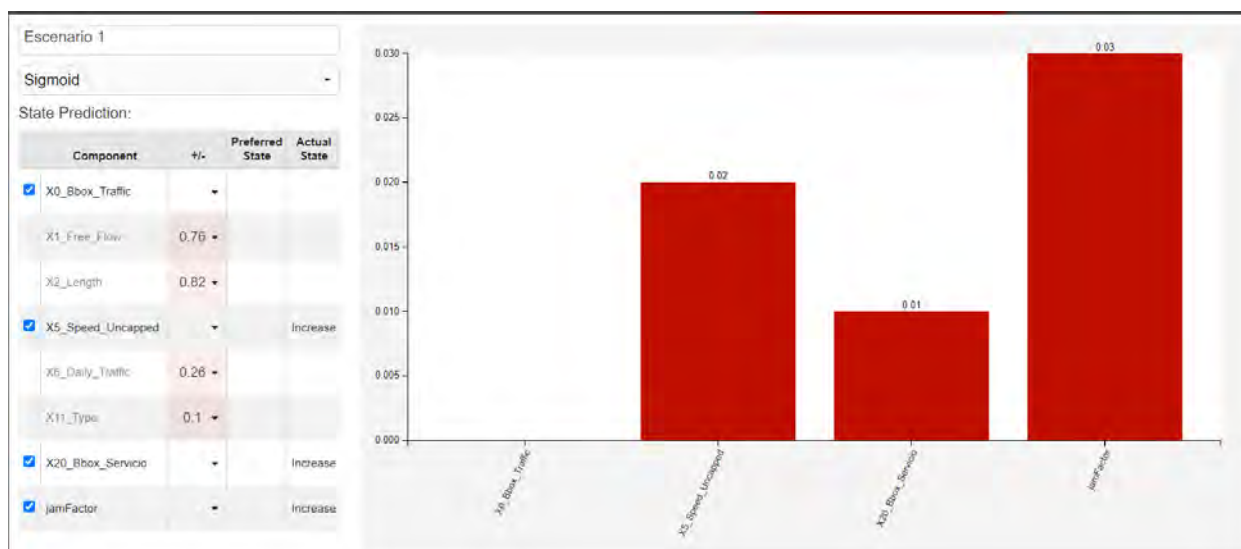


Figura 3.40. Sensibilidad de las variables del IM

A partir de las actividades desarrolladas en el marco del método de evaluación y ajuste continuo del modelo de movilidad vehicular, se ha definido un conjunto de actividades sugeridas (ver Tabla 3.34) que permiten la validación, calibración y mejora progresiva del modelo, asegurando su capacidad de adaptación a la dinámica del sistema complejo, en este caso, la movilidad vehicular. Este enfoque posibilita la identificación de discrepancias entre las caracterizaciones del modelo y los datos observados, lo que permite la implementación de ajustes paramétricos y estructurales que optimicen su precisión y robustez.

El proceso de evaluación y ajuste continuo se fundamenta en la monitorización constante del desempeño del modelo, garantizando que su representación del tráfico vehicular sea coherente con las condiciones reales del sistema. Para ello, se llevan a cabo diversas actividades, como el seguimiento del modelo en tiempo real, la comparación de los resultados con datos observados, y la identificación de errores o desviaciones en las predicciones del modelo. Una vez detectadas inconsistencias, se procede a ajustar los parámetros del modelo, incorporando nuevos datos para mejorar su capacidad predictiva. En este proceso, se pueden emplear técnicas de reentrenamiento, basadas en aprendizaje continuo, que permiten actualizar el modelo dinámicamente conforme evolucionan las condiciones del tráfico. Adicionalmente, se generan reportes regulares sobre la precisión y confiabilidad del modelo, lo que facilita su evaluación constante y la implementación de estrategias de optimización.

Para la correcta ejecución de este método, se requiere la participación de roles especializados. La colaboración interdisciplinaria de estos roles permite asegurar que el modelo sea preciso, flexible y ajustable a las variaciones del sistema de movilidad vehicular. Como salidas del proceso, se obtienen informes regulares sobre el desempeño del modelo, proporcionando un análisis detallado de su precisión y posibles desviaciones. Asimismo, se generan ajustes y optimización de parámetros, lo que permite mejorar su capacidad predictiva. También se desarrollan nuevas

versiones de modelos reentrenados y validados, adaptándose a los cambios en la movilidad. Además, se actualizan métricas de precisión y error, asegurando que el modelo refleje con fidelidad la realidad de la movilidad vehicular.

Tabla 3.34. Cuadro descriptivo del método evaluación y ajuste continuo del modelo

Asistencia al modelo	
Descripción	Evaluar continuamente los modelos implementados para identificar desviaciones, validar su precisión y realizar ajustes dinámicos con base en nuevos datos o cambios en las condiciones del sistema.
Actividades	- Monitorear el desempeño de los modelos en tiempo real. - Comparar los resultados de los modelos con datos observados. - Identificar errores o desviaciones en las predicciones del modelo. - Ajustar los parámetros de los modelos conforme se recopilan nuevos datos. - Reentrenar modelos utilizando técnicas de aprendizaje continuo. - Generar reportes regulares sobre la precisión y confiabilidad de los modelos.
Roles	- Analistas de modelos dinámicos. - Ingenieros de datos. - Matemáticos y estadísticos. - Especialistas en validación de modelos. - Desarrolladores de sistemas de monitoreo.
Salidas	- Informes regulares de desempeño del modelo. - Parámetros ajustados y optimizados en los modelos. - Nuevas versiones de modelos reentrenados y validados. - Métricas actualizadas de precisión y error del modelo. - Documentación de ajustes realizados y recomendaciones futuras.

3.8 Interpretación, representación e implementación continua

Esta fase se enfoca en la interpretación de los modelos de sistemas complejos, específicamente en el modelo de movilidad vehicular. Su correcta interpretación es importante para garantizar que el modelo sea comprensible y útil para los tomadores de decisiones. Resulta esencial asegurar que los hallazgos obtenidos del modelo sean analizados en el contexto del sistema complejo estudiado, lo que permitirá identificar patrones, tendencias y relaciones decisivos entre las variables involucradas, facilitando así una comprensión integral del comportamiento del sistema.

La representación de los resultados incluye la generación de visualizaciones, como gráficos, diagramas y mapas interactivos, que permitan transmitir los resultados con claridad y accesibilidad. Estas representaciones están diseñadas para ser dinámicas, de forma que puedan actualizarse en tiempo real a medida que se incorporan nuevos datos o cambian las condiciones del sistema.

En cuanto a la implementación continua, esta etapa asegura que los modelos y las representaciones obtenidas puedan incorporarse con efectividad en herramientas o sistemas operativos, promoviendo su uso en escenarios prácticos y adaptativos. Esto incluye el desarrollo de interfaces de usuario, sistemas de monitoreo en tiempo real y módulos de actualización automática que garanticen la relevancia y precisión del modelo ante cambios en el entorno. Esta fase no solo cierra el ciclo del proceso metodológico, sino que también lo retroalimenta, generando información valiosa que puede ser utilizada para refinar los modelos, ajustar parámetros o explorar nuevas dinámicas en el sistema complejo estudiado.

La Fase 6 incluye tres métodos principales: representación dinámica del modelo, implementación adaptativa y actualización continua del modelo, los cuales se ilustran en la Figura 3.41. Su propósito es interpretar, representar y mantener actualizados los modelos continuos, asegurando su relevancia y efectividad en el análisis y la gestión de sistemas complejos altamente dinámicos.



Figura 3.41. Fase 6 Interpretación, representación e implementación continua de la metodología propuesta

Para comprender las acciones necesarias en la Fase 6 Interpretación, representación e implementación continua, es determinante delinear el procedimiento aplicado en el caso de estudio sobre movilidad vehicular. Este análisis presenta una estructura detallada, la ejecución de cada actividad dentro de los métodos empleados, permitiendo una visión clara de su implementación en el sistema complejo de movilidad vehicular.

3.8.1 Representación dinámica del modelo de movilidad vehicular

En esta etapa de representación dinámica del modelo de movilidad vehicular, se integraron diversas técnicas avanzadas provenientes de sistemas complejos, ciencia de datos y teoría de control difuso. Desde la teoría de sistemas complejos, se aplicaron modelos de lógica difusa para manejar la incertidumbre inherente al comportamiento vehicular, permitiendo la categorización continua entre estados como movilidad baja, media y alta. En ciencia de datos, se empleó el procesamiento de grandes volúmenes de datos y la generación automatizada de intervalos difusos, utilizando funciones de pertenencia gaussianas que facilitaron la transformación de datos numéricos en clasificaciones interpretables. La construcción del indicador (IM) y su representación escalar emplearon además técnicas de normalización estadística, modelado simbólico y asignación dinámica de rangos. Desde la minería de datos, se utilizaron métodos de clustering difuso, ordenamiento y segmentación para construir una lógica de clasificación escalable.

Para llevar la representación escalar de movilidad vehicular, se implementó un proceso, que a continuación, se detallan los pasos seguidos.

Paso 1, Representación del índice con lógica difusa: Para representar la gama de colores del índice de movilidad vehicular se realizó el uso de la lógica difusa, a continuación, se describe el proceso:

- Se utilizó una muestra de 1, 048, 575 registros (nuevos registros que no forman parte de la muestra que se utilizó para el desarrollo del modelo) con 7 variables que pertenecen al IM y una variable el cual tiene los valores de movilidad vehicular, esos valores permiten determinar los intervalos de movilidad, como se muestra en la Figura 3.42.

$=2*B2 + 3*RAIZ(A2) + RAIZ(LOG(E2)) - 5*LOG(F2) + SENO(LOG(D2)) + SENO(COS(SENO(B2 + RAIZ(C2) - LOG(G2) + SENO(LOG(D2)) - 0.729))) - 1.458$

A	B	C	D	E	F	G	H	I	J	K	L	M
Bbox_Traffic	Free_Flow	Length	Speed_Uncapped	Daily_Traffic	Type	Bbox_Servicio	IM					
7	2.777778	480	5.555556	2	401	7	0.778378					
7	2.777778	480	5	3	401	7	0.881052					
1	5.555556	979	2.777778	3	401	1	1.330425					
3	4.444447	24	3.611112	2	401	3	1.387254					
1	5.555556	47	2.5	3	401	1	1.491851					
1	5.555556	340	6.111111	4	401	1	1.692591					
1	5.555556	340	8.611112	3	401	1	1.731646					
1	5.555556	340	6.666667	5	401	1	1.785871					
1	5.555556	340	9.166667	4	401	1	1.837424					
1	5.555556	340	8.611112	5	401	1	1.876951					
1	5.833335	223	3.611112	3	401	1	1.934933					
3	4.444447	24	3.055556	30	401	3	1.970117					
1	5.833335	223	4.166667	4	401	1	2.078073					
1	5.833335	41	3.055556	3	401	1	2.099952					
1	5.833335	41	3.055556	6	401	1	2.291341					
1	5.833335	223	6.111111	6	401	1	2.333669					
1	5.833335	41	4.722223	4	401	1	2.384217					

Figura 3.42. Fragmento de conjunto de datos de 1, 048, 575 registros, donde cada renglón tiene su valor de IM

- Se realizó la definición de conjuntos difusos de baja, media y alta, a partir de la función de membresía de tipo Gaussiano, que le corresponde la siguiente ecuación.

$$\mu_A(x) = \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right) \quad (3.15)$$

Donde:

$\mu_A(x)$ es el grado de pertenencia del valor x al conjunto difuso A .

c es el valor alrededor del cual la función de membresía alcanza su máximo (normalmente igual a 1).

σ controla la anchura de la curva: valores más grandes de σ hacen que la curva sea más ancha y menos pronunciada.

Se definieron tres conjuntos difusos para los valores del IMC utilizando funciones de membresía gaussianas: "bajo", "medio" y "alto". Los valores de media y desviación estándar se obtuvieron dividiendo el conjunto de datos de 1, 048,575 registros, ordenado, en tres conjuntos de datos: el primer conjunto abarca desde el registro 0 hasta el 349,524, el segundo conjunto abarca desde el registro 349,525 hasta el 699,049, y el tercer conjunto abarca desde el registro 699,050 hasta el 1, 048,575.

- o IM baja: Media $c = 16.71$ y desviación estándar $\sigma=4.15$ conjunto 1
- o IM media: Media $c = 26.33$ y desviación estándar $\sigma=2.5$ conjunto 2
- o IM alta: Media $c = 36.15$ y desviación estándar $\sigma=5.30$ conjunto 3

Se implementó la biblioteca de scikit-fuzzy en Python para visualizar los valores de pertenencia en la función de membresía gaussiana, como se muestra en la Figura 3.43.

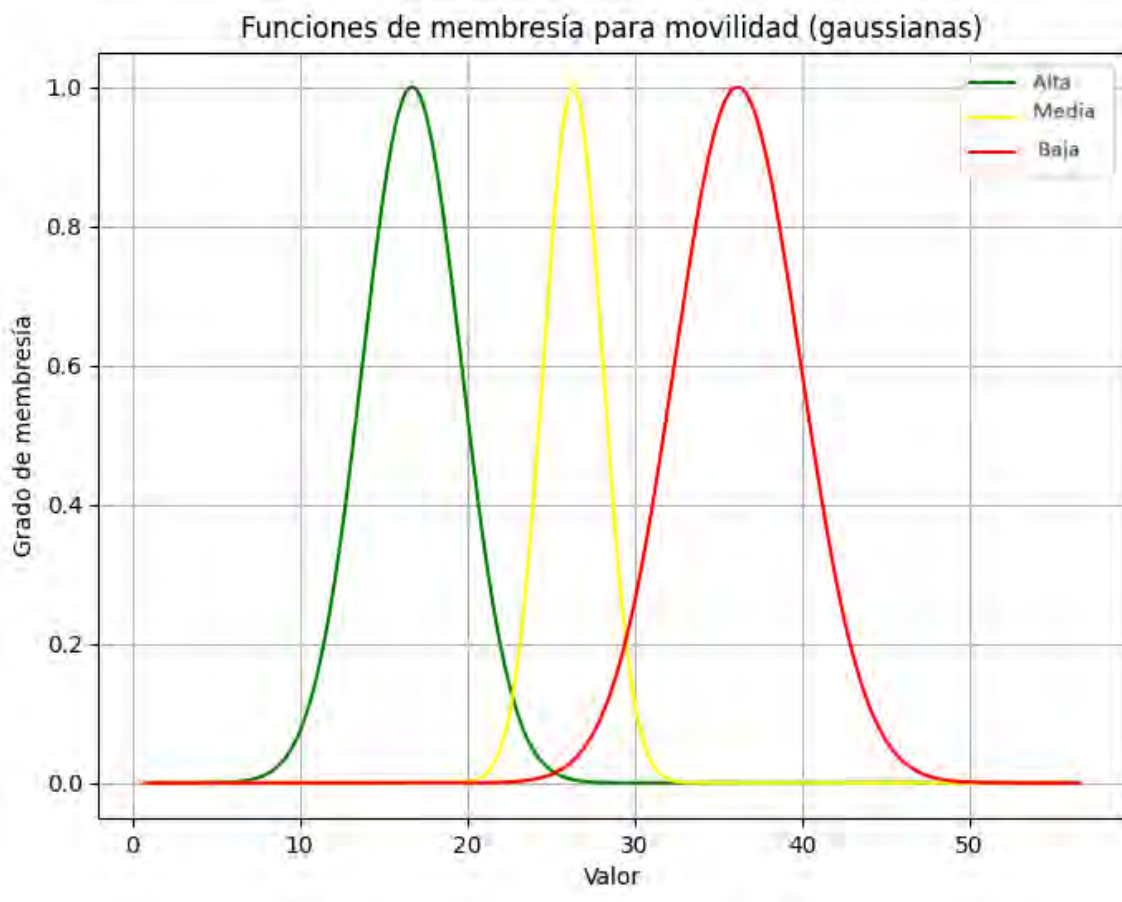


Figura 3.43. Función de membresía de tipo gaussiano para obtener los valores de pertenencia

Los límites de la función de pertenencia encontrados corresponden los siguientes valores:

- Movilidad alta (Verde): 4.12 - 29.30: Este rango indica una alta movilidad vehicular.
- Movilidad media (Amarillo): 18.74 - 33.92: Este rango representa una movilidad vehicular moderada.
- Movilidad baja (Rojo): 20.07 - 52.23: Este rango indica una baja movilidad vehicular.

La superposición de estos rangos puede causar confusión en la interpretación de los datos. Idealmente, cada rango debería ser mutuamente exclusivo para permitir una distinción clara. Por lo tanto, se proponen los siguientes intervalos:

- Movilidad Alta: $[0, 21]$
- Movilidad Media: $(21, 29]$

- Movilidad Baja: (29, 56]

Con las acciones anteriormente descritas se obtuvieron tres intervalos de valores para representar cuantitativamente el nivel de desempeño de la movilidad vehicular. Estos intervalos han sido asignados a una escala de colores, lo que permite una visualización intuitiva y estructurada de los distintos niveles de movilidad, facilitando el análisis y la interpretación de los datos.

Mediante la aplicación de lógica difusa, se logró una asignación precisa de los valores de pertenencia asociados a la movilidad vehicular. Este enfoque permitió modelar la incertidumbre y la variabilidad inherente al sistema, facilitando una representación flexible y realista de los niveles de desempeño de la movilidad.

A partir de las actividades desarrolladas en el marco del método de representación dinámica del modelo de movilidad vehicular, se ha definido un conjunto de actividades sugeridas (ver Tabla 3.35) que permiten la visualización y análisis interactivo de la evolución del sistema en distintos escenarios. Este enfoque posibilita la implementación de técnicas avanzadas de visualización de datos y modelado gráfico, lo que facilita la interpretación de los patrones del sistema complejo y su relación con variables. La representación dinámica del modelo es imprescindible para comprender la interacción entre factores como la densidad vehicular, velocidad, congestión y patrones de tráfico, permitiendo además evaluar la efectividad de estrategias de gestión y optimización de la movilidad vehicular.

El proceso de representación dinámica se basa en la formulación matemática y visual del comportamiento del sistema, asegurando una captura precisa de las relaciones entre las variables centrales. Para ello, se desarrollan diversas actividades, como la identificación de variables y parámetros vitales, el diseño de ecuaciones diferenciales y representaciones gráficas, y la validación de los modelos mediante datos históricos y nuevos datos. Además, los modelos dinámicos se ajustan en tiempo real mediante la integración de datos en tiempo real y la retroalimentación del sistema, lo que permite una actualización continua del modelo de movilidad vehicular.

Para la implementación efectiva de este método, se requiere la participación de roles especializados. Como salidas del proceso, se obtienen representaciones matemáticas detalladas de las dinámicas del sistema, junto con visualizaciones gráficas interactivas que ilustran las relaciones entre las variables de movilidad vehicular. Además, se generan informes validados sobre los modelos dinámicos, incluyendo análisis de precisión de escenarios. También se establecen métricas de ajuste y desempeño del modelo, permitiendo evaluar su eficacia en la representación del tráfico y su evolución en el tiempo.

Tabla 3.35. Cuadro descriptivo del método representación dinámica del modelo

Caracterización del modelo	
Descripción	Representar de forma matemática y visual las dinámicas de un sistema complejo, capturando las relaciones entre las variables necesarios para predecir y comprender su comportamiento.
Actividades	- Identificar las variables y parámetros principales del sistema. - Diseñar ecuaciones y representaciones gráficas que modelan las relaciones entre variables. - Validar las representaciones dinámicas mediante datos históricos y simulaciones. - Incorporar técnicas como análisis difuso y algoritmos genéticos para mejorar la precisión del modelo. - Ajustar los modelos dinámicos basándose en datos en tiempo real y retroalimentación del sistema. - Generar visualizaciones dinámicas para la interpretación de los modelos.
Roles	- Modeladores matemáticos. - Ingenieros de sistemas complejos. - Analistas de datos. - Especialistas en visualización de datos. - Investigadores en sistemas dinámicos.
Salidas	- Representaciones matemáticas de las dinámicas del sistema. - Visualizaciones gráficas de las interacciones entre variables. - Informes validados de modelos dinámicos. - Métricas de ajuste y desempeño del modelo. - Documentación técnica de los modelos representados.

3.8.2 Implementación adaptativa del modelo de movilidad vehicular

La implementación adaptativa del modelo asegura que este se ajuste a las variaciones dinámicas inherentes al sistema complejo, como lo es la movilidad vehicular. En esta etapa se integraron técnicas de sistemas complejos como el modelado adaptativo no lineal y el análisis de sensibilidad dinámica, lo cual permitió ajustar el comportamiento del modelo en función de condiciones cambiantes en tiempo real. Desde la ciencia de datos, se aplicaron estrategias de aprendizaje supervisado y aprendizaje continuo, utilizando redes neuronales tipo feedforward para validar predicciones del modelo y feature importance para identificar variables clave. En el ámbito de la construcción de indicadores, se utilizó una estructura modular para evaluar el Índice de Movilidad (IM) bajo distintos escenarios, integrando técnicas de análisis multivariado y escalado normalizado. Desde la minería de datos, se aplicaron algoritmos de comparación de series temporales, así como detección de patrones mediante ventanas móviles y análisis de error. En el área de cómputo evolutivo, se adaptaron parámetros del algoritmo de programación genética en función de la retroalimentación del sistema y los datos más recientes. Complementariamente, se incorporaron conceptos de sistemas de control adaptativo, teoría de la información (para evaluar la ganancia informativa de variables) y teoría de la estabilidad de Lyapunov, proporcionando una base teórica sólida para garantizar la convergencia del modelo bajo condiciones dinámicas.

Para llevar a cabo la implementación adaptativa del modelo previamente construido de movilidad vehicular, se siguió un proceso estructurado compuesto por las siguientes etapas. Este procedimiento tiene como objetivo integrar el modelo en un entorno operativo, ajustando dinámicamente sus parámetros conforme a las condiciones reales observadas en el sistema.

Paso 1, Selección de la muestra: Se utilizó una muestra de 1, 000, 000 registros (Datos nuevos a los anteriores experimentos) con 7 variables que pertenecen al IM.

Paso 2, Análisis de características: Se analizó cómo el modelo adaptativo pondera las diferentes características para realizar caracterizaciones. Este análisis proporciona información importante sobre cuáles variables tienen un mayor impacto en la precisión del modelo y cómo estas influyen en los resultados. La importancia de las características se mide en términos relativos, asignando un peso a cada variable en función de su contribución al desempeño del modelo.

La gráfica de la Figura 3.44 ilustra estos pesos asignados. Se observa que la velocidad_promedio tiene la mayor relevancia, indicando que el modelo se apoya significativamente en esta variable para realizar sus predicciones. Otras variables como hora y día también muestran una importancia moderada, reflejando cómo las condiciones temporales (hora del día o día de la semana) afectan el comportamiento del tráfico. Por último, la variable incidencias, aunque tiene un menor peso relativo, aporta información que puede ser relevante en situaciones específicas, como alteraciones por eventos inesperados.

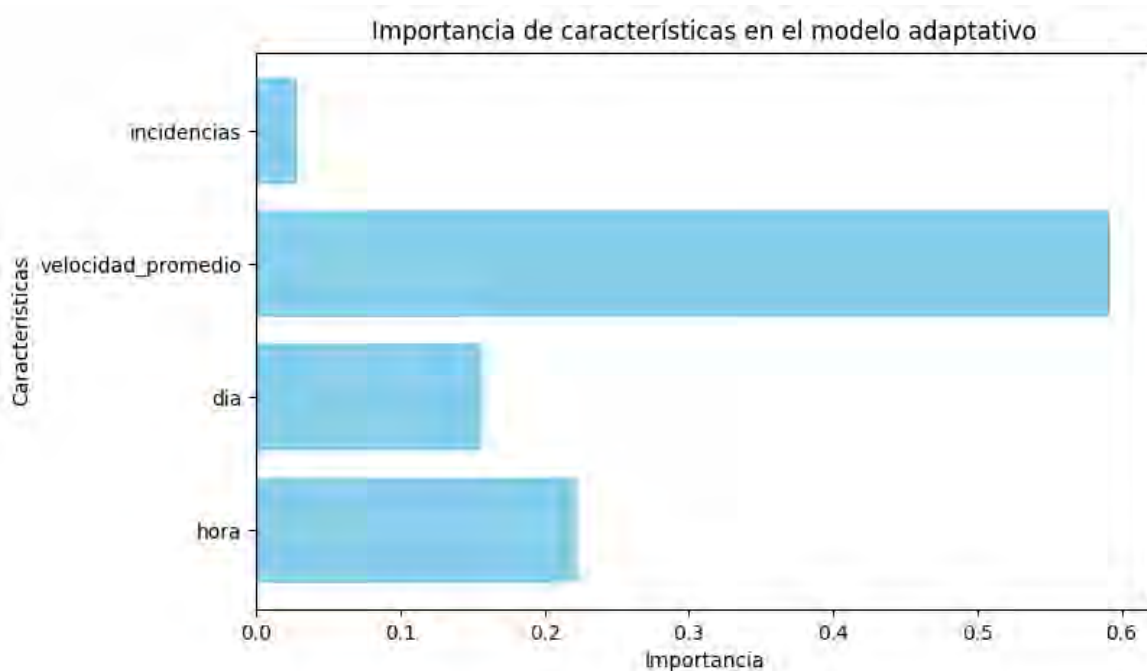


Figura 3.44. Importancia de las características en el modelo adaptativo

Paso 3, Comparación del modelo adaptativo y estático con los datos reales: Se realizó un análisis comparativo entre las caracterizaciones del modelo adaptativo, el modelo estático y los datos reales. Este análisis permitió evaluar cómo cada modelo se ajusta a la dinámica observada en los datos y determinar cuál proporciona una mejor representación del sistema.

La comparación se ilustra mediante una gráfica de la Figura 3.45 en la que los datos reales se representan con una línea azul, las predicciones del modelo adaptativo con una línea verde y las del modelo estático con una línea roja discontinua. El modelo adaptativo se caracteriza por su capacidad para capturar con mayor precisión las

fluctuaciones de los datos reales, reflejando las variaciones temporales del sistema. Por otro lado, el modelo estático, al basarse en un valor promedio, muestra un comportamiento más uniforme, lo que puede resultar útil en sistemas con menor variabilidad.

Esta etapa permite identificar las ventajas y limitaciones de cada modelo. Mientras el modelo adaptativo destaca en escenarios dinámicos y con alta variabilidad, el modelo estático puede ser preferible en sistemas estables y con menor complejidad. Este análisis es de vital importancia para seleccionar el enfoque más adecuado, considerando tanto las características del sistema estudiado como el objetivo planteado en el análisis.

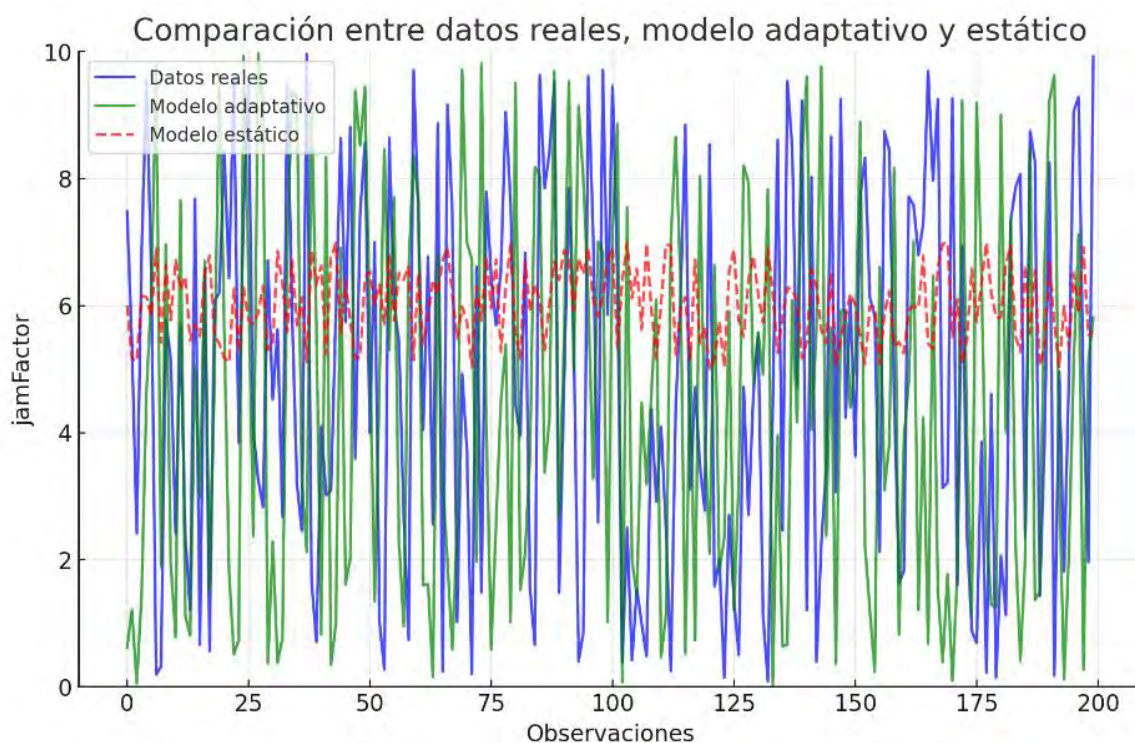


Figura 3.45. Comportamiento del *jamFactor* a lo largo de las observaciones

Paso 4, Representación gráfica del error cuadrático medio (MSE) para ambos modelos: Se comparó el rendimiento de los modelos adaptativo y estático utilizando el Error Cuadrático Medio (MSE) como métrica de evaluación. El MSE mide la diferencia promedio al cuadrado entre las predicciones de los modelos y los valores reales, siendo una métrica sensible a los errores grandes. Un menor valor de MSE indica un mejor ajuste del modelo a los datos reales.

La gráfica de barras de la Figura 3.46 ilustra el MSE de ambos modelos, donde se observa que el modelo estático tiene un MSE menor en comparación con el modelo adaptativo. Este resultado podría indicar que, en este caso particular, el modelo estático presenta una menor desviación con respecto a los datos reales. Sin embargo, esta métrica debe interpretarse junto con otros factores, como la naturaleza del

sistema y la variabilidad de los datos, ya que un modelo más simple como el estático puede no captar adecuadamente las dinámicas complejas.

Este análisis resalta la importancia de combinar métricas de evaluación y análisis visual para entender las fortalezas y debilidades de cada modelo. Además, permite identificar escenarios en los que uno de los modelos podría ser efectivo, considerando las características del sistema y el objetivo del análisis. La representación gráfica facilita la comparación visual de los resultados, ayudando a tomar decisiones informadas sobre la selección del modelo adecuado.

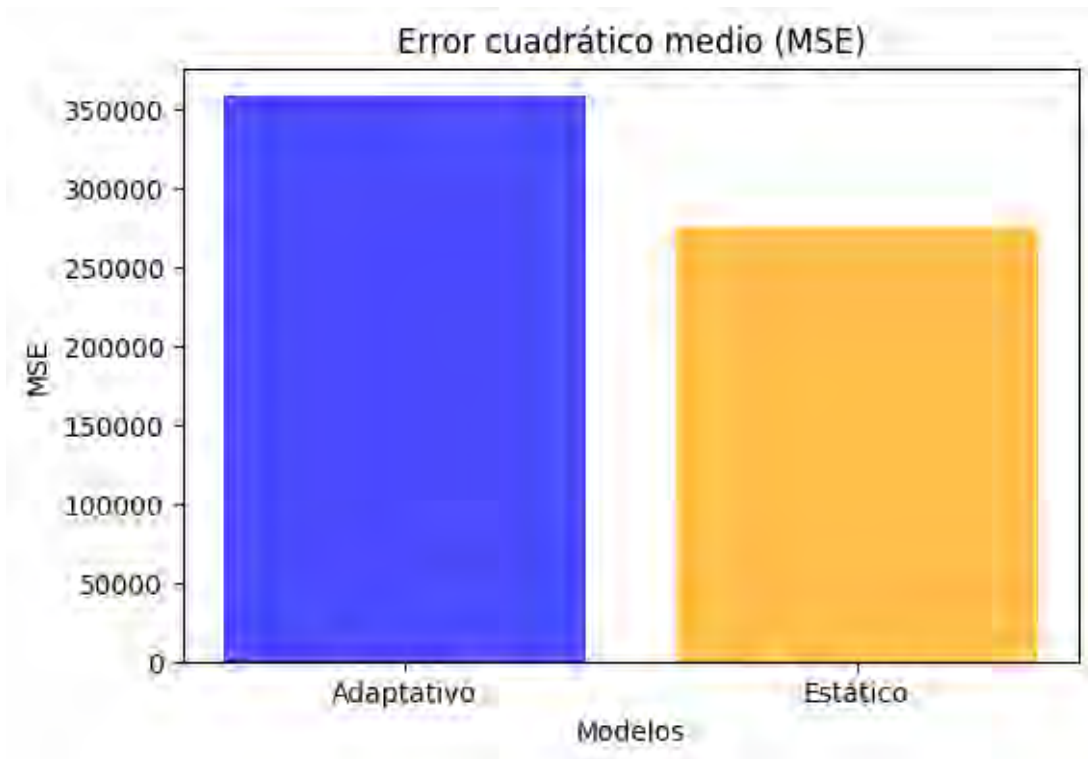


Figura 3.46. Comparación del Error Cuadrático Medio (MSE) entre el modelo adaptativo y el modelo estático

La implementación del modelo adaptativo de movilidad vehicular permitió evaluar su capacidad de ajuste a las variaciones dinámicas del sistema en comparación con un modelo estático. A través de un proceso estructurado, se analizaron distintos aspectos del modelo, incluyendo la importancia relativa de las variables, la precisión en la predicción de la dinámica vehicular y la comparación del Error Cuadrático Medio (MSE) entre ambos enfoques. Los resultados indican que la velocidad promedio es la variable con mayor peso en la predicción del modelo, seguida por variables temporales como la hora del día y el día de la semana, lo que confirma la influencia de estos factores en la movilidad.

