Revista Científica Multidisciplinaria LEMACIENCIAS

ISSN: 3103-1390

Geolocalización vehicular y análisis de movilidad urbana en ciudades cubanas: una aproximación metodológica

Vehicle geolocation and urban mobility analysis in Cuban cities: a methodological approach

Carlos Lázaro Jiménez Puerto[1*] https://orcid.org/0000-0001-8967-2935
carloslazaro1989@gmail.com
Gerardo Yero Morejón[1] https://orcid.org/0009-0004-4338-9228
gyero1993@gmail.com
Gabriel Ernesto Rodríguez Carmenate[1] https://orcid.org/0009-0001-7799-3544
gabrielerc64@gmail.com
Brayan Lazval Bello[1] https://orcid.org/0009-0003-4002-7929
brayanlasvalbello@gmail.com
Leandro Enrique González Ramos[1] https://orcid.org/0009-0004-7474-267X
legonzalezramos@gmail.com
[1]Universidad de Sancti Spíritus "José Martí Pérez", Facultad de Ciencias Técnicas y Económicas, Departamento de Ingeniería Informática, Cuba, Sancti Spíritus.

* Autor para correspondencia

Enviado: 05/11/2026
Aceptado: 23/11/2026
Publicado: 23/11/2026

RESUMEN

La disponibilidad de datos de geolocalización ha transformado el estudio de la movilidad urbana, pero su aprovechamiento en ciudades cubanas carece de un marco metodológico que reconozca las restricciones locales de datos, conectividad y cartografía. El objetivo de este trabajo fue proponer y someter a contraste experimental una aproximación metodológica para analizar la movilidad urbana mediante geolocalización vehicular. Se realizó una investigación de diseño metodológico con revisión analítica de literatura indexada y tres experimentos computacionales replicables, con semilla fija y datos sintéticos, sobre la asociación de posiciones a la red vial, la cobertura de flotas instrumentadas y el riesgo de reidentificación de trazas. Se formuló un procedimiento de cinco fases y seis dimensiones de indicadores. En los experimentos, la asociación basada en un modelo oculto de Markov superó al criterio del tramo más cercano solo cuando el ruido de posicionamiento era elevado (3,1 puntos porcentuales con 30 m), una flota restringida a corredores arteriales no superó el 20 % de cobertura de la red con ningún tamaño de flota, y cuatro puntos espaciotemporales bastaron para singularizar más del 97 % de las trazas incluso con resolución de 800 m. Se concluye que los límites principales son la composición de la flota, la calidad del posicionamiento y la privacidad, y que la validación con datos reales es la siguiente etapa.

Palabras clave: geolocalización vehicular; movilidad urbana; trayectorias GNSS; ciudades cubanas; simulación.

ABSTRAT

The availability of geolocation data has transformed the study of urban mobility, yet its use in Cuban cities lacks a methodological framework that acknowledges local constraints on data, connectivity and cartography. The aim of this study was to propose and experimentally test a methodological approach for analyzing urban mobility through vehicle geolocation. A methodological design study was carried out, combining an analytical review of indexed literature with three replicable computational experiments, using fixed seeds and synthetic data, on map matching, coverage of instrumented fleets and the re-identification risk of traces. A five-phase procedure and six

indicator dimensions were formulated. In the experiments, a hidden Markov model matched positions to the road network better than the nearest-segment criterion only under high positioning noise (3.1 percentage points at 30 m); a fleet restricted to arterial corridors never exceeded 20% network coverage regardless of fleet size; and four spatiotemporal points singled out more than 97% of traces even at 800 m resolution. It is concluded that fleet composition, positioning quality and privacy are the main limits, and that validation with real data is the next stage.

Keywords: vehicle geolocation; urban mobility; GNSS trajectories; Cuban cities; simulation.

I. INTRODUCCIÓN

La movilidad urbana es uno de los procesos que con mayor claridad expresan el funcionamiento de una ciudad, pues articula la distribución espacial de las actividades con los recursos materiales, energéticos e institucionales disponibles para sostenerlas. Durante las últimas décadas, la creciente disponibilidad de datos georreferenciados sobre el desplazamiento de personas y vehículos modificó la manera de estudiarla. Barbosa et al. (2018) describen cómo la expansión de conjuntos de datos geolocalizados permitió analizar de forma cuantitativa los patrones individuales y colectivos de movimiento y construir modelos capaces de reproducir sus regularidades espaciotemporales, con aplicaciones que abarcan la previsión del tráfico, la planificación urbana y la modelación epidemiológica. Desde la ingeniería informática, este cambio se tradujo en el desarrollo de técnicas para adquirir, depurar, almacenar y minar trayectorias (Zheng, 2015), de modo que la geolocalización vehicular dejó de ser una herramienta exclusivamente operativa para convertirse en una fuente de conocimiento sobre la ciudad.

Los desarrollos más recientes confirman esa tendencia. Wang et al. (2022) señalan que los avances en sensores y dispositivos móviles han incrementado de manera sin precedentes la disponibilidad de datos urbanos de trayectoria y, con ello, la demanda de procedimientos más eficientes para gestionarlos y analizarlos; su revisión abarca el preprocesamiento, el almacenamiento, las herramientas analíticas habituales y, por primera vez en la literatura de este tipo, el aprendizaje profundo sobre trayectorias. Sin embargo, el aprovechamiento de estas técnicas no es uniforme entre contextos. Gran parte de la literatura procede de entornos con abundancia de datos, cobertura cartográfica detallada y acuerdos comerciales o institucionales que facilitan el acceso a las trazas de posicionamiento. Markovic et al. (2019) lo ilustran al señalar que las agencias de transporte suelen adquirir los datos de trayectoria a proveedores externos y, por tanto, deben valorar de antemano su utilidad frente al costo de adquisición. Esta condición no puede trasladarse sin examen a ciudades donde la instrumentación es parcial, los recursos son limitados y la información debe producirse, en buena medida, dentro de las propias instituciones. En consecuencia, resulta pertinente preguntarse cómo adaptar el conocimiento acumulado en el análisis de trayectorias a las ciudades cubanas sin perder rigor metodológico.

La literatura científica aborda la movilidad desde dos tradiciones complementarias. La primera, de raíz física y computacional, la concibe como un fenómeno empíricamente observable cuyas leyes pueden describirse y modelarse. Barbosa et al. (2018) organizan este campo distinguiendo entre movilidad individual y poblacional, y entre movilidad de corto y de largo alcance, y reúnen bajo esa estructura tanto modelos generales como aplicaciones a escalas intraurbanas. La segunda tradición, de raíz urbanística y de políticas públicas, desplaza el énfasis del desplazamiento en sí hacia aquello que el desplazamiento permite alcanzar. Esta orientación se aprecia en la propuesta de Morris et al. (2015) para La Habana, que plantea una visión basada en el acceso y la habitabilidad en lugar de la simple expansión de la capacidad de circulación. Ambas tradiciones son relevantes para el presente trabajo, porque la geolocalización vehicular permite observar el desplazamiento, mientras que su interpretación para la decisión pública exige vincularlo con el acceso a servicios y oportunidades.

