UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS

Facultad de Ciencias Matemáticas y Naturales

Programa Académico de Física

Simulación computacional del sistema de transporte masivo TransMilenio

Un enfoque desde la física estadística

Laura Daniela Torres Torres

20212107038

Director externo: PhD. J. Alfonso Leyva

Director interni: PhD. J. Miguel Espitia

Modalidad: **Pasantía**

Septiembre 2026

Resumen

Este trabajo desarrolla un modelo físico-estadístico del sistema de transporte masivo TransMilenio de Bogotá, con el objetivo de caracterizar la congestión desde la perspectiva de la mecánica estadística y de formular estrategias de optimización fundamentadas. El modelo integra dos componentes: la estimación de la demanda de viajes mediante el principio de máxima entropía que conduce a un modelo gravitacional doblemente restringido, formalmente análogo a la distribución de Boltzmann, en el que el parámetro de sensibilidad al costo se interpreta como una temperatura inversa y la simulación del flujo resultante sobre la red real de la ciudad, reconstruida a partir de datos geográficos y operacionales, mediante el simulador de código abierto UXsim. A partir de un barrido del factor de demanda se caracteriza la transición entre el régimen de flujo libre y el régimen congestionado como una transición de fase, tomando la proporción de viajes completados como parámetro de orden y localizando la demanda crítica que separa ambos regímenes. El análisis de distintos escenarios revela que únicamente la demanda del día laboral supera el umbral crítico, y que el efecto de las condiciones de lluvia depende de la proximidad del sistema a dicho umbral. Finalmente, la evaluación cuantitativa de propuestas de optimización demuestra que, si bien las intervenciones individuales resultan insuficientes, una estrategia integral que combina el refuerzo de la capacidad, la priorización semafórica y la gestión de la demanda desplaza la demanda crítica en un orden del veintisiete por ciento, aunque sin eliminar un límite físico irreductible bajo demandas extremas.

Palabras clave: física estadística, máxima entropía, transición de fase, transporte masivo, TransMilenio, simulación de tráfico, congestión.

Abstract

This work develops a statistical-physical model of Bogotá's TransMilenio mass transit system, aiming to characterize congestion from the perspective of statistical mechanics and to formulate well-founded optimization strategies. The model integrates two components: the estimation of travel demand through the maximum entropy principle —which yields a doubly-constrained gravity model, formally analogous to the Boltzmann distribution, in which the cost sensitivity parameter is interpreted as an inverse temperature— and the mesoscopic simulation of the resulting flow over the city's real network, reconstructed from geographic and operational data, using the open-source simulator UXsim. By sweeping the demand factor, the transition between the free-flow and congested regimes is characterized as a phase transition, taking the proportion of completed trips as the order parameter and locating the critical demand that separates both regimes. The analysis of different scenarios reveals that only the weekday demand exceeds the critical threshold, and that the effect of rain conditions depends on the system's proximity to that threshold. Finally, the quantitative evaluation of optimization proposals shows that, while individual interventions are insufficient, an integral strategy combining capacity reinforcement, signal prioritization, and demand management shifts the critical demand by around twenty-seven percent, although without eliminating an irreducible physical limit under extreme demand

Keywords: statistical physics, maximum entropy, phase transition, mass transit, TransMilenio, traffic simulation, congestion.

Índice

1. Introducción **4**
1.1. Planteamiento del problema 4
1.2. Antecedentes 4
1.2.1. Los modelos hidrodinámicos 4
1.2.2. El modelo de autómata celular de Nagel y Schreckenberg 5
1.2.3. El modelo de máxima entropía de Wilson 6

2. Fundamentos Termodinámicos **7**
2.1. Las leyes de la termodinámica 7

3. Mecánica Estadística **9**
3.1. Microestados, macroestados y espacio de fases 9
3.2. El postulado fundamental y la entropía de Boltzmann 9
3.3. El ensamble canónico y la distribución de Boltzmann 10

4. El Principio de Máxima Entropía **11**
4.1. Entropía de la información e inferencia 11
4.2. Justificación del principio 11
4.3. Formulación matemática mediante multiplicadores de Lagrange 11

5. La analogía física **13**
5.1. Fundamento conceptual 13
5.2. Correspondencia formal 13
5.3. Alcance y límites de la analogía 14

6. Aplicación a la Distribución de Viajes **15**
6.1. Conteo de configuraciones 15
6.2. Aproximación de Stirling y entropía 15
6.3. Restricciones y maximización 15