El modelo adaptativo muestra una mayor capacidad para ajustarse a variaciones temporales y responder a cambios en las condiciones de movilidad, su desempeño depende de la calidad y actualización de los datos de entrada. Por otro lado, el modelo

estático, al operar con valores promedio, puede ofrecer una estabilidad que resulta útil en sistemas con menor variabilidad. La evaluación del MSE evidencia que la complejidad del modelo no siempre garantiza un mejor ajuste, por lo que es esencial considerar métricas adicionales y análisis cualitativos en la selección del modelo más adecuado. En consecuencia, la combinación de ambos enfoques puede ser una estrategia óptima para llevar a cabo la modelización de la movilidad vehicular en distintos contextos.

A partir de las actividades desarrolladas en el marco del método de implementación adaptativa del modelo de movilidad vehicular, se ha definido un conjunto de actividades sugeridas (ver Tabla 3.36) que permiten la integración del modelo en entornos operativos, asegurando su capacidad de ajuste dinámico ante nuevas condiciones del sistema complejo. Este enfoque posibilita la automatización de procesos de calibración y actualización, permitiendo que el modelo evolucione en función de los cambios en la movilidad vehicular y la incorporación de nuevos datos.

La implementación adaptativa de modelos dinámicos se basa en la integración de sistemas de monitoreo en tiempo real, que facilitan la reconfiguración automática de parámetros y la incorporación de estrategias de aprendizaje continuo para mejorar el desempeño del modelo. Para ello, se desarrollan diversas actividades, tales como la integración del modelo en plataformas de monitoreo en tiempo real, la configuración de sistemas automatizados que ajusten parámetros en función de datos recientes, y la implementación de algoritmos de aprendizaje continuo que optimicen la precisión del modelo conforme se recopila nueva información.

Para la correcta ejecución de este método, es básico la participación de roles especializados. La colaboración interdisciplinaria entre estos roles garantiza la integración eficiente del modelo en entornos operacionales, permitiendo su ajuste continuo mediante técnicas de calibración automatizada. Como salidas del proceso, se obtiene un modelo implementado y adaptativo, funcional en plataformas reales de monitoreo y análisis de movilidad vehicular. Se generan también informes de monitoreo y desempeño del modelo, proporcionando métricas detalladas sobre su capacidad de adaptación.

Tabla 3.36. Cuadro descriptivo del método implementación adaptativa

Adaptación del modelo	
Descripción	Implementar modelos dinámicos en sistemas reales, garantizando flexibilidad y adaptabilidad a los cambios en tiempo real mediante ajustes automáticos y técnicas de aprendizaje continuo.
Actividades	- Integrar modelos dinámicos en plataformas de monitoreo en tiempo real. - Configurar sistemas automatizados para ajustar parámetros basados en datos recientes. - Implementar algoritmos de aprendizaje continuo para mejorar el desempeño del modelo. - Monitorear el funcionamiento del modelo y evaluar su adaptabilidad a condiciones cambiantes. - Incorporar retroalimentación del sistema para optimizar la implementación. - Realizar pruebas regulares para validar la efectividad de los ajustes automáticos.
Roles	- Desarrolladores de software en tiempo real. - Ingenieros de datos. - Especialistas en aprendizaje automático. - Analistas de desempeño de sistemas. - Investigadores en adaptabilidad de modelos.
Salidas	- Modelos implementados y adaptativos en plataformas reales. - Algoritmos de ajuste automático y aprendizaje continuo. - Informes de monitoreo y desempeño del modelo. - Métricas de efectividad de la implementación adaptativa. - Documentación de retroalimentación y ajustes realizados.

3.8.3 Actualización del modelo de movilidad vehicular

Este método tiene el propósito de evaluar el impacto de la actualización dinámica en un modelo de del sistema complejo, incorporando datos recientes en tiempo real para analizar su desempeño frente a un modelo estático. En la actualización del modelo de movilidad vehicular, se integraron múltiples técnicas de diversas disciplinas. Desde la perspectiva de sistemas complejos, se aplicaron mecanismos de retroalimentación adaptativa y actualización continua basada en observaciones del entorno, propios de modelos autoorganizados que permiten capturar la naturaleza cambiante de los sistemas urbanos. En ciencia de datos, se utilizaron estrategias de validación cruzada, aprendizaje incremental y entrenamiento continuo, garantizando la precisión del modelo ante nuevas condiciones. En el ámbito de la construcción de indicadores, se ajustó el Índice de Movilidad mediante la recalibración funcional con base en derivadas parciales simbólicas, manteniendo la coherencia del modelo con la realidad dinámica. Desde la minería de datos, se implementaron técnicas de análisis de series temporales y detección de desviaciones, fundamentales para identificar puntos críticos que motivan la actualización del modelo. Además, el enfoque incorporó principios del cómputo evolutivo, al ajustar la estructura del modelo simbólico mediante operadores de mutación dirigidos y estrategias de búsqueda local.

Para realizar la Actualización del modelo de movilidad vehicular, se implementó un proceso estructurado que comprende los siguientes pasos detallados a continuación.

Paso 1. Obtención del modelo base adaptativo: Se seleccionó la ecuación 9 $IM = 2X_1 + 3\sqrt{X_0} + \sqrt{\log(X_6)} - 5\log(X_{11}) + \sin(\log(X_5)) + \sin(\cos(\sin(X_1 + \sqrt{X_2} - \log(X_{20}) + \sin(\log(X_5)) - 0.729))) - 1.458$ de la Tabla 3.29, como el modelo base inicial para las actualizaciones y análisis posteriores, haciendo uso de su capacidad para representar con efectividad las dinámicas de movilidad vehicula.

Paso 2. Actualización del modelo: El modelo de movilidad vehicular actualizado se expresa mediante la siguiente función: $IM \approx 2 \cdot X_1 + 3 \cdot \sqrt{X_0} + \sqrt{\log(X_6)} - 5 \cdot \log(X_{11}) - 0.308$. Esta expresión representa una aproximación optimizada del índice de movilidad (IM) a partir de variables clave relacionadas con la dinámica del tráfico. La actualización del modelo consistió en ajustar sus parámetros e integrar nuevos datos provenientes de condiciones recientes de movilidad vehicular, previamente observadas y estructuradas en versiones anteriores del sistema. Este proceso no solo mejoró la precisión del modelo, sino que también potenció su capacidad de adaptación frente a cambios en la estructura del tráfico urbano. Como resultado, se incrementó su desempeño predictivo, logrando una representación más realista y eficiente del comportamiento vehicular actual.

A partir de las actividades desarrolladas en el marco del método de actualización del modelo de movilidad vehicular, se ha definido un conjunto de actividades sugeridas (ver Tabla 3.37) que permiten la integración continua de nuevos datos y la recalibración del modelo, asegurando que refleje con mayor precisión la dinámica del sistema complejo de movilidad vehicular. Este enfoque posibilita la incorporación de información en tiempo real, lo que permite mejorar la parametrización del modelo, optimizar su capacidad de aproximación y ajuste, y garantizar que se mantenga alineado con las condiciones actuales del sistema de movilidad.

La actualización del modelo dinámico es un proceso para mantener su validez y aplicabilidad en un entorno en constante cambio. Para ello, se implementan diversas actividades, tales como el monitoreo del desempeño del modelo en tiempo real, lo que permite detectar desviaciones o pérdida de precisión en las predicciones. Asimismo, se lleva a cabo la incorporación de nuevos datos, para ajustar los parámetros del modelo. Una vez integrados los datos, se procede a reentrenar los modelos mediante técnicas avanzadas de aprendizaje automático, lo que mejora su capacidad predictiva y adaptativa.

Para la correcta ejecución de este método, se requiere la participación de roles especializados. La colaboración de estos roles garantiza que el modelo se actualice eficientemente, manteniendo su precisión, relevancia y capacidad de adaptación ante cambios en la movilidad vehicular. Como salidas del proceso, se obtiene un modelo actualizado y optimizado, con parámetros ajustados para reflejar las condiciones actuales del tráfico. Se generan informes de desempeño comparativo, que permiten evaluar las mejoras obtenidas en relación con versiones previas. Asimismo, se producen nuevas métricas de precisión y confiabilidad, asegurando que el modelo mantenga su capacidad predictiva con datos recientes.

Tabla 3.37. Cuadro descriptivo del método Actualización del modelo

Retroalimentación del modelo	
Descripción	Actualizar periódicamente los modelos dinámicos para garantizar su precisión, relevancia y adaptabilidad a cambios en las condiciones del sistema mediante la incorporación de nuevos datos. - Monitorear el desempeño de los modelos existentes en tiempo real. - Incorporar nuevos datos recolectados al sistema para ajustar los modelos.
Actividades	- Reentrenar los modelos utilizando técnicas avanzadas de aprendizaje automático. - Comparar el desempeño de los modelos actualizados con versiones anteriores. - Implementar validaciones cruzadas para asegurar la confiabilidad de los modelos actualizados. - Documentar los cambios realizados y las métricas de mejora obtenidas.
Roles	- Ingenieros de datos. - Especialistas en aprendizaje automático. - Analistas de modelos dinámicos. - Investigadores en validación de modelos. - Administradores de sistemas en tiempo real.
Salidas	- Modelos actualizados y optimizados. - Informes de desempeño comparativo entre versiones del modelo. - Nuevas métricas de precisión y confiabilidad del modelo. - Algoritmos ajustados para incorporaciones futuras. - Documentación técnica sobre los procesos de actualización.

3.9 Conclusiones

La construcción de la metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos ha permitido establecer un enfoque estructurado y adaptable para el modelado matemático de fenómenos dinámicos en entornos de alta variabilidad. A lo largo de este capítulo, se han definido seis fases, cada una diseñada para analizar aspectos trascendentales en el proceso de modelado, asegurando una construcción progresiva y coherente del sistema complejo.

Cada fase ha sido diseñada mediante la integración de un conjunto de métodos específicos, los cuales permiten capturar, analizar y modelar la dinámica del sistema bajo estudio. Para cada método, se han propuesto actividades genéricas sugeridas, con el propósito de proporcionar un marco flexible y adaptable a diferentes dominios de aplicación. Inicialmente, se llevó a cabo la identificación de entidades y relaciones dinámicas, estableciendo las interacciones críticas que definen el comportamiento del sistema. Posteriormente, la definición de variables críticas permitió seleccionar los factores necesarios que determinan su evolución temporal.

En la fase de niveles de agregación del sistema complejo dinámico, se estableció una estructura jerárquica que facilita el análisis a distintas escalas espaciales y temporales, mientras que el análisis de incertidumbre del sistema permitió cuantificar la variabilidad inherente en los datos y minimizar sesgos en las representaciones matemáticas. La identificación de las variables decisivas fue un paso esencial para garantizar que los modelos reflejen con precisión las dinámicas del sistema.

La recopilación continua de datos y la integración de fuentes dinámicas fueron esenciales para consolidar un entorno de análisis basado en información en tiempo real, lo que llevó al diseño de almacenamiento eficiente, garantizando la disponibilidad y accesibilidad de los datos. En paralelo, el preprocesamiento en tiempo real y el muestreo dinámico permitieron mejorar la calidad de la información, asegurando que el modelo recibiera datos consistentes y representativos de la dinámica del sistema.

El modelado dinámico fue la fase en la que se formalizó la representación matemática del sistema mediante técnicas de aprendizaje evolutivo y modelado simbólico, permitiendo la aproximación de su comportamiento en distintos escenarios. La evaluación y ajuste continuo del modelo consolidó un proceso iterativo de optimización, mediante la recalibración de parámetros y la comparación con datos observacionales. La combinación de una representación dinámica, una implementación flexible y una actualización constante hizo posible la integración del modelo en entornos variables y su uso en plataformas interactivas.

El resultado final de esta metodología fue la construcción de un modelo matemático denominado Índice de Movilidad, el cual se diseñó específicamente para aproximarse al comportamiento de la movilidad vehicular. Este modelo, basado en la estructura metodológica propuesta, refleja con alta fidelidad la relación entre variables base de la movilidad vehicular y facilita una estimación precisa de la dinámica de la movilidad en contextos urbanos de alta variabilidad.

Capítulo 4. Validación de la metodología

4.1 Introducción

El modelado de sistemas complejos altamente dinámicos exige metodologías capaces de capturar su naturaleza no lineal, adaptativa y multiescalar, al tiempo que deben ser sometidas a una evaluación rigurosa que garantice su fiabilidad para la toma de decisiones y su contribución al avance del conocimiento científico. La validación de la metodología desarrollada implica poner a prueba su capacidad para representar problemas desde una perspectiva sistémica, asegurando que refleje con precisión las interacciones entre variables, la evolución del sistema en distintos escenarios y los patrones emergentes que caracterizan este tipo de entornos. El principal desafío en la modelación de estos sistemas radica en la necesidad de representar tanto su estructura subyacente como su comportamiento dinámico y contingente, utilizando datos históricos y en tiempo real. En este contexto, la validación no es solo un proceso técnico, sino un paso crítico que permite identificar las fortalezas, limitaciones y condiciones de aplicabilidad del enfoque propuesto, asegurando que los modelos generados mantengan coherencia con la realidad observada.

Mediante el análisis de casos de estudio concretos, se evaluó la respuesta de la metodología bajo condiciones variables, su sensibilidad ante la disponibilidad y calidad de los datos, así como su capacidad de ajuste y predicción de patrones temporales y espaciales en diferentes dominios. Más allá de comprobar su viabilidad, esta validación permitió establecer criterios concretos para su mejora y adaptación, ampliando así su utilidad en el estudio y gestión de sistemas complejos. En este trabajo, la validación empírica de la metodología propuesta se llevó a cabo mediante su aplicación a dos sistemas complejos altamente dinámicos: la calidad del aire en la Ciudad de México y la profundidad de los sismos en el territorio nacional. La metodología fue evaluada de acuerdo con criterios específicos, entre los cuales destacan:

- La representación precisa de dinámicas multiescales mediante esquemas de agregación jerárquica, como la segmentación urbana para el caso del aire y la agregación geoespacial nacional para los sismos.
- La capacidad de los modelos generados por programación genética para capturar relaciones no lineales entre variables e identificar patrones robustos y coherentes con el conocimiento del dominio.
- La integración automatizada de datos en tiempo real e históricos, lo que permitió mejorar la capacidad de respuesta frente a eventos críticos, como episodios de contaminación o sismos atípicos.
- La identificación de variables clave en cada fenómeno: PM_{2.5}, PM₁₀ y O₃ en la calidad del aire; latitud, longitud y minuto del evento en la profundidad sísmica.

- La interpretabilidad de los modelos obtenidos y su potencial de aplicación práctica, mediante plataformas interactivas que facilitan la toma de decisiones en tiempo real.

Aunque los modelos generados han mostrado una alta capacidad de representación y predicción, también se identificaron áreas de mejora, como la incorporación de nuevas variables geofísicas en el caso de los sismos o factores socioeconómicos en el análisis del aire. La optimización de los parámetros evolutivos de los algoritmos mediante estrategias adaptativas también se considera una vía prometedora para incrementar la precisión y eficiencia computacional. En conjunto, la validación realizada confirma la versatilidad de la metodología como una herramienta integral para el modelado de sistemas complejos altamente dinámicos en distintos contextos de aplicación.

4.2 Proceso de validación de la metodología propuesta para el desarrollo de modelos de comportamiento en sistemas complejos altamente dinámicos

Este capítulo presenta la aplicación de la metodología propuesta en la captura y modelado de sistemas altamente dinámicos, para representar las interacciones complejas y no lineales que definen estos fenómenos. A través de su implementación en dos casos de estudio significativos, la calidad del aire en la Ciudad de México y la profundidad de los sismos en México (ver Tabla 4.1), se analiza cómo la estructura metodológica propuesta permite modelar las relaciones causales y dinámicas emergentes dentro de cada sistema, como se ilustra en la Figura 4.1.



Figura 4.1 Validación de la metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos

El primer caso de estudio se centra en la calidad del aire en la Ciudad de México, un sistema influenciado por múltiples factores ambientales, meteorológicos y urbanos. Mediante la metodología propuesta, se integran datos de contaminación atmosférica

y condiciones climáticas, lo que permite modelar la evolución de los niveles de contaminantes en función de variables como la concentración de partículas PM2.5 y PM10, la temperatura, la humedad y los patrones de viento. A través de técnicas de modelado y análisis de datos, se logran representaciones matemáticas que facilitan la caracterización de escenarios y la identificación de estrategias de mitigación, optimizando la gestión ambiental y la toma de decisiones en políticas públicas.

El segundo caso de estudio se centra en el análisis de la profundidad de los sismos en México, un fenómeno que involucra procesos tectónicos complejos con variabilidad espacial y temporal significativa. La metodología permite estructurar un modelo que incorpora factores geológicos, parámetros sísmicos y distribución espacial de epicentros, posibilitando la identificación de tendencias y correlaciones entre la profundidad de los eventos sísmicos y su magnitud. La aplicación de técnicas de modelado de sistemas complejos en este contexto contribuye a la comprensión de los procesos sísmicos en la región, proporcionando información sustancial para el desarrollo de estrategias de gestión del riesgo y prevención de desastres.

Tabla 4.1. *Aplicación de cada fase de la metodología propuesta en los casos de estudio de calidad del aire y profundidad de sismos*

Fase Metodológica	Aplicación en Calidad del Aire	Aplicación en Sismos
Fase 1: Análisis contextual y dinámicas del sistema	Identificación de contaminantes clave (PM2.5, PM10, O ₃) y su relación con variables meteorológicas.	Análisis de parámetros sísmicos (magnitud, profundidad, localización) y patrones geológicos.
Fase 2: Agregación multinivel y medición de la dinámica compleja	Segmentación por zonas urbanas y niveles de contaminación para análisis multiescalar.	Agregación espacial a nivel nacional y por regiones sísmicas; clasificación por profundidad.
Fase 3: Recopilación y monitoreo continuo de datos en tiempo real	Integración de datos desde APIs de monitoreo ambiental y estaciones en tiempo real.	Recopilación de datos históricos del Servicio Sismológico Nacional y fuentes geofísicas.
Fase 4: Preprocesamiento y actualización dinámica de datos	Preprocesamiento de series temporales, control de calidad y muestreo horario.	Filtrado, estandarización y actualización continua de registros sísmicos.
Fase 5: Modelado iterativo y desarrollo continuo	Modelado de escenarios de dispersión de contaminantes y ajuste de ecuaciones mediante programación genética.	Generación de un Índice de Profundidad Sísmica (IPS) mediante programación genética.
Fase 6: Interpretación, representación e implementación continua	Representación gráfica del ICA, implementación en plataformas interactivas y actualización adaptativa.	Visualización de mapas sísmicos, identificación de zonas críticas y soporte a estrategias de prevención.

El primer caso de estudio, la calidad del aire en la Ciudad de México, se enfoca en un problema de gran relevancia para la salud pública y la sostenibilidad ambiental. Este fenómeno, caracterizado por la interacción de dinámicas atmosféricas, emisiones vehiculares y condiciones climáticas, requiere de herramientas analíticas que permitan identificar patrones complejos y cuantificar su impacto mediante un Índice de Calidad del Aire. Este índice busca sintetizar información crítica para su

interpretación y contribuir a la toma de decisiones informadas en la gestión de la calidad del aire.

El segundo caso de estudio se enfoca en un tema igualmente complejo: la profundidad de los sismos en México. Dado que México se encuentra en una de las regiones tectónicamente más activas del mundo, comprender los factores que influyen en la profundidad de los sismos es importante para mejorar la evaluación de riesgos sísmicos y diseñar estrategias de mitigación. En este caso, se construye un índice que caracteriza las interacciones entre variables geológicas y dinámicas tectónicas, permitiendo una representación más clara y estructurada del fenómeno.

Este capítulo muestra cómo se implementaron los métodos de cada fase de la metodología en ambos casos de estudio, desde la identificación de variables importantes y la integración de datos dinámicos hasta la construcción y validación de los índices. A través de estos ejemplos, se busca mostrar la capacidad de la metodología para tratar problemas reales con altos niveles de complejidad, mostrando su utilidad como un marco de trabajo robusto para el análisis, modelado y toma de decisiones en sistemas complejos.

4.3 Caso de estudio 1: Calidad de aire en la Ciudad de México

La calidad del aire en la Ciudad de México constituye un desafío estructural de alta complejidad, debido a su impacto directo en la salud pública, el bienestar social y la sostenibilidad ambiental. Analizar este fenómeno desde un enfoque de sistemas complejos permite identificar, representar y modelar las múltiples interacciones no lineales que lo configuran, considerando tanto variables físicas como socioeconómicas que actúan de manera interdependiente y dinámica en el espacio urbano. Este enfoque no solo enriquece la comprensión integral del problema, sino que también habilita la formulación de estrategias de intervención más eficaces y adaptativas.

El propósito central de este caso de estudio es implementar la metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos, con el fin de diseñar un Índice de Calidad del Aire. Este índice busca integrar y relacionar las variables base que influyen en la calidad del aire, representándolas a través de un modelo matemático para caracterizar su comportamiento en la Ciudad de México. La meta es proporcionar un modelo matemático como una herramienta analítica clara y funcional, que facilite la comprensión de las dinámicas del fenómeno y sirva como soporte para la toma de decisiones informadas en la gestión ambiental.

Este caso de estudio no solo verifica la validez de la metodología propuesta, sino que también destaca su utilidad práctica para enfrentar problemas de gran relevancia social y ambiental. Al aplicar la metodología, se desarrolla una herramienta sólida y efectiva para comprender y gestionar un sistema complejo como la calidad del aire en la Ciudad de México. El análisis se lleva a cabo con una estructura que sigue cada una de las fases de la metodología, descritas a continuación.

Fase 1: Análisis contextual y de la dinámica del sistema

En esta primera fase, se llevó a cabo un análisis integral del contexto y de las dinámicas que determinan la calidad del aire en la Ciudad de México, con el propósito de establecer los fundamentos para su modelado. Dado que se trata de un sistema altamente complejo y dinámico, se identificaron los principales factores que influyen en su comportamiento, incluyendo tanto contaminantes atmosféricos como variables meteorológicas que afectan su dispersión y concentración.

Para ello, se realizó una recopilación de datos de estaciones de monitoreo, destacando el uso de información proveniente de la estación Merced en la Ciudad de México, la cual proporciona mediciones horarias de diversos contaminantes y variables meteorológicas. Durante el periodo de junio a diciembre de 2023, se recopilaron un total de 5,040 registros, permitiendo un análisis detallado de las fluctuaciones en la calidad del aire (**Figura 4.2**). En términos de estructura de datos, el conjunto de información analizado incluyó 10 variables principales, entre ellas PM2.5 (Partículas finas en suspensión), PM10 (Partículas en suspensión menores), O3 (Ozono, contaminante secundario), NO2 (Dióxido de nitrógeno), SO2 (Dióxido de azufre), CO (Monóxido de carbono), Temperatura (Grado de calor del aire), Presión atmosférica (Fuerza ejercida por el aire sobre la superficie), Humedad relativa (Cantidad de vapor de agua en el aire), Velocidad del viento (Movimiento del aire medido en metros por segundo) ver Figura 4.3. Estas variables fueron utilizadas para evaluar las relaciones dinámicas dentro del sistema, permitiendo determinar patrones de variabilidad y su influencia en la calidad del aire.

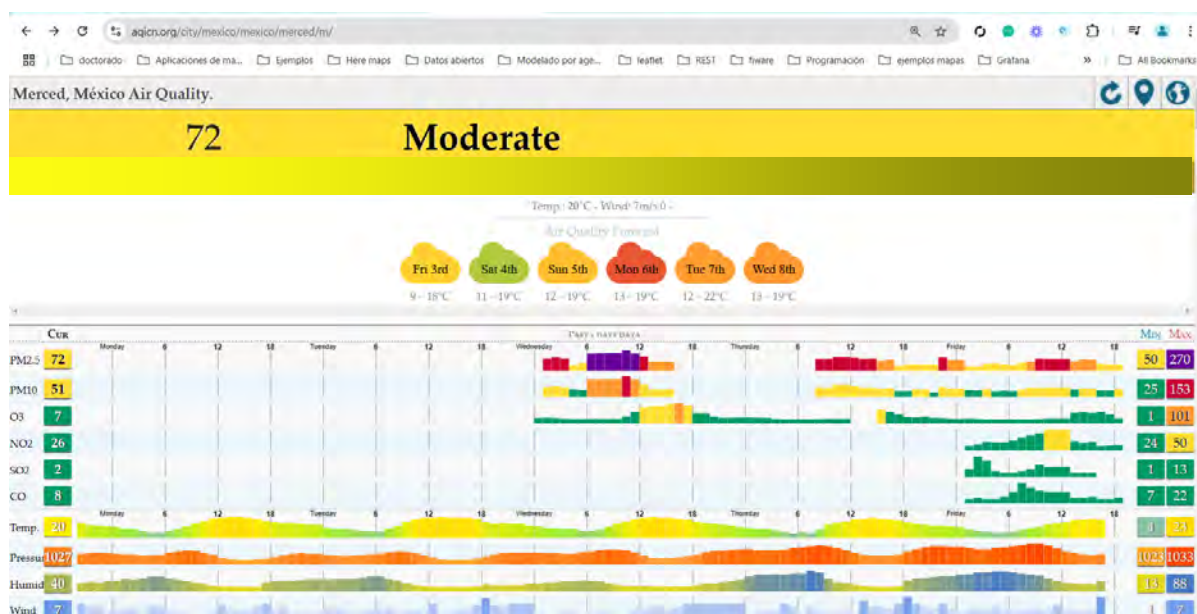


Figura 4.2. Índice de calidad del aire (AQI)

	PM2.5	PM10	O3	NO2	SO2	CO	Temperatura	Presión	Humedad	Viento
1	132.40	122.56	59.84	38.80	8.36	14.42	22.19	1024.07	33.51	5.87
2	259.16	100.74	56.70	24.93	12.35	14.09	10.58	1028.83	13.60	3.45
3	211.04	49.58	66.58	37.35	9.10	15.69	20.49	1031.48	85.79	4.66
4	181.70	110.34	50.70	39.85	10.39	10.29	22.23	1032.54	74.82	3.73
5	84.32	104.79	4.21	28.97	5.40	14.89	21.92	1031.61	59.17	6.42
...	...	...	...	...	...	...	...	...	...	...
5036	133.54	60.09	83.72	28.95	5.98	7.78	5.41	1028.59	25.09	2.94
5037	185.78	50.94	68.06	35.93	1.29	17.03	14.66	1029.05	14.77	2.88
5038	263.36	81.60	90.44	39.95	5.16	11.10	21.48	1027.40	29.45	2.49
5039	210.08	32.28	6.72	30.46	10.68	7.30	21.37	1031.84	57.25	4.03
5040	252.97	150.25	60.70	31.82	10.57	8.00	4.21	1027.27	26.05	2.43

5040 rows × 10 columns

Figura 4.3. Conjunto de datos de calidad del aire

Adicionalmente, se realizó un análisis estadístico detallado para comprender la distribución y variabilidad de los contaminantes y las condiciones meteorológicas. Se generó un resumen estadístico con medidas descriptivas como la media, la desviación estándar y los rangos de cada variable (Figura 4.4), seguido de una matriz de correlación (Figura 4.5) que permitió identificar interacciones lineales entre los distintos factores. Si bien los coeficientes de correlación obtenidos indicaron una relación débil entre la mayoría de las variables, esto sugiere la necesidad de emplear métodos no lineales para capturar interacciones más complejas.

Resumen estadístico:					
	PM2.5	PM10	O3	NO2	SO2
count	5040.000000	5040.000000	5040.000000	5040.000000	5040.000000
mean	159.337504	87.927774	51.015296	37.208575	7.120732
std	63.697822	36.620171	29.045699	7.469400	3.472351
min	50.000000	25.010000	1.020000	24.010000	1.000000
25%	103.905000	56.580000	25.512500	30.877500	4.130000
50%	160.040000	87.240000	50.485000	37.340000	7.190000
75%	214.497500	119.177500	76.872500	43.522500	10.120000
max	269.940000	152.940000	100.980000	50.000000	13.000000

	CO	Temperatura	Presión	Humedad	Viento
count	5040.000000	5040.000000	5040.000000	5040.000000	5040.000000
mean	14.357718	13.507069	1027.986230	50.409109	3.957200
std	4.271436	5.465236	2.904504	21.680825	1.737101
min	7.000000	4.000000	1023.000000	13.000000	1.000000
25%	10.680000	8.690000	1025.480000	31.457500	2.430000
50%	14.325000	13.550000	1027.980000	49.965000	3.950000
75%	18.000000	18.130000	1030.502500	69.532500	5.470000
max	22.000000	23.000000	1032.990000	88.000000	7.000000

Figura 4.4. Resumen estadístico de las variables de calidad del aire

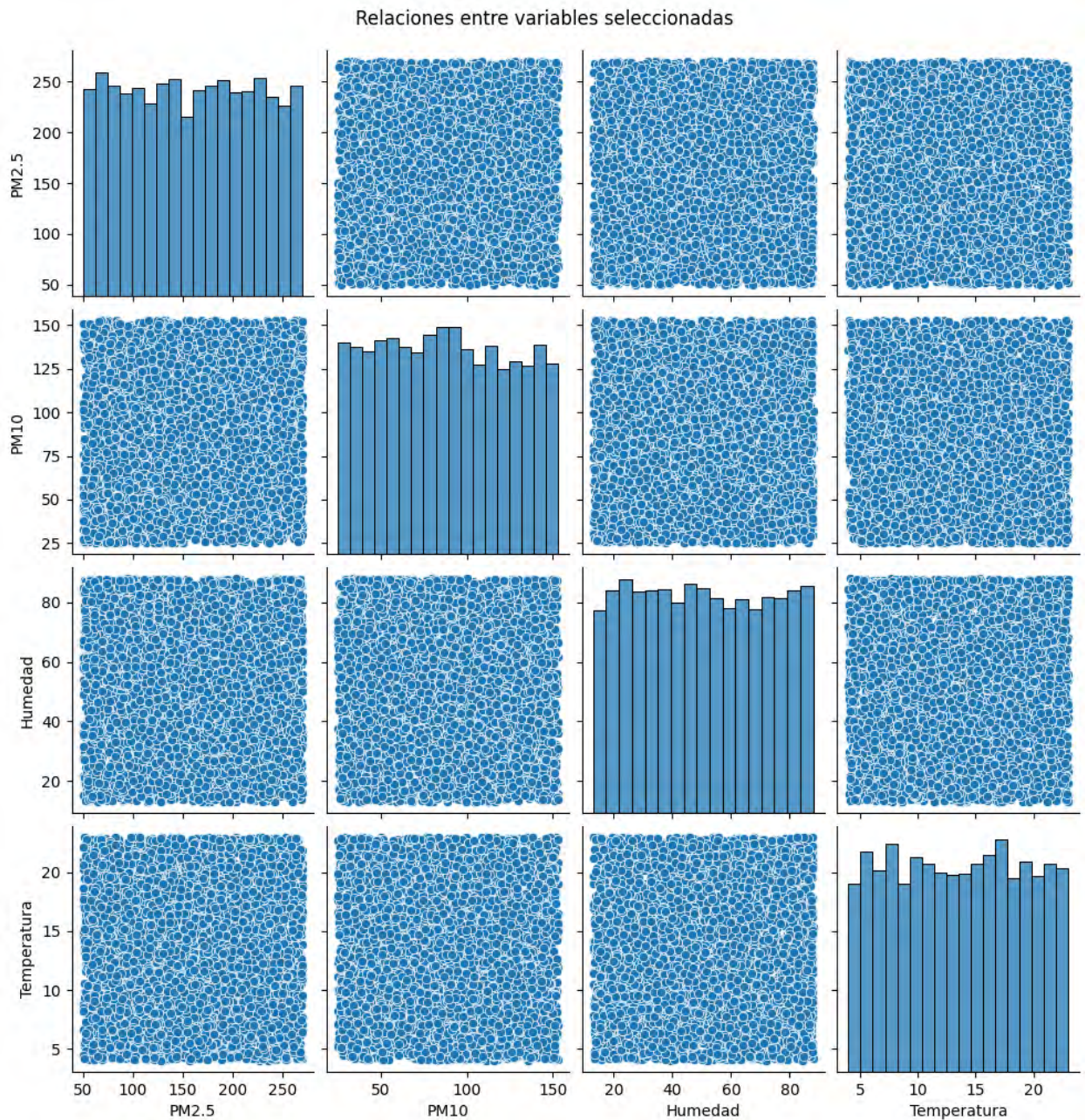


Figura 4.5. Matriz de correlación entre las variables de calidad del aire

Para complementar estos hallazgos, se realizaron gráficos de dispersión y pairplots de variables seleccionadas, específicamente PM2.5, PM10, humedad y temperatura, con el fin de evaluar visualmente la existencia de patrones en la distribución de los datos (Figura 4.6). Posteriormente, se aplicaron técnicas de análisis multivariable, como el Análisis de Componentes Principales (PCA) y el Análisis Multivariado de la Varianza (MANOVA). La matriz de cargas obtenida en el PCA mostró la contribución relativa de cada variable a los componentes principales (Figura 4.7), mientras que el MANOVA

permitió evaluar la existencia de diferencias significativas entre grupos definidos dentro del conjunto de datos (Figura 4.8 y Figura 4.9).