En el plano teórico, el trabajo de González et al. (2008) constituye un punto de inflexión. Al analizar las trayectorias de cien mil usuarios anonimizados de telefonía móvil durante seis meses, los autores mostraron que el movimiento humano exhibe un alto grado de regularidad espacial y temporal, con una distancia característica de desplazamiento propia de cada individuo y una probabilidad significativa de retorno a pocos lugares muy frecuentados, lo que contradecía los modelos de vuelo de Lévy y de caminata aleatoria que dominaban hasta entonces. Esta regularidad es la base que hace viable predecir y resumir la movilidad a partir de muestras parciales. Junto a ella, Barbosa et al. (2018) sistematizan los modelos de movilidad poblacional, entre los que figuran los modelos gravitacionales y los de oportunidades

intervinientes, que explican los flujos entre zonas a partir de la distancia y de la distribución de las oportunidades. Estos enfoques aportan los fundamentos para interpretar matrices de origen y destino derivadas de trayectorias vehiculares.

Con respecto a las dimensiones de análisis, Markovic et al. (2019) identifican seis áreas de la ingeniería del transporte en las que los datos de trayectoria han demostrado utilidad: la estimación de la demanda, la modelación del comportamiento humano, el diseño del transporte público, la medición y predicción del desempeño del tráfico, el impacto ambiental y la seguridad vial. Tomando esta clasificación como referencia y añadiendo una dimensión institucional y ética, imprescindible por las razones que se discuten más adelante, en este trabajo se organizan los indicadores de movilidad en seis dimensiones: espacial, temporal, operativa, de demanda, ambiental e institucional. Se trata de una síntesis analítica propia, y no de una taxonomía consolidada en la literatura, por lo que su pertinencia se somete a discusión en el apartado correspondiente.

Una trayectoria puede entenderse como la huella que un objeto en movimiento deja en el espacio a lo largo del tiempo (Wu y Xia, 2022). Zheng (2015) precisa que los avances en la adquisición de ubicación y en la computación móvil han generado volúmenes masivos de datos de trayectoria espacial que representan la movilidad de objetos tan diversos como personas, vehículos y animales, y propone una hoja de ruta que parte de la obtención de las trayectorias, continúa con su preprocesamiento y gestión, y desemboca en tareas de minería como la detección de patrones, la identificación de valores atípicos y la clasificación. El mismo autor destaca que las trayectorias pueden transformarse en otras representaciones, como grafos, matrices y tensores, a las que es posible aplicar técnicas adicionales de minería de datos y aprendizaje automático. Esta secuencia, que va del dato bruto al conocimiento, es el esqueleto conceptual sobre el cual se construye la propuesta de este artículo.

Las trazas vehiculares presentan características que condicionan su análisis. Markovic et al. (2019) señalan que los conjuntos de trayectoria incluyen trazas de GPS, registros de detalle de llamadas y datos de detectores Bluetooth y Wi-Fi, con diferente resolución espacial y temporal, y documentan el análisis visual de veinte millones de trazas de GPS en el estado de Maryland para ilustrar aplicaciones existentes y sugerir otras nuevas. Sin embargo, la calidad del posicionamiento no es homogénea. Kubička et al. (2018) explican que el ajuste de la posición de un vehículo a la red vial, conocido como map matching, es necesario porque los sistemas de posicionamiento resultan poco fiables en áreas urbanizadas, donde la visibilidad satelital es limitada y las señales reflejadas introducen ambigüedad en el proceso de asociación. En la misma línea, Gao et al. (2024) subrayan que los errores inherentes a los datos de GPS hacen indispensable este paso de preprocesamiento y que los edificios altos atenúan la exactitud del posicionamiento, por lo que proponen un sistema de puntuación multiobjetivo que integra criterios de penalización, geometría, topología y tiempo. Aunque ese estudio se refiere a datos de ciclismo, el problema que aborda es transferible al contexto vehicular, y su énfasis en la complejidad de la red vial sugiere que las zonas urbanas densas demandan procedimientos de ajuste más cuidadosos.

A estas consideraciones técnicas se suma la necesidad de interpretar y comunicar los resultados. Wu y Xia (2022) sostienen que el aprovechamiento de los datos de trayectoria requiere no solo métodos de análisis automáticos y eficientes, sino también visualizaciones intuitivas que faciliten su comprensión. Esta observación es relevante en un contexto institucional como el cubano, donde los resultados de un estudio de movilidad deben ser comprensibles para decisores que no necesariamente poseen formación en ciencia de datos.

El enfoque contemporáneo sitúa el análisis de movilidad en el marco más amplio de la ciudad instrumentada. Kitchin (2014) describe cómo las ciudades son equipadas con dispositivos e infraestructuras digitales que producen grandes volúmenes de datos, los cuales, según los defensores de la ciudad inteligente, permiten el análisis en tiempo real de la vida urbana y nuevas formas de gobernanza. Ese mismo autor advierte, sin embargo, sobre cinco preocupaciones que condicionan cualquier uso de datos urbanos, entre ellas la política de los datos masivos, la gobernanza tecnocrática y el riesgo de una ciudad panóptica. Asimismo, De Montjoye et al. (2013) demostraron, con quince meses de datos de movilidad de un millón y medio de personas, que bastan cuatro puntos espaciotemporales para identificar de manera única al 95 % de los individuos, incluso con una resolución temporal horaria y espacial equivalente a la de las antenas de una operadora, y que reducir la resolución ofrece escasa protección. Estos hallazgos obligan a tratar las trazas de posicionamiento como datos sensibles, aun cuando se hayan retirado los identificadores directos.

A la privacidad se suma el problema de la representatividad. Li et al. (2024) examinaron los sesgos de los datos de ubicación móvil de un proveedor comercial en Estados Unidos y advirtieron que los paneles de dispositivos rastreados no están bien documentados en cuanto a su representatividad, de modo que las interpretaciones basadas en ellos pueden ser cuestionables; además, hallaron que grupos como la población hispana, los hogares de bajos ingresos y las personas con menor nivel educativo mostraban niveles más altos de subrepresentación, con variaciones en el espacio, el tiempo y el grado de urbanización. Aunque ese estudio se refiere a dispositivos móviles y no a flotas vehiculares, su lección metodológica es transferible: cualquier conjunto de trazas es una muestra de quienes están instrumentados, y su composición condiciona lo que puede afirmarse sobre la ciudad.

En relación con el contexto cubano, la evidencia revisada se concentra en La Habana y en el plano estratégico. Warren et al. (2015) reportan un proyecto con casi cien participantes de Cuba y el Reino Unido, desarrollado durante doce meses a partir de junio de 2013, en el que se halló un alto grado de acuerdo en la comunidad de políticas de transporte sobre tres problemas clave de la capital: una elevada demanda insatisfecha, la falta de financiamiento y el mal estado del sistema de transporte. El estudio registró además un consenso casi general en que cualquier política futura debería mejorar la calidad del transporte público mediante mayor inversión, ahorros de eficiencia y una gestión más inteligente, aunque ese consenso se ve cuestionado por el deseo de desregular el mercado de automóviles. Morris et al. (2015) abordan esta tensión al señalar que las reformas económicas y la normalización de relaciones con Estados Unidos alimentaron expectativas de aumento de la propiedad de automóviles, y que La Habana puede evitar los errores observados en otras ciudades. Por su parte, Warren y Ortegon-Sanchez (2015) modelaron una ruta de autobús como caso de un sistema de autobuses de tránsito rápido, evaluando emisiones y capacidad, y concluyeron que un diseño adecuado podría mejorar la accesibilidad, la eficiencia en el consumo de combustible y la calidad del aire.