7. La Congestión como Fenómeno Crítico **17**
7.1. Transiciones de fase y parámetro de orden 17
7.2. La transición de fase en el flujo de tráfico 17
7.3. El parámetro de orden del sistema de transporte 18

8. El Simulador UXsim **19**
8.1. El modelo de ondas cinemáticas 19
8.2. El modelo de Newell y la resolución mesoscópica 19
8.3. Modelo de nodos, semáforos y elección de ruta 20
8.4. Herramientas de análisis y visualización 20

9. Objetivos **22**
9.1. Objetivo General 22
9.2. Objetivos Específicos 22

10.Plan de trabajo **23**

11. Metodología **24**
11.1. Fuentes de datos . . . 24
11.2. Proyección geográfica . . . 25
11.3. Construcción de la malla física de la red . . . 25
11.3.1. Reconstrucción de los corredores . . . 26
11.3.2. Composición de la malla y conectividad . . . 26
11.3.3. Expansión de los recorridos . . . 27
11.4. Incorporación de los semáforos . . . 28
11.5. Modelo de demanda por máxima entropía . . . 29
11.5.1. Matriz de costos . . . 29
11.5.2. Totales de origen y destino . . . 29
11.5.3. Calibración del parámetro de sensibilidad . . . 29
11.6. Acoplamiento demanda-simulación y escenarios . . . 30

12. Resultados y Discusión **32**
12.1. Calibración del modelo de demanda . . . 32
12.2. Escenario de referencia: día laboral en condiciones secas . . . 32
12.2.1. Evolución espacial de la congestión . . . 33
12.2.2. Análisis de un tramo congestionado . . . 33
12.3. Comparación de escenarios . . . 34
12.4. Transición de fase y demanda crítica . . . 36

13. Optimización del Sistema **39**
13.1. Metodología de la evaluación . . . 39
13.2. Evaluación individual de las propuestas . . . 40
13.3. Escenario de optimización integral . . . 43
13.4. Interpretación de la optimización . . . 45
13.5. Líneas de investigación futura . . . 46

14. Conclusiones **48**

A. Anexo **52**
A.1. Configuración, geometría y lectura de datos . . . 52
A.2. Construcción de la malla física . . . 54
A.3. Modelo de demanda por máxima entropía . . . 56
A.4. Construcción del mundo, demanda y simulación . . . 56
A.5. Barrido de la transición de fase y optimización . . . 57

1. Introducción

La movilidad urbana constituye uno de los sistemas complejos más determinantes para el funcionamiento de las ciudades contemporáneas, en tanto articula la actividad económica, el acceso a servicios y la calidad de vida de millones de personas. Bogotá, con una población que casi supera los ocho millones de habitantes en el área urbana, enfrenta un desafío de movilidad a diario, en el cual el sistema de transporte masivo TransMilenio opera como la columna vertebral del transporte público de la ciudad. Inaugurado el 18 de diciembre del año 2000, este sistema de buses de tránsito rápido (*Bus Rapid Transit*, BRT) moviliza en la actualidad más de dos millones de pasajeros diarios en su componente troncal, a través de aproximadamente 1.100 kilómetros de vías exclusivas, once troncales en operación, 155 estaciones y nueve portales que conectan las distintas zonas de la ciudad. En conjunto, el sistema integrado de transporte público realiza cerca de cuatro millones de viajes diarios.

1.1. Planteamiento del problema

A pesar de su cobertura y de su papel estratégico, el sistema TransMilenio enfrenta de manera persistente problemáticas asociadas a la congestión, los retrasos, el sobrecupo y las fallas operativas, especialmente durante las horas de mayor demanda. Estas dificultades tienen un impacto directo sobre el bienestar de los usuarios: incremento en los tiempos de viaje, pérdida de confiabilidad del servicio y un costo social y económico considerable para la ciudad en su conjunto.Los periodos de saturación, concentrados en las franjas pico de la mañana y de la tarde, constituyen el escenario donde estas problemáticas se manifiestan con mayor severidad, y donde un tramo cuyo recorrido en condiciones normales toma pocos minutos puede prolongarse considerablemente.

Estos fenómenos no obedecen a una causa única, sino a la interacción de múltiples variables como la variabilidad temporal de la demanda, la capacidad finita de las troncales, la operación de los cruces semaforizados y las condiciones ambientales que dan lugar a comportamientos colectivos difíciles de anticipar mediante el análisis tradicional.