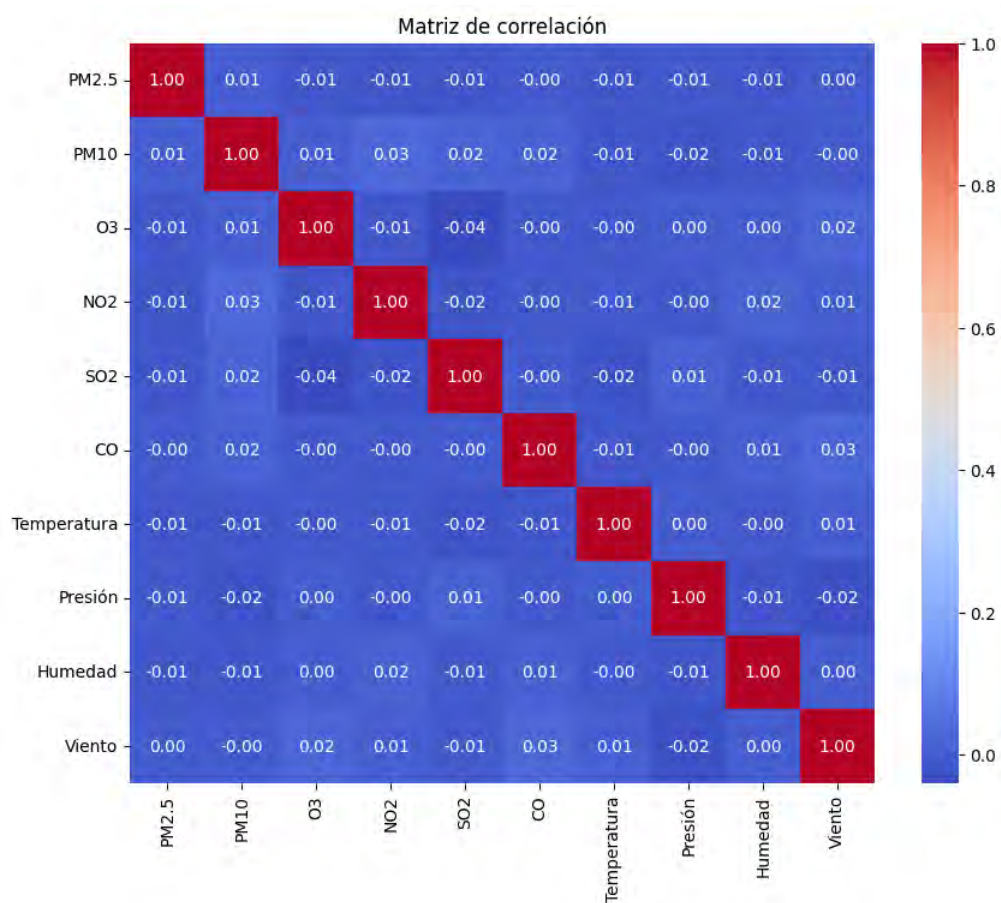


Figura 4.6. *Pairplot de las variables seleccionadas*

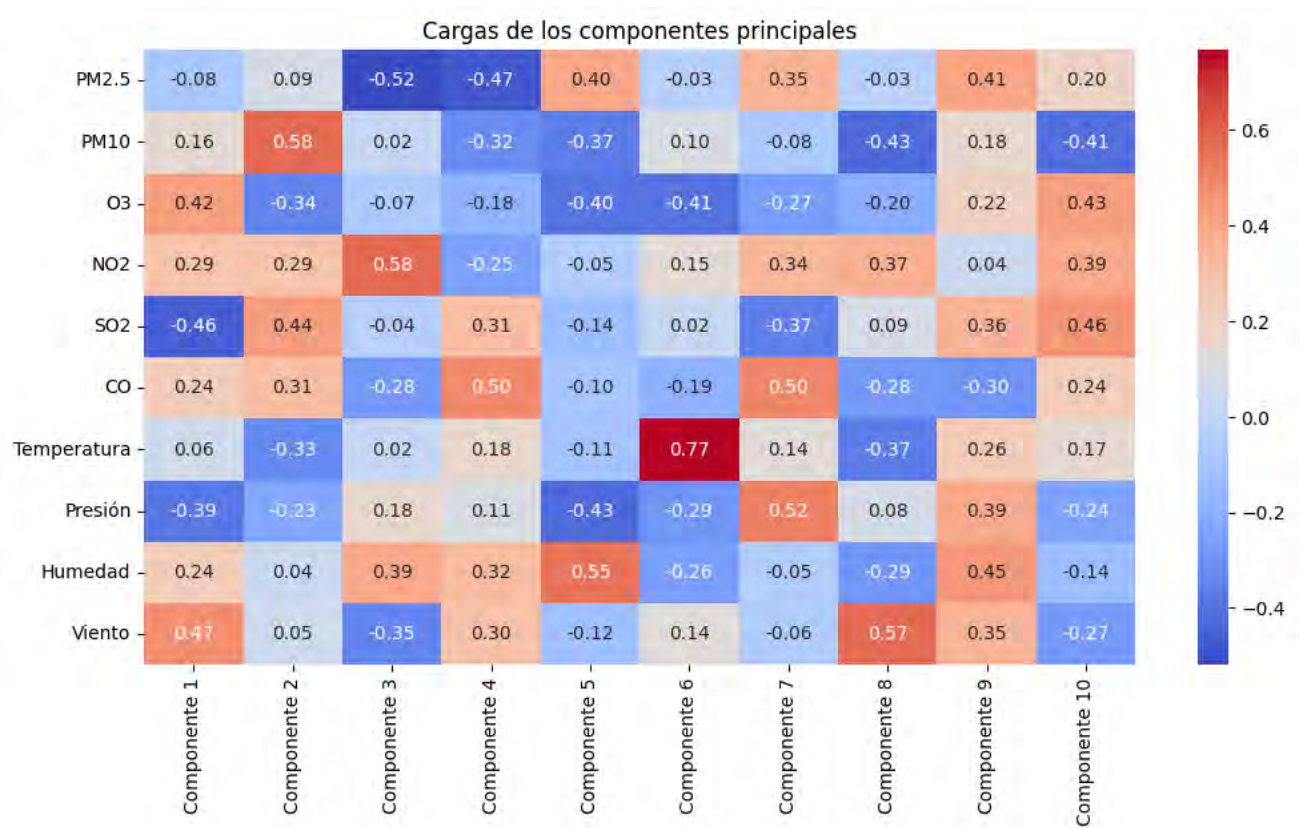


Figura 4.7. Matriz de cargas de los componentes principales que muestra la contribución de cada variable original a los componentes principales

Resultados del MANOVA:

Multivariate linear model

Intercept	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.9982	10.0000	5028.0000	0.8893	0.5423
Pillai's trace	0.0018	10.0000	5028.0000	0.8893	0.5423
Hotelling-Lawley trace	0.0018	10.0000	5028.0000	0.8893	0.5423
Roy's greatest root	0.0018	10.0000	5028.0000	0.8893	0.5423

Grupo	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.9968	20.0000	10056.0000	0.8092	0.7051
Pillai's trace	0.0032	20.0000	10058.0000	0.8091	0.7052
Hotelling-Lawley trace	0.0032	20.0000	8504.9831	0.8093	0.7049
Roy's greatest root	0.0027	10.0000	5029.0000	1.3400	0.2025

Figura 4.8. Resumen de los resultados del MANOVA

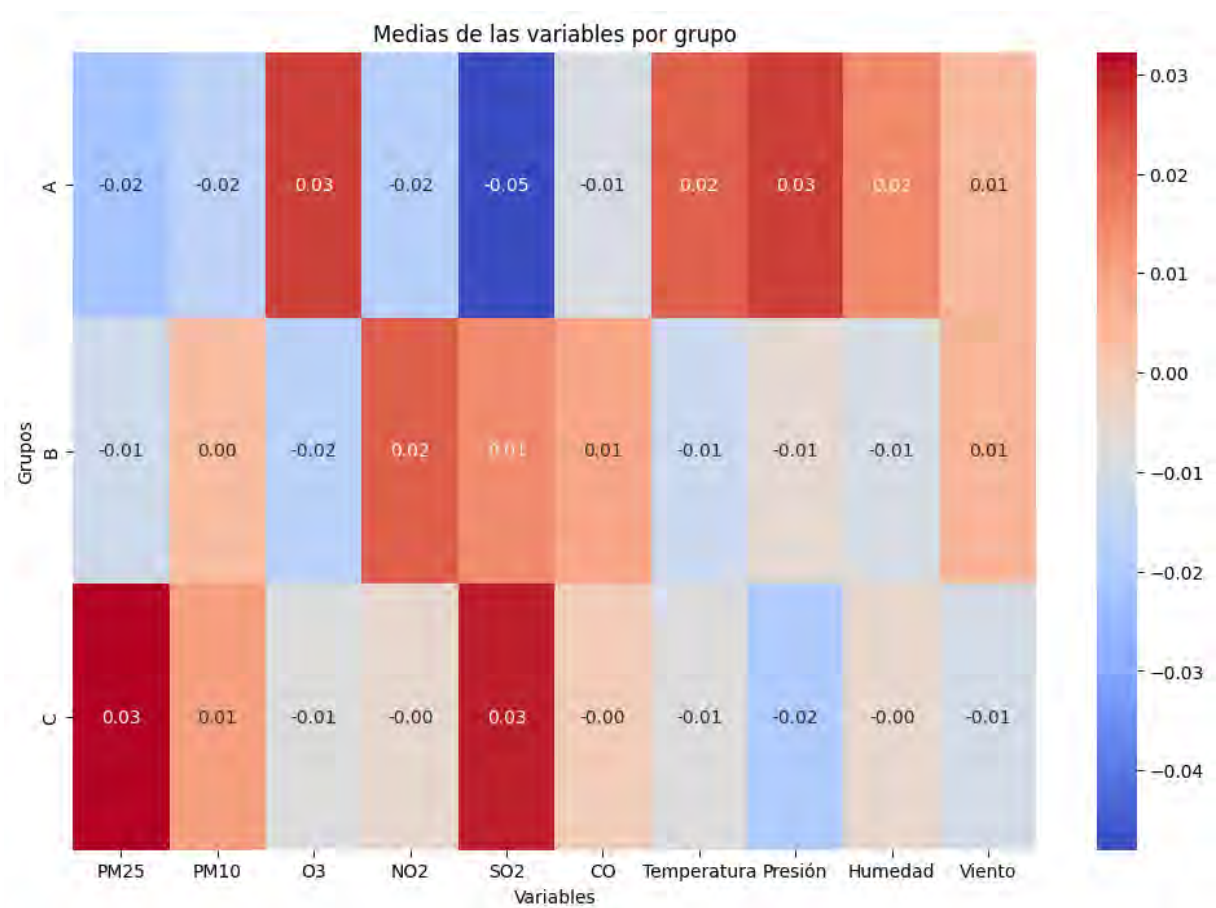


Figura 4.9. Mapa de calor que representa las medias estandarizadas de las variables del aire por grupo (A, B y C)

Se empleó el Modelo de Ecuaciones Estructurales (SEM) con el objetivo de examinar las relaciones entre variables meteorológicas (humedad, presión, temperatura y velocidad del viento) y diversos contaminantes atmosféricos (CO, SO₂, NO₂, O₃, PM₁₀ y PM_{2.5}). Esta técnica permite modelar simultáneamente relaciones directas e indirectas entre múltiples variables observadas, lo cual resulta especialmente adecuado para sistemas complejos donde las interacciones no son necesariamente lineales ni unidimensionales. El modelo fue estimado utilizando el método de máxima verosimilitud, y fue validado con indicadores de ajuste global (como CFI y RMSEA), los cuales arrojaron valores dentro de los rangos aceptables, lo que respalda la validez del modelo especificado. Previo a la estimación, se realizó un análisis exploratorio para verificar la adecuación de los datos, y posteriormente se identificaron rutas de influencia relevantes entre factores meteorológicos y contaminantes.

Los resultados se muestran en la Figura 4.10, donde se representan los coeficientes estandarizados de las relaciones estimadas, junto con sus respectivos valores p. Entre los hallazgos destacados, se observa una relación positiva y estadísticamente significativa entre la temperatura y el CO (coeficiente = 0.025, p = 0.007), lo que indica que el incremento de temperatura puede favorecer la concentración de este contaminante, posiblemente por su efecto en la actividad fotoquímica y la movilidad

vehicular. Asimismo, la presión atmosférica muestra una relación negativa significativa con el ozono (coeficiente = -0.027, $p = 0.03$), evidenciando su papel regulador en la dispersión vertical de contaminantes.

También se identificó una asociación positiva significativa entre NO_2 y O_3 (coeficiente = 0.030, $p = 0.03$), reflejando el papel del dióxido de nitrógeno en la formación de ozono troposférico. Aunque otras rutas no alcanzaron significancia estadística, como las relacionadas con humedad y partículas PM, sus signos sugieren tendencias importantes. Por ejemplo, la humedad muestra una relación negativa con PM_{10} y $\text{PM}_{2.5}$, consistente con la hipótesis de que una mayor humedad favorece la deposición de partículas suspendidas.

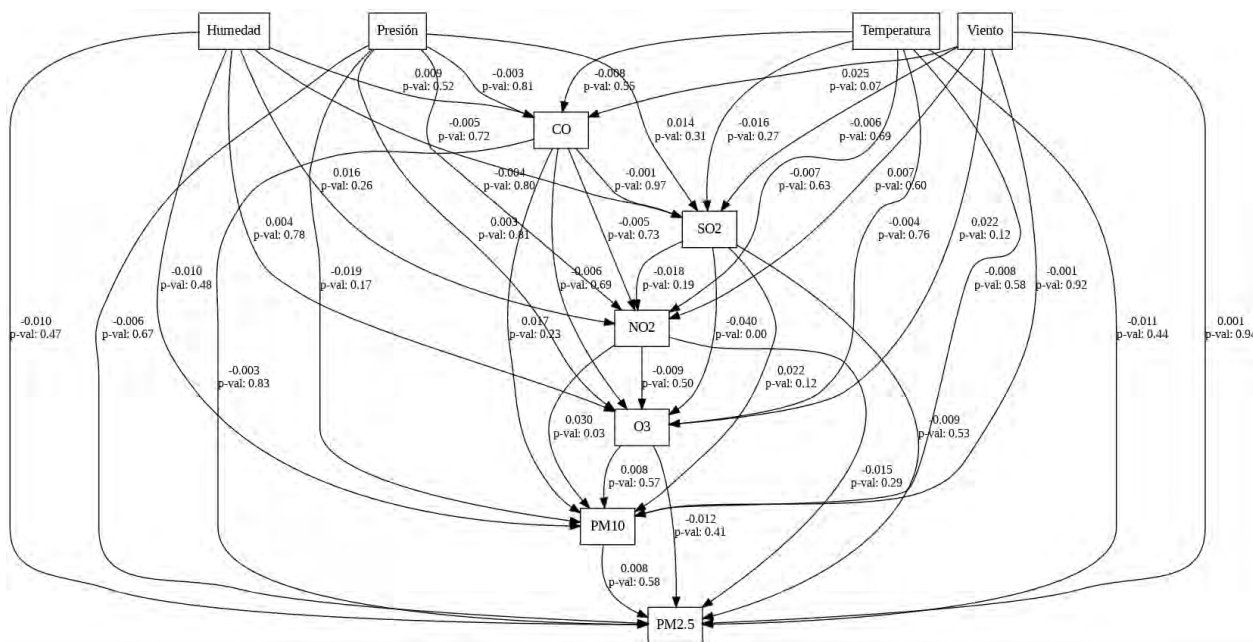


Figura 4.10. Diagrama estructural del Modelo de Ecuaciones Estructurales (SEM) que ilustra las relaciones estimadas entre variables meteorológicas y contaminantes

A partir de los resultados obtenidos, se identificaron las variables con mayor impacto en la calidad del aire, destacando $\text{PM}_{2.5}$, PM_{10} , NO_2 , O_3 , temperatura, presión atmosférica y humedad relativa. Para asegurar la correcta selección de estas variables en el modelado, se aplicaron técnicas de regresión LASSO y Random Forest, cuyos resultados fueron consistentes en señalar a $\text{PM}_{2.5}$ y PM_{10} como los principales indicadores de contaminación (Figura 4.11 y Figura 4.12).

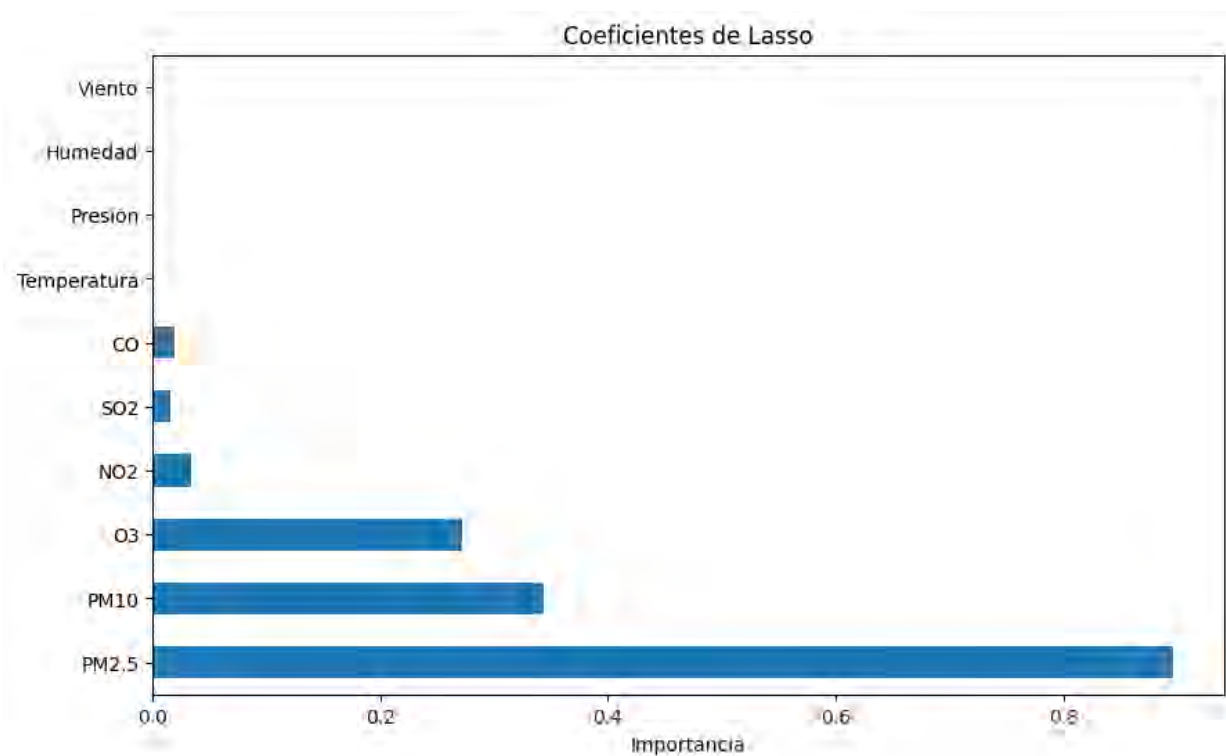


Figura 4.11. Distribución de los coeficientes obtenidos mediante la regresión LASSO, destacando a PM2.5 como la variable de mayor importancia, seguida por PM10 y O3

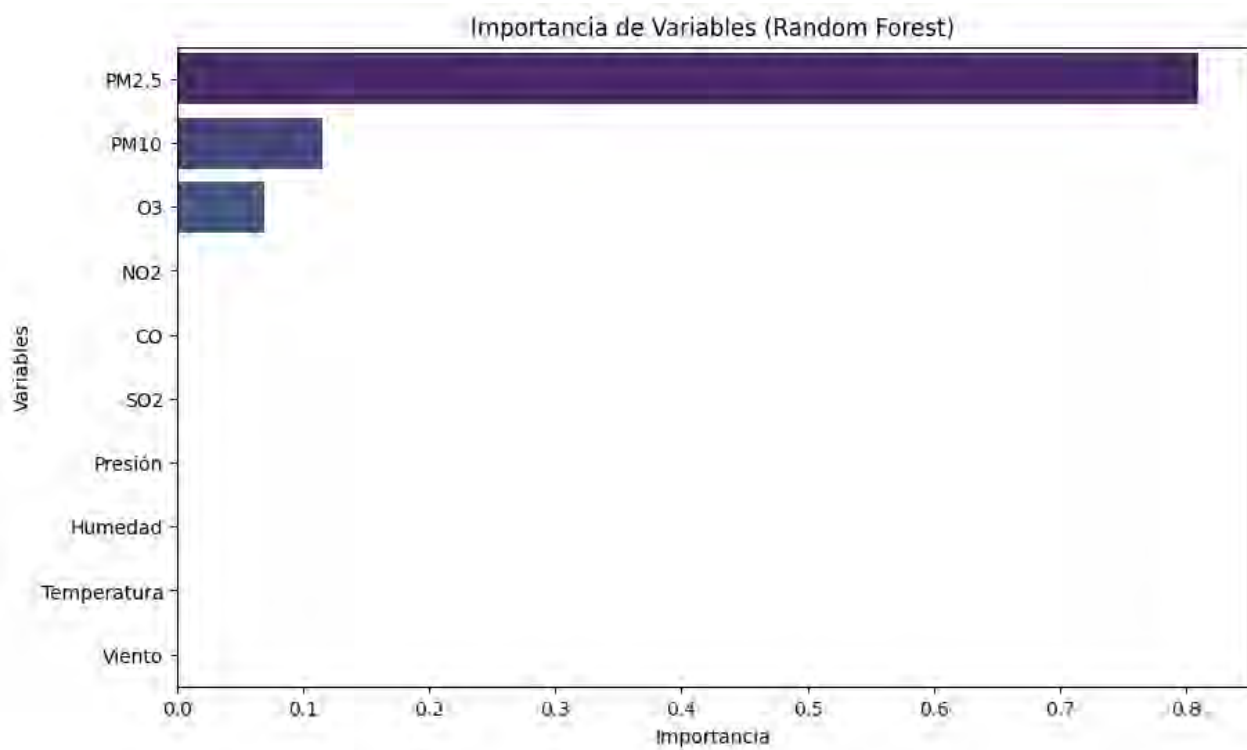


Figura 4.12. *Importancia relativa de las variables determinadas por el modelo Random Forest, en el que PM2.5 tiene la mayor influencia, seguida por PM10 y O3*

La fase de análisis contextual y de la dinámica del sistema ha permitido establecer un marco de referencia para la comprensión de la calidad del aire en la Ciudad de México como un sistema complejo altamente dinámico. La identificación de los factores que intervienen en la calidad del aire, tanto en términos de contaminantes atmosféricos como de variables meteorológicas, ha sido primordial para estructurar el modelo de análisis.

La descripción estadística de las mediciones, ha permitido cuantificar la variabilidad en la calidad del aire, evidenciando patrones y tendencias relevantes. Asimismo, el análisis de correlaciones entre variables, ha mostrado que los contaminantes no presentan relaciones lineales significativas con las variables meteorológicas, lo que indica la necesidad de emplear modelos avanzados basados en técnicas no lineales para capturar interacciones complejas. Además, la aplicación de métodos multivariados, como el Análisis de Componentes Principales (PCA) y el Análisis Multivariado de la Varianza (MANOVA), ha sido significativo para determinar la contribución relativa de cada variable a la dinámica del sistema. Por otro lado, la integración del Modelo de Ecuaciones Estructurales (SEM) ha permitido una representación detallada de las relaciones subyacentes entre contaminantes y condiciones atmosféricas.

Fase 2: Agregación multinivel y medición de la dinámica compleja

En esta fase, se abordó la agregación multinivel y la medición de la dinámica compleja del sistema de calidad del aire en la Ciudad de México. Este enfoque permitió establecer los niveles jerárquicos adecuados para el análisis del sistema, facilitando la comprensión de las interacciones y la transferencia de información entre distintos niveles espaciales. Se definieron criterios específicos para la selección de los niveles de agregación apropiados, asegurando que las mediciones reflejaran con precisión las condiciones locales y regionales de calidad del aire.

El análisis incluyó la identificación de agregación que abarcan desde zonas urbanas específicas hasta regiones metropolitanas amplias, considerando criterios como la cobertura geográfica y la distribución espacial de las estaciones de monitoreo ambiental. Bajo esta perspectiva, se estableció un nivel de agregación óptimo para la recopilación y análisis de datos, tomando en cuenta la influencia de fuentes emisoras relevantes, como el tráfico vehicular, la actividad industrial y las condiciones topográficas, en la distribución espacial de los contaminantes atmosféricos (ver Figura 4.1). La selección adecuada de estos niveles de agregación permitió garantizar que los datos obtenidos fueran representativos y útiles para el modelado del sistema, maximizando la calidad informativa sin distorsionar la dinámica espacial del fenómeno. Además, se realizó un análisis de incertidumbre, componente fundamental en el estudio de sistemas complejos dinámicos, ya que la variabilidad inherente en los datos puede afectar directamente la precisión, estabilidad y confiabilidad de los modelos generados.

Se cuantificó la incertidumbre asociada a cada variable mediante una evaluación comparativa basada en entropía de Shannon, lo que permitió identificar aquellas variables con mayor grado de variabilidad e influencia estocástica. Los resultados, presentados en la Tabla 4.2, indicaron que la variable PM2.5 presentó el valor de entropía más alto (12.077546), lo que sugiere un comportamiento altamente variable y complejo, resultado probablemente de la interacción simultánea de múltiples factores, tales como las emisiones vehiculares, la quema de biomasa, las condiciones meteorológicas (viento, humedad relativa, inversión térmica), y fenómenos de resuspensión de polvo en zonas urbanas densas. Por el contrario, la variable correspondiente a la velocidad del viento mostró la menor entropía (9.145596), lo que indica una menor variabilidad en su comportamiento y, por ende, una mayor previsibilidad. Esta menor variabilidad puede explicarse por la estabilidad general de los patrones de circulación atmosférica a escala regional, especialmente en comparación con los procesos más erráticos asociados a otras fuentes de emisión.

Estos análisis permitieron caracterizar la complejidad del sistema de calidad del aire. Los hallazgos obtenidos en esta fase fueron esenciales para el desarrollo de estrategias de modelado más precisas y robustas en las siguientes etapas del estudio.

Tabla 4.2. Nivel de incertidumbre de las variables del conjunto de datos de aire

Variable	Entropía
Viento	9.145596
Presión	9.809236
SO2	10.061120
CO	10.325877
Temperatura	10.600189
NO2	10.909382
Humedad	11.728907
O3	11.835681
PM10	11.963626
PM2.5	12.077546

La fase de agregación multinivel y medición de la dinámica compleja del sistema ha permitido identificar los niveles de agregación para analizar la calidad del aire en la Ciudad de México, respaldada por criterios de granularidad espacial, ha facilitado la descomposición del sistema en zonas urbanas, suburbanas y regionales, garantizando que la información recopilada represente adecuadamente la dinámica ambiental en diferentes escalas.

Fase 3: Recopilación y monitoreo continuo de datos en tiempo real

En esta fase, se desarrolló e implementó un sistema automatizado para la recopilación, almacenamiento y monitoreo en tiempo real de datos relacionados con la calidad del aire en la Ciudad de México. El objetivo principal fue la obtención de información actualizada que permitiera analizar las concentraciones de contaminantes, identificar tendencias y detectar posibles anomalías en las condiciones ambientales.

Se diseñó un sistema de almacenamiento en tiempo real, basado en una base de datos NoSQL, específicamente MongoDB, para garantizar la correcta estructuración y gestión de la información recopilada por las estaciones de monitoreo. Este diseño permite manejar datos generados continuamente, por sensores ambientales y asegurar su disponibilidad tanto para análisis detallados como para su visualización dinámica. Para el diseño de este sistema de almacenamiento, se estableció un proceso estructurado y sistemático, el cual se detalla en la Figura 4.13.



Figura 4.13. Diagrama del flujo de procesos para el diseño de sistemas de almacenamiento en tiempo real

Adicionalmente, se implementó un proceso de recopilación continua de datos, mediante la automatización de solicitudes a una API de calidad del aire, lo que permite la extracción sistemática de información en tiempo real desde las estaciones de monitoreo de la Ciudad de México (Figura 4.14). Para ello, se desarrollaron implementaciones en Python que ejecutan solicitudes periódicas, ver en Figura 4.15, formatean los datos obtenidos y los almacenan con una estructura organizada en la base de datos, facilitando la eficiencia del flujo de información.