Este conjunto de estudios aporta diagnósticos y escenarios, pero en la revisión realizada para este trabajo no se identificó un marco metodológico que oriente la obtención y el procesamiento sistemático de datos de geolocalización vehicular para producir indicadores de movilidad en ciudades cubanas, ni evidencia comparable para ciudades distintas de la capital. Esta brecha motiva la pregunta de investigación: ¿qué configuración metodológica permite derivar indicadores de movilidad urbana confiables a partir de la geolocalización vehicular en ciudades cubanas, considerando las restricciones de datos, conectividad e infraestructura cartográfica? Para responderla se formulan tres hipótesis operativas, sometidas a contraste experimental en este artículo. La primera sostiene que, bajo ruido de posicionamiento elevado y muestreo disperso, un modelo probabilístico de asociación a la red que incorpora la topología alcanza mayor exactitud que el criterio del tramo más cercano, aunque la ventaja disminuye cuando el ruido es bajo. La segunda plantea que la cobertura de la red depende de la composición de la flota instrumentada y no solo de su tamaño, de modo que una flota restringida a corredores arteriales tiene un techo de cobertura que no se supera añadiendo vehículos. La tercera establece que las trazas vehiculares siguen siendo singulares con pocos puntos incluso con resolución espacial baja, por lo que la agregación espacial por sí sola no basta como protección.

El objetivo general es proponer y someter a contraste experimental una aproximación metodológica para el análisis de la movilidad urbana mediante geolocalización vehicular en ciudades cubanas. Como objetivos específicos se plantean: fundamentar conceptualmente las fases, dimensiones e indicadores del procedimiento; cuantificar, en un entorno simulado y controlado, el efecto del ruido y del muestreo sobre la asociación a la red vial; estimar el efecto del tamaño y la composición de la flota sobre la cobertura y el error de los indicadores; y evaluar el riesgo de reidentificación de las trazas según su resolución. La estrategia combina una revisión analítica de la literatura con un diseño por fases y con experimentos computacionales replicables.

II. METODOLOGÍA

Se desarrolló una investigación de diseño metodológico con enfoque mixto: un componente cualitativo-documental, orientado a fundamentar el procedimiento, y un componente experimental computacional, orientado a contrastar tres hipótesis mediante simulación. Se optó por datos sintéticos porque el estudio no contó con trazas reales de ciudades cubanas, y porque la simulación permite conocer la verdad de referencia, es decir, el tramo y la velocidad reales de cada vehículo, lo cual es imposible con datos de campo sin un sistema de verificación independiente. Esta elección implica limitaciones que se discuten explícitamente y condiciona el alcance de las conclusiones: los experimentos evalúan la

sensibilidad de las decisiones metodológicas a condiciones controladas y no estiman parámetros de ninguna ciudad concreta.

Se realizó una revisión analítica no sistemática de la literatura científica indexada, con prioridad para artículos de revisión, estudios metodológicos y trabajos fundacionales sobre cuatro ejes: la movilidad humana y sus modelos, la minería y el preprocesamiento de trayectorias, el ajuste de posiciones a la red vial y los problemas de privacidad, representatividad y gobernanza de los datos urbanos. Se incorporaron, además, estudios sobre transporte en Cuba y en La Habana. Se incluyeron únicamente trabajos con identificador verificable, ya sea un DOI o un enlace de acceso estable, y se excluyeron los documentos sin revisión científica. Dado que la búsqueda no siguió un protocolo sistemático, los resultados se interpretan como una síntesis orientadora y no como un mapa exhaustivo del conocimiento disponible.

A partir de la secuencia propuesta por Zheng (2015), que va de la obtención de la trayectoria a su minería, y de las áreas de aplicación sistematizadas por Markovic et al. (2019), se estructuraron las fases del proceso, se definieron las dimensiones e indicadores y se identificaron los requisitos de datos y las limitaciones previsibles de cada uno. Las decisiones sobre preprocesamiento se fundamentaron en la literatura sobre map matching (Gao et al., 2024; Kubička et al., 2018), y las relativas a la gobernanza, en los trabajos de Kitchin (2014) y De Montjoye et al. (2013). Los supuestos de diseño referidos a las condiciones locales se formularon como hipótesis de trabajo explícitas y no como hechos establecidos, de modo que puedan contrastarse en estudios posteriores.

Se construyó una ciudad sintética de trama regular con 15 × 15 intersecciones y tramos de 200 m, lo que equivale a una superficie aproximada de 2,8 km por lado y a 420 tramos no dirigidos. Cada sexta línea de la red, es decir, las líneas 0, 5 y 10 en cada sentido, se definió como arteria con velocidad de flujo libre de 40 km/h, y el resto de la red se definió con 25 km/h. A cada tramo se le asignó un factor persistente de congestión extraído de una distribución lognormal de parámetros −0,25 y 0,30, acotado entre 0,25 y 1,3, y a cada recorrido individual se le añadió una variación lognormal de parámetros 0 y 0,15, acotada entre 0,5 y 1,5. Los vehículos se desplazan entre orígenes y destinos aleatorios mediante rutas en forma de L, y las observaciones de posicionamiento se generan interpolando la posición real a intervalos regulares y añadiendo ruido gaussiano bidimensional de desviación típica σ. Se trata de una abstracción deliberadamente simple, que prescinde de calles irregulares, sentidos de circulación, semáforos y cañones urbanos, y cuyas consecuencias para la validez externa se discuten más adelante.

El primer experimento contrastó la hipótesis sobre la asociación a la red vial. Se compararon dos procedimientos: el criterio del tramo más cercano y un modelo oculto de Markov resuelto mediante el algoritmo de Viterbi, con probabilidad de emisión gaussiana según la distancia al tramo candidato, un radio de candidatos igual al mayor valor entre 60 m y 3,5σ, y una probabilidad de transición que penaliza la diferencia entre la distancia por la red, estimada como el número de saltos entre tramos multiplicado por la longitud del tramo, y la distancia entre observaciones consecutivas. Esta formulación sigue la lógica de los métodos probabilísticos revisados por Kubička et al. (2018), pero es una versión simplificada y no una réplica de ningún algoritmo publicado. Se evaluaron dieciséis condiciones, resultado de combinar σ de 5, 10, 20 y 30 m con intervalos de muestreo de 5, 15, 30 y 60 s, con seis réplicas por condición, veinte vehículos por réplica y 900 s de simulación por vehículo. La variable de respuesta fue la exactitud de asociación, definida como la proporción de observaciones asignadas al tramo en el que el vehículo se encontraba realmente.

El segundo experimento contrastó la hipótesis sobre cobertura y representatividad. Se simuló durante 3600 s una población de 600 vehículos con rutas aleatorias, cuyas velocidades por tramo constituyeron la verdad de referencia, calculadas mediante la velocidad media armónica de la Ecuación (1). Se extrajeron flotas instrumentadas de 5, 10, 20, 40, 80 y 160 vehículos bajo dos escenarios: una muestra aleatoria de la población y una flota de corredores, cuyos vehículos recorren de extremo a extremo las arterias, como lo haría un servicio de transporte público de ruta fija. Se consideró que un tramo estaba cubierto cuando acumulaba al menos tres observaciones, y el error se calculó como el error porcentual absoluto medio de la velocidad estimada respecto a la de referencia, únicamente en los tramos cubiertos. Se ejecutaron treinta réplicas independientes. Para aislar el efecto de la composición de la flota se supuso asociación perfecta a la red, de modo que el experimento no se contamina con los errores del primero.