Una característica central de la congestión es su carácter *emergente* y *no lineal*, esto quiere decir que el sistema no se degrada de manera gradual y proporcional al aumento de la demanda, sino que puede transitar de forma abrupta de un régimen de flujo libre a un régimen de saturación cuando ciertos parámetros de control superan valores críticos; este comportamiento, en el cual un cambio pequeño en las condiciones de entrada produce un cambio cualitativo en el estado global del sistema, es análogo a las transiciones de fase que se observan en los sistemas físicos, como el paso del agua líquida al vapor o la pérdida de magnetización de un material al superar la temperatura de Curie.

En este contexto, la física estadística ofrece un marco conceptual y metodológico adecuado, al proporcionar herramientas para describir sistemas de muchos cuerpos fuera del equilibrio y para caracterizar las transiciones entre sus distintos regímenes dinámicos.

1.2. Antecedentes

1.2.1. Los modelos hidrodinámicos

El primer enfoque, de carácter macroscópico, corresponde a los modelos hidrodinámicos o de medios continuos, cuyo referente es el modelo de Lighthill, Whitham y Richards (LWR), formulado en la década de 1950 [6]. En este marco, el tráfico se describe como

un fluido continuo caracterizado por una densidad $k(x,t)$ y un flujo $q(x,t)$, sin hacer referencia a los vehículos individuales, el modelo se sustenta en la ecuación de continuidad, que expresa la conservación del número de vehículos,

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0, \tag{1}$$

y se cierra mediante una relación entre el flujo y la densidad, $q = Q(k)$, denominada diagrama fundamental, la resolución de esta ecuación predice que las perturbaciones de la densidad se propagan a lo largo de la vía como ondas cinemáticas, cuya velocidad es la derivada del diagrama fundamental, dQ/dk

Este resultado permite interpretar la congestión como una onda que se propaga a lo largo del tráfico hacia atras, y explica fenómenos como la formación de embotellamientos aparentemente sin causa

1.2.2. El modelo de autómata celular de Nagel y Schreckenberg

El segundo enfoque, de carácter microscópico y discreto, fue introducido por el modelo de autómata celular de Nagel y Schreckenberg en 1992 [3], que constituye a uno de los cimientos que seran utilizados para los modelos que se implementan en el trafico. En este modelo, la vía se discretiza en una sucesión de celdas, cada una de las cuales puede estar vacía u ocupada por un único vehículo, y el tiempo avanza en pasos discretos. Cada vehículo posee una velocidad entera v comprendida entre cero y una velocidad máxima $v_{\text{máx}}$, y en cada paso temporal el estado del sistema se actualiza simultáneamente para todos los vehículos según cuatro reglas sucesivas:

1. **Aceleración:** si la velocidad del vehículo es menor que la máxima, se incrementa en una unidad, $v \to \text{mín}(v+1, v_{\text{máx}})$, lo que representa la tendencia de los conductores a avanzar lo más rápido posible.

2. **Frenado:** si la distancia d al vehículo precedente es menor o igual que la velocidad, esta se reduce para evitar la colisión, $v \to \text{mín}(v, d-1)$, lo que representa la interacción entre vehículos.

3. **Aleatorización:** con una probabilidad p, la velocidad se reduce en una unidad, $v \to \text{máx}(v-1, 0)$. Esta regla estocástica, que introduce las fluctuaciones características de la conducción real —la desaceleración espontánea e injustificada—, es el ingrediente esencial del modelo y el responsable de la formación espontánea de embotellamientos.

4. **Movimiento:** cada vehículo avanza un número de celdas igual a su velocidad actualizada, $x \to x + v$.

A pesar de la simplicidad de estas reglas, el modelo reproduce fenómenos observados en el tráfico real, y exhibe una *transición de fase* entre un régimen de flujo libre a densidades bajas, en el que los vehículos circulan sin interferencia y un régimen congestionado a densidades altas, en el que emergen atascos que se propagan hacia atrás [4]. La densidad a la que ocurre esta transición separa las dos fases, y en sus proximidades el sistema exhibe el comportamiento crítico característico de los fenómenos de fase. Este resultado estableció el tráfico vehicular como un sistema paradigmático de la física estadística de no equilibrio, estrechamente emparentado con procesos de exclusión como el *Asymmetric*

Simple Exclusion Process (ASEP) —un modelo de partículas que se desplazan sobre una red unidimensional con exclusión mutua, para el cual existen soluciones analíticas exactas y diagramas de fase completamente caracterizados—, y motivó un extenso cuerpo de investigación sobre los diagramas fundamentales, los exponentes críticos y la naturaleza de dicha transición [5].