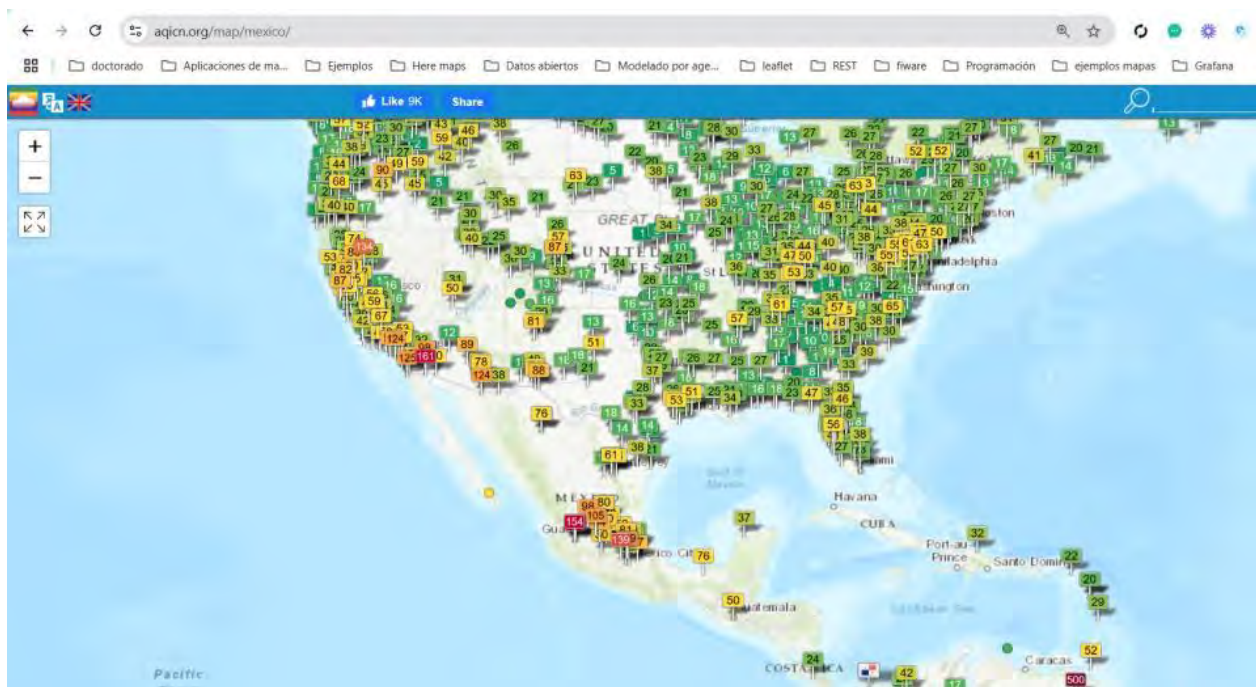


Figura 4.14. Mapa que muestra las estaciones de monitoreo de calidad del aire en México y otros países

```

1 import requests
2
3 def fetch_air_quality_data():
4     API_URL = "https://api.waqi.info/feed/merced/?token=tu_api_key"
5     try:
6         response = requests.get(API_URL)
7         if response.status_code == 200:
8             data = response.json()
9             if data["status"] == "ok":
10                 return data["data"]
11             else:
12                 return None
13         else:
14             return None
15     except Exception:
16         return None

```

Figura 4.15. Fragmento de código en Python que utiliza un endpoint de la API de AQICN para solicitar datos de calidad del aire en tiempo real

Una vez recopilados, los datos fueron estructurados y almacenados en la base de datos utilizando un modelo optimizado que asegura una organización eficiente (ver Figura 4.16). Esta estructura, diseñada para la consulta y el análisis posterior, permite una recuperación ágil y precisa de la información, como se muestra en la Figura 4.17. La implementación de este esquema de almacenamiento asegura la integridad y

disponibilidad de los registros, optimizando su procesamiento para futuras evaluaciones y el modelado de la calidad del aire.

```
1 {
2   "status": "ok",
3   "data": {
4     "aqi": 85,
5     "idx": 137,
6     "attributions": [
7       {
8         "url": "https://www.aire.cdmx.gob.mx/",
9         "name": "Secretaría del Medio Ambiente - Ciudad de México"
10      },
11      {
12        "url": "https://waqi.info/",
13        "name": "World Air Quality Index Project"
14      }
15    ],
16    "city": {
17      "geo": [19.42847, -99.12766],
18      "name": "Merced, Ciudad de México",
19      "url": "https://aqicn.org/city/mexico/mexico/merced/"
20    },
21    "iaqi": {
22      "pm25": {"v": 45},
23      "pm10": {"v": 72},
24      "o3": {"v": 19},
25      "no2": {"v": 26},
26      "so2": {"v": 8},
27      "co": {"v": 10}
28    },
29    "time": {
30      "s": "2025-01-07 10:00:00",
31      "tz": "+06:00",
32      "v": 1673085600,
33      "iso": "2025-01-07T10:00:00+06:00"
34    }
35  }
36 }
```

Figura 4.16. Estructura de los datos en formato JSON proporcionados por API de AQICN

```

{
  "_id": ObjectId("64d5f3b2734e2b2e2c3d8d12"),
  "iaqi": {
    "pm25": { "v": 45 },
    "pm10": { "v": 72 },
    "o3": { "v": 19 },
    "no2": { "v": 26 },
    "so2": { "v": 8 },
    "co": { "v": 10 }
  }
}

```

Figura 4.17. Datos de calidad del aire en organizados bajo la estructura de MongoDB

La implementación de un sistema automatizado de recopilación, almacenamiento y monitoreo en tiempo real representa un avance en la gestión de datos ambientales para el análisis de la calidad del aire en la Ciudad de México. La integración de una arquitectura basada en bases de datos NoSQL, en particular MongoDB, ha permitido el procesamiento eficiente de grandes volúmenes de información con mínima latencia, garantizando una disponibilidad continua para su análisis y visualización en tiempo real.

Fase 4: Preprocesamiento y actualización dinámica de datos

En esta fase, se implementaron estrategias de preprocesamiento de datos con el objetivo de asegurar la calidad, integridad y disponibilidad de la información recopilada en tiempo real sobre la calidad del aire en la Ciudad de México. Para ello, se establecieron procesos de actualización dinámica y muestreo eficiente que permiten mantener una sincronización constante con los registros más recientes. Los datos fueron estandarizados a un rango entre 0 y 1, lo que garantiza su uniformidad y facilita el análisis posterior (ver Figura 4.18).

	PM2.5	PM10	O3	NO2	SO2	CO	Temperatura	Presión	Humedad	Viento	ICA
0	0.374648	0.762526	0.588435	0.569065	0.613333	0.494667	0.957368	0.107107	0.273467	0.811667	0.504630
1	0.950987	0.591964	0.557023	0.035398	0.945833	0.472667	0.348316	0.583584	0.008000	0.408333	0.803208
2	0.732200	0.192058	0.655862	0.513274	0.675000	0.579333	0.867895	0.848849	0.970533	0.610000	0.601623
3	0.598800	0.667005	0.496999	0.609465	0.782500	0.219333	0.959474	0.954955	0.824267	0.455000	0.601744
4	0.156043	0.623622	0.031913	0.190843	0.366667	0.526000	0.943158	0.861862	0.615600	0.903333	0.221711

Figura 4.18. Conjunto de datos de aire estandarizado entre 0 y 1

Se diseñó un esquema de preprocesamiento en tiempo real que permite la detección y corrección de errores, la eliminación de valores atípicos y el manejo de datos faltantes. Este enfoque permite que la información procesada sea confiable y representativa de las condiciones ambientales actuales. La aplicación de técnicas de transformación de datos, incluyendo normalización, cálculo del z-score y transformación cúbica, contribuyó significativamente a mejorar la distribución de las variables, optimizando su uso en modelos analíticos y predictivos. Las ecuaciones (4.1), (4.2), (4.3) y (4.4) fueron aplicadas en este proceso, asegurando que los datos sean coherentes y estén listos para su correspondiente análisis.

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (4.1)$$

Donde:

x_{norm} es el valor original que se desea normalizar.

$x_{\min}$ es el valor mínimo en el conjunto de datos original.

$x_{\max}$ es el valor máximo en el conjunto de datos original.

$$\text{Variable normalizada} = \sqrt{(\text{Variable original})} \quad (4.2)$$

$$z = \frac{x - \mu}{\sigma} \quad (4.3)$$

Donde:

x Es el valor individual de la variable.

μ Es la media de la distribución de la variable.

σ Es la desviación estándar de la distribución de la variable.

$$\text{Variable normalizada} = \sqrt[3]{(\text{Variable original})} \quad (4.4)$$

Esta fase ha establecido una infraestructura de datos en tiempo real eficiente, asegurando que la información procesada sea precisa, representativa y adecuada para la detección de patrones dinámicos y eventos críticos en la calidad del aire. La combinación de automatización de la actualización, muestreo optimizado y transformación de datos avanzados refuerza la solidez del sistema y su aplicabilidad en modelos predictivos.

Fase 5: Modelado iterativo y desarrollo continuo

En esta fase, se desarrolló y refinó un modelo computacional adaptativo para caracterizar el comportamiento de la calidad del aire en la Ciudad de México. La implementación de un enfoque basado en programación genética permitió la construcción de un modelo matemático que permita integrar múltiples variables ambientales y contaminantes en una representación simbólica optimizada y aproximarse al comportamiento de la calidad del aire. A través de un proceso iterativo, se logró mejorar la precisión del modelo, ajustándose dinámicamente con base en nuevos datos y condiciones emergentes del entorno urbano.

La construcción del modelo comenzó con la identificación y selección de variables, tal como se muestra en la Tabla 4.3, donde se definieron los contaminantes atmosféricos y las variables meteorológicas que tienen un impacto significativo en la calidad del aire. Como complemento, se establecieron los parámetros del programa evolutivo, los cuales fueron cuidadosamente configurados para optimizar la generación de expresiones simbólicas y mejorar la precisión del modelo, como se presenta en la Tabla 4.4.

Tabla 4.3. Descripción de variables del conjunto de datos de aire

Id	Variable	Descripción
X0	PM2.5	Partículas en suspensión con diámetro menor o igual a 2.5 micrómetros
X1	PM10	Partículas en suspensión con diámetro menor o igual a 10 micrómetros.
X2	O3	Ozono, un gas reactivo que forma parte del smog.
X3	NO2	Dióxido de nitrógeno, generado por la quema de combustibles fósiles.
X4	SO2	Dióxido de azufre, emitido principalmente por procesos industriales.
X5	CO	Monóxido de carbono.
X6	Temperatura	Medida del calor en el aire.
X7	Presión	Presión atmosférica medida en hectopascales (hPa).
X8	Humedad	Porcentaje de vapor de agua en el aire.
X9	Viento	Velocidad y dirección del viento.
X10	ICA	Resumen de la calidad del aire en un valor numérico.

Tabla 4.4. Parámetros utilizados en el programa evolutivo

PARÁMETRO	VALOR
Conjunto De Datos	Aire con 24,192 registros y 10 variables
Número De Generaciones	1,000
Población	10,000
Número De Ejecuciones	30
Tournament Size	2000

Posteriormente, se definieron funciones primitivas como add, sub, mul, div, sqrt, log, abs, neg, inv, sin, cos, tan, arcsin, arccos, arctan, sinh, cosh, tanh, square, cube, exp, mod, pow, clip, step, relu, heaviside, sigmoid, min, max, mean, round, floor y ceil. adecuadas para representar la relación entre las variables, permitiendo la generación de expresiones simbólicas que mejor reflejan la dinámica del sistema. Entre las múltiples expresiones obtenidas, se seleccionó la mejor ecuación $f(X) = \sqrt{X_0 \cdot \sin(\sin(\sqrt{X_1 + X_2}))} + \cos(X_5 + X_6)$, con un MSE de 0.006401768 y un MAE de 0.06239935 con base en métricas de error como el Error Cuadrático Medio (MSE) y el Error Absoluto Medio (MAE), cuyos valores más bajos indicaron un alto grado de ajuste del modelo (Tabla 4.5).

Tabla 4.5. Expresiones matemáticas generadas mediante un algoritmo evolutivo para el comportamiento de la calidad del aire

Función	MSE	MAE
$\sqrt{X_0} \cdot \sin\left(\sin\left(\sqrt{X_1 + X_2}\right)\right) + \cos(X_5 + X_6)$	0.006401768	0.06239935
$\frac{\sqrt{X_0} \cdot \sqrt{\sin\left(\sin\left(X_1 + X_2 + X_3\right)\right)} + X_6 + X_7}{\sin(0.727 \cdot X_0 + 0.173 + X_3) \cdot \log(X_9 + X_4 + 1)}$	0.007548552	0.068885555
$0.745351374856182 \cdot \sqrt{X_0 + X_6} + \cos\left(\frac{X_4 \cdot X_8}{X_2 + 1}\right)$	0.00787591	0.07277676
$0.742967024840268 \cdot \sqrt{X_0 + X_7} - \tan(X_9 + X_{10})$	0.009025975	0.077446104
$0.737 \cdot \sqrt{X_0 + X_2} - (X_{7^2} + X_4)$	0.00906162	0.07758607
$\sqrt{X_0 \cdot \sqrt{X_1 + X_4}} - \exp(-X_3 - X_7)$	0.009196394	0.077689239
$\sin\left(\sin(\sin(X_0 + 0.141 + X_{10}))\right) - \cosh(X_4 + X_7)$	0.009085067	0.077874681
$\sin\left(\sin(\sin(X_0 + 0.127 + X_2))\right) \cdot (X_2 + X_6)$	0.009111996	0.077905121
$0.719104 \cdot \sqrt{X_0 + X_8} + \sin(X_3 + X_5)$	0.009380821	0.0788256
$0.712741187248218 \cdot \sqrt{X_0 + X_3}$	0.009571044	0.079538172
$\frac{1 + X_{10} + X_5}{0.83 \cdot \sin(\sqrt{X_0 + X_4}) + \tan(X_8 + X_9)}$	0.010322026	0.082878153
$\sin(X_0 + X_1) + \cos(X_1 \cdot X_9 \cdot X_3)$	0.014563309	0.096469904
$\sin(X_0 + X_5) \cdot \log(X_2 + X_6 + 1)$	0.014563309	0.096469904
$0.684835746730557 \cdot \sqrt{-(X_0 + X_8)} + \left(\frac{X_5}{1 + X_6 + X_9}\right)$	0.284681434	0.495242576

El modelo fue evaluado mediante su comparación con el Índice de Calidad del Aire (ICA), aplicando la ecuación (4.5), lo que permitió evidenciar una alta correlación con la tendencia de los datos reales. Aunque presenta buen desempeño general, el análisis reveló discrepancias en ciertos puntos extremos, observadas durante la revisión de los resultados visuales y cuantitativos (ver Figura 4.19 y Tabla 4.6.). Estas diferencias pusieron de manifiesto la necesidad de realizar ajustes adicionales en los parámetros del modelo para optimizar su precisión y capacidad predictiva. Con este fin, se llevó a cabo un proceso sistemático de ajuste mediante GridSearchCV, evaluando múltiples combinaciones de parámetros evolutivos con el propósito de minimizar el error y mejorar el rendimiento general del modelo.

$$ICA_P = \frac{I_{high} - I_{low}}{C_{high} - C_{low}} \times (C - C_{low}) + I_{low} \quad (4.5)$$

Donde:

ICA_P es el Índice de Calidad del Aire para el contaminante

C es la concentración medida del contaminante

C_{low} y C_{high} son los límites inferiores y superiores de la concentración del contaminante

I_{low} y I_{high} son los valores del ICA correspondientes a C_{low} y C_{high}

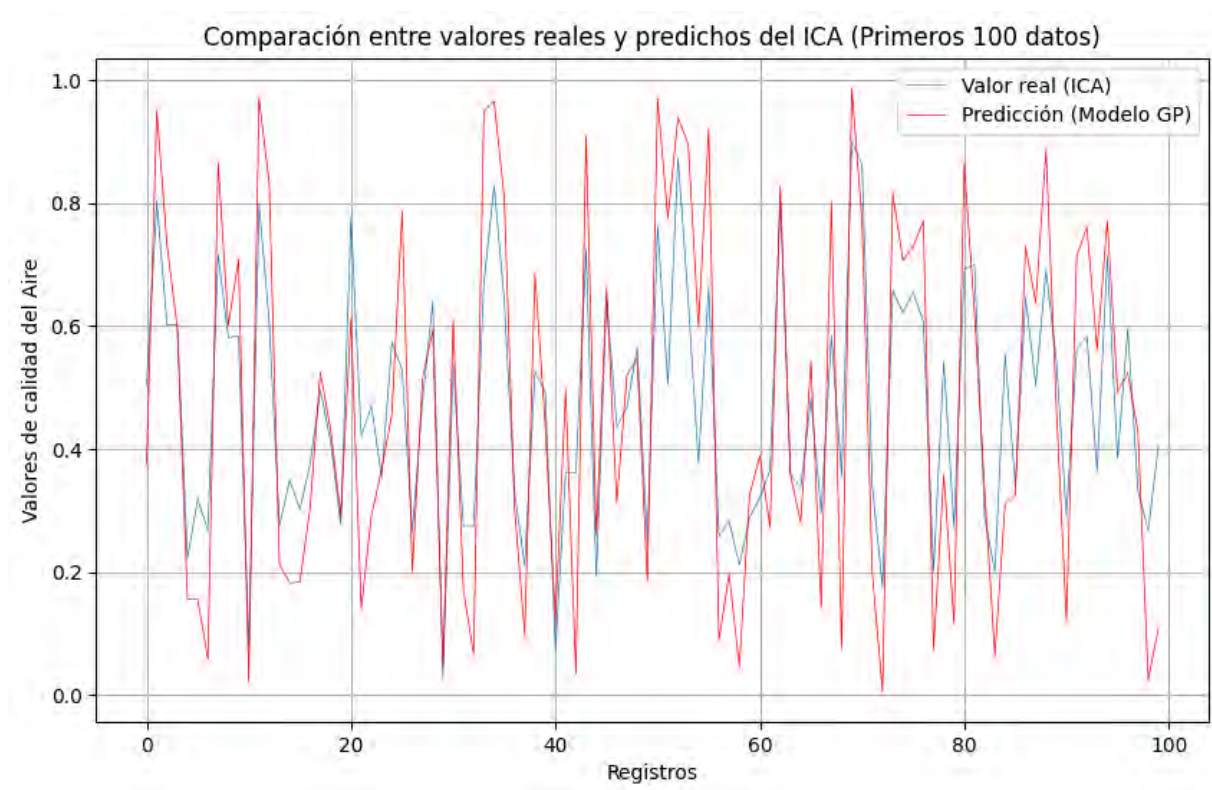


Figura 4.19. Comparación entre los valores reales del Índice de Calidad del Aire (ICA) y las predicciones generadas por el modelo de programación genética (GP)

Tabla 4.6. Evaluación del desempeño del modelo mediante métricas de error y ajuste

Métrica	Valor
MAE	0.005312939
MSE	0.05244191

Como se muestra en la Tabla 4.7, la evaluación del modelo refinado reflejó una reducción significativa en los errores de predicción, con un MAE de 0.0049 y un MSE de 0.0487, consolidando un enfoque altamente eficiente para la estimación de la calidad del aire. Estos resultados subrayan la relevancia de la optimización adaptativa en modelos de aprendizaje evolutivo, permitiendo realizar ajustes dinámicos y progresivos que mejoran continuamente la precisión, estabilidad y capacidad explicativa del sistema.

Tabla 4.7. Evaluación del desempeño del modelo ajustado

Métrica	Valor
MAE	0.0049
MSE	0.0487

Esta fase ha establecido un proceso para la modelización de la calidad del aire mediante técnicas de programación genética. La combinación de aprendizaje

evolutivo, ajuste dinámico de parámetros y validación con respecto a otro modelo de la literatura ha permitido desarrollar un modelo computacional robusto y adaptable. Estos avances sientan las bases para futuras mejoras en la predicción de la calidad del aire, fortaleciendo la toma de decisiones basada en evidencia y consolidando un enfoque de computación evolutiva aplicada a sistemas ambientales complejos.

Fase 6: Interpretación, representación e implementación continua

En esta fase, se llevó a cabo la interpretación, representación e implementación continua de los resultados obtenidos a lo largo del estudio sobre la calidad del aire en la Ciudad de México. El objetivo central fue transformar los datos procesados en información comprensible y útil para la toma de decisiones en la gestión ambiental. Para ello, se analizaron los patrones y tendencias identificados, evaluando la influencia de diversos factores ambientales y urbanos en la dinámica de la contaminación.

Uno de los aspectos primordiales de esta fase fue la representación visual del modelo, que facilitó una exploración intuitiva de la relación entre los contaminantes atmosféricos y las variables meteorológicas. La generación de mapas de calor de correlaciones permitió identificar interacciones significativas entre diversos contaminantes, destacando una alta correlación positiva entre PM_{2.5} y PM₁₀, lo que indica que sus concentraciones tienden a aumentar o disminuir de forma conjunta (ver Figura 4.20). Además, el análisis mostró que el Índice de Calidad del Aire (ICA) está fuertemente influenciado por la presencia de partículas y gases contaminantes, mientras que variables como la humedad relativa presentaron correlaciones negativas, lo que sugiere su papel en la reducción de partículas en suspensión.

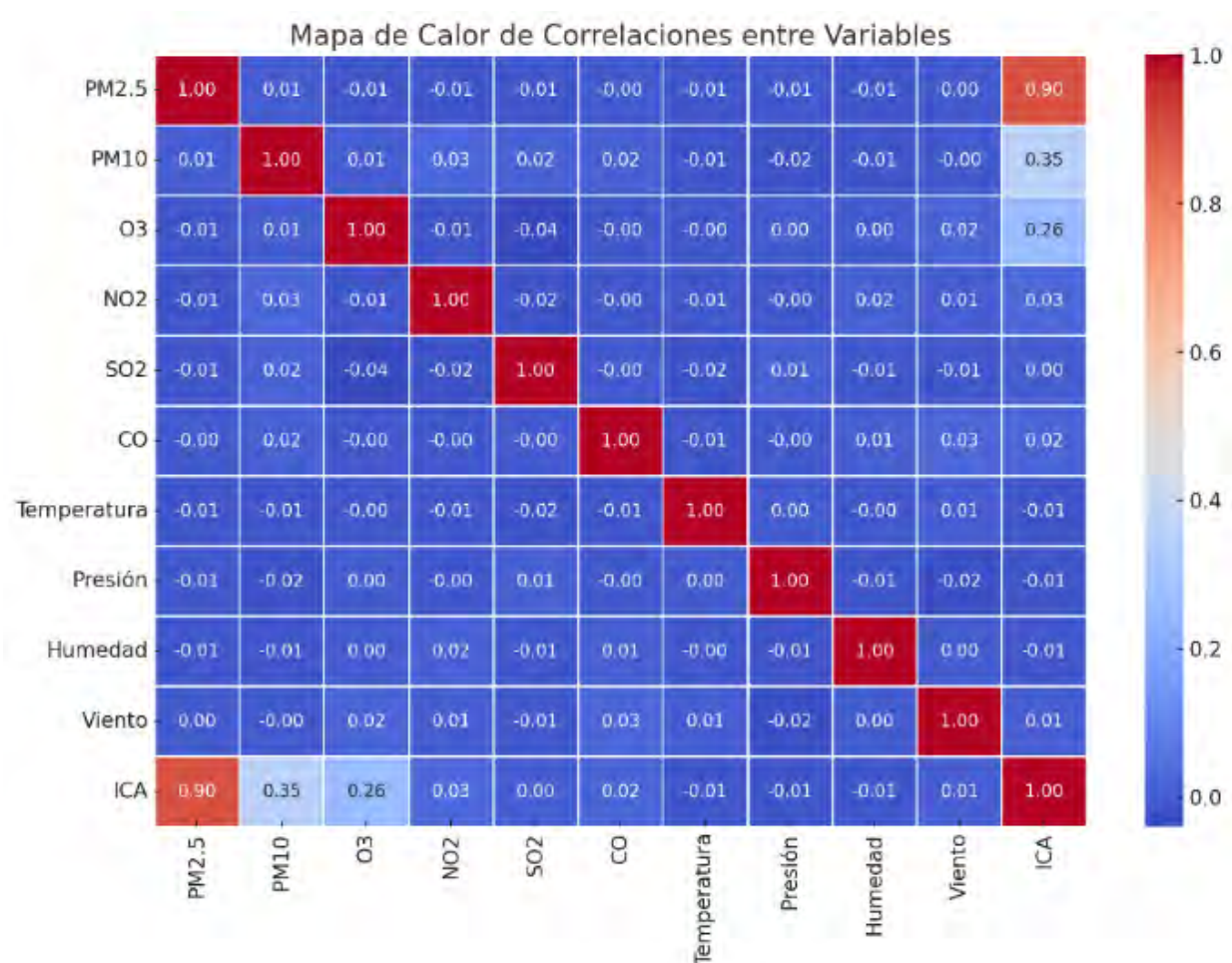


Figura 4.20. Mapa de calor de correlaciones entre variables

Asimismo, en esta fase se modificó el modelo $\sqrt{X0 * \sin(\sin(\sqrt{X1 + X2}))} + \cos(X5 + X6)$ con un enfoque adaptativo, garantizando su actualización y ajuste continuo en función de los datos en tiempo real. Este enfoque dinámico permite la integración de nueva información en el sistema, optimizando su capacidad predictiva y asegurando que el modelo refleje con mayor precisión la evolución de la calidad del aire. El proceso de actualización del modelo se basa en la incorporación de datos recientes, la recalibración de los algoritmos y la optimización de sus parámetros, lo que permite mejorar su desempeño y robustez en entornos dinámicos. Este procedimiento se representa en la Figura 4.21.

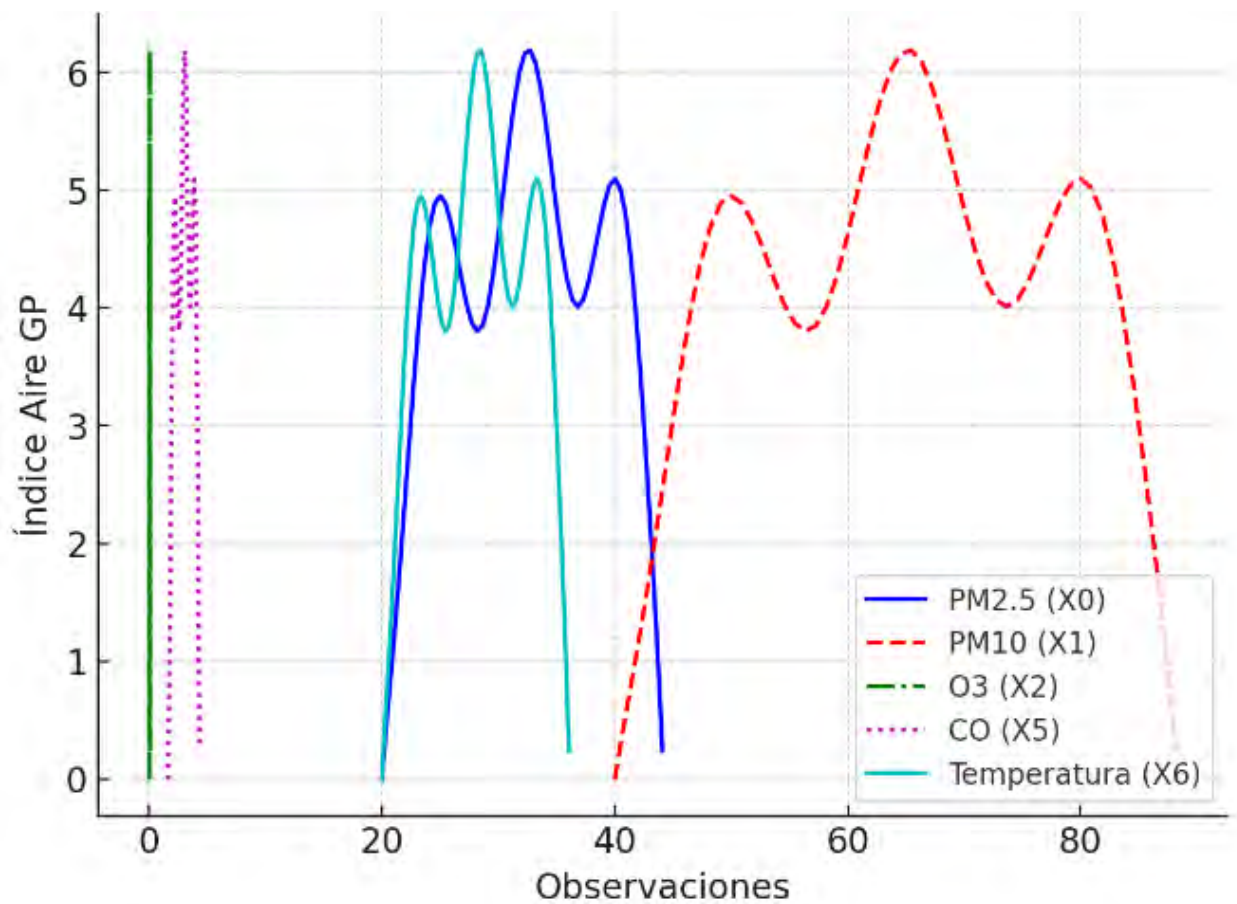


Figura 4.21. *Modelo adaptativo de calidad del aire*

La adaptación del modelo a los datos en tiempo real, donde la línea azul representa las predicciones ajustadas del índice de calidad del aire, mientras que los puntos rojos indican las estimaciones individuales en cada iteración. Este comportamiento dinámico refleja la capacidad del modelo para actualizarse continuamente con base en nuevas mediciones, optimizando su precisión en la estimación de la calidad, como se puede visualizar en la Figura 4.22.

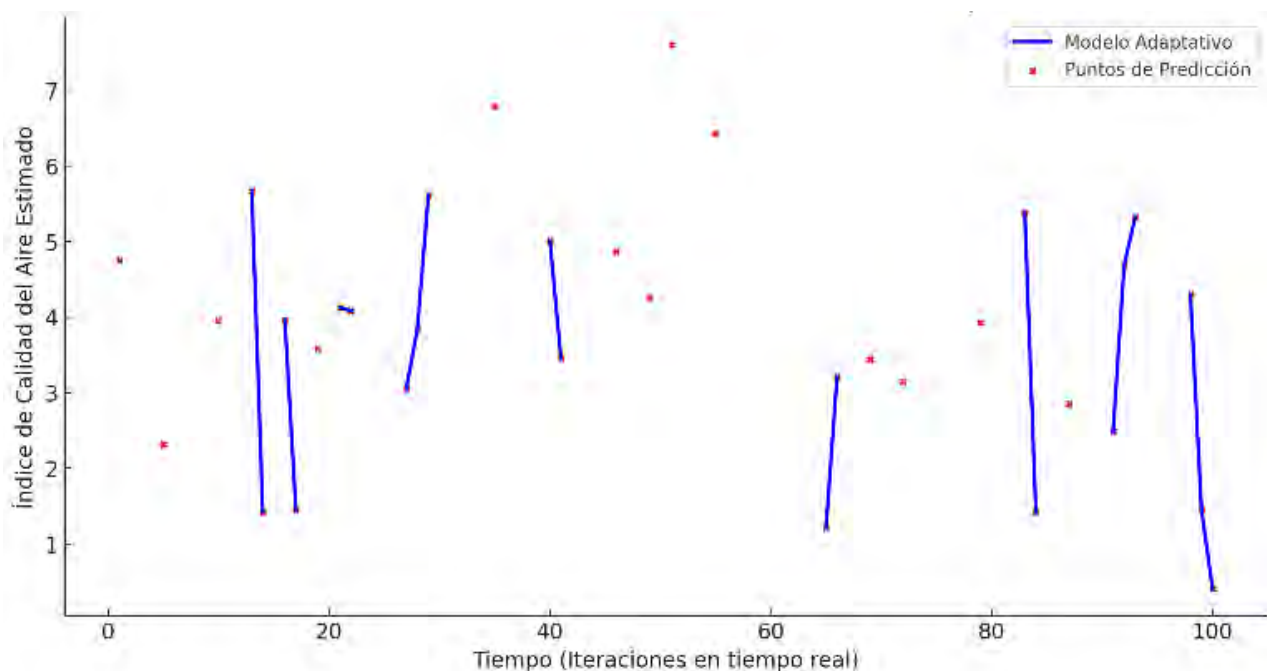


Figura 4.22. Evolución del modelo de calidad del aire en tiempo real

Se desplegó el índice de calidad del aire de manera interactiva como se muestra en la Figura 4.23, permitiendo la consulta dinámica de información y mejorando la accesibilidad para distintos sectores involucrados en la gestión ambiental. Esta integración facilita la detección temprana de cambios en la calidad del aire y permite una respuesta rápida ante episodios críticos de contaminación.