El tercer experimento contrastó la hipótesis sobre privacidad. Se generaron las posiciones horarias de 800 vehículos durante cinco días, con doce horas diarias, bajo un modelo en el que cada vehículo posee cuatro nodos ancla, que visita con probabilidad de 0,6 en cada hora, y se mueve a un nodo aleatorio en caso contrario. Las posiciones se agregaron en

celdas de 200, 400 y 800 m, con resolución temporal horaria. Para cada resolución espacial y para un número p de puntos conocidos de 1 a 4 se extrajeron 200 vehículos al azar, y de cada uno se seleccionaron p registros espaciotemporales también al azar; la traza se consideró singular cuando ningún otro vehículo coincidía con todos esos puntos. El procedimiento se repitió diez veces. Este diseño reproduce, de forma deliberadamente simplificada, la lógica de la medida de singularidad de De Montjoye et al. (2013), y los valores absolutos dependen del modelo de anclas, por lo que solo la tendencia es interpretable.

Todos los experimentos se ejecutaron con generadores pseudoaleatorios de semilla fija, y el código completo, escrito en Python con la biblioteca NumPy, se entrega como material suplementario para permitir la reproducción exacta de los resultados. En cuanto a las consideraciones éticas, el estudio no implicó recolección de datos personales, pues todas las trazas son sintéticas. Se establecieron, no obstante, principios de diseño para la aplicación futura del procedimiento a datos reales, entre ellos la minimización de los datos, la agregación espacial y temporal antes de cualquier difusión y la obtención del consentimiento institucional de los operadores de los vehículos, responsabilidad que corresponde a quienes ejecuten la investigación empírica.

III. RESULTADOS Y DISCUSIÓN

El primer resultado es un procedimiento de cinco fases encadenadas con retroalimentación, representado en la Figura 1. La primera fase, de definición de objetivos y unidades de análisis, establece qué decisión pública se pretende informar, qué tipo de vehículos conformarán la muestra y qué escala territorial y temporal es pertinente. Esta fase es determinante porque, como muestran los resultados del segundo experimento, el alcance de las conclusiones depende de quién esté instrumentado. La segunda fase, de adquisición, contempla el registro de posiciones mediante receptores del Sistema Global de Navegación por Satélite (GNSS) instalados en vehículos de flotas institucionales o de transporte público, con una frecuencia de muestreo acordada de forma explícita entre resolución analítica y capacidad de almacenamiento y transmisión, y con el registro de los metadatos necesarios para evaluar posteriormente la calidad de la traza.



Figura 1. *Procedimiento metodológico de cinco fases con retroalimentación*.
Fuente: *Elaboración propia.*

La tercera fase agrupa el preprocesamiento, que sigue la lógica de Zheng (2015) de depurar y estructurar la trayectoria antes de minarla. Incluye la eliminación de posiciones imposibles, la segmentación de la traza en viajes y paradas y el ajuste de las posiciones a la red vial mediante algoritmos de map matching. La elección del algoritmo no es trivial, dado que Kubička et al. (2018) clasifican los métodos en familias con diferentes supuestos y los evalúan con una metodología unificada que permite comparar sus resultados; por ello, la propuesta establece que la selección se realice con una muestra de validación local. La cuarta fase deriva los indicadores, organizados en seis dimensiones que se resumen en la Tabla 1, y la quinta fase se ocupa de la validación y la gobernanza del dato, incluyendo la contrastación de los indicadores con mediciones independientes, la documentación de la calidad y la definición de reglas de acceso.

Tabla 1. *Dimensiones, indicadores y requisitos de la aproximación propuesta*.

Dimensión	Indicador propuesto	Requisitos de datos	Limitaciones previsibles
Espacial	Radio de giro y zonas de actividad por vehículo	Trayectorias con muestreo regular durante varios días	Sesgo hacia las flotas instrumentadas

Revista Científica Multidisciplinaria LEMACIENCIAS
ISSN: 3103-1390

Dimensión	Indicador propuesto	Requisitos de datos	Limitaciones previsibles
Temporal	Variación horaria de la velocidad por tramo	Marcas de tiempo sincronizadas y red vial digital	Sensibilidad a la frecuencia de muestreo
Operativa	Tiempo de recorrido y variabilidad por ruta	Identificador de ruta y de paradas	Requiere un inventario de rutas actualizado
Demanda	Matrices de origen y destino por zonas agregadas	Segmentación de viajes y zonificación	Describe viajes de vehículos, no de personas
Ambiental	Estimación de emisiones a partir de perfiles de velocidad	Perfiles de velocidad y factores de emisión externos	Dependencia de factores de emisión no locales
Institucional	Cobertura, calidad y trazabilidad del dato; riesgo de reidentificación	Metadatos y protocolo de gobernanza	Exige acuerdos interinstitucionales

Fuente: *Elaboración propia a partir de Markovic et al. (2019).*

A manera de ilustración de la fase cuatro, la velocidad media de un tramo s en una franja horaria h se calcula como el cociente entre la suma de las longitudes recorridas y la suma de los tiempos empleados en las observaciones ajustadas a ese tramo, según la Ecuación (1). Esta definición, basada en sumas y no en promedios de velocidades instantáneas, atenúa el sesgo que introducen las observaciones muy breves o con paradas, y es la utilizada como estimador en el segundo experimento.

$$v_{s,h} = \Sigma\, l_{s,i} / \Sigma\, t_{s,i} \quad (1)$$

Los indicadores propuestos no son novedosos en sí mismos, pues retoman áreas de aplicación ya documentadas (Markovic et al., 2019), y la contribución radica en su organización, en la explicitación de los requisitos y límites de cada uno y en la cuantificación experimental de tres de esos límites, que se presenta a continuación. El radio de giro y las zonas de actividad por vehículo encuentran respaldo en la regularidad del movimiento descrita por González et al. (2008), aunque dicha regularidad se documentó en usuarios de telefonía móvil y no en flotas vehiculares. Las matrices de origen y destino pueden interpretarse con ayuda de los modelos gravitacionales y de oportunidades intervinientes sistematizados por Barbosa et al. (2018), siempre que se reconozca que describen desplazamientos de vehículos y no de personas.

Para que los indicadores sean replicables, cada uno requiere una definición operativa que no deje margen a interpretaciones. La cobertura de la red se define como la proporción de tramos que acumulan al menos m observaciones válidas en la franja analizada, con m igual a tres en los experimentos, y se reporta siempre junto con el indicador al que acompaña. La variabilidad del tiempo de recorrido por ruta se expresa mediante el coeficiente de variación de los tiempos observados en una franja horaria, de modo que una ruta con tiempo medio aceptable pero alta dispersión se distinga de una ruta estable, distinción especialmente relevante para la fiabilidad del servicio. El radio de giro de un vehículo, empleado por González et al. (2008) para caracterizar la extensión espacial de la movilidad individual, se calcula como la raíz cuadrada de la distancia cuadrática media entre las posiciones observadas y su centro de masa, y permite distinguir vehículos de operación local de vehículos de operación extendida. Las matrices de origen y destino se construyen asignando el inicio y el final de cada viaje segmentado a zonas de análisis predefinidas, y se agregan por franja horaria para evitar la exposición de recorridos individuales. Finalmente, el indicador institucional combina la cobertura, la proporción de observaciones descartadas en el preprocesamiento y la documentación del riesgo de reidentificación, de forma que el lector pueda ponderar la fiabilidad del resultado.

La implementación del procedimiento exige, además, una decisión sobre la arquitectura de gestión de los datos, que en contextos de recursos limitados debe ser deliberadamente modesta. Wang et al. (2022) revisan los componentes de un sistema de gestión y análisis de trayectorias, que abarcan el preprocesamiento, el almacenamiento, las consultas y el análisis, y concluyen señalando las cualidades esenciales que un sistema de este tipo debería poseer para maximizar su flexibilidad. Trasladado a la escala de una flota institucional, ello sugiere un diseño en capas: una capa de datos brutos de acceso restringido, una capa de trayectorias depuradas y asociadas a la red, y una capa de indicadores agregados que es la única que se comparte. Esta separación no es solo una decisión de ingeniería, pues materializa en la arquitectura el principio de minimización de datos y reduce la superficie de riesgo señalada por De Montjoye et al. (2013). Cada capa debería conservar los metadatos de calidad de la capa anterior, incluida la incertidumbre de las posiciones, un aspecto que Zheng (2015) incluye entre los temas centrales de la minería de trayectorias.

La justificación de las seis dimensiones merece una precisión adicional. Las cinco primeras corresponden a lo que los datos de trayectoria permiten observar y se alinean con las áreas de aplicación sistematizadas por Markovic et al. (2019), mientras que la sexta responde a lo que los datos exigen de quienes los usan. Esta asimetría es deliberada: en la mayoría de los estudios la calidad y la gobernanza del dato se tratan como una sección de limitaciones, y aquí se tratan como un conjunto de indicadores con definición operativa, para que puedan medirse, compararse entre ciudades y exigirse en los acuerdos de cooperación. Que esta decisión sea acertada es una cuestión empírica que solo el estudio piloto puede responder, y es probable que la experiencia con datos reales aconseje fusionar o subdividir alguna de las dimensiones.

Por último, la secuencia de fases no debe entenderse como un proceso lineal. La retroalimentación representada en la Figura 1 indica que los resultados de validación pueden obligar a revisar la definición de objetivos, como ocurriría si la cobertura obtenida fuese insuficiente para el nivel de desagregación inicialmente previsto, o a modificar la frecuencia de muestreo, si el error de asociación resultara incompatible con el uso pretendido de los indicadores. La propuesta prevé, por tanto, que cada ciclo documente las decisiones tomadas y los motivos, de manera que la trazabilidad metodológica sea tan explícita como la trazabilidad del dato.

Experimento 1: asociación de posiciones a la red vial

La exactitud de asociación disminuyó de forma sistemática al aumentar el ruido de posicionamiento (Tabla 2 y Figura 2). Promediada sobre los cuatro intervalos de muestreo, la exactitud del criterio del tramo más cercano fue de 95,7 % con σ = 5 m, de 91,1 % con 10 m, de 82,1 % con 20 m y de 72,4 % con 30 m, mientras que la del modelo de Markov fue de 95,9 %, 91,5 %, 83,4 % y 75,5 %, respectivamente. La ventaja del modelo probabilístico fue, por tanto, de 0,2, 0,4, 1,4 y 3,1 puntos porcentuales, lo que indica que crece con el ruido. En el conjunto de las dieciséis condiciones, la exactitud media fue de 85,3 % para el tramo más cercano y de 86,6 % para el modelo de Markov. La desviación típica entre réplicas osciló entre 0,002 y 0,020.

Tabla 2. Exactitud de asociación a la red vial (tramo más cercano / modelo de Markov) según σ e intervalo de muestreo.

σ (m)	5 s	15 s	30 s	60 s
5	0,954 / 0,968	0,959 / 0,964	0,958 / 0,948	0,957 / 0,956
10	0,909 / 0,935	0,912 / 0,914	0,912 / 0,894	0,911 / 0,917
20	0,816 / 0,868	0,822 / 0,817	0,820 / 0,822	0,824 / 0,831
30	0,723 / 0,776	0,725 / 0,732	0,729 / 0,760	0,719 / 0,752

Fuente: *Elaboración propia a partir de simulación (medias de seis réplicas).*

El efecto del intervalo de muestreo fue menos nítido que el del ruido. La mayor ventaja del modelo de Markov se observó con muestreo de 5 s, con diferencias de 5,2 puntos porcentuales con σ = 20 m y de 5,3 con σ = 30 m, y el modelo resultó inferior al criterio del tramo más cercano en cuatro de las dieciséis condiciones, con una diferencia máxima desfavorable de 1,7 puntos porcentuales (σ = 10 m, intervalo de 30 s). Estos resultados respaldan solo parcialmente la primera hipótesis: la ventaja del modelo con topología existe y aumenta con el ruido, pero es modesta en la mayor parte del rango explorado y no es uniforme.

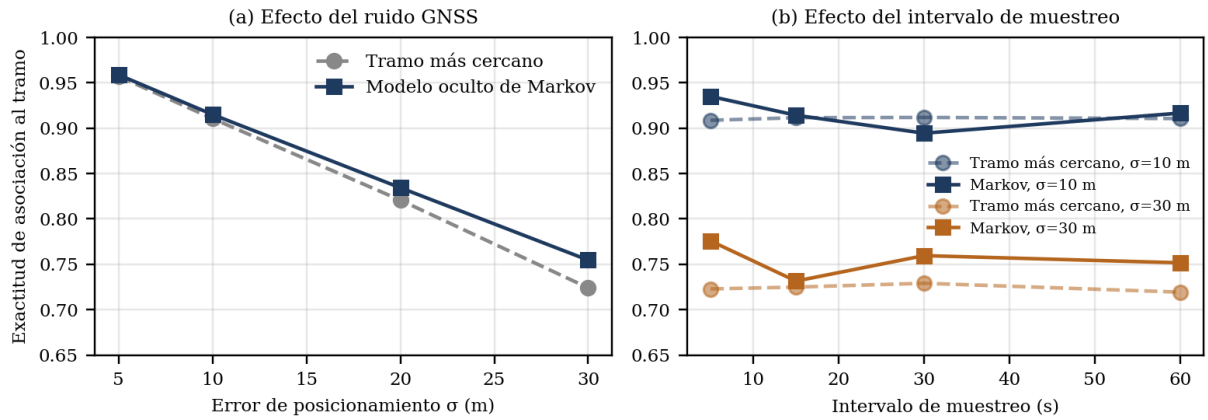


Figura 2. *Exactitud de asociación a la red vial según el ruido de posicionamiento y el intervalo de muestreo*. (a) *Promedio sobre intervalos de muestreo*; (b) *efecto del intervalo con σ = 10 y 30 m*.
Fuente: *Elaboración propia a partir de simulación.*

Experimento 2: tamaño y composición de la flota instrumentada

La muestra aleatoria de vehículos alcanzó una cobertura creciente con el tamaño de la flota: 14,2 % de los tramos con cinco vehículos, 41,3 % con diez, 74,9 % con veinte, 93,3 % con cuarenta, 99,2 % con ochenta y 100 % con 160 (Tabla 3 y Figura 3). El error porcentual absoluto medio en los tramos cubiertos descendió simultáneamente de 6,3 % a 1,8 %. La flota de corredores arteriales mostró un comportamiento diferente: su cobertura fue de 19,9 % con cinco vehículos y de 20,0 % con diez o más, es decir, la fracción de tramos que pertenecen a las arterias, y no aumentó al añadir vehículos. El error en los tramos cubiertos fue, en cambio, menor que el de la muestra aleatoria con tamaños pequeños, y descendió de 4,3 % con cinco vehículos a 1,3 % con 160.