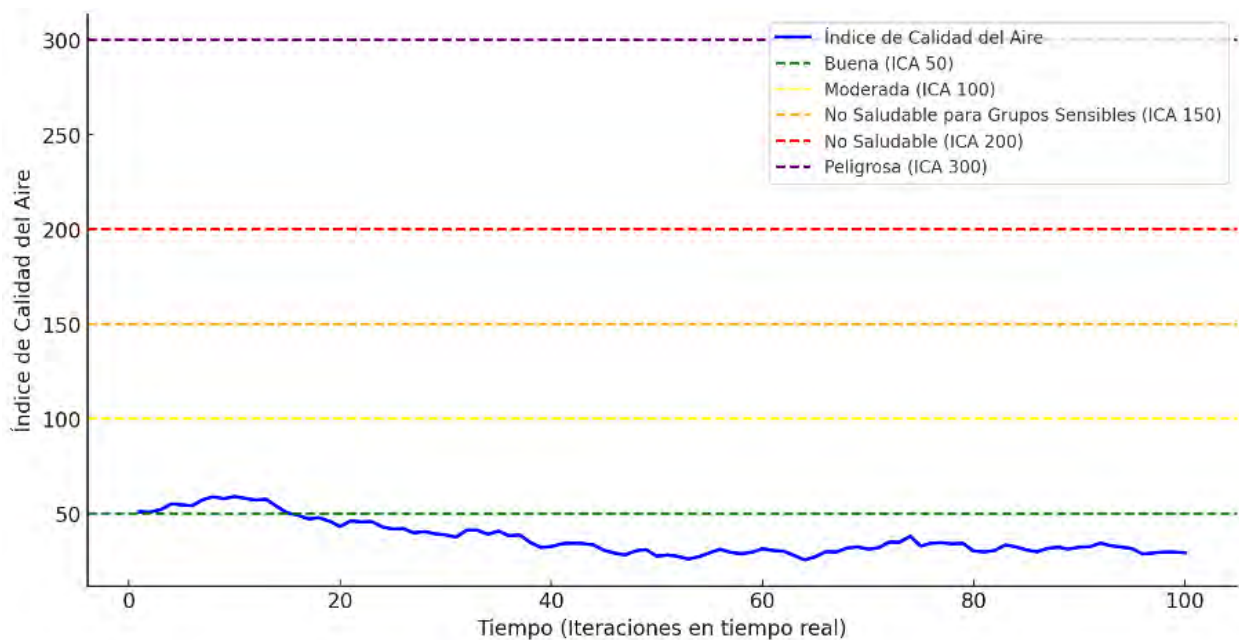


Figura 4.23. Monitoreo del Índice de Calidad del Aire en Plataforma Interactiva

La fase de interpretación, representación e implementación continua permitió una visualización precisa y estructurada de la dinámica de la calidad del aire, al tiempo que optimizó la capacidad del modelo para actualizarse y adaptarse en tiempo real. La integración de técnicas de visualización de datos, implementación y recalibración dinámica fortaleció la robustez del modelo, facilitando la captura precisa de la variabilidad y los patrones emergentes en la contaminación atmosférica. Este enfoque adaptativo mejora significativamente su capacidad de aproximación al comportamiento complejo de la calidad del aire, asegurando una representación más fiel de los procesos ambientales y proporcionando herramientas analíticas más eficaces para la toma de decisiones informadas en la gestión de la contaminación.

4.3 Caso de estudio 2: Profundidad de los sismos en México

México es un país con una alta actividad sísmica a partir de su ubicación en el Cinturón de Fuego del Pacífico, donde interactúan diversas placas tectónicas, incluyendo la placa de Cocos, la placa de Norteamérica, la placa del Pacífico y la placa del Caribe. Esta interacción da lugar a sismos de diversas magnitudes y profundidades, lo que convierte el análisis de la profundidad sísmica en un factor necesario para comprender la dinámica geológica del país y su impacto en la infraestructura y la población.

El estudio de la profundidad de los sismos es esencial, ya que determina la forma en que las ondas sísmicas se propagan y afectan la superficie terrestre. Sismos superficiales (< 70 km) suelen generar más daños en estructuras y edificaciones, mientras que sismos profundos (> 300 km) pueden tener una menor percepción en la superficie, pero ofrecen información valiosa sobre la dinámica del manto terrestre.

Fase 1: Análisis contextual y de la dinámica del sistema

Para el caso de estudio sobre la profundidad de los sismos en México, el primer paso fue la comprensión del sistema sísmico como un sistema complejo y dinámico, en el que múltiples factores interactúan para determinar la profundidad de los sismos. En el contexto de México, este fenómeno está influenciado principalmente por la interacción de placas tectónicas, destacando la subducción de la placa de Cocos bajo la placa de Norteamérica, lo que da lugar a sismos de diversas profundidades.

Para realizar la comprensión del problema de la profundidad de los sismos como un problema bajo la perspectiva

Los sismos en la zona de subducción del Pacífico tienden a ser más profundos que los sismos en fallas corticales dentro del territorio continental. De acuerdo a la literatura los factores que influyen en la profundidad pueden ser diversos, pero se puede considerar la composición del manto, la rigidez de las rocas y la energía liberada en el hipocentro determinan la propagación de las ondas sísmicas.

Este análisis inicial es vital para reconocer que la profundidad sísmica no es un fenómeno aislado, sino el resultado de múltiples interacciones geofísicas y tectónicas. La identificación de entidades y relaciones dinámicas permiten identificar los

componentes trascendentales del sistema sísmico y sus interacciones. Para el caso de estudio, las principales entidades y sus relaciones incluyen:

Placas tectónicas: Determinan la localización y el tipo de sismos, hipocentro del sismo: Punto donde ocurre la liberación de energía, determinando la profundidad, Epicentro: Relacionado con la distribución de daños y percepción del sismo en la superficie, Magnitud del sismo: Puede influir en la propagación de ondas y el impacto en la superficie, Propagación de ondas sísmicas: Depende de la profundidad y de la composición del medio geológico y Estaciones sismológicas: Proveen datos en tiempo real sobre los eventos sísmicos y permiten el modelado de la dinámica sísmica.

Se definieron las variables para el análisis y modelado de la profundidad sísmica, seleccionadas con base en su relevancia teórica y empírica dentro de la geofísica y el estudio de sistemas tectónicos. Estas variables fueron elegidas por su capacidad para explicar o influir en la variabilidad observada en la profundidad de los eventos sísmicos, permitiendo una representación más completa de su comportamiento espacial y temporal. Entre ellas se incluyen la magnitud del evento, la ubicación geográfica (latitud y longitud), el tipo de evento sísmico, la zona tectónica y el tiempo de ocurrencia. La elección de estas variables está respaldada por estudios recientes que muestran su influencia en la caracterización de sismos profundos y su relación con procesos de subducción, estructuras litosféricas y patrones de acoplamiento tectónico (Zhao et al., 2001; Graham et al., 2021; Qin et al., 2022). Su incorporación permite construir modelos que capturan de manera realista la dinámica sísmica, integrando tanto la dimensión espacial como temporal del fenómeno. Para este caso, algunas variables principales incluyen:

- Profundidad del sismo (D): Variable dependiente principal, medida en kilómetros.
- Magnitud (M): Para evaluar si existe correlación entre magnitud y profundidad.
- Latitud y longitud (L): Ubicación del sismo para analizar patrones espaciales.
- Tipo de evento sísmico (T): Diferenciar sismos superficiales, intermedios y profundos.
- Zona tectónica (Z): Si el sismo ocurre en una zona de subducción o en el interior del continente.
- Tiempo de ocurrencia (T_{sismo}): Para analizar tendencias temporales en la profundidad sísmica.

Fase 2: Agregación multinivel y medición de la dinámica compleja

El propósito de la agregación multinivel es identificar patrones emergentes en los datos de profundidad sísmica a través de distintos niveles de resolución espacial y temporal. En este estudio, se emplean datos históricos de sismos registrados en México, considerando variables como latitud, longitud, profundidad (km), magnitud (M_w), fecha y hora del evento, organizados a nivel estatal. Para este caso de estudio, se opta por una agregación a nivel nacional (México) con el fin de desarrollar el índice de profundidad de sismos (ver Figura 4.24).

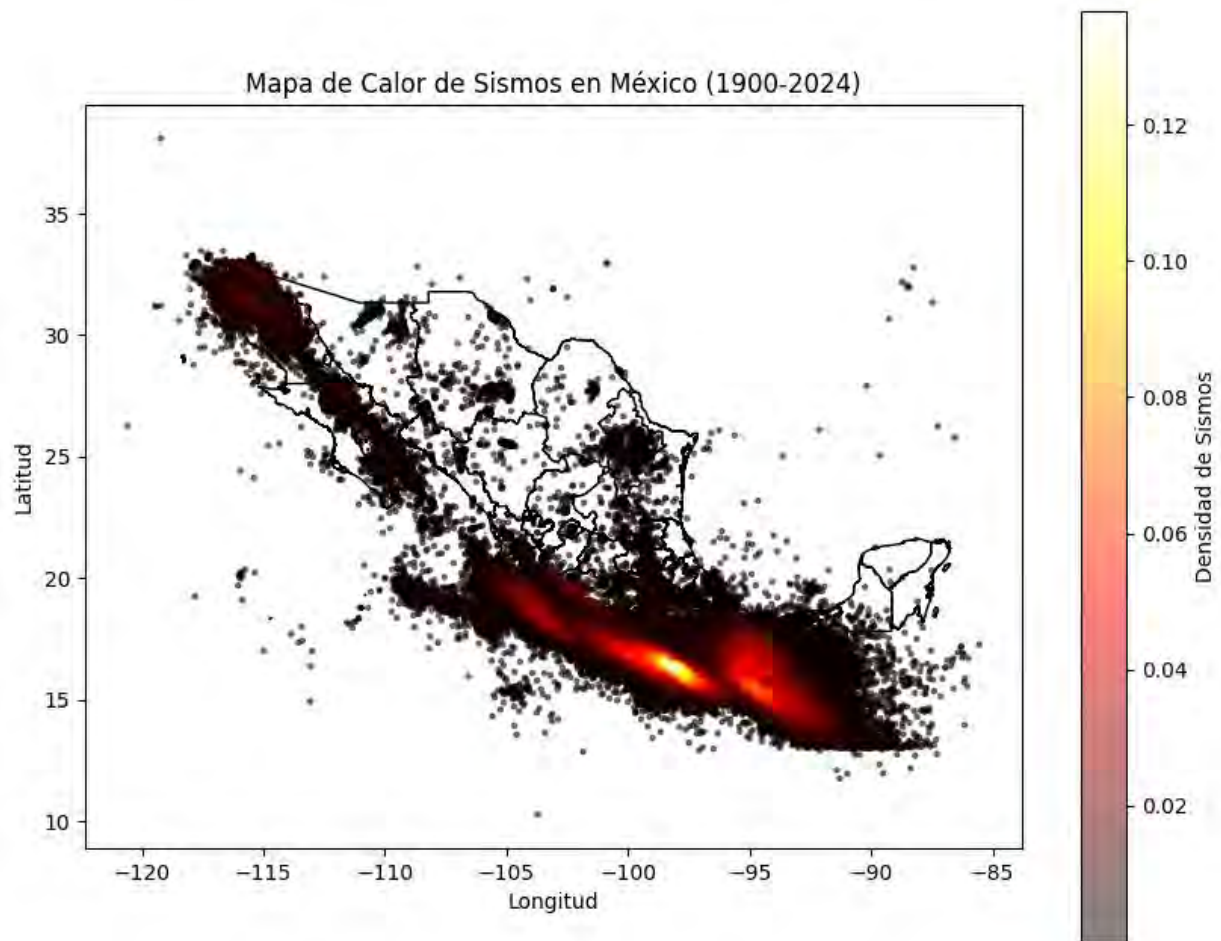


Figura 4.24. Nivel de agregación por país para el análisis de sismos

Fase 3: Recopilación y monitoreo continuo de datos en tiempo real

Los datos utilizados en este estudio provienen del Servicio Sismológico Nacional (SSN), accesibles a través de su catálogo en línea (SSN). La base de datos comprende un registro histórico de sismos ocurridos en México desde 1900 hasta 2024, proporcionando una visión de largo plazo sobre la actividad sísmica en el país. Abarca información de todos los estados de México, con un total de 294,900 registros organizados en 12 variables, incluyendo latitud, longitud, profundidad, magnitud (Mw), fecha y hora del evento, entre otras. Este extenso conjunto de datos permite analizar la evolución de la actividad sísmica a lo largo de más de un siglo, facilitando la identificación de tendencias, patrones emergentes y posibles cambios en el comportamiento tectónico. Un fragmento representativo de estos registros se ilustra en la Figura 4.25, proporcionando un ejemplo de la estructura y el contenido del dataset.

	Magnitud	Latitud	Longitud	Profundidad	Referencia de localizacion	Estatus	dia	mes	anio	hora	minuto	segundo
0	7.4	20.00	-105.00	33.0	71 km al NOROESTE de AUTLAN DE NAVARRO, JAL	1	20	1	1900	0	33	30
1	6.9	20.00	-105.00	33.0	71 km al NOROESTE de AUTLAN DE NAVARRO, JAL	1	16	5	1900	14	12	0
2	6.9	25.00	-110.00	33.0	100 km al NORESTE de LA PAZ, BCS	1	5	3	1901	4	45	0
3	7.0	26.00	-110.00	33.0	83 km al OESTE de AHOME, SIN	1	8	12	1901	20	17	0
4	7.0	17.62	-99.72	33.0	21 km al OESTE de ZUMPANGO DEL RIO, GRO	1	16	1	1902	17	19	0
...	...	...	...	...	...	...	...	...	...	...	...	...
294895	3.5	17.13	-94.11	164.1	63 km al NOROESTE de CINTALAPA, CHIS	2	1	1	2024	21	31	37
294896	4.0	18.20	-100.25	52.1	13 km al SURESTE de ARCELIA, GRO	2	1	1	2024	21	33	4
294897	3.5	27.30	-112.53	24.3	27 km al OESTE de SANTA ROSALIA, BCS	2	1	1	2024	22	6	9
294898	3.1	19.41	-103.86	25.6	21 km al NOROESTE de VILLA DE ALVAREZ, COL	2	1	1	2024	22	33	39
294899	3.3	16.77	-98.71	30.1	34 km al NOROESTE de OMETEPEC, GRO	2	1	1	2024	23	25	49

294900 rows x 12 columns

Figura 4.25. Fragmento del conjunto de datos sísmicos utilizados en el estudio, recopilados por el Servicio Sismológico Nacional (SSN)

Fase 4: Preprocesamiento y actualización dinámica de datos

El preprocesamiento del conjunto de datos sísmicos se centró en garantizar la calidad, coherencia y estructuración adecuada de la información para su análisis posterior. Inicialmente, se llevó a cabo una inspección exhaustiva del dataset, compuesto por 294,900 registros y 12 columnas, que incluyen variables clave como magnitud, ubicación geográfica (latitud y longitud), profundidad, fecha y hora del evento, así como una clasificación en clústeres y una referencia codificada. Se verificó la inexistencia de valores nulos, asegurando así la integridad de los datos. Como parte de la transformación, se consolidaron las variables de fecha y hora en una única columna con formato datetime, lo que optimiza la manipulación y análisis de series temporales. En cuanto a la clasificación en clústeres, se aplicó un algoritmo de agrupamiento no supervisado (como K-means) que permitió segmentar los eventos sísmicos en seis clústeres distintos, definidos con base en patrones espaciales y de profundidad. Esta segmentación facilita el análisis exploratorio de la distribución y comportamiento de los sismos, permitiendo identificar zonas sísmicamente activas con características particulares.

Asimismo, se validaron los rangos y la coherencia de los valores, confirmando que la magnitud oscila entre 0.0 y 7.4, que la profundidad varía entre 1 km y más de 100 km, y que las coordenadas geográficas están dentro de los límites esperados para México. Se realizó un análisis de registros duplicados, sin hallar instancias repetidas, lo que garantiza que cada sismo ha sido registrado de forma única. Además, se estandarizaron las variables categóricas, como estatus y clúster, para asegurar su correcta interpretación en estudios posteriores.

Se aplicaron técnicas de normalización y escalado de datos a las variables continuas, como magnitud, latitud, longitud y profundidad, utilizando el método min-max scaling. Como resultado, se obtuvo un conjunto de datos estructurado para su uso en análisis exploratorios, modelado predictivo y visualización de tendencias sísmicas en México.

El análisis exploratorio del dataset de sismos revela un total de 294,900 registros distribuidos en 12 columnas, que incluyen variables para el estudio de la actividad sísmica. La magnitud, expresada en la escala de Richter, permite evaluar la energía liberada en cada evento. Las coordenadas de latitud y longitud indican la ubicación geográfica precisa de los sismos, facilitando el análisis espacial. La profundidad, medida en kilómetros, aporta información sobre la ubicación del hipocentro en la corteza terrestre. Además, la variable de fecha y hora, desglosada en día, mes, año, hora, minuto y segundo, posibilita el estudio de patrones temporales en la ocurrencia de los eventos. El clúster asignado a cada sismo sugiere una clasificación basada en patrones identificados, lo que puede ser útil para detectar tendencias o comportamientos específicos en la actividad sísmica.

El análisis de estadística descriptiva proporciona una visión detallada de la distribución de los datos sísmicos. En cuanto a la magnitud, se observa un valor medio de 3.42 con una desviación estándar de 0.95, indicando que la mayoría de los sismos son de baja intensidad, con valores comprendidos entre 0.0 y 7.4. La profundidad de los eventos presenta una media de 32.55 km, aunque con una desviación estándar de 36.22 km, lo que refleja una alta variabilidad. El rango varía desde 1 km hasta más de 100 km, evidenciando la presencia de sismos tanto superficiales como profundos. En términos de ubicación geográfica, los valores promedio de latitud (17.76°) y longitud (-98.91°) sugieren que los eventos sísmicos están mayormente concentrados en el centro y sur de México, donde se ubican diversas fallas tectónicas activas. Se contempló la distribución temporal de los sismos ocurridos entre los años 1900 y 2024, con una distribución uniforme a lo largo de los días y meses, lo que indica una recurrencia sísmica constante en el tiempo (ver Figura 4.26).

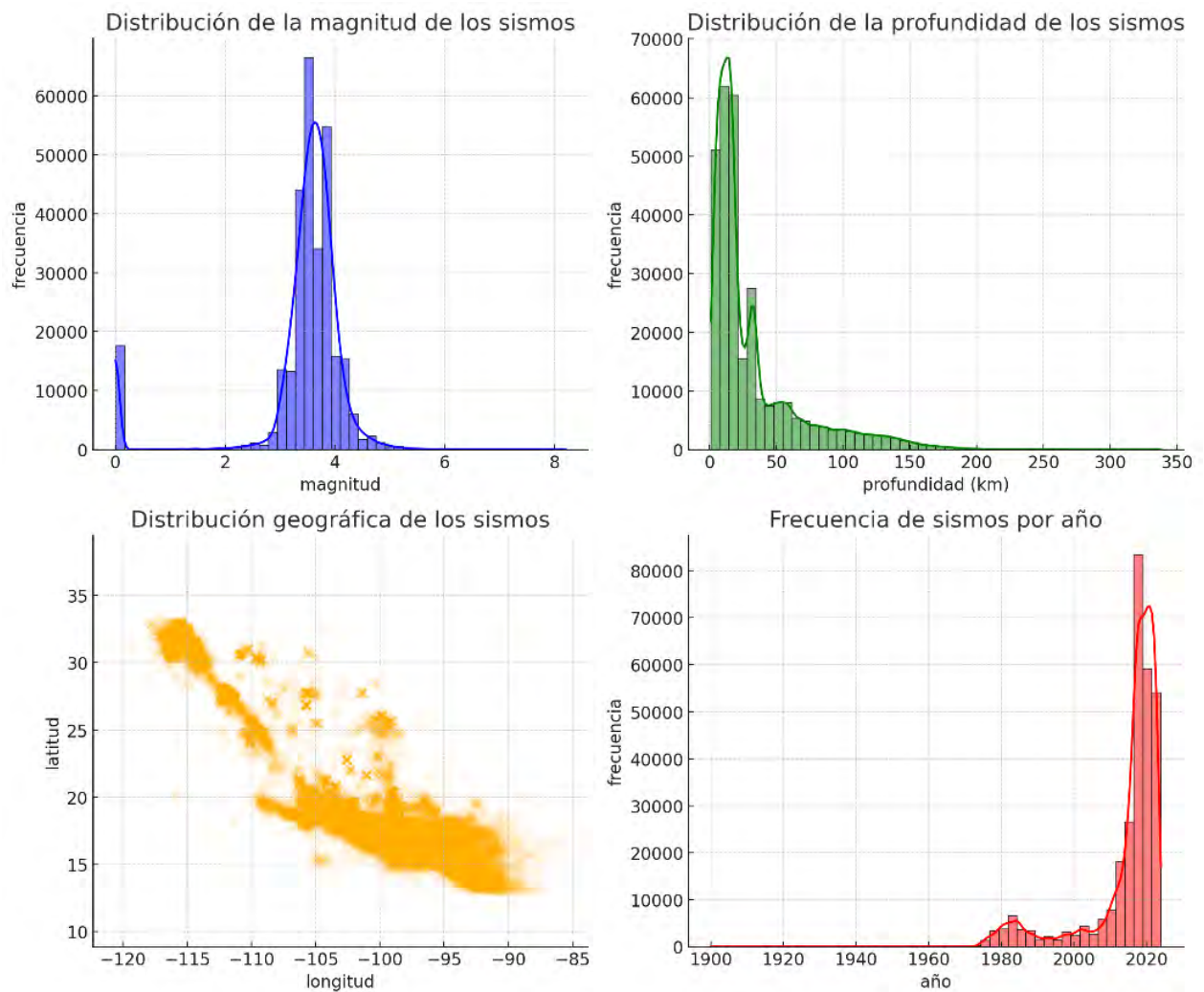


Figura 4.26. Análisis exploratorio de los datos sísmicos en México (1900-2024)

El análisis de correlación entre magnitud, profundidad y ubicación geográfica permite identificar patrones y dependencias en los datos sísmicos. La matriz de correlación revela que la relación entre magnitud y profundidad es débil (0.05), lo que indica que la profundidad a la que ocurre un sismo no influye significativamente en su magnitud, sugiriendo que otros factores, como la dinámica tectónica local, podrían desempeñar un papel más relevante en la generación de eventos de gran intensidad.

Por otro lado, se observa una correlación negativa entre profundidad y latitud (-0.18) y una correlación positiva entre profundidad y longitud (0.36). Esto sugiere que los sismos más profundos tienden a concentrarse en regiones longitudinales específicas, posiblemente en zonas de subducción, mientras que la distribución en latitud presenta cierta variabilidad, lo que podría estar relacionado con la complejidad geológica del territorio mexicano.

La relación entre latitud y longitud muestra una fuerte correlación negativa (-0.90), lo que confirma que la distribución espacial de los sismos sigue un patrón bien definido,

alineado con la estructura de las placas tectónicas en México. Este resultado es consistente con la localización de los principales sistemas de fallas y zonas de subducción, reforzando la hipótesis de que la actividad sísmica en el país está fuertemente condicionada por su configuración geológica. Además, la correlación entre otras variables, como la magnitud y el año de registro, sugiere un aumento en la detección y documentación de eventos en las últimas décadas, lo que podría deberse a mejoras en la instrumentación y la expansión de la red de monitoreo sísmico.

La relación entre referencia codificada y clúster muestra una correlación moderada (0.32), lo que sugiere que la clasificación de los sismos en distintos grupos puede estar parcialmente influenciada por la ubicación geográfica de los eventos. Esto indica que ciertos patrones espaciales podrían estar siendo capturados en la segmentación de los datos, posiblemente a las diferencias en la actividad sísmica entre regiones.

Por otro lado, las variables temporales como día, mes, hora, minuto y segundo presentan correlaciones bajas con el resto de las variables, lo que indica que los sismos ocurren con una distribución relativamente uniforme a lo largo del tiempo. Esto sugiere que la variabilidad en la frecuencia anual de eventos, no hay una dependencia fuerte entre el momento exacto de ocurrencia de un sismo y otras características como su magnitud o profundidad, lo que concuerda con la naturaleza impredecible de la actividad sísmica (ver el mapa de calor de correlación de la Figura 4.27).

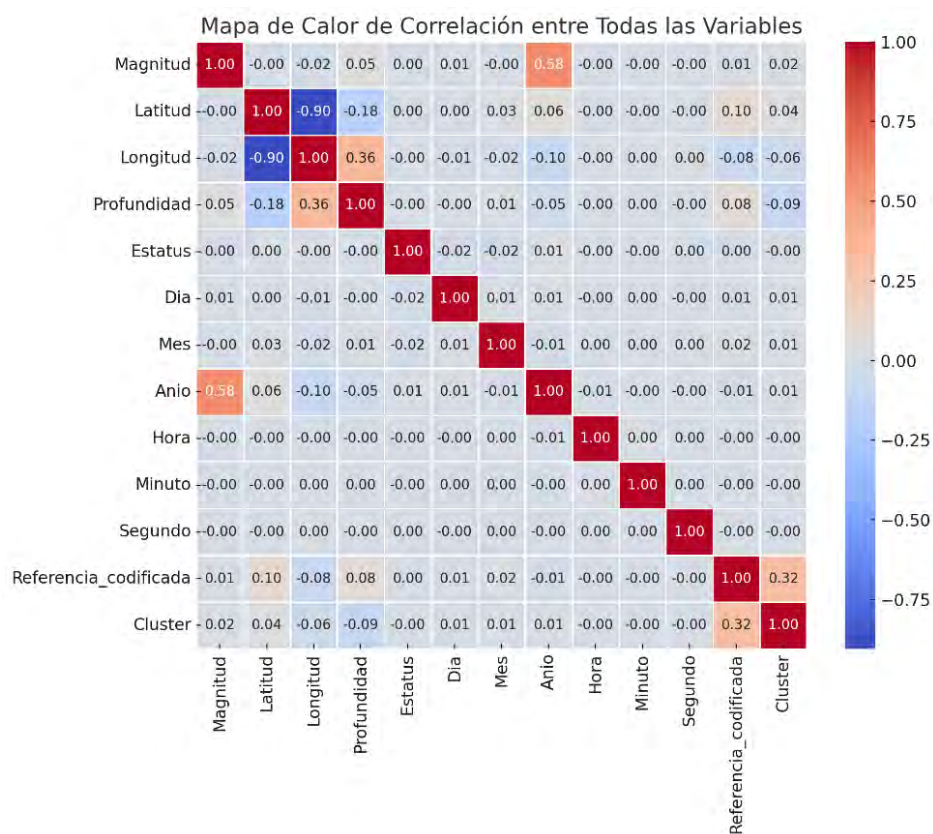


Figura 4.27. Correlación de variables del conjunto de datos de sismos

Se realizó el Análisis de Componentes Principales del conjunto de datos para analizar las variables de mayor impacto en la profundidad de los sismos. La Figura 4.28 muestra las cargas de cada variable en los 13 componentes principales, reflejando la contribución de cada variable en la estructura del conjunto de datos sísmicos. Los valores resaltados en amarillo indican cargas mayores o iguales a 0.5, lo que sugiere una fuerte influencia de la variable correspondiente en la respectiva componente principal. Se observa que la latitud (0.908) y la longitud (-0.957) tienen una alta carga en la Componente 1, indicando que la variabilidad en los datos está fuertemente influenciada por la ubicación geográfica de los sismos. Otras variables como estatus (0.723 en la Componente 9) y referencia codificada (0.809 en la Componente 3) también presentan valores significativos, lo que sugiere su relevancia en la diferenciación de patrones sísmicos.

Variable	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6	Componente 7	Componente 8	Componente 9	Componente 10	Componente 11	Componente 12	Componente 13
Magnitud	0.109255524	-0.884404792	0.017271485	-0.022950616	0.019756131	0.008517965	0.028129286	0.004543185	-0.017779534	-0.018666089	0.012558032	0.450497044	0.005016681
Latitud	0.908929644	0.129293133	-0.082826388	-0.048663321	0.009067301	0.075216637	0.063884648	0.06127407	0.040862673	-0.287546816	0.125735248	0.002212092	0.184831153
Longitud	-0.957620193	-0.108719407	0.113409134	0.013681449	-0.003674127	-0.023399986	-0.020762599	-0.020678962	-0.01329951	0.121419051	-0.048590818	-0.019570857	0.197521953
Profundidad	-0.470626693	-0.118887129	0.176650603	-0.139170167	0.026956951	0.198530333	0.179391827	0.165546815	0.118542344	-0.702422698	0.317146893	-0.065389615	-0.042070819
Estatus	0.006182272	-0.019669874	-0.012936251	0.622529036	0.121200216	0.170288559	0.078170195	0.164382461	0.723653223	0.106156483	0.001925334	0.010086383	0.000180829
dia	0.013379807	-0.027809886	0.031531975	-0.500176491	-0.281256001	-0.246362679	-0.439045868	-0.210691269	0.608139068	-0.032705478	0.004035638	0.003074591	0.000298269
mes	0.042328662	0.033207413	0.070224258	-0.58019937	0.150210539	0.452043321	0.434931729	0.201758763	0.192413406	0.402099832	-0.005777739	-0.005095087	-0.001035981
anio	0.214595079	-0.860455387	-0.050849661	0.009985934	0.007298962	-0.011981482	-0.005085386	-0.007299036	-0.015100376	0.063111938	-0.047369477	-0.451819248	0.003443315
hora	-0.003501905	0.018717549	-0.002748922	-0.087469522	0.537961013	-0.603675544	0.462138859	-0.31282449	0.152824706	-0.05769405	0.00180296	-0.006086255	0.000556545
minuto	-0.002051443	0.0058491	-0.006447034	-0.132501353	0.528157926	-0.217258486	-0.44428529	0.676126999	-0.035002928	0.02123705	0.000245548	0.001153694	8.63E-05
segundo	-0.006846526	0.002695913	-0.004501649	-0.011674222	0.565616728	0.4849516	-0.38775458	-0.538293557	-0.023368633	-0.06348064	7.17E-05	-0.001697805	-0.000137281
Referencia_codificada	0.170403119	0.021589418	0.809151474	0.002633078	0.01132606	0.038358112	0.035252131	0.042289176	0.021761291	-0.227359633	-0.508884322	-0.002615751	0.000510077
Cluster	0.175232425	-0.004570769	0.769251373	0.105553887	-0.007033791	-0.104817442	-0.09403874	-0.079056215	-0.074558266	0.322205233	0.48014186	-0.018603576	0.00010768

Figura 4.28. Matriz de Cargas de Componentes en el Análisis de Componentes Principales del conjunto de datos de sismos

Fase 5: Modelado iterativo y desarrollo continuo

En esta fase, se implementa un enfoque iterativo para la construcción del modelo que caracteriza el comportamiento de los sismos en México. Utilizando la metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos, se ha obtenido un modelo que captura las interacciones y patrones subyacentes en los datos sísmicos.

Este proceso involucra ajustes y refinamientos sucesivos, donde cada iteración permite mejorar la precisión y capacidad explicativa del modelo. A través de la integración de técnicas avanzadas de análisis de datos, modelado estadístico y enfoques computacionales, se busca no solo representar la dinámica sísmica, sino también identificar factores relevantes que influyen en la ocurrencia y características de los eventos sísmicos en México.

Para la construcción del modelo matemático del Índice de Profundidad de Sismos (IPS), se empleó programación genética, una técnica de optimización basada en algoritmos evolutivos. Este enfoque permitió modelar el comportamiento de la profundidad sísmica a partir de un conjunto de variables relevantes, identificando aquellas que tienen una influencia significativa en la distribución de la profundidad de los sismos.

El uso de programación genética facilita la exploración de relaciones no lineales en los datos, proporcionando expresiones matemáticas que explican la variabilidad observada en la profundidad de los eventos sísmicos. A través de este método, se

pueden identificar patrones ocultos y mejorar la interpretación de los factores que contribuyen a la ocurrencia de sismos a diferentes profundidades.

En la Tabla 4.8 se presenta una descripción detallada de las variables utilizadas en el modelado, mientras que la Tabla 4.9 resume los parámetros del algoritmo evolutivo empleado para optimizar la función que define el IPS. Estos parámetros incluyen el tamaño de la población, el número de generaciones y la configuración del torneo de selección, los cuales son necesarios para garantizar la convergencia del modelo hacia una solución representativa del fenómeno sísmico.

Tabla 4.8. Descripción de variables del conjunto de datos de sismos

ID	Variable	Descripción
X1	Magnitud	Medida de la energía liberada en un sismo, expresada en la escala de Richter.
X2	Latitud	Coordenada geográfica que indica la posición del evento en el eje norte-sur.
X3	Longitud	Coordenada geográfica que indica la posición del evento en el eje este-oeste.
X4	Profundidad	Distancia en kilómetros desde la superficie terrestre hasta el hipocentro del sismo.
X5	Estatus	Indicador de validación del evento sísmico en la base de datos.
X6	Día	Día del mes en que ocurrió el sismo.
X7	Mes	Mes en que ocurrió el sismo.
X8	Año	Año en que ocurrió el sismo.
X9	Hora	Hora exacta en que se registró el evento sísmico.
X10	Minuto	Minuto exacto en que ocurrió el sismo.
X11	Segundo	Segundo exacto en que ocurrió el sismo.
X12	Referencia codificada	Código numérico que representa la localización geográfica del sismo.
X13	Clúster	Clasificación del sismo en grupos en base a patrones espaciales y temporales.