Tabla 3. *Cobertura de la red y error de la velocidad estimada según el tamaño y la composición de la flota.*

Vehículos	Aleatoria: cobertura (%)	Aleatoria: error (%)	Corredores: cobertura (%)	Corredores: error (%)
5	14,2	6,3	19,9	4,3
10	41,3	6,0	20,0	3,1
20	74,9	5,1	20,0	2,3
40	93,3	4,0	20,0	1,8
80	99,2	2,9	20,0	1,5
160	100,0	1,8	20,0	1,3

Fuente: *Elaboración propia a partir de simulación (medias de treinta réplicas).*

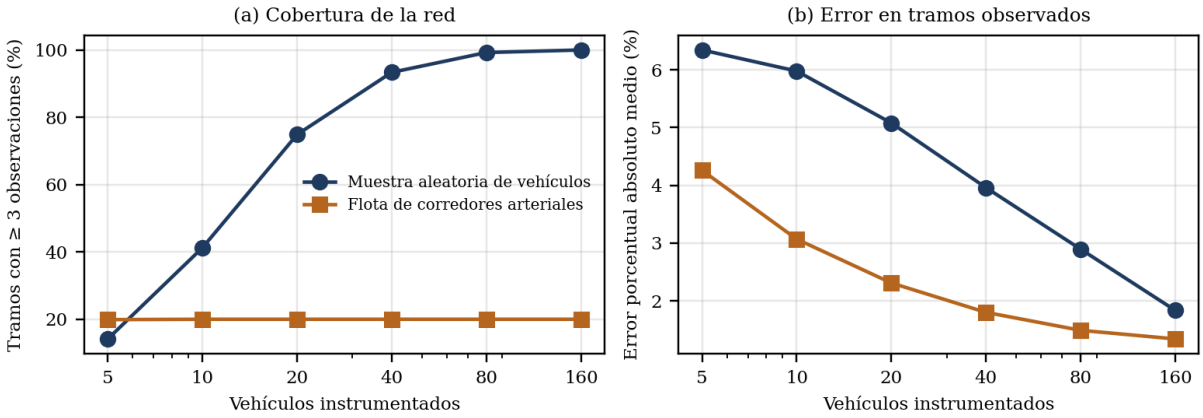


Figura 3. *Cobertura de la red y error de la velocidad estimada según el tamaño y la composición de la flota. (a) Cobertura; (b) error en los tramos cubiertos.*
Fuente: *Elaboración propia a partir de simulación.*

Estos resultados respaldan la segunda hipótesis. La precisión de los indicadores en los tramos observados mejora con el tamaño de la flota en ambos escenarios, pero solo en el escenario de muestra aleatoria la flota llega a observar toda la red. Una flota de corredores produce estimaciones aparentemente más precisas, porque concentra sus observaciones en pocos tramos, a costa de dejar sin información al 80 % de la red.

Experimento 3: singularidad de las trazas

Con resolución espacial de 200 m, un solo punto singularizó al 2,4 % de las trazas, dos puntos al 96,0 % y tres o cuatro puntos a más del 99,9 % (Tabla 4). Con 400 m, un punto no singularizó ninguna traza, dos puntos singularizaron el 74,7 % y tres puntos el 99,3 %. Con 800 m, dos puntos singularizaron el 5,1 %, tres puntos el 75,2 % y cuatro puntos el 97,4 %. La reducción de la resolución espacial aumentó, por tanto, el número de puntos necesario para singularizar una traza, pero no impidió que cuatro puntos bastaran para identificar a la inmensa mayoría de los vehículos en todas las resoluciones.

Tabla 4. *Porcentaje de trazas singularizadas según la resolución espacial y el número de puntos conocidos*.

Resolución espacial	1 punto	2 puntos	3 puntos	4 puntos
200 m	2,4	96,0	100,0	99,9
400 m	0,0	74,8	99,3	99,9

Resolución espacial	1 punto	2 puntos	3 puntos	4 puntos
800 m	0,0	5,1	75,2	97,4

Fuente: *Elaboración propia a partir de simulación (medias de diez réplicas).*

Este resultado respalda la tercera hipótesis en cuanto a la tendencia, aunque con la salvedad ya señalada de que los valores absolutos dependen del modelo de anclas empleado, que supone rutinas muy marcadas y una población pequeña de 800 vehículos.

Antes de interpretar los resultados conviene delimitar con precisión la contribución que este trabajo reclama. No se afirma haber descubierto leyes de la movilidad de las ciudades cubanas, algo que exigiría datos reales. La contribución es de tres órdenes:

- Un procedimiento por fases que integra, en una secuencia única y explícita, la adquisición, el preprocesamiento, la derivación de indicadores y la gobernanza del dato, con una dimensión institucional que la literatura de trayectorias rara vez trata como parte del propio método.
- Una cuantificación de sensibilidad que muestra, bajo condiciones controladas, hasta qué punto tres decisiones de diseño, la calidad del posicionamiento, la composición de la flota y la resolución de los datos, determinan la validez de los indicadores.
- Un simulador abierto y reproducible que permite a otros equipos repetir los experimentos con parámetros propios, lo que convierte las hipótesis aquí contrastadas en puntos de partida y no en conclusiones cerradas.

El conjunto de resultados permite una valoración matizada de las tres hipótesis. La primera se confirma parcialmente, la segunda se confirma y la tercera se confirma en su tendencia. Más allá del contraste formal, lo que interesa discutir es qué implican estos hallazgos para la propuesta metodológica en el contexto cubano y qué límites tienen.

En cuanto a la asociación a la red vial, el hallazgo más relevante no es la superioridad del modelo de Markov, sino su magnitud limitada. Con ruido bajo, la diferencia respecto del criterio más simple es prácticamente nula, y con ruido alto, aunque aumenta, la exactitud absoluta sigue siendo baja: con $\sigma = 30$ m, aun el mejor procedimiento asoció correctamente apenas tres de cada cuatro observaciones. La conclusión práctica es que, en la trama simulada, la calidad del receptor y de su entorno de señal pesa más que la sofisticación del algoritmo. Esto tiene una consecuencia de ingeniería concreta: antes de invertir en algoritmos complejos, conviene caracterizar el error real de los receptores disponibles mediante pruebas con posiciones conocidas. Esta lectura es coherente con la advertencia de Kubička et al. (2018) sobre la degradación de los sistemas de posicionamiento en áreas urbanizadas y con la de Gao et al. (2024) sobre el efecto de los edificios altos, aunque ninguno de esos trabajos permite extrapolar magnitudes a las ciudades cubanas.

El resultado no monótono con el intervalo de muestreo merece una explicación cautelosa. Que el modelo de Markov rinda peor que el criterio simple en algunas condiciones de muestreo intermedio sugiere que la penalización de transición empleada puede estar mal calibrada para esos intervalos, ya que a 30 s un vehículo recorre una distancia comparable a la longitud de un tramo y la estimación de saltos es más ambigua. Se trata de una hipótesis explicativa y no de un resultado, y la propia sensibilidad observada refuerza el argumento metodológico de seleccionar y calibrar el algoritmo con una muestra de validación local en lugar de adoptarlo por su difusión en la literatura. También limita la generalidad de los resultados el hecho de que el algoritmo evaluado sea una versión simplificada, de modo que los resultados no deben interpretarse como un juicio sobre los métodos probabilísticos publicados.

Los resultados del segundo experimento muestran que la representatividad no se resuelve con volumen. Añadir vehículos a una flota de corredores redujo el error en los tramos que ya observaba, pero no extendió la cobertura más allá de las arterias, de modo que los indicadores resultantes describen el desempeño de una parte de la red, y su aparente precisión puede inducir a una falsa sensación de conocimiento del conjunto. Este hallazgo dialoga con la advertencia de Li et al. (2024) acerca de que cualquier panel de dispositivos o vehículos rastreados es una muestra no necesariamente representativa y de que sus sesgos varían en el espacio y en el tiempo. En el contexto estudiado, la implicación operativa es que una flota de transporte público de ruta fija, que es la fuente de datos más probable en una ciudad con instrumentación limitada, debe complementarse con otras fuentes, como vehículos de servicios que recorran zonas no arteriales, para estimar el desempeño de la red completa; y que todo indicador debe reportarse junto con su cobertura.

Conviene precisar un límite adicional que los experimentos no abordan. Un sistema de geolocalización vehicular observa únicamente a los vehículos instrumentados y, por definición, no observa a quienes se desplazan a pie o en bicicleta. La proporción de viajes no motorizados en las ciudades cubanas debe medirse mediante encuestas y no puede suponerse, y la propuesta no pretende sustituirlas. Los indicadores resultantes describen el desempeño del sistema vial y del transporte motorizado instrumentado, y no la movilidad urbana en su conjunto, por lo que la dimensión de demanda se formula como aproximación a los viajes de vehículos. Quien utilice estos indicadores para decidir sobre accesibilidad peatonal incurriría en una extrapolación injustificada.

El tercer experimento confirma, en un entorno controlado, la lógica de De Montjoye et al. (2013): la singularidad de las trazas es tan elevada que la agregación espacial retrasa, pero no evita, la identificación. Debe subrayarse que las magnitudes obtenidas no son comparables con las de ese estudio, pues aquí se simuló una población pequeña con rutinas muy marcadas, y que lo transferible es la dirección del efecto y la forma de la relación entre resolución y número de puntos. De ello se desprende que la agregación, aun siendo necesaria, debe complementarse con controles institucionales sobre quién accede a los datos brutos y con qué finalidad. Kitchin (2014) aporta el marco crítico de esta discusión, al advertir sobre la gobernanza tecnocrática y la ciudad panóptica. En un país donde los datos se generan principalmente en el seno de entidades estatales, el riesgo no es tanto la corporatización de la información urbana como la ausencia de reglas explícitas de uso y de rendición de cuentas. La inclusión de una dimensión institucional en el conjunto de indicadores responde a la convicción de que la trazabilidad del dato y la gestión del riesgo de reidentificación forman parte del resultado científico y no de un trámite administrativo posterior.

Desde la perspectiva de las políticas, los resultados del análisis solo adquieren valor si dialogan con las prioridades de decisión. Warren et al. (2015) hallaron un amplio acuerdo en que la mejora del transporte público requiere inversión, ahorros de eficiencia y una gestión más inteligente. Los indicadores operativos de la propuesta, como el tiempo de recorrido y su variabilidad por ruta, podrían aportar evidencia cuantitativa a la segunda y la tercera de esas líneas, al permitir identificar tramos y franjas horarias donde la pérdida de eficiencia es mayor. De manera análoga, el modelado de una ruta de autobús de tránsito rápido realizado por Warren y Ortegon-Sanchez (2015) requirió supuestos sobre emisiones y capacidad que, en principio, podrían calibrarse con perfiles de velocidad observados, aunque esta posibilidad es una inferencia de este trabajo y no un resultado de dichos autores. Debe subrayarse, no obstante, que los datos de geolocalización no resuelven por sí solos el problema que Warren et al. (2015) identifican como primero, la demanda insatisfecha, pues una demanda que no se materializa en viajes no deja traza alguna.

Existe también una tensión de fondo entre la visión de acceso y habitabilidad propuesta por Morris et al. (2015) y la lógica de medición de la velocidad y el flujo vehicular. Un indicador de velocidad media creciente podría interpretarse como mejora, aun cuando refleje un cambio hacia un modelo más dependiente del automóvil, cuyos elevados costos para las ciudades los propios autores advierten. Para evitar esa lectura, la propuesta recomienda no presentar los indicadores de velocidad de forma aislada, sino acompañarlos de medidas de cobertura del transporte público y de la proporción de viajes que no pueden observarse. Esta precaución se alinea con la crítica de Kitchin (2014) sobre la veracidad de los datos y la manera en que se interpretan y se actúa a partir de ellos.

Respecto de la transferibilidad, Markovic et al. (2019) construyeron su revisión desde la perspectiva de una agencia que compra datos a proveedores y dispone de decenas de millones de trazas, un contexto que difiere del cubano en la escala y en la fuente de los datos. Una flota instrumentada de tamaño modesto no ofrece la cobertura estadística que sustenta muchas de las técnicas de minería descritas por Zheng (2015) y Wang et al. (2022), de modo que parte de ellas, particularmente las que requieren grandes volúmenes para entrenar modelos de aprendizaje profundo, pueden resultar inaplicables o insuficientes. La regularidad individual identificada por González et al. (2008) sugiere que pocas trayectorias bien observadas pueden ser informativas, pero la naturaleza de aquel estudio, con cien mil usuarios, advierte contra la extrapolación directa a muestras pequeñas. La propuesta toma entonces una posición prudente: privilegia indicadores robustos ante muestras reducidas, como la velocidad media y la fiabilidad del tiempo de recorrido por ruta, sobre aquellos que exigen modelos de mayor complejidad.

Un último aspecto de la discusión concierne a las condiciones institucionales de aplicación. La propuesta presupone que las entidades que operan los vehículos pueden y desean compartir sus trazas con un equipo de investigación, lo cual exige acuerdos interinstitucionales que no son un detalle administrativo sino una condición de viabilidad. En esos acuerdos conviene explicitar los usos permitidos de los datos, los productos que se entregarán a la entidad operadora, como informes de desempeño por ruta, y las salvaguardas para los conductores, cuyas trayectorias son datos personales

aun cuando el vehículo pertenezca a una institución. La experiencia internacional recogida por Kitchin (2014) muestra que la legitimidad de los sistemas de datos urbanos depende en buena medida de que los beneficios sean visibles para quienes los alimentan, y de que exista transparencia sobre su finalidad. Trasladado a este contexto, ello sugiere que los primeros productos del sistema deberían ser útiles para los propios operadores, por ejemplo mediante la identificación de tramos de pérdida sistemática de tiempo, antes de ampliarse a usos de planificación más generales.

Finalmente, deben reconocerse las limitaciones del estudio, que son sustanciales. Los experimentos se realizaron sobre una ciudad sintética de trama regular, sin calles irregulares, sentidos de circulación, semáforos ni cañones urbanos, de modo que es esperable que la ambigüedad de asociación sea mayor en redes reales, y en particular en centros históricos de calles estrechas, circunstancia que debe verificarse empíricamente y no darse por supuesta. La verdad de referencia se define por construcción dentro del mismo modelo que genera los datos, lo que favorece la coherencia interna a costa de la validez externa. El algoritmo de asociación es una versión simplificada, el número de réplicas es moderado y no se emplearon datos reales de ninguna ciudad cubana. La revisión de la literatura fue analítica y no sistemática, de manera que pueden existir trabajos relevantes, incluidos estudios cubanos de acceso restringido, que no fueron considerados. Por todo ello, los resultados deben leerse como evidencia sobre la sensibilidad de las decisiones metodológicas, y no como estimaciones aplicables a una ciudad concreta.

Las recomendaciones para investigaciones futuras se derivan de esas limitaciones:

1. Implementar un estudio piloto en una ciudad intermedia con una flota de transporte público, que permita contrastar los resultados sintéticos con trazas reales.
2. Caracterizar el error de posicionamiento de los receptores disponibles mediante pruebas con posiciones conocidas, para alimentar el simulador con parámetros locales.
3. Comparar al menos dos algoritmos de asociación a la red publicados con una muestra de verdad terreno local.
4. Evaluar la cuantificación de la parte de la movilidad no observada mediante encuestas complementarias.