Tabla 4.9. Parámetros utilizados en el programa evolutivo, para determinar el índice de profundidad de sismos

Parámetro	Valor
Conjunto de datos	Sismos con 294,900 registros y 13 variables
Número de generaciones	1,000
Población	10,000
Número de ejecuciones	30
Tournament size	2,000

El modelo final obtenido, tras un proceso iterativo de optimización y ajuste, se representa mediante la ecuación 4.6. Esta expresión ha sido seleccionada por su capacidad para capturar las relaciones subyacentes entre las variables y su influencia en la profundidad de los sismos. Su rendimiento ha sido evaluado utilizando métricas de error como el Error Cuadrático Medio (MSE) y el Error Absoluto Medio (MAE), lo que garantiza un alto grado de ajuste a los datos observados. A continuación, se presenta la ecuación que define el comportamiento del Índice de Profundidad de Sismos (IPS) en función de las variables seleccionadas.

$$IPS = \sqrt{X_9 \cdot \sqrt{\sqrt{((-X_{10} - X_5)(-X_1 + \sin(X_3) \cdot |X_{10}|) \sin(X_3))^{1/4} \cdot (-X_{10} - X_9) \cdot \sin(X_1) \cdot \sin(X_3) \cdot |X_1|^{11/32}}}} \quad (4.6)$$

Donde:

IPS es el índice de profundidad del sismo.

X1 (Magnitud) representa la magnitud del sismo, que es una medida de la energía liberada en el epicentro del temblor.

X3 (Longitud) se refiere a la longitud geográfica del epicentro del sismo. Es una coordenada para determinar la ubicación exacta del sismo.

X5 (Día) representa el día del mes en que ocurrió el sismo. Esta variable temporal es esencial para analizar patrones temporales en la ocurrencia de sismos y puede ayudar a identificar tendencias estacionales o ciclos sísmicos en una región determinada.

X9 (Minuto) indica el minuto exacto en que ocurrió el sismo. Esta variable de tiempo proporciona una resolución más fina para el análisis temporal de la ocurrencia de sismos. Es importante para estudiar la distribución temporal de los sismos a lo largo del día y para identificar eventos sísmicos cercanos en el tiempo.

X10 (Segundo) refiere al segundo exacto en que ocurrió el sismo. Al igual que la variable de minuto, proporciona una resolución temporal adicional para el análisis detallado de la ocurrencia de sismos.

El comportamiento del IPS para ajustarse al comportamiento de la profundidad del sismo se muestra en la Figura 4.29. En la gráfica se puede apreciar en color naranja el comportamiento de la profundidad de los sismos y en color azul con la etiqueta de nueva función que representa al IPS que se ajusta al comportamiento de la profundidad de los sismos.

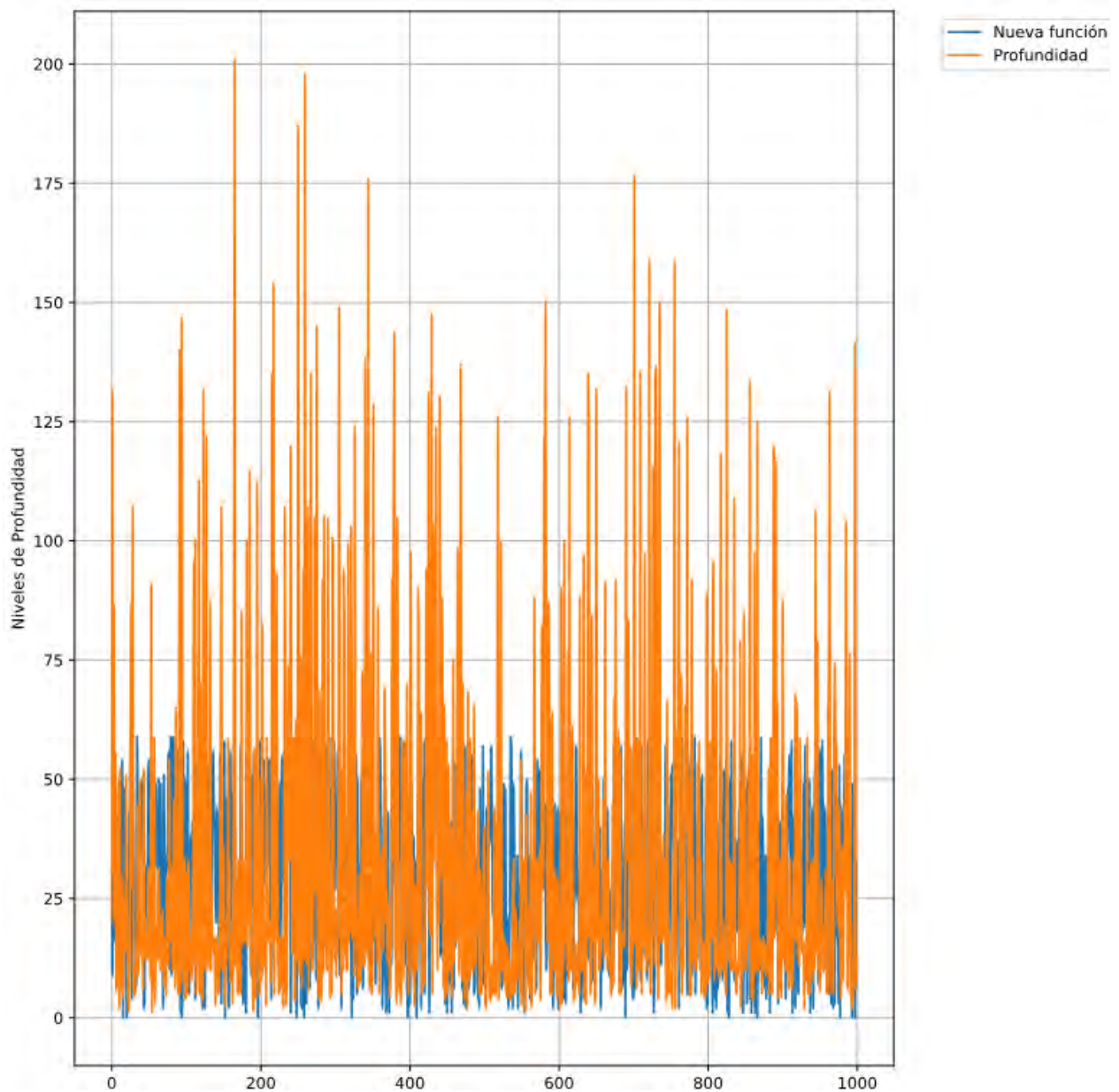


Figura 4.29. Comportamiento del IP (color azul) con respecto al comportamiento de la profundidad del sismo (color naranja)

Para mejorar la precisión del modelo del Índice de Profundidad del Sismo (IPS), es decisivo considerar variables adicionales que influyen en la determinación de la profundidad del sismo. Entre estas variables se encuentran el tipo de falla, el tipo de placa tectónica, los mecanismos de deformación y las características de la roca. La naturaleza de la falla sísmica, ya sea inversa, normal o de desgarre, impacta la profundidad del sismo, al igual que la interacción entre las placas tectónicas y los procesos de deformación en la corteza terrestre. Además, las propiedades físicas y geológicas de la roca circundante, como densidad, rigidez y composición mineralógica, desempeñan un papel crucial en la propagación de las ondas sísmicas

y, por lo tanto, en la profundidad del sismo. La inclusión de estas variables adicionales en el modelo del IPS fortalecerá su capacidad de caracterización.

Fase 6: Interpretación, representación e implementación continua

La última fase del desarrollo del Índice de Profundidad de Sismos (IPS) se centra en la interpretación de los resultados obtenidos, la representación de los patrones sísmicos detectados y la implementación continua del modelo para su mejora y actualización.

El modelo presentado en la ecuación 4.6 permite evaluar cómo distintas variables influyen en la profundidad de los sismos. La baja correlación entre magnitud y profundidad indica que los eventos sísmicos más intensos no necesariamente ocurren a mayores profundidades, lo que concuerda con estudios previos sobre sismicidad en México. Además, la relación positiva entre profundidad y longitud sugiere que los sismos más profundos están concentrados en ciertas zonas geográficas, probablemente asociadas a la subducción de la Placa de Cocos bajo la Placa de Norteamérica.

Las métricas de error obtenidas para el modelo de profundidad sísmica específicamente un Error Cuadrático Medio (MSE) de 41.28 y un Error Absoluto Medio (MAE) de 30.58 reflejan que, si bien el modelo logra capturar ciertas tendencias generales en los datos, su nivel de precisión es limitado en comparación con otros casos modelados, como el de calidad del aire. Esta diferencia puede atribuirse a la naturaleza inherentemente caótica y menos predecible de los fenómenos sísmicos, los cuales dependen de múltiples factores geofísicos de difícil observación directa, como la estructura del subsuelo, las propiedades de las fallas tectónicas y la velocidad de propagación de las ondas sísmicas (Koulakov et al., 2023; Rawlinson & Blom et al., 2022). En este contexto, si bien los valores de error pueden parecer elevados desde una perspectiva comparativa, podrían considerarse aceptables dadas las condiciones del fenómeno, especialmente en ausencia de variables profundas como la composición geológica detallada o registros sísmicos tridimensionales. No obstante, los resultados también sugieren que existe un margen significativo de mejora. La inclusión de datos geofísicos adicionales, tales como la velocidad de ondas P y S, mapas de tomografía sísmica o clasificaciones detalladas de tipo de falla, podría contribuir a una representación más precisa del comportamiento del sistema y, en consecuencia, a una reducción sustancial del error.

Para visualizar los resultados, se han generado mapas de calor de correlación, gráficos de distribución de profundidad y análisis espacio-temporales de los sismos en México. Estas representaciones permiten identificar zonas de alta actividad sísmica a diferentes profundidades y evaluar cómo las características del subsuelo afectan la propagación de ondas sísmicas.

Un hallazgo primordial es la existencia de diferentes regímenes de profundidad, donde los sismos superficiales (≤ 50 km) se concentran en el Eje Neovolcánico y la Sierra Madre del Sur, mientras que los eventos más profundos (> 100 km) ocurren predominantemente en el cinturón de subducción del Pacífico. Estos resultados

refuerzan la hipótesis de que la profundidad sísmica está influenciada por la tectónica de placas y la estructura litosférica del país.

El modelo desarrollado no es estático, sino que forma parte de un proceso de implementación continua. Se propone una integración con bases de datos sísmicas en tiempo real, permitiendo la actualización del modelo a medida que se registran nuevos eventos. Además, la optimización del modelo puede incluir refinamiento de la selección de variables: Incorporando parámetros adicionales como la energía liberada o el tipo de falla geológica. Ajuste de hiperparámetros del algoritmo evolutivo: Modificando el tamaño de la población, número de generaciones y funciones primitivas para mejorar la generalización del modelo.

4.5 Conclusiones

La validación de la metodología propuesta para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos mediante los casos de estudio de calidad del aire en la Ciudad de México y profundidad de los sismos en México ha mostrado su efectividad en la representación de problemas complejos. A través de un enfoque estructurado y basado en fases, se logró capturar la naturaleza dinámica de ambos sistemas, permitiendo una caracterización precisa y una mejor comprensión de sus interacciones y patrones emergentes.

La aplicación de un esquema de agregación multinivel permitió identificar escalas adecuadas para el análisis de cada sistema, asegurando una representación precisa de la dinámica de los problemas bajo una perspectiva de sistemas complejos. En el caso de la calidad del aire, la segmentación por zonas urbanas y niveles de contaminación facilitó el modelado de la dispersión de contaminantes. En el caso de los sismos, la agregación a nivel nacional permitió detectar patrones geoespaciales en la distribución de la profundidad sísmica.

La implementación de procesos automatizados para la recopilación y actualización continua de datos garantiza la disponibilidad de información actualizada y de alta calidad. En el estudio de calidad del aire, el uso de bases de datos NoSQL y la integración con API de monitoreo en tiempo real fortalecieron la capacidad de detección temprana de eventos críticos. En el análisis de los sismos, la conexión con registros históricos permitió evaluar tendencias a largo plazo y su relación con factores geofísicos.

La utilización de programación genética (PG) resultó en modelos computacionales robustos capaces de capturar la complejidad de ambos problemas complejos. Para la calidad del aire, se obtuvo una ecuación que correlaciona los niveles de contaminación con variables meteorológicas, alcanzando una reducción del error en la estimación del Índice de Calidad del Aire (ICA). Para la profundidad de los sismos, el modelo generó un Índice de Profundidad de Sismos (IPS) que reflejó patrones espaciales y temporales, demostrando la influencia de variables como magnitud, latitud y longitud en la determinación de la profundidad de los eventos sísmicos.

La interpretación de los modelos generados permitió identificar las variables más influyentes en cada sistema. En el caso del aire, PM_{2.5}, PM₁₀ y O₃ fueron las variables más determinantes para el ICA. En los sismos, se evidenció que la ubicación geográfica (latitud y longitud) y el minuto exacto del evento son factores vitales en la variabilidad de la profundidad. Además, la implementación de plataformas interactivas de monitoreo y la visualización de patrones ayudó a la aplicación de estos modelos en la toma de decisiones.

Aunque los modelos generados capturan tendencias generales y permiten una representación funcional de los sistemas analizados, existen áreas de oportunidad para su refinamiento. La inclusión de nuevas variables geofísicas en el análisis de los sismos, como el tipo de falla tectónica y la composición del manto, podría mejorar la precisión del modelo. En calidad del aire, la incorporación de factores socioeconómicos y el uso de técnicas de aprendizaje profundo podrían optimizar la capacidad predictiva del modelo.

La optimización de los parámetros evolutivos en los algoritmos de programación genética mediante estrategias adaptativas permitiría mejorar la convergencia del modelo y reducir errores en la predicción. La validación de la metodología confirma su eficacia en el modelado de sistemas complejos altamente dinámicos. La combinación de enfoques de agregación multinivel, recopilación de datos en tiempo real y modelado basado en programación genética permitió capturar con precisión las interacciones y comportamientos emergentes en los dos casos de estudio. La aplicabilidad de estos modelos en la gestión ambiental y en la evaluación del riesgo sísmico refuerza la utilidad de la metodología propuesta como una herramienta versátil para el análisis de sistemas dinámicos en diversos contextos (ver Tabla 4.10).

Tabla 4.10. Comparación metodológica entre los casos de estudio de calidad del aire y profundidad de sismos

Aspecto	Calidad del aire (CDMX)	Profundidad de sismos (México)
Tipo De Sistema	Ambiental, urbano, afectado por factores meteorológicos y sociales	Geofísico, dependiente de variables tectónicas y espacio-temporales
Escala De Agregación	Zonas urbanas y niveles de contaminación	Agregación a nivel nacional con análisis espacial
Variables Clave	PM2.5, PM10, O ₃ , temperatura, humedad, viento	Magnitud, latitud, longitud, minuto del evento
Base De Datos Utilizada	NoSQL + API de monitoreo en tiempo real	Registros históricos oficiales (servicios sismológicos)
Técnica De Modelado	Programación genética (PG)	Programación genética (PG)
Resultado Principal	Ecuación predictiva del Índice de Calidad del Aire (ICA)	Ecuación del Índice de Profundidad de Sismos (IPS)
Capacidad Del Modelo	Reducción significativa del error en estimación del ICA	Detección de patrones espacio-temporales en la profundidad sísmica
Visualización	Plataforma interactiva con mapas de contaminación	Visualización geoespacial de profundidad y epicentros
Aplicaciones Prácticas	Detección temprana de eventos críticos, gestión ambiental urbana	Evaluación del riesgo sísmico, planificación territorial
Mejoras Propuestas	Incluir variables socioeconómicas, usar aprendizaje profundo	Incorporar tipo de falla tectónica, análisis del manto geológico
Impacto Metodológico	Valida la capacidad de la metodología para modelar fenómenos ambientales complejos	Valida la aplicabilidad en fenómenos geofísicos dinámicos

Capítulo 5. Prototipo de Movilidad vehicular en la Ciudad de México

5.1 Introducción

El desarrollo de prototipo de movilidad vehicular en la Ciudad de México presenta el diseño y construcción de un prototipo orientado a modelar y analizar patrones de movilidad vehicular en un entorno urbano que implementa el índice de movilidad desarrollado en esta investigación. Este prototipo busca capturar las dinámicas de tráfico y las interacciones entre distintos factores que influyen en el flujo vehicular, como el cuadro delimitador de donde se origina el tráfico, la velocidad de referencia, la longitud del segmento (calle), velocidad esperada, día del monitoreo de tráfico el tipo de incidente (accidente, construcción, congestión, avería de vehículo, clima entre otros) y el cuadro delimitador de donde se ubica el servicio. El desarrollo del prototipo implica la integración de datos en tiempo real y el modelo matemático que representa el comportamiento del tráfico, proporcionando una herramienta práctica para evaluar posibles escenarios y soluciones en la gestión de la movilidad. Esta sección detalla las etapas base del desarrollo, incluyendo la recolección de datos, el diseño de algoritmos, la implementación del prototipo y las pruebas de validación, con el objetivo de ofrecer un modelo funcional que contribuya a mejorar la eficiencia y sostenibilidad del sistema de transporte urbano.

5.2 Prototipo de Movilidad Vehicular

El desarrollo del prototipo de movilidad vehicular en la Ciudad de México ha requerido la integración de múltiples funcionalidades que permitan analizar, visualizar e interpretar la dinámica del tráfico urbano. Para ello, se han diseñado diversas secciones interactivas, cada una enfocada en proporcionar información importante para el estudio del flujo vehicular en distintas escalas geográficas.

Este prototipo implementa un enfoque modular y escalable, lo que facilita su aplicación en la gestión de la movilidad urbana. Las diferentes secciones han sido desarrolladas con el propósito de ofrecer herramientas de análisis en tiempo real, permitiendo la evaluación de indicadores Imprescindible como el Índice de Movilidad (IM) y el Índice de Movilidad de la Red Vial (IMRV). Dichos índices cuantifican el nivel de congestión en una calle o una zona urbana en función de variables como la velocidad de referencia, el tipo de incidente y la distribución de la infraestructura vial.

A continuación, se detallan las principales secciones del prototipo, describiendo sus características, funcionalidades y su contribución al análisis integral de la movilidad vehicular en entornos urbanos de alta densidad.

a. Página principal

Se diseñó y desarrolló un menú de control con diferentes opciones desplegables que permite a los usuarios interactuar con el portal de movilidad, incluyendo opciones como "Polígono de la CDMX," "Seleccionar Alcaldía," "Semáforos," "Servicios," "Tráfico e Incidentes," "Validades," "Índice de Movilidad" e "Índice de Red Vial." Estas opciones permiten filtrar la información visualizada en el mapa o centrarse en aspectos específicos de la movilidad, como los semáforos o los incidentes de tráfico, como se muestra en la Figura 5.7.

La aplicación está diseñada para proporcionar un índice de movilidad de las calles en zonas de interés. Este portal resulta útil para estudios de tráfico, planificación urbana y para mejorar la eficiencia del sistema de transporte tanto público como privado en una ciudad densamente poblada.

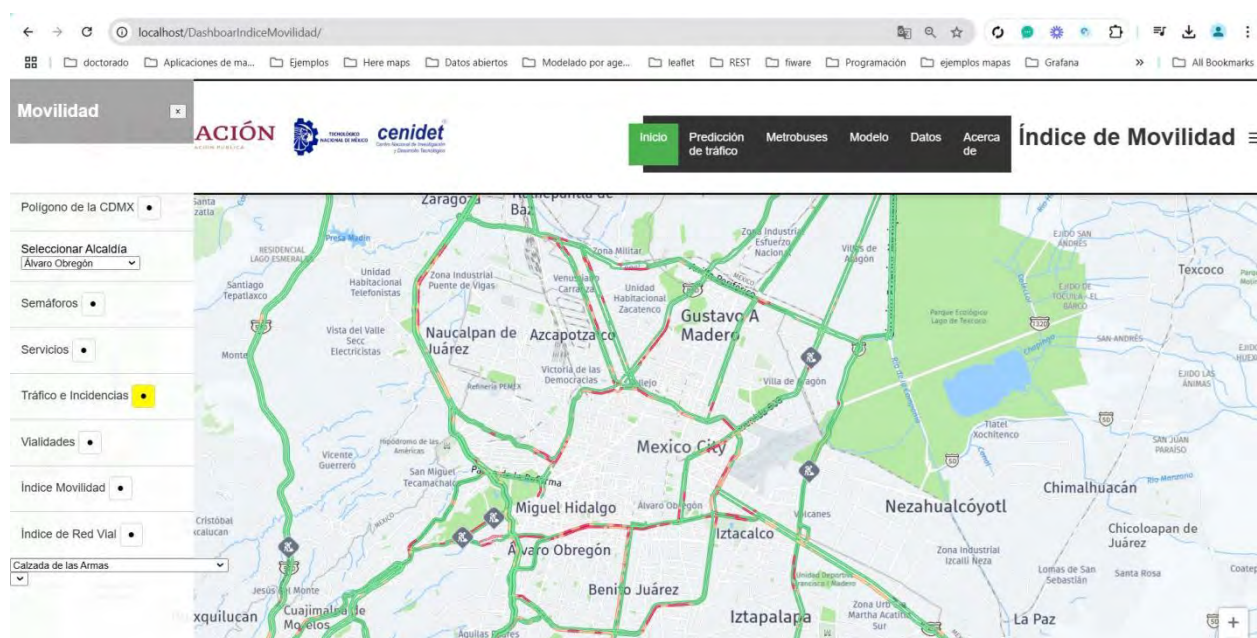


Figura 5.1. Página principal del prototipo del índice de Movilidad

b. Opción de polígono de la CDMX

Se diseñó y desarrolló esta opción para resaltar en colores distintos las diferentes alcaldías o polígonos de la Ciudad de México como se muestra en la Figura 5.2. Cada sección de color representa una demarcación territorial, permitiendo identificar visualmente los límites y la distribución de las alcaldías dentro de la ciudad. La asignación de un color específico para cada alcaldía facilita la rápida identificación del tráfico y las incidencias viales en las zonas de interés, lo cual resulta útil para el análisis de movilidad y estudios territoriales específicos.

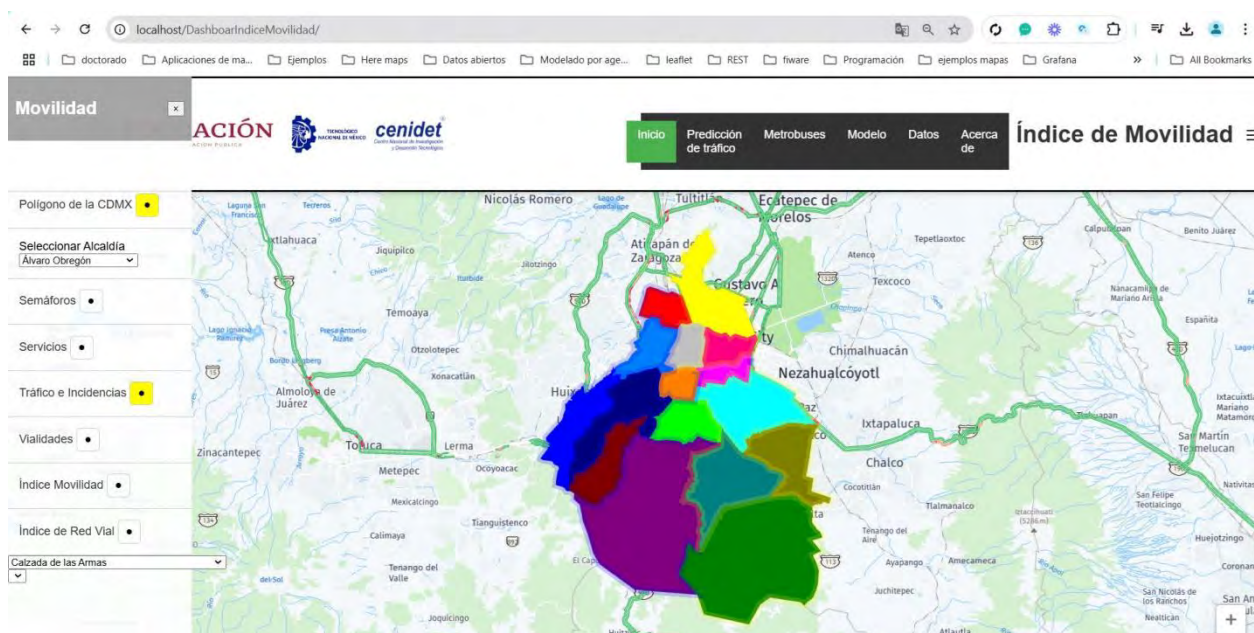


Figura 5.2. Demarcación de las 16 alcaldías de la Ciudad de México

c. Opción de Semáforos

Se diseñó y desarrolló esta opción con el propósito de visualizar la distribución geográfica de los semáforos en el mapa, lo cual permite monitorear su ubicación y analizar su impacto potencial en el flujo vehicular, como se ilustra en la Figura 5.3. Los colores sobre las rutas principales indican el nivel de congestión en cada tramo, facilitando una evaluación detallada del estado del tránsito en tiempo real. La información sobre los semáforos fue obtenida mediante una consulta en la plataforma Overpass API, extrayendo las coordenadas geográficas desde la base de datos colaborativa OpenStreetMap (OSM). La consulta utilizada permitió filtrar los elementos etiquetados como `highway=traffic_signals`, recuperando así únicamente aquellos nodos que representan semáforos de las zonas geográficas de interés.

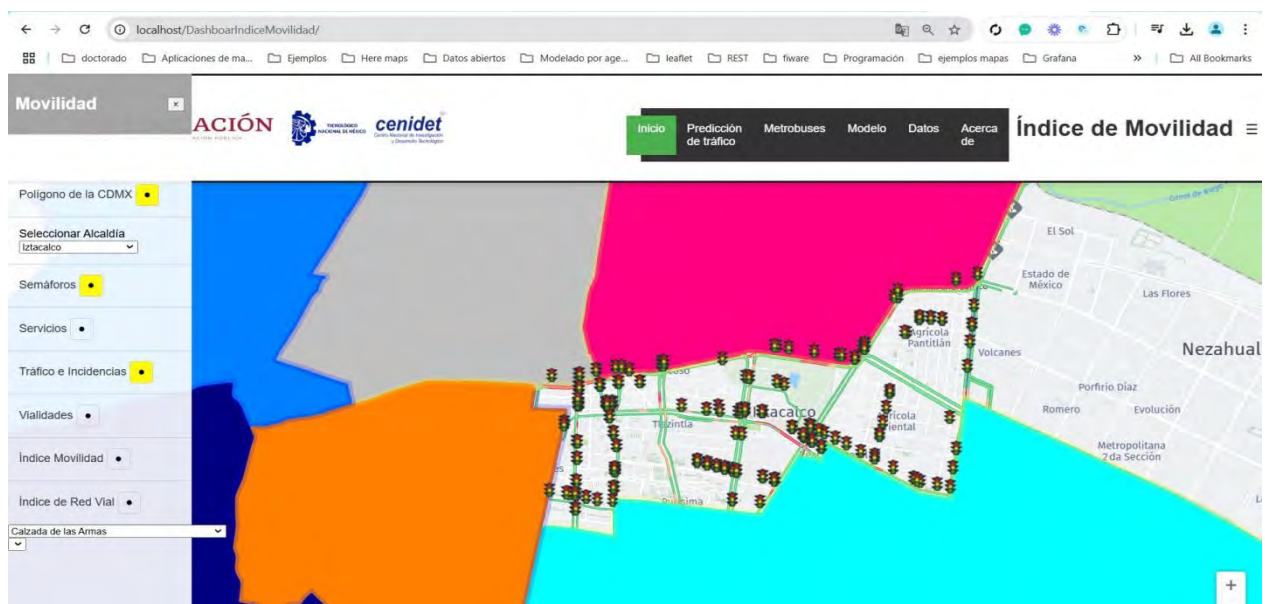


Figura 5.3. Distribución de semáforos en las vías principales de las alcaldías de la Ciudad de México, en este caso se muestran los de la alcaldía Iztacalco

d. Opción de Vialidades

Se diseñó y desarrolló esta opción para mostrar el conjunto de vialidades correspondientes a una alcaldía, proporcionando una visualización detallada de sus características. Entre los atributos destacados se incluyen el nombre de la calle, el tipo de vía (principal, secundaria, etc.), el sentido de circulación, y el tiempo estimado de tránsito, como se muestra en la Figura 5.4. Esta funcionalidad facilita un análisis exhaustivo de la movilidad en la alcaldía, brindando información útil tanto para la gestión del tráfico como para la planificación urbana y la toma de decisiones en temas de transporte.

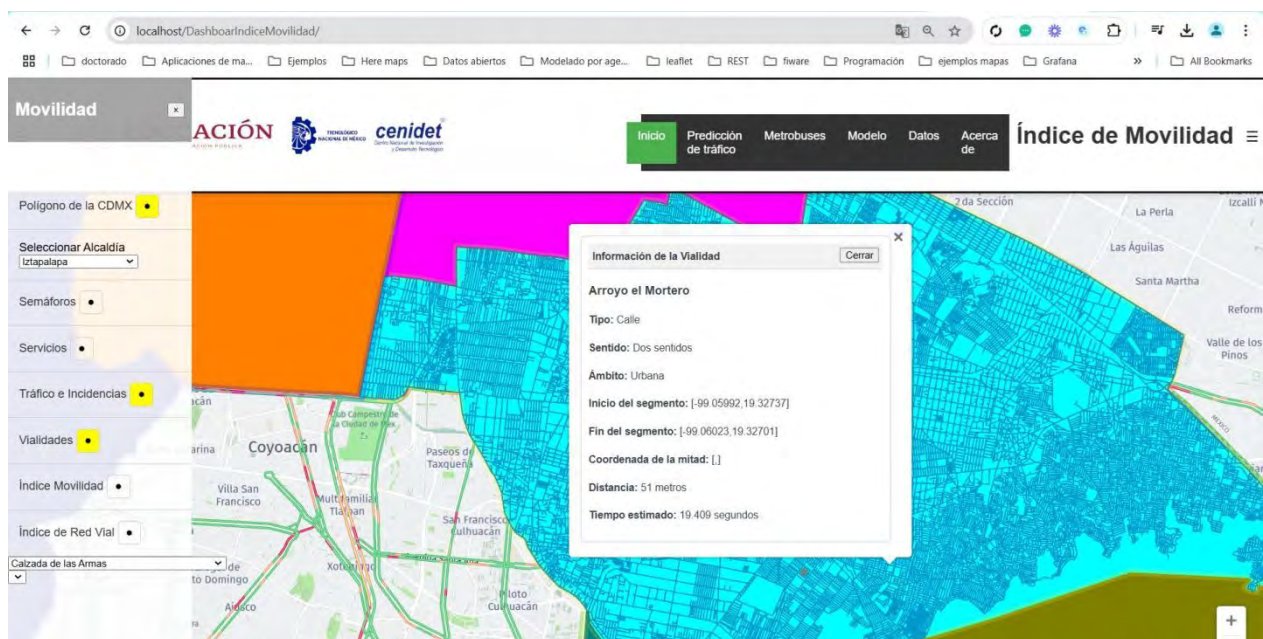


Figura 5.4. Conjunto de vías de una alcaldía de interés (en este caso la alcaldía Iztapalapa)

e. Opción de Servicios

Se diseñó y desarrolló una funcionalidad que permite seleccionar cualquier zona geográfica o calle de interés dentro de una alcaldía, con el objetivo de visualizar con detalle todos los puntos de interés (servicios) que la rodean. Al seleccionar un área específica, el sistema despliega servicios esenciales y puntos relevantes, tales como escuelas, empresas, hospitales y espacios de recreación, dentro de la zona delimitada (ver Figura 5.5). Esta herramienta facilita un análisis integral de los servicios disponibles en áreas específicas, resultando útil para la planificación urbana, el estudio de la accesibilidad a servicios básicos y recreativos, así como para la toma de decisiones informadas en materia de desarrollo y movilidad. Los datos de los servicios se obtienen en tiempo real, mediante la combinación de consultas a la Overpass API (para recuperar elementos desde la base de datos abierta de OpenStreetMap) y la plataforma Geoapify, que complementa la información con datos adicionales estructurados por categorías y niveles de relevancia.

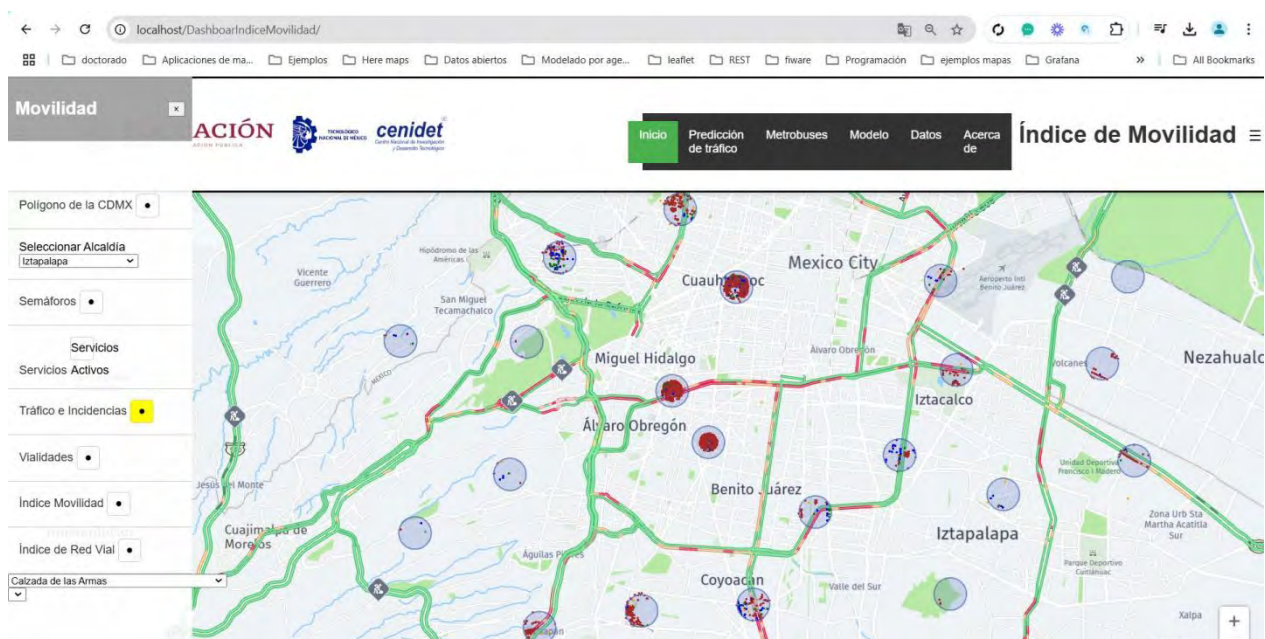


Figura 5.5. Conjunto de puntos de interés(servicios) que rodean a cada calle de interés

f. Opción Índice de Movilidad

Se diseñó y desarrolló esta opción para implementar el Índice de Movilidad (IM) utilizando el siguiente modelo: $IM \approx 2 \cdot X_1 + 3 \cdot \sqrt{X_0} + \sqrt{(\log(X_6)) - 5 \cdot \log(X_{11}) - 0.308}$.

Este modelo permite representar cuantitativamente el grado de movilidad de una calle, considerando la interacción entre variables como la velocidad de referencia (X_1), la longitud del segmento (X_0), el día del monitoreo (X_6), y el tipo de incidente reportado (X_{11}). Al integrar este índice, se facilita el análisis de las condiciones de tránsito y movilidad, permitiendo identificar patrones y comportamientos diferenciados en distintas vías urbanas, como se muestra en la Figura 5.6. Desde una perspectiva interpretativa, el valor del Índice de Movilidad (IM) ofrece una medida cuantitativa de la fluidez del tránsito en una vía específica. Un IM bajo refleja condiciones de movilidad restringida, generalmente asociadas con alta congestión vehicular, presencia de incidentes viales, o características físicas limitantes de la infraestructura. En este escenario, variables como un valor reducido de X_1 (velocidad de referencia) o un valor elevado de X_{11} (tipo de incidente) ejercen una influencia negativa significativa sobre el índice, reduciéndolo y evidenciando cuellos de botella en la red vial.

En contraste, un IM alto indica una situación de tránsito favorable, caracterizada por trayectos fluidos, velocidades cercanas a las ideales, y mínima interferencia causada por eventos disruptivos. Este comportamiento se ve reforzado por valores elevados de X_1 y moderados de X_0 (longitud del segmento), así como por una baja incidencia de eventos registrados (X_{11}), lo que reduce el impacto de los términos negativos del modelo.

Este análisis permite entender el aporte relativo de cada variable en la conformación del índice, similar a una ablación funcional, en la que la modificación o exclusión de

g. Opción Índice de Red Vial

177

5.3 Conclusiones

El desarrollo del prototipo de movilidad vehicular en la Ciudad de México constituye un avance significativo en la modelización y análisis de la dinámica de la movilidad vehicular, especialmente en entornos altamente congestionados como la capital mexicana y otras megaciudades del mundo. Esta herramienta computacional integra un conjunto de variables elementales para evaluar la movilidad vehicular a nivel de calles o zonas geográficas específicas, proporcionando un enfoque analítico que facilita la toma de decisiones informadas en la gestión del tránsito.

La implementación del Índice de Movilidad (IM) y del Índice de Movilidad de la Red Vial (IMRV) ha permitido la construcción de una representación cuantitativa precisa del comportamiento de la movilidad vehicular, lo que posibilita la identificación de patrones de movilidad y el desarrollo de estrategias orientadas a la mitigación de congestión vehicular.

El diseño modular del prototipo incorpora funcionalidades de mayor relevancia, tales como la visualización de vialidades, la identificación de semáforos y puntos de interés, el monitoreo de incidentes de tráfico y la segmentación geográfica de la ciudad. La integración de datos en tiempo real y la estructuración de información en una interfaz interactiva permite la accesibilidad y usabilidad del sistema, facilitando su aplicación en estudios de movilidad, planificación urbana y gestión del tráfico vehicular.

Asimismo, la capacidad del prototipo para mostrar de forma visual la movilidad y sus variables asociadas permiten una mejora considerable en la interpretación de la información, fortaleciendo el análisis de la infraestructura vial y el impacto de diversos factores sobre el desempeño del sistema de transporte. Su diseño escalable permite adaptarse a diferentes niveles de granularidad, desde el análisis detallado de una calle específica hasta la evaluación de zonas urbanas completas, lo que proporciona un marco flexible y adaptable a múltiples aplicaciones.

Este desarrollo valida la metodología propuesta para la construcción de modelos de comportamiento en sistemas complejos altamente dinámicos, estableciendo un referente en la integración de herramientas computacionales para la gestión inteligente de la movilidad. La combinación de modelos matemáticos avanzados con representaciones visuales interactivas fortalece la capacidad del sistema para caracterizar, analizar y predecir el comportamiento del tráfico, asegurando un ajuste continuo a la realidad de la movilidad vehicular. Además, abre nuevas oportunidades para la implementación de estrategias de optimización y planificación, contribuyendo al desarrollo de soluciones innovadoras para la movilidad urbana en la Ciudad de México y otras metrópolis con desafíos similares.

Capítulo 6. Resultados

6.1 Introducción

En este capítulo se presentan los resultados obtenidos durante la investigación, se describen los hallazgos relevantes alcanzados en cada fase de la metodología, desde la contextualización y análisis del sistema hasta la validación de los modelos desarrollados. Además, se analiza cómo estos resultados reflejan los enfoques implementados, destacando su capacidad para analizar dinámicas no lineales, fenómenos emergentes y alta sensibilidad a las condiciones iniciales.

6.2 Desarrollo de la Metodología para la Modelación de Sistemas Complejos Altamente Dinámicos

El primer resultado significativo de esta investigación es el desarrollo de una metodología estructurada en 6 fases y 16 métodos para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos. Esta metodología ofrece un proceso sistemático que permite enfrentar la complejidad inherente a estos sistemas y facilita la obtención de modelos precisos y robustos.

- Contextualización y análisis del sistema: Se identifican las variables críticas y las interacciones entre las entidades del sistema, proporcionando una comprensión detallada de su comportamiento.
- Agregación multinivel y análisis de entropía: La complejidad del sistema se organiza en niveles jerárquicos, permitiendo establecer indicadores cuantitativos que facilitan el análisis dinámico en diferentes escalas.
- Obtención y monitoreo continuo de datos: Se asegura un flujo constante de datos en tiempo real, integrando diversas fuentes para garantizar la calidad y disponibilidad de la información.
- Preprocesamiento y muestreo de datos: Los datos se limpian, normalizan y transforman para garantizar su homogeneidad y calidad, permitiendo obtener resultados confiables en las etapas de modelado.
- Modelado iterativo: Se desarrollan expresiones simbólicas que capturan el comportamiento del sistema. Estas expresiones se refinan mediante ajustes y validaciones continuas para optimizar su capacidad predictiva.
- Interpretación y representación de resultados: Se diseñan herramientas visuales que facilitan la comprensión de los resultados obtenidos y su aplicación práctica en la toma de decisiones estratégicas.

Desde una perspectiva crítica, la fase de modelado iterativo emergió como la más desafiante y significativa dentro del proceso metodológico, no solo por su complejidad técnica, sino por su papel estructural en la consolidación de los resultados. En términos de aporte individual al desempeño general de la metodología, su eliminación o implementación deficiente tendría un impacto sustancial en la calidad y funcionalidad de los modelos desarrollados, lo que evidencia su alta sensibilidad dentro del conjunto

de fases. Esta etapa implicó representar fenómenos no lineales, multivariantes y altamente dinámicos mediante expresiones simbólicas obtenidas a través de programación genética. La dificultad radicó en seleccionar las variables más representativas, evitar el sobreajuste (overfitting) y aplicar validaciones cruzadas, todo ello en un entorno con gran variabilidad e incertidumbre inherente a los sistemas analizados.

6.3 Construcción del Índice de Movilidad Vehicular (IMV)

El segundo resultado de la investigación es el desarrollo del Índice de Movilidad Vehicular (IMV) para la Ciudad de México. Este índice fue diseñado como un modelo matemático para evaluar el comportamiento de la movilidad vehicular y fue definido por la siguiente expresión aproximada:

$$IM \approx 2 \cdot X1 + 3 \cdot \sqrt{X0} + \sqrt{\log(X6)} - 5 \cdot \log(X11) - 0.308 \quad (6.1)$$

La descripción detallada de las variables utilizadas en el modelo propuesto para la evaluación del Índice de Movilidad (IM) se presenta en la Tabla 6.7. Estas variables representan distintos componentes del sistema de tránsito urbano, incluyendo factores espaciales, dinámicos, temporales y contextuales que influyen directamente en el comportamiento de la movilidad vehicular.

Tabla 6.1. Descripción de las variables empleadas en el modelo de movilidad (IM)

Variable	Significado
X0	Área geográfica utilizada para identificar zonas críticas de congestión vehicular.
X1	Velocidad del flujo vehicular, actúa como indicador directo de las condiciones de movilidad.
X6	Día del monitoreo, introduce un componente temporal que refleja variaciones por tipo de día (laboral, fin de semana, eventos especiales).
X11	Incidencias que afectan la movilidad, como accidentes, cierres viales, obras o fenómenos meteorológicos.

El IMV permite evaluar el desempeño de la movilidad vehicular en diferentes zonas y horarios, identificar patrones de congestión y diseñar estrategias de gestión del tráfico, como se visualiza en la Figura 6.7.

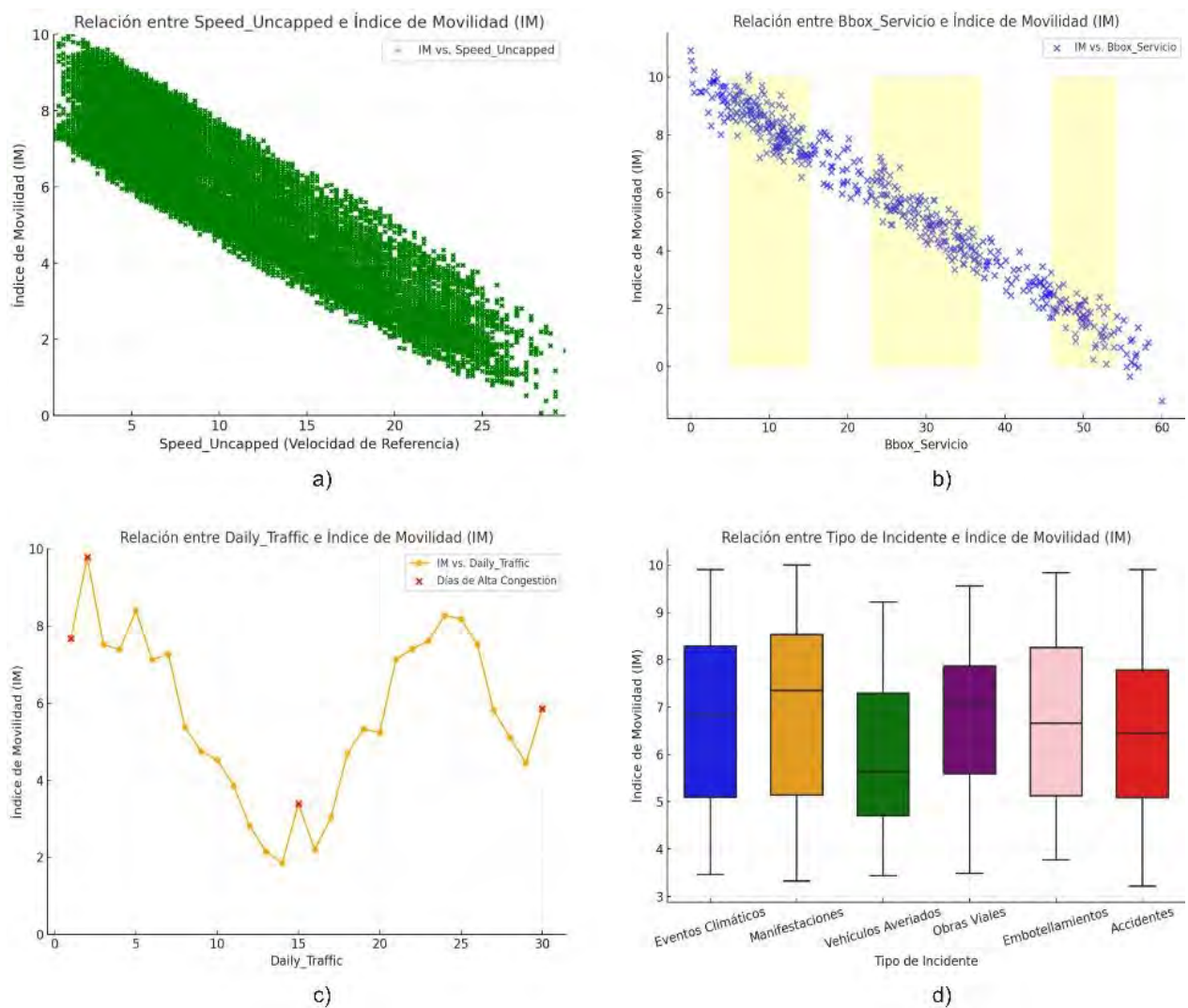


Figura 6.1. Relación entre el Índice de Movilidad (IM) y diferentes factores

Análisis del índice de Movilidad Vehicular (IMV) de acuerdo a los gráficos de la Figura 6.1.

- (a) Una menor velocidad de referencia se asocia con mayor congestión.
- (b) Ciertas zonas geográficas muestran menor desempeño en la movilidad.
- (c) Se observan fluctuaciones diarias que destacan los días con alta congestión.
- (d) Los tipos de incidentes, como accidentes y embotellamientos, tienen mayor impacto en el IMV.

El comportamiento de la movilidad vehicular se clasificó en distintos niveles utilizando valores escalares, basados en el análisis de desviaciones estándar, como se muestra en la Figura 6.2.

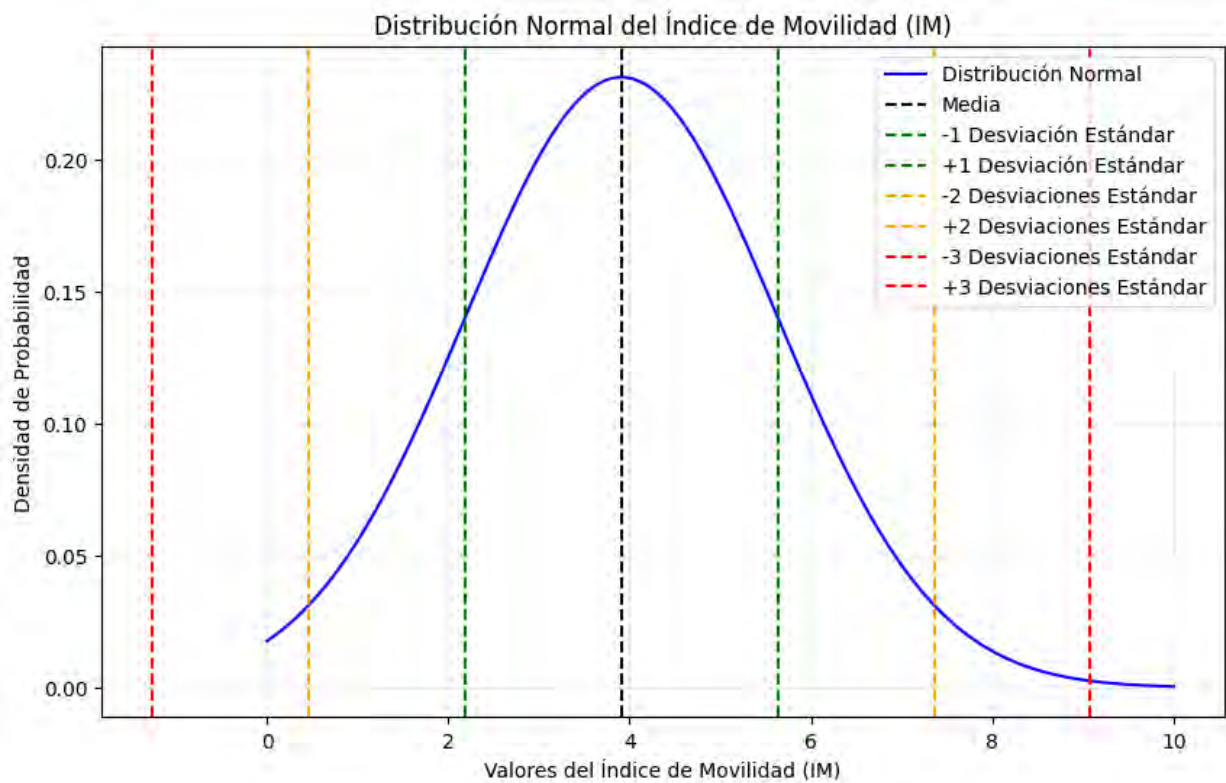


Figura 6.2. *Distribución normal del Índice de Movilidad (IM)*

La distribución normal del IMV se utilizó con una única desviación estándar (ver Figura 6.3) para clasificar la movilidad en:

- Alta movilidad (color verde): Valores por debajo de 2.19 (media - 1 desviación estándar).
- Movilidad media (color amarillo): Valores entre 2.19 y 5.64 (± 1 desviación estándar).
- Baja movilidad (color rojo): Valores por encima de 5.64 (media + 1 desviación estándar).

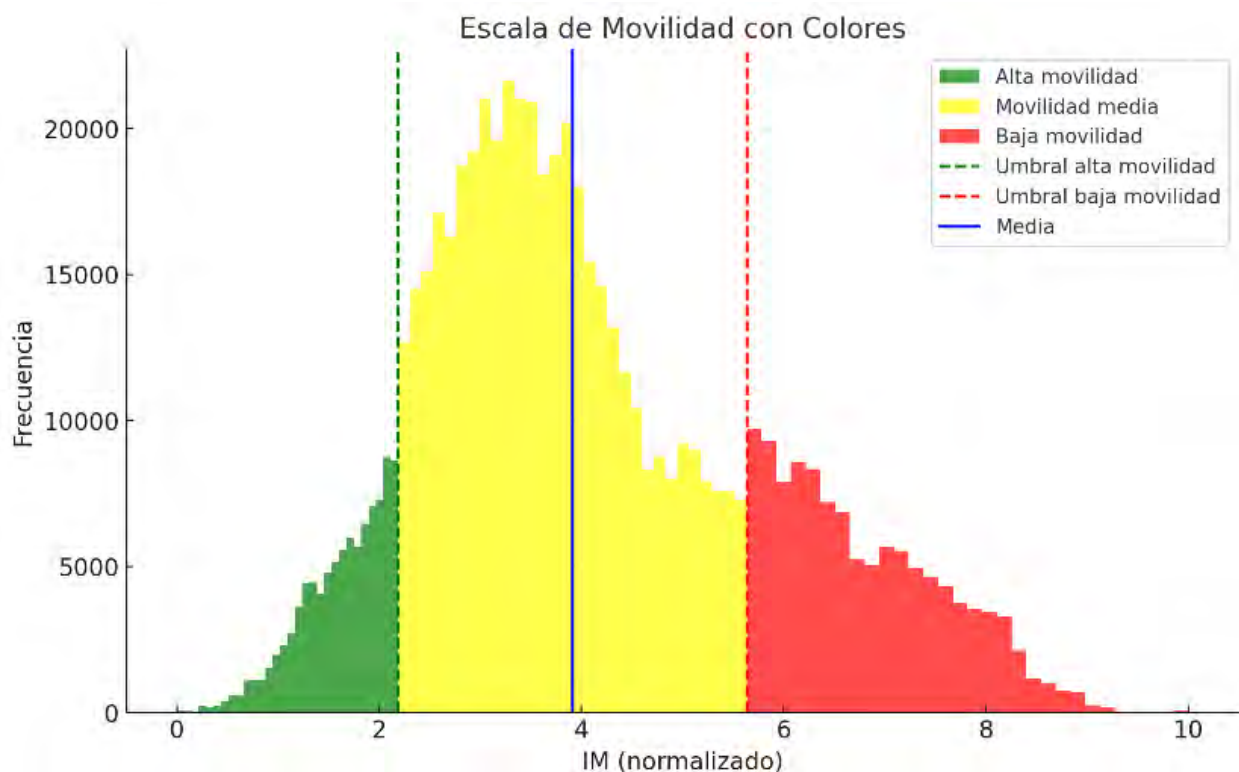


Figura 6.3. Histograma que muestra la distribución de movilidad de acuerdo al Índice de Movilidad (IM)

6.4 Desarrollo de Índices Complementarios: ICA e IPS

El tercer resultado de esta investigación es la construcción de dos índices complementarios que amplían el alcance de la metodología:

- Índice de Calidad del Aire (ICA): Integra las concentraciones de contaminantes atmosféricos (PM_{2.5}, PM₁₀, O₃, NO₂, SO₂ y CO) en un único valor, permitiendo evaluar la calidad del aire en áreas específicas. Este índice facilita la identificación de zonas con altos niveles de contaminación y el diseño de estrategias para mitigar riesgos en salud pública y medio ambiente.
- Índice de Profundidad de los Sismos (IPS): Representa la profundidad de eventos sísmicos, proporcionando información crítica para comprender los riesgos asociados a los movimientos tectónicos. Este índice permite evaluar patrones sísmicos y diseñar estrategias de prevención y respuesta ante desastres naturales.

Ambos índices, aunque inicialmente surgieron como parte del proceso de validación de la metodología, se consolidan como herramientas analíticas independientes con aplicaciones prácticas en sus respectivos campos.

6.5 Desarrollo de un Prototipo Web para Evaluación de la Movilidad

El cuarto resultado es el desarrollo de un prototipo en ambiente web que integra el Índice de Movilidad Vehicular (IMV), permitiendo evaluar la movilidad en tiempo real y de forma interactiva (ver video funcional en la Figura 6.4).

- Características del prototipo: Visualización geoespacial: Utiliza la API de HERE Maps para representar datos en mapas interactivos.
- Actualización en tiempo real: Integra datos dinámicos para mantener el IMV actualizado, mejorando la precisión del análisis.
- Interfaz intuitiva: Ofrece herramientas visuales que permiten explorar la movilidad en distintos segmentos viales y analizar el impacto de variables como velocidad, congestión e incidencias.
- Escalabilidad: La arquitectura del sistema permite futuras mejoras, como la incorporación de modelos predictivos y análisis avanzados.
- Aplicaciones prácticas: Este prototipo facilita la gestión del tráfico y la planificación urbana, proporcionando información para la toma de decisiones estratégicas. Además, puede ser utilizado por autoridades de transporte, urbanistas y tomadores de decisiones para diseñar políticas públicas enfocadas en mejorar la movilidad urbana.



Figura 6.4. Código QR que enlaza al video funcional del prototipo Web para Evaluación de la Movilidad

6.6 Conclusiones

Los resultados obtenidos en esta investigación resaltan la eficacia de la metodología propuesta para modelar sistemas complejos altamente dinámicos, demostrando su capacidad para analizar diferentes problemas desde la perspectiva de los sistemas complejos. El desarrollo del Índice de Movilidad Vehicular (IMV), junto con los índices complementarios como el Índice de Calidad del Aire (ICA) y el Índice de Profundidad de los Sismos (IPS), válida la versatilidad del marco de trabajo adoptado y su aplicabilidad en contextos diversos.

Estos resultados no solo demuestran la capacidad de la metodología para desarrollar modelos matemáticos de aproximación al comportamiento de sistemas complejos, sino que también destacan su potencial para generar herramientas analíticas que facilitan la interpretación de fenómenos dinámicos y la toma de decisiones estratégicas. El IMV, por ejemplo, permite evaluar la movilidad urbana considerando variables críticas como la velocidad del tráfico, la ubicación geográfica y la incidencia de eventos disruptivos, mientras que el ICA y el IPS amplían el alcance de la metodología a ámbitos como la gestión ambiental y la sismología.

Asimismo, el desarrollo del prototipo web, que incorpora el Índice de Movilidad Vehicular (IMV), amplía significativamente las posibilidades de aplicación de los resultados obtenidos en esta investigación. Esta herramienta no solo permite la visualización y análisis en tiempo real de las condiciones de movilidad vehicular, sino que también facilita su integración en sistemas de gestión urbana y planificación estratégica, optimizando la toma de decisiones en contextos complejos.

En el caso específico del problema del sistema complejo de movilidad vehicular, el prototipo actúa como una plataforma interactiva que ayuda a identificar patrones de congestión, evaluar el impacto de incidentes y planificar intervenciones eficientes. Además, ofrece soporte para el diseño de políticas públicas orientadas a mejorar la fluidez del tránsito y reducir los tiempos de traslado. Su flexibilidad y escalabilidad permiten que esta herramienta se adapte a diferentes escenarios urbanos, potenciando su uso en la gestión de la movilidad en ciudades con dinámicas altamente variables y contribuyendo a la creación de entornos urbanos más sostenibles y eficientes.

Esta investigación no solo contribuye al entendimiento de los fenómenos analizados, sino que también ofrece soluciones concretas que pueden ser utilizadas en la mejora de servicios urbanos y en la toma de decisiones informadas. En definitiva, los hallazgos presentados consolidan el valor práctico de la metodología y abren nuevas oportunidades para futuras investigaciones en el ámbito del modelado de sistemas complejos.

Capítulo 7. Conclusiones generales y trabajos futuros

7.1 Introducción

Este capítulo resume los principales logros a lo largo de esta investigación, subrayando su importancia y aporte al análisis de sistemas complejos altamente dinámicos. Aquí se reflexiona sobre los objetivos alcanzados, los retos superados y las aplicaciones prácticas de los resultados obtenidos. Asimismo, se plantean posibles direcciones para futuras investigaciones, enfocadas en profundizar y expandir los conocimientos adquiridos, con la finalidad de perfeccionar y adecuar los modelos y metodologías propuestas a diferentes contextos y desafíos emergentes.

7.2 Conclusiones generales

Esta sección describe un análisis de los principales hallazgos y aportes generados a lo largo de esta investigación. En este apartado, se reflexiona sobre el cumplimiento de los objetivos planteados, destacando la contribución de la metodología desarrollada y los resultados obtenidos en el contexto del estudio de sistemas complejos altamente dinámicos. Asimismo, se examinan las implicaciones prácticas de los modelos y enfoques propuestos, subrayando su utilidad en escenarios reales y su potencial para orientar investigaciones futuras.

- El modelado de sistemas complejos altamente dinámicos es de vital importancia en ingeniería y ciencias aplicadas, ya que su comportamiento influye directamente en la toma de decisiones estratégicas en campos como la movilidad urbana, los sistemas financieros y los ecosistemas naturales. Sin embargo, el análisis de estos sistemas presenta retos significativos a causa de su complejidad intrínseca, como las interacciones no lineales, la multiescalaridad, la evolución temporal y la interdependencia de componentes. La metodología propuesta para el desarrollo de modelos de comportamiento de estos sistemas analiza estos desafíos con un enfoque integral, proporcionando un marco estructurado que facilita la identificación de variables, la integración de datos multiescalares, la adaptación a la evolución temporal y la representación precisa de las interacciones dinámicas entre componentes. Esto permite desarrollar modelos robustos, precisos y aplicables en diversos contextos.
- El trabajo de investigación ha logrado desarrollar una metodología integral para el modelado de sistemas complejos altamente dinámicos, basada en un enfoque multidisciplinario que combina ciencia de datos, minería de datos, indicadores compuestos y técnicas de desarrollo de software. Esta metodología no solo permite una representación precisa y eficiente del comportamiento sistémico, sino que también establece un marco robusto para la construcción y validación de modelos.
- La vocación de la metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos radica en su propósito como un marco de trabajo especializado, orientado a la construcción de modelos aproximados que permitan caracterizar y comprender los comportamientos de estos sistemas. Su enfoque está dirigido a representar con precisión las dinámicas y las interacciones entre las variables, ofreciendo herramientas prácticas que

favorecen tanto la caracterización de desempeño de los sistemas, identificación de patrones como la predicción de comportamientos en contextos altamente dinámicos y en constante cambio.

- La metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos actúa como una entrada estratégica para el desarrollo de nuevas líneas de investigación que exploren, por ejemplo, la automatización de procesos de modelado, la mejora en la capacidad de adaptación de los modelos a través de sistemas autoajustables o el diseño de algoritmos optimizados para manejar incertidumbres y cambios abruptos en los sistemas. Además, su enfoque en la caracterización de comportamientos a partir de datos reales abre la posibilidad de construir modelos más generalizables, transferibles entre dominios, y con un impacto directo en áreas como la gestión urbana, la sostenibilidad ambiental y la resiliencia ante eventos críticos.
- A partir del análisis estadístico de los resultados experimentales, se determinó que la hipótesis nula H_{01} debe ser rechazada, aceptándose en su lugar la hipótesis alternativa H_{11} , al evidenciarse una relación significativa entre el uso de marcos de trabajo, programación genética y minería de datos para el modelado sistemático de sistemas complejos altamente dinámicos. De igual forma, la hipótesis nula H_{02} fue rechazada, y se aceptó la hipótesis alternativa H_{12} , ya que los modelos construidos lograron caracterizar matemáticamente la interrelación entre variables latentes y observables, reflejando distintos niveles de comportamiento del sistema conforme a la variabilidad y dinámica de los datos.
- En conclusión, la metodología no solo se limita a su aplicabilidad en los casos de estudio abordados, sino que trasciende como una herramienta que inspira y guía la próxima generación de investigaciones, permitiendo la creación de paradigmas más robustos y avanzados en el modelado de sistemas complejos altamente dinámicos.
- El Índice de Movilidad Vehicular (IMV) se posiciona como una herramienta efectiva para analizar y comprender la dinámica del tráfico en entornos urbanos complejos. Este índice integra diversas variables, como la zona donde ocurre el tráfico, la velocidad, el día en que se registra la congestión y las incidencias relacionadas, ofreciendo una representación simplificada pero relevante del estado del tránsito vehicular. A través del análisis del IMV, es posible identificar patrones de congestión, zonas críticas y periodos de mayor afectación, proporcionando información valiosa para apoyar la planificación estratégica y la gestión eficiente del tráfico.
- Los modelos obtenidos y presentados en este trabajo, junto con los diversos experimentos realizados a lo largo de la investigación, muestran que una metodología integral permite construir modelos de comportamiento capaces de representar, con precisión y eficiencia, la dinámica e interacción entre las entidades de sistemas complejos altamente dinámicos, de forma verificable mediante criterios objetivos.