Dado que los resultados proceden de simulación, la propuesta incluye un protocolo para su validación con datos reales, cuyo diseño es parte de la contribución metodológica. El primer paso es la selección de una ciudad intermedia y de una flota con acceso institucional garantizado, preferiblemente un conjunto de rutas de transporte público complementado con vehículos de servicios que recorran calles no arteriales, para que el resultado del segundo experimento pueda contrastarse con una composición de flota distinta. Se recomienda documentar desde el inicio las características de la flota, su número, su distribución por ruta y su horario de operación, porque son los factores que, según los experimentos y según lo señalado por Li et al. (2024), condicionan la representatividad.

El segundo paso es la caracterización del error de posicionamiento. Antes de procesar trazas de operación, los receptores deben someterse a pruebas con posiciones conocidas, tanto en puntos estáticos como en recorridos de referencia, en entornos abiertos y en calles estrechas o con edificaciones continuas. Estas pruebas proporcionan estimaciones locales del parámetro σ, que reemplazan los valores hipotéticos de 5 a 30 m empleados en la simulación, y permiten identificar zonas donde la asociación a la red será menos fiable. Con esos valores se recalibra el simulador y se vuelve a ejecutar el primer experimento, de modo que la selección del algoritmo de asociación se apoye en condiciones representativas del lugar.

El tercer paso es la construcción de una verdad terreno local. Una muestra de recorridos, que puede obtenerse con vehículos de prueba que registren simultáneamente la secuencia de tramos recorridos mediante anotación manual o un receptor de mayor calidad, permite calcular la exactitud de asociación real de al menos dos algoritmos, entre ellos uno de la familia probabilística descrita por Kubička et al. (2018). Los criterios de aceptación de la exactitud deben fijarse antes de analizar los resultados y justificarse en función del uso previsto de los indicadores, pues una decisión sobre la regulación horaria de una ruta tolera menos error que un diagnóstico exploratorio.

El cuarto paso es la derivación de los indicadores y su contraste con mediciones independientes, como aforos manuales de velocidad o tiempos de recorrido cronometrados en una muestra de tramos. Esta comparación permite estimar el sesgo y la variabilidad de los estimadores en condiciones reales, y se acompaña de un análisis de sensibilidad al intervalo de muestreo y al tamaño de la flota similar al realizado en este trabajo. El quinto paso es la evaluación de la gobernanza: antes de la difusión de cualquier resultado debe documentarse quién accede a los datos brutos, con qué finalidad y durante cuánto tiempo, y debe aplicarse la agregación espacial y temporal, reconociendo, a la luz del tercer experimento

y de los hallazgos de De Montjoye et al. (2013), que esta medida es necesaria pero no suficiente. Los resultados del protocolo deben reportarse con sus limitaciones, incluidos los tramos y franjas en los que la cobertura no alcanzó el mínimo definido, para evitar que la ausencia de datos se interprete como ausencia de problema.

IV. CONCLUSIONES

La geolocalización vehicular ofrece un camino viable para producir conocimiento sobre la movilidad urbana en ciudades cubanas, pero su valor depende de que el procedimiento reconozca explícitamente las condiciones en que se aplica. La aproximación propuesta, estructurada en cinco fases y seis dimensiones de indicadores, articula el flujo que va del dato de posicionamiento al conocimiento, descrito por la literatura de minería de trayectorias, con exigencias específicas de calidad, representatividad y gobernanza.

Los experimentos computacionales permiten formular tres conclusiones acotadas a las condiciones simuladas. La calidad del posicionamiento condiciona la asociación a la red más que la sofisticación del algoritmo, pues un modelo probabilístico con topología solo mejoró de forma apreciable al criterio simple cuando el ruido era elevado. La composición de la flota condiciona la cobertura más que su tamaño, ya que una flota restringida a corredores arteriales no superó el 20 % de cobertura con ningún número de vehículos. Y la agregación espacial no protege por sí sola la privacidad, dado que cuatro puntos espaciotemporales singularizaron a la inmensa mayoría de las trazas con todas las resoluciones evaluadas.

En consecuencia, los indicadores deben presentarse siempre acompañados de su cobertura y calidad, y los datos de trayectoria deben someterse a controles institucionales de acceso. La limitación principal del estudio es que sus resultados proceden de simulación y no de datos reales, por lo que la validación empírica mediante un estudio piloto constituye el requisito indispensable para convertir esta aproximación metodológica en una herramienta de apoyo a la decisión.

Disponibilidad de datos y código

Los datos del estudio son sintéticos y se generan íntegramente con el código suplementario (sim.py), escrito en Python con NumPy y ejecutable con las semillas indicadas. No se utilizaron datos personales ni trazas reales.

V. REFERENCIAS BIBLIOGRÁFICAS

Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., & Tomasini, M. (2018). Human mobility: Models and applications. Physics Reports, 734, 1–74. https://doi.org/10.1016/j.physrep.2018.01.001

De Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. Scientific Reports, 3, Article 1376. https://doi.org/10.1038/srep01376

Gao, T., Daamen, W., Krishnakumari, P., & Hoogendoorn, S. (2024). Map-matching for cycling travel data in urban area. IET Intelligent Transport Systems, 18(11), 2178–2203. https://doi.org/10.1049/itr2.12567

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. Nature, 453(7196), 779–782. https://doi.org/10.1038/nature06958

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

Kubička, M., Mounier, H., & Niculescu, S.-I. (2018). Comparative study and application-oriented classification of vehicular map-matching methods. IEEE Intelligent Transportation Systems Magazine, 10(2), 150–166. https://xplorestaging.ieee.org/document/8344804

Li, Z., Ning, H., Jing, F., & Lessani, M. N. (2024). Understanding the bias of mobile location data across spatial scales and over time: A comprehensive analysis of SafeGraph data in the United States. PLOS ONE, 19(1), Article e0294430. https://doi.org/10.1371/journal.pone.0294430

Marković, N., Sekuła, P., Vander Laan, Z., Andrienko, G., & Andrienko, N. (2019). Applications of trajectory data from the perspective of a road transportation agency: Literature review and Maryland case study. IEEE Transactions on Intelligent Transportation Systems, 20(5), 1858–1869. https://arxiv.org/pdf/1708.07193

Morris, E., Ortegon-Sanchez, A., & Warren, J. P. (2015). Movilidad sostenible: Alternativa viable para las ciudades cubanas. Temas, 83, 36–43. https://discovery.ucl.ac.uk/id/eprint/1520821/

Wang, S., Bao, Z., Culpepper, J. S., & Cong, G. (2022). A survey on trajectory data management, analytics, and learning. ACM Computing Surveys. https://doi.org/10.1145/3440207

Warren, J. P., Morris, E., Enoch, M., Padilla Magdaleno, I., Parra Arias, Z., & Guanche, J. (2015). Developing an equitable and sustainable mobility strategy for Havana. Cities, 45, 133–141. https://doi.org/10.1016/j.cities.2015.02.007

Warren, J. P., & Ortegon-Sanchez, A. (2015). Designing and modelling Havana's future bus rapid transit. Proceedings of the Institution of Civil Engineers – Urban Design and Planning, 169(2), 104–119. https://doi.org/10.1680/jurdp.15.00015

Wu, X., & Xia, L. (2022). Review and analysis on trajectory big data. E3S Web of Conferences, 338, Article 01039. https://doi.org/10.1051/e3sconf/202233801039

Zheng, Y. (2015). Trajectory data mining: An overview. ACM Transactions on Intelligent Systems and Technology, 6(3), Article 29. https://doi.org/10.1145/2743025