- La metodología desarrollada para modelar el comportamiento de sistemas complejos altamente dinámicos ha alcanzado un Nivel de Madurez Tecnológica TRL 6, conforme a los criterios establecidos por la escala de la NASA. Este nivel implicó la demostración de sistemas en entornos relevantes, lo cual quedó evidenciado mediante su aplicación en casos de estudio realizados en contextos reales. En estos escenarios, se verificó su capacidad para construir modelos precisos, eficientes y verificables que representan adecuadamente la dinámica no lineal e interactiva de los sistemas analizados. Los resultados obtenidos confirman su viabilidad técnica y su aplicabilidad práctica, consolidando un avance significativo respecto a las fases de validación teórica y en laboratorio. Este resultado posiciona a la metodología como una herramienta madura, confiable y con alto potencial de escalabilidad, transferencia y adaptación a distintos dominios donde el modelado de sistemas complejos represente un desafío estratégico.

7.3 Trabajos futuros

Esta sección ofrece una visión prospectiva sobre las posibilidades de investigación y desarrollo derivadas de este trabajo. Se destacan áreas prioritarias que demandan una mayor exploración, tanto para optimizar los enfoques propuestos como para extender su aplicabilidad a contextos y problemáticas más diversos. Entre los temas de interés resaltan la incorporación de tecnologías emergentes, la automatización de los procesos de modelado y el perfeccionamiento en la gestión de incertidumbres propias de los sistemas complejos altamente dinámicos.

- Incorporar en la metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos el concepto de transferencia de modelos, permitiendo aprovechar la capacidad de los modelos existentes para ser reutilizados o adaptados a otros sistemas complejos con características similares. Esto no solo optimizaría los recursos y esfuerzos en el desarrollo de nuevos modelos, sino que también ampliará la aplicabilidad y versatilidad de la metodología en diversos contextos.
- Evaluar los modelos generados mediante la metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos a través de la construcción de nuevas métricas. Estas métricas no solo deben medir la precisión de los modelos, sino también su capacidad para capturar dinámicas emergentes y su eficiencia computacional, garantizando un análisis integral de su desempeño y aplicabilidad.
- Diseñar y desarrollar modelos resilientes capaces de mantener su robustez y funcionalidad ante interrupciones, cambios drásticos en las dinámicas del sistema o la presencia de datos incompletos, asegurando su adaptabilidad y confiabilidad en escenarios complejos.
- En la etapa de desarrollo de modelos de comportamiento de la metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos, se propone construirlos utilizando diversas técnicas que generen expresiones simbólicas interpretables. Entre estas técnicas destacan la regresión

simbólica basada en aprendizaje automático, redes neuronales simbólicas, modelos de regresión mediante expansión funcional, optimización evolutiva simbólica, sistemas basados en gramáticas y métodos algebraicos computacionales.

- Transformar el Índice de Movilidad Vehicular en un proceso automatizado que pueda ejecutarse en segundo plano en aplicaciones de mapas, como Google Maps, Waze, Moovit, Here Maps y otras. Este proceso actuaría como un complemento que proporcione información sobre el nivel de congestión en zonas de interés, utilizando valores escalares. Dichos valores permiten interpretar el grado de congestión, basado en los niveles de las variables definitorias del Índice de Movilidad Vehicular. Esto facilita una integración efectiva en plataformas de navegación, mejorando la experiencia del usuario y apoyando la toma de decisiones en tiempo real.
- Para construir modelos de movilidad eficientes y robustos, es necesario integrar diversas fuentes de datos que proporcionen información completa y precisa sobre la movilidad. Estas fuentes incluyen cámaras de tráfico, sensores en carreteras, crowdsourcing, GPS de vehículos, teléfonos móviles, sistemas de transporte público, drones, estaciones meteorológicas, bases de datos históricas, redes sociales, sistemas de conteo de vehículos, sensores LIDAR y RADAR, imágenes satelitales, dispositivos IoT (Internet de las Cosas), aplicaciones móviles, sensores ambientales, datos de operadores logísticos, sensores de pasos peatonales y ciclovías, sistemas de videovigilancia urbana y sensores de estacionamiento inteligente. La combinación de estas herramientas permite comprender la complejidad del análisis de movilidad con mayor precisión y adaptabilidad.

7.4 Publicaciones

En esta sección se detallan las publicaciones de artículos en revistas de alto impacto indexadas en el JCR, así como en revistas indizadas y congresos especializados.

Publicación de artículos en JCR

- De la Cruz-Nicolás, E., Estrada-Esquivel, H., Martínez-Rebollar, A., Clemente, E., & Pliego-Martínez, O. A. (2025). Identification of variables influencing the origin of traffic congestion points. *Revista Facultad de Ingeniería Universidad de Antioquia*. doi: 10.17533/udea.redin.20250261
- De la Cruz-Nicolás, E., Estrada-Esquivel, H., Martínez-Rebollar, A., Pliego-Martínez, O. A., & Clemente, E. (2024). Index for assessing the performance level of vehicular traffic on urban streets. *Urban Science*, 8(4), 204. doi:10.3390/urbansci8040204
- De la Cruz-Nicolás, E., Esquivel, H. E., Rebollar, A. M., Pliego Martínez, O. A., & Hernández Pérez, M. Y. (2024). Representation of traffic congestion: A multicriteria analysis approach based on distances. *Programming and Computer Software*, 50(8), 599–611. doi:10.1134/s0361768824700646

- Pliego-Martínez, O., Martínez-Rebollar, A., Estrada-Esquivel, H., & De la Cruz-Nicolás, E. (2024). An integrated attribute-weighting method based on PCA and entropy: Case of study marginalized areas in a city. *Applied Sciences (Basel, Switzerland)*, 14(5), 2016. <https://doi.org/10.3390/app14052016>

Publicación de artículos en revistas indexadas

- De la Cruz-Nicolás, E., Estrada-Esquivel, H., Martínez-Rebollar, A., Pliego-Martínez, O. A., & Clemente, E. (2024). Modeling Vehicular Flow Using an Approach Based on the Navier-Stokes Equations.
- De la Cruz-Nicolás, E., Estrada-Esquivel, H., Martínez-Rebollar, A., Pliego-Martínez, O. A. (2024). Impact of the proximity of Points of Interest on traffic Congestion: A Kernel Density Approach. *Revista Digital de Tecnologías Informáticas y Sistemas*, 8(1), 2024. <https://doi.org/10.61530/redtis.vol8.n1.2024>
- Ernesto De la Cruz-Nicolás, Hugo Estrada Esquivel, Alicia Martínez-Rebollar, Odette Alejandra Pliego Martínez. (2024). Identificación de variables causales de la congestión de tráfico mediante la prueba de Granger. *Congreso Mexicano de Inteligencia Artificial 2024*.
- De la Cruz-Nicolás, E., Martínez-Rebollar, A., Estrada-Esquivel, H., Pliego-Martínez, O.A. (2024). Methodology to Obtain Traffic Data and Road Incidents Through Maps Applications. In: Nesmachnow, S., Hernández Callejo, L. (eds) *Smart Cities. ICSC-Cities 2023. Communications in Computer and Information Science*, vol 1938. Springer, Cham. https://doi.org/10.1007/978-3-031-52517-9_1.
- De-la-Cruz-Nicolás, E., Martínez-Rebollar, A., Estrada-Esquivel, H., & Ortiz-Hernández, J. (2022, October). Monitoring of traffic incidents in real time for decision making in road infrastructures. In *2022 10th International Conference in Software Engineering Research and Innovation (CONISOFT)* (pp. 112-120). 10.1109/CONISOFT55708.2022.00024
- De la Cruz Nicolás, Ernesto., Martínez Rebollar, Alicia, & Estrada Esquivel, Hugo. (2023). Analysis of the Behavior of Vehicular Traffic Flow through Classical Statistical Techniques. *European Modern Studies Journal*, Volumen 7, Páginas 348-359., ISSN 2522-9400, <https://doi.org/10.59573/emsj.7>.
- Lara, V. M. A., Miranda, M. L. A., De la Cruz Nicolás, E., & Barrera, J. H. J. (2022). Los repositorios digitales institucionales como un recurso educativo. *ANFEI Digital*, (14).
- Ernesto de la Cruz Nicolás, Mónica Leticia Acosta Miranda, Víctor Manuel Ayala Lara y C. Jesús Hazael Jiménez Barrera. (2021). Accesibilidad del Repositorio Digital de Contenidos Científicos y Tecnológicos para la inclusión con UX Research. *Academia Journals Hidalgo* 2021, 13 (9), 400-405.

Publicación de artículos de congreso

- Ernesto De la cruz-Nicolás, Odette Alejandra Pliego-Martínez, Jonathan Villanueva-Tavira, José Arnulfo Corona Calvario, Carlos Manuel Romano Acevedo. (2024). Recolección y almacenamiento de datos ambientales en la nube con ESP32 para Modelos Predictivos. 3er Congreso Internacional de Tecnología y Ciencias Aplicadas (CITCA).
- De la Cruz-Nicolás, E., Estrada-Esquivel, H., Martínez-Rebollar, A., Pliego-Martínez, O. A. (2024). Modelado de velocidades vehiculares en función de la congestión de tráfico utilizando la Función Logística. Congreso Internacional de Investigación e Innovación Multidisciplinaria (CIIIM 2024).
- De la Cruz-Nicolás, E., Estrada-Esquivel, H., Martínez-Rebollar. 2024. Presentación de la tesis Metodología para el desarrollo de modelos de comportamiento de sistemas complejos altamente dinámicos. International Doctoral Consortium at MICAI 2024.
- Pliego-Martínez, O., Martínez-Rebollar, A., Estrada-Esquivel, H., De la Cruz-Nicolás, E., & Neftalí-Escobar Gómez, E. (2024). Comparative analysis using distance measures in GIS: Accessibility to hospitals. 3er Congreso Internacional de Tecnología y Ciencias Aplicadas (CITCA).
- Pliego-Martínez, O., Martínez-Rebollar, A., Estrada-Esquivel, H., & De la Cruz-Nicolás, E. (2024). Identificación de patrones en la selección de variables mediante el análisis factorial: Marginación en localidades. Ponencia presentada en el Congreso Internacional de Investigación e Innovación Multidisciplinaria (CIIIM 2024).
- Ernesto De la Cruz-Nicolás, Hugo Estrada-Esquivel, Alicia Martínez-Rebollar, Odette Alejandra Pliego-Martínez, Rosa Olivia Maquinay-Díaz. (2024). Ponderación de variables causales de congestión de tráfico utilizando el método de entropía. 2º Congreso Internacional de Tecnología y Ciencia Aplicada 2024.
- Pliego Martínez, O.; Martínez Rebollar, A.; Estrada Esquivel, H.; De la Cruz-Nicolás, E. (2024). Visualización de extensos conjuntos de datos espaciales con QGIS: Caso de estudio CDMX. Tecnología y Ciencia Aplicada. ISSN: 2992-8338.
- Víctor Manuel Ayala Lara, Mónica Leticia Acosta Miranda, Ernesto De la Cruz Nicolás, Jesús Hazael Jiménez Barrera, Kevin Manuel Victoria Rodríguez, Hugo Armando Rodríguez Villalba. (2022). Etapas de Desarrollo del Repositorio Digital en el TecNM Campus Cuautla. Jornada de Ciencia y Tecnología Aplicada, 5(1).

Referencias

Abarbanel, H. D. I. (2013). Predicting the Future: Completing Models of Observed Complex Systems. Physical Review

Adams, K. M. (2012). Systems theory: a formal construct for understanding systems. *International Journal of System of Systems Engineering*, 3(3/4), 209. <https://doi.org/10.1504/ijssse.2012.052684>

Albert, R., Jeong, H., & Barabási, A. L. (2000). Error and attack tolerance of complex networks. *Nature*, 406(6794), 378–382. <https://doi.org/10.1038/35019019>

Albert-László Barabási, R. A. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509–512. <https://doi.org/10.1126/science.286.5439.509>

Andres, A., & Sandoval, G. (2024). Aplicación del aprendizaje supervisado para reconocer patrones de datos de ciberamenazas y aprendizaje no supervisado para analizar y agrupar conjuntos de datos.

Arti, C., G., Pradeep, K., P., & Kumar, S. S. (2022). Urban traffic congestion: its causes-consequences-mitigation. *Research Journal of Chemistry and Environment*, 26(12), 164–176. <https://doi.org/10.25303/2612rjce1640176>

Axelrod, R., & Tesfatsion, L. (2006). Appendix A A guide for newcomers to agent-based modeling in the social sciences. En *Handbook of Computational Economics* (pp. 1647–1659). Elsevier.

Bagdasaryan, A. (2012). Systems theoretic techniques for modeling, control and decision support in complex dynamic systems. En *Artificial Intelligence Resources in Control and Automation Engineering* (pp. 15–72). BENTHAM SCIENCE PUBLISHERS.). <https://doi.org/10.2174/978160805126711201010015>

Barabási, A.-L. (2012). *Network science*. Cambridge University Press Battiston, F., Amico, E., Barrat, A., Bianconi, G., Ferraz de Arruda, G., Franceschiello, B., Iacopini, I., Kéfi, S., Latora, V., Moreno, Y., Murray, M. M., Peixoto, T. P., Vaccarino, F., & Petri, G. (2021). The physics of higher-order interactions in complex systems. *Nature Physics*, 17(10), 1093–1098. <https://doi.org/10.1038/s41567-021-01371-4>

Bar-yam, Y. (1997). *Dynamics of Complex Systems*. Westview Press.

Baskerville, R., Georgia State University, Baiyere, A., Gergor, S., Hevner, A., Rossi, M., University of Turku, Finland, Australian National University, University of South Florida, & Aalto University. (2018). Design science research contributions: Finding a balance between artifact and theory. *Journal of the Association for Information Systems*, 19(5), 358–376. <https://doi.org/10.17705/1jais.00495>

Bila, J., Reshak, A. H., & Chysky, J. (2021). Modeling Complex Systems by Structural Invariants Approach. *Complexity*, 2021(1), 1–17. <https://doi.org/10.1155/2021/6650619>

Blasco-Blasco, O., Rodríguez-Castro, M., & Túnñez-López, M. (2020). Composite indicators as an innovative methodology for communication sciences: Implementation for the assessment of european public service media. *El Profesional de la Informacion*, 29(4), 1–13. <https://doi.org/10.3145/epi.2020.jul.37>

Blom, N., Hardalupas, P.-S., & Rawlinson, N. (2022). Mitigating the effect of errors in source parameters on seismic (waveform) tomography. *Geophysical Journal International*, 232(2), 810–828. <https://doi.org/10.1093/gji/ggac314>

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424(4–5), 175–308. <https://doi.org/10.1016/j.physrep.2005.10.009>

Bonaño, J. M. C. (2002). Indicadores de desarrollo sostenible urbano. una aplicación para andalucía . Universidad de Málaga.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5a ed.). John Wiley & Sons.

Ceiih, R. (2015). Complexity. A Guided Tour. *Inter disciplina*, 3(6). <https://doi.org/10.22201/ceiih.24485705e.2015.6.49482>

Cochran, W. G. (1977). *Sampling Techniques* (3a ed.). John Wiley & Sons

Crucitti, P., Latora, V., Marchiori, M., & Rapisarda, A. (2003). Efficiency of scale-free networks: Error and attack tolerance. *Physica a Statistical Mechanics and Its Applications*, 320, 622–642. [https://doi.org/10.1016/S0378-4371\(02\)01545-5](https://doi.org/10.1016/S0378-4371(02)01545-5)

Cruz, N. A. G. (2019). Simulación basada en agentes: Una metodología para el estudio de sistemas complejos. En *Construcción de problemas de investigación* (pp. 230–268). Fondo Editorial FCSH, Facultad de Ciencias Sociales y Humanas de la Universidad de Antioquia.

Dahia, S. S., & Szabo, C. (2024). Detecting Emergent Behavior in Complex Systems: A Machine Learning Approach. *ACM International Conference Proceeding Series*, 81–87. https://doi.org/10.1145/3615979.3656064_CONFERENCIA

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE transactions on evolutionary computation: a publication of the IEEE Neural Networks Council*, 6(2), 182–197. <https://doi.org/10.1109/4235.996017>

Dirk Helbing, S. B. (2012). Agent-based modeling. En *Understanding Complex Systems* (pp. 25–70). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24004-1_2

Dorogovtsev, S. N., & Mendes, J. F. F. (2002). Evolution of networks. *Advances in Physics*, 51(4), 1079–1187. <https://doi.org/10.1080/00018730110112519>

Doss, R., & Sharma, S. (Eds.). (2021). *The essence of chaos the essence of chaos*. Ink Gladiators Press.

Duan, J., Zeng, G., Serok, N., Li, D., Lieberthal, E. B., Huang, H. J., & Havlin, S. (2023). Spatiotemporal dynamics of traffic bottlenecks yields an early signal of heavy

congestions. *Nature Communications*, 14(1),8002. <https://doi.org/10.1038/s41467-023-43591-7>

Duncan J. Watts*, S. H. S. (2018). Collective dynamics of » small- world « networks. *Nature* 393, S. 440 – 442. En *Schlüsselwerke der Netzwerkforschung* (pp. 551–553). Springer Fachmedien Wiesbaden.

Duranton, G., & Turner, M. A. (2011). The fundamental law of road congestion: Evidence from US cities. *American Economic Review*, 101(6), 2616–2652. <https://doi.org/10.1257/aer.101.6.2616>

Espinosa Zúñiga, J. J. (2020). Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. *Ingeniería Investigación y Tecnología*, 21(1), 1–13. <https://doi.org/10.22201/fi.25940732e.2020.21n1.008>

Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. *Nature*, 460(7256), 685–686. <https://doi.org/10.1038/460685a>

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3–5), 75–174. <https://doi.org/10.1016/j.physrep.2009.11.002>

Gallardo-Pérez, H. de J., Vergel-Ortega, M., & Rojas-Suárez, J. P. (2020). Análisis dinámico de series temporales multivariadas. *Mundo FESC*, 10(20), 41–49. <https://doi.org/10.61799/2216-0388.618>

Gao, J., & Xu, B. (2021). Complex systems, emergence, and multiscale analysis: A tutorial and brief survey. *Applied Sciences (Switzerland)*, 11(12), 1–62. <https://doi.org/10.3390/app11125736>

George-Williams, H., Santhosh, T. V., & Patelli, E. (2022). Simulation methods for the analysis of complex systems. En *Uncertainty in Engineering* (pp. 95–113). Springer International Publishing.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley

Graham, S. E., Loveless, J. P., & Meade, B. J. (2021). A global set of subduction zone earthquake scenarios and recurrence intervals inferred from geodetically constrained block models of interseismic coupling distributions. *Geochemistry, Geophysics, Geosystems: G* (3), 22(11). <https://doi.org/10.1029/2021gc009802>

Gray, S., Gray, S., De Kok, J., Helfgott, A., O'Dwyer, B., & Jordan, R. (2013). *Mental Modeler*. Retrieved from <http://www.mentalmodeler.org>

Gregor, S., The Australian National University, Hevner, A. R., & University of South Florida. (2013). Positioning and presenting design science research for maximum impact. *MIS quarterly: management information systems*, 37(2), 337–355. <https://doi.org/10.25300/misq/2013/37.2.01>

Griffin, T. (1987). *Semiology of graphics: diagrams, networks, maps*. Bertin, Jacques (trans. W. j. Berg). The University of Wisconsin Press, Madison, 1983. 416 pages. ISBN 0 299 09060 4. US\$75.00, cloth. *Cartography*, 16(1), 81–82. <https://doi.org/10.1080/00690805.1987.10438353>

Grznár, P., Gregor, M., Mozol, Š., Mozolová, L., Krump, H., Mizerák, M., & Trojan, J. (2024). A Comprehensive Analysis of Sensitivity in Simulation Models for Enhanced System Understanding and Optimisation. *Processes*, 12(4), 716. <https://doi.org/10.3390/pr12040716>

Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction, second edition (2a ed.)*. Springer.

Helbing, D. (2001). Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*, 73(4), 1067–1141. <https://doi.org/10.1103/revmodphys.73.1067>

Helbing, D. (2014). *Social self-organization: Agent-based simulations and experiments to study emergent social behavior*. Springer.

Hevner, March, Park, & Ram. (2004). Design Science in Information Systems Research. *MIS quarterly: management information systems*, 28(1), 75. <https://doi.org/10.2307/25148625>

Hey, T. (2009). *The fourth paradigm: Data-intensive scientific discovery*. Microsoft Research.

Hofkirchner, W., & Schafranek, M. (2011). *General System Theory*. En *Philosophy of Complex Systems*, 177–194. Elsevier eBooks

Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press

Huang, Z., & Loo, B. P. Y. (2023). Urban traffic congestion in twelve large metropolitan cities: A thematic analysis of local news contents, 2009–2018. *International Journal of Sustainable Transportation*, 17(6), 592–614. <https://doi.org/10.1080/15568318.2022.2076633>

Iyengar, S. S., & Rao, M. S. (1983). Statistical Techniques in Modeling of Complex Systems: Single and Multiresponse Models. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-13(2), 175–189. <https://doi.org/10.1109/TSMC.1983.6313111>

Jin, G., Liu, L., Li, F., & Huang, J. (2023). Spatio-Temporal Graph Neural Point Process for Traffic Congestion Event Prediction. *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, AAAI 2023, 37, 14268–14276. <https://doi.org/10.1609/aaai.v37i12.26669>

John Guckenheimer, & Holmes, P. (1983). Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. In Home Healthcare Nurse (Vol. 5, Issue 1). <https://doi.org/10.1097/00004045-198701000-00010>

Kalnay, E. (2005). Atmospheric modeling, data assimilation and predictability. Cambridge University Press.

Kaplan, D., Glass, L., & Berger, S. A. (1996). Understanding Nonlinear Dynamics. Springer. <https://doi.org/10.1063/1.2807512>

Kates, R. W., & Wilbanks, T. J. (2003). A grand query: how scale matters in global change research. In *Global Change and Local Places* (pp. 3–26). Cambridge University Press.

Kerner, B. S. (2004). The Physics of Traffic: Empirical Freeway Pattern Features. Springer. <https://link.springer.com/book/10.1007/978-3-540-40986-1>

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1), 1–14. <https://doi.org/10.1007/s10708-013-9516-8>

Koulakov, I., D'Auria, L., Prudencio, J., Cabrera-Pérez, I., Barrancos, J., Padilla, G. D., Abramnikov, S., Pérez, N. M., & Ibáñez, J. M. (2023). Local earthquake seismic tomography reveals the link between crustal structure and volcanism in Tenerife (Canary Islands). *Journal of Geophysical Research. Solid Earth*, 128(3). <https://doi.org/10.1029/2022jb025798>

Koza, J. R. (1994). Genetic programming as a means for programming computers by natural selection. *Statistics and Computing*, 4(2), 87–112. <https://doi.org/10.1007/BF00175355>

Kutz, J. N. (2013). Data-driven modeling & scientific computation: Methods for complex systems & big data. Oxford University Press.

Ladyman, J., Lambert, J., & Wiesner, K. (2013). What is a complex system? *European Journal for Philosophy of Science*, 3(1), 33–67. <https://doi.org/10.1007/s13194-012-0056-8>

Levin, S. A. (1998). Ecosystems and the biosphere as complex adaptive systems. *Ecosystems* (New York, N.Y.), 1(5), 431–436. <https://doi.org/10.1007/s100219900037>

Li, C., Wang, D., Chen, H., & Liu, E. (2024). Analysis of Urban Congestion Traceability: The Role of the Built Environment. *Land*, 13(2), 255. <https://doi.org/10.3390/land13020255>

Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20(2), 130–141. [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2)

Ludwig von Bertalanffy, J. W. S. (1974). General System Theory: Foundations, Development, Applications. George Braziller Inc

Lyapunov, A. M. (1992). The general problem of motion stability. *International journal of control*, 55(3), 531–534. <https://doi.org/10.1080/00207179208934253>

- MacLeod, M. (2020). Model-Based Inferences in Modeling of Complex Systems. *Topoi*, 39(4), 915–925. <https://doi.org/10.1007/s11245-018-9569-x>
- Malberti, M. A., Beguerí, G. E., Klenzi, R. O., Ortega, M. O., Olguin, L. A., Amaya, F., & Cortez, J. (2021). Visualizando la información en ciencia de datos. XXIII Workshop de Investigadores, 296–300. <http://sedici.unlp.edu.ar/handle/10915/120126>
- May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, 261(1), 459–467. <https://doi.org/10.1038/261459a0>
- Méndez Gurrola, I. I., Ramírez Reyes, A., & Mora Gutiérrez, R. A. (2020). Aprendizaje automático aplicado en física: Una revisión de la literatura científica. *Research in Computing Science*, 149(8), 803–816. <https://cathi.uacj.mx/handle/20.500.11961/18429>
- Miller, J. G. (1975). The nature of living systems. *Behavioral Science*, 20(6), 343–365. <https://doi.org/10.1002/bs.3830200602>
- Mitchell, M. (2009) *Complexity: A guided tour*. Oxford England: Oxford University Press.
- Mohammed Almatar, K. (2023). Traffic congestion patterns in the urban road network: (Dammam metropolitan area). *Ain Shams Engineering Journal*, 14(3), 101886. <https://doi.org/10.1016/j.asej.2022.101886>
- Murray, D. (2002). *Mathematical Biology I: An Introduction*. Springer
- Nelson, E. S. (1996). How maps work: Representation, visualization, and design. *Cartographic perspectives*, 24, 27–30. <https://doi.org/10.14714/cp24.757>
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2), 167–256. <https://doi.org/10.1137/S003614450342480>
- Nguyen, L. K. N., Howick, S., & Megiddo, I. (2021). A hybrid simulation modelling framework for combining system dynamics and agent-based models. *Operational Research Society 10th Simulation Workshop, SW 2021 - Proceedings*, April, 385–394. <https://doi.org/10.36819/SW21.042>
- Nicolis, G., & Prigogine, I. (1989). *Exploring complexity: An introduction*. W.H. Freeman & Company.
- Oberkampff, W. L., & Roy, C. J. (2025). *Verification, validation, and uncertainty quantification in scientific computing (2a ed.)*. Cambridge University Press.
- OECD. (2008). *Handbook on Constructing Composite Indicators Methodology and user guide*. In OECD eBooks publishing. <https://doi.org/10.1787/9789264043466-en>
- Omari Firas. (2023). A combination of SEMMA & CRISP-DM models for effectively handling big data using formal concept analysis-based knowledge discovery: A data mining approach. *World Journal of Advanced Engineering Technology and Sciences*, 8(1), 009–014. <https://doi.org/10.30574/wjaets.2023.8.1.0147>

Orallo José Hernández, Ramírez Quintana M.José, & Ramírez César Ferri Ramírez. (2004). *Introducción a la Minería de Datos* (1ª ed.). Prentice Hall/Pearson

Ott, E. (2002). *Chaos in Dynamical Systems*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511803260>

Pardi, A. L., & Burina, E. (2024). Democratising complex system modelling. *Scientific Reports*, 14(1), 1–11. <https://doi.org/10.1038/s41598-024-61318-6>

Perveen, S., Yigitcanlar, T., Kamruzzaman, M., & Hayes, J. (2017). Evaluating transport externalities of urban growth: a critical review of scenario-based planning methods. *International Journal of Environmental Science and Technology*, 14(3), 663–678. <https://doi.org/10.1007/s13762-016-1144-7>

Peters, T. (2017). The structure of complex networks: theory and applications, by Ernesto Estrada. *Contemporary Physics*, 58(3), 287–287. <https://doi.org/10.1080/00107514.2017.1308969>

Poincaré, H. (1892). *Les méthodes nouvelles de la mécanique céleste*. Paris: Gauthier-Villars.

Prokopenko, M. (2014). *Guided self-organization: Inception*. Springer. <https://doi.org/https://doi.org/10.1007/978-3-642-53734-9> CON DOI

Qin, Y., Nakamura, Y., Kodaira, S., & Fujie, G. (2022). Seismic imaging of subsurface structural variations along the Japan trench south of the 2011 Tohoku earthquake rupture zone. *Earth and Planetary Science Letters*, 594(117707), 117707. <https://doi.org/10.1016/j.epsl.2022.117707>

Ramana, K., Srivastava, G., Kumar, M. R., Gadekallu, T. R., Lin, J. C. W., Alazab, M., & Iwendi, C. (2023). A Vision Transformer Approach for Traffic Congestion Prediction in Urban Areas. *IEEE Transactions on Intelligent Transportation Systems*, 24(4), 3922–3934. <https://doi.org/10.1109/TITS.2022.3233801>

Ramírez, A., Martínez, A., Quesada, C., & Jenkins, M. (2020). Uso de técnicas de minería de datos y aprendizaje automático para la detección de fraudes en estados financieros: un mapeo sistemático de literatura. *Revista Iberica de Sistemas y Tecnologías de Información*, 28(April), 97–109.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195–204. <https://doi.org/10.1038/s41586-019-0912-1>

Ruey S. Tsay, R. C. (2019). *Nonlinear Time Series Analysis*. Wiley. https://www.biodyn.ro/course/literatura/Nonlinear_Time_Series_Analysis_2019_Ruey_S._Tsay,_Rong_Chen.pdf

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global sensitivity analysis: The primer* (1a ed.). Wiley-Interscience.

- Sayama, H. (2015). Introduction to the Modeling and Analysis of Complex Systems. Open Suny. <https://open.umn.edu/opentextbooks/textbooks/233>
- Schuschny, A., & Soto, H. (2009). Guía metodológica Diseño de indicadores compuestos de desarrollo sostenible. Cepal. <https://hdl.handle.net/11362/3661>
- Simon, H. A. (1991). The architecture of complexity. En Facets of Systems Science (pp. 457–476). Springer US. https://doi.org/10.1007/978-1-4899-0718-9_31
- Spencer, D. (2009). Cities and complexity: Understanding cities with cellular automata, agent-based models, and fractals. *Journal of Architecture*, 14(3), 446–450. <https://doi.org/10.1080/13602360903028044>
- Sterman, J. D. (2006). Business Dynamics, System Thinking and Modeling for a Complex World. Irwin Professional Pub
- Strogatz, S. (2008). Nonlinear Dynamics and Chaos. Westview Press.
- Strogatz, S. H. (2015). Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering. CRC Press.
- Thompson, J. M. T., & Stewart, H. B. (2002). Nonlinear dynamics and chaos (2ª ed.). Taylor & Francis Group
- Turner, S., Hanel, R., & Klimekl, P. (2018). Introduction to the theory of complex systems. Oxford University Press. <https://doi.org/10.1093/oso/9780198821939.001.0001>
- Urbina-Nájera, A. B., Camino-Hampshire, J. C., & Cruz Barbosa, R. (2020). University dropout: Prevention patterns through the application of educational data mining. RELIEVE - Revista Electronica de Investigacion y Evaluacion Educativa, 26(1), 1–19. <https://doi.org/10.7203/relieve.26.1.16061>
- Wiener, N. (1985). Cybernetics: Or Control and Communication in the Animal and the Machine. In M.I.T. PRESS.
- Wieringa, R. J. (2014). Abductive Inference Design. En Design Science Methodology for Information Systems and Software Engineering (pp. 177–199). Springer Berlin Heidelberg.
- Yang, J., Purevjav, A. O., & Li, S. (2020). The marginal cost of traffic congestion and road pricing: Evidence from a natural experiment in Beijing. *American Economic Journal: Economic Policy*, 12(1), 418–453. <https://doi.org/10.1257/pol.20170195>
- Zhang, Q. X., Zhao, B. Y., He, W., Zhu, H. L., & Zhou, G. H. (2024). A behavior prediction method for complex system based on belief rule base with structural adaptive. *Applied Soft Computing*, 151, 111118. <https://doi.org/10.1016/j.asoc.2023.111118>

Zhao, D. (2001). Seismological structure of subduction zones and its implications for arc magmatism and dynamics. *Physics of the Earth and Planetary Interiors*, 127(1–4), 197–214. [https://doi.org/10.1016/s0031-9201\(01\)00228-x](https://doi.org/10.1016/s0031-9201(01)00228-x)

Zhao, T. (2024). Artificial Intelligence in Mathematical Modeling of Complex Systems., *ICST Transactions on e-Education and e-Learning*, 1–12. <https://doi.org/10.4108/eetel.5256>

Zhao, X., Hu, L., Wang, X., & Wu, J. (2022). Study on Identification and Prevention of Traffic Congestion Zones Considering Resilience-Vulnerability of Urban Transportation Systems. *Sustainability (Switzerland)*, 14(24), 16907. <https://doi.org/10.3390/su142416907>

Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2014). Urban computing: Concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology*, 5(3), 1–55. <https://doi.org/10.1145/2629592>