1.2.3. El modelo de máxima entropía de Wilson

Para hablar del modelo del principio de maxima entropia debemos reomantarnos a Wilson que introdujo en 1970 la aplicación del principio de máxima entropía a la estimación de la demanda de viajes [1]. A diferencia de los enfoques anteriores, que describen la dinámica del flujo, el modelo aborda el problema de la distribución de la demanda, que consiste en determinar cuántos viajes se realizan entre cada par de zonas de una ciudad a partir de información parcial

Su formulación parte de considerar el número de maneras en que un total de viajes puede distribuirse entre los distintos pares origen-destino, una multiplicidad combinatoria idéntica a la de la mecánica estadística, y demuestra que la distribución más probable es aquella que maximiza una función de entropía sujeta a las restricciones conocidas del sistema: los totales de viajes originados en cada zona, los totales atraídos por cada zona y el costo total de transporte.

La resolución de este problema de maximización mediante multiplicadores de Lagrange conduce al modelo gravitacional doblemente restringido,

$$T_{ij} = A_i B_j O_i D_j \, e^{-\beta c_{ij}}, \tag{2}$$

en el que el número de viajes T_{ij} entre el origen i y el destino j depende de los totales de origen O_i y destino D_j, de unos factores de balance A_i y B_j que garantizan el cumplimiento de las restricciones, y de un factor exponencial $e^{-\beta c_{ij}}$ que penaliza los viajes de mayor costo c_{ij}. La estructura de este resultado es formalmente idéntica a la de la distribución de Boltzmann de la mecánica estadística, con el costo de viaje desempeñando el papel de la energía y el parámetro β el de una temperatura inversa. El modelo de Wilson establece así un puente directo entre la termodinámica y la movilidad urbana, y constituye el fundamento del modelo de demanda empleado en este trabajo.

2. Fundamentos Termodinámicos

El presente capítulo expone dichos fundamentos necesarios para la aplicación del modelo computacionla. Se comienza por los fundamentos de la termodinámica; se prosigue con el puente que la mecánica estadística establece entre las escalas microscópica y macroscópica; se aborda el principio de máxima entropía como principio de inferencia; se justifica la legitimidad de su aplicación a sistemas no físicos; y se concluye con la derivación del modelo de demanda, la caracterización de la congestión como fenómeno crítico y la descripción del simulador empleado.

2.1. Las leyes de la termodinámica

La termodinámica es la rama de la física que estudia las transformaciones de la energía en los sistemas macroscópicos, y que describe el estado de dichos sistemas mediante un número reducido de variables como la presión, el volumen, la temperatura o el número de partículas sin referencia a la estructura microscópica.

La *ley cero* establece la existencia de la temperatura como propiedad de estado: si dos sistemas se encuentran en equilibrio térmico con un tercero, entonces lo están entre sí. Esta ley fundamenta la posibilidad misma de definir una escala de temperatura y de tratarla como una variable macroscópica bien definida.

La *primera ley* expresa el principio de conservación de la energía. Para un sistema que intercambia energía con su entorno en forma de calor δQ y de trabajo δW, la variación de su energía interna U se escribe

$$dU = \delta Q - \delta W. \tag{3}$$

La energía interna es una función de estado, cuyo valor depende únicamente de la configuración macroscópica del sistema y no de la trayectoria seguida para alcanzarla. El calor y el trabajo, en cambio, son formas de energía en tránsito, dependientes del proceso.

La *segunda ley* introduce la entropía S como función de estado y establece la dirección en la que evolucionan los procesos naturales. En su formulación mediante la entropía, la segunda ley afirma que, para un sistema aislado, la entropía no puede disminuir:

$$dS \geq 0, \tag{4}$$

alcanzando su máximo en el estado de equilibrio. La igualdad corresponde a los procesos reversibles y la desigualdad estricta a los irreversibles. Para un proceso reversible, la variación de entropía se relaciona con el calor intercambiado a la temperatura T mediante $dS = \delta Q_{\text{rev}}/T$. La segunda ley establece una direccionalidad al tiempo termodinámico y confirma que los sistemas aislados evolucionan espontáneamente hacia estados de mayor entropía.

De la combinación de la primera y la segunda ley, para un sistema con número variable de partículas, se obtiene la relación fundamental de la termodinámica,

$$dU = T\,dS - p\,dV + \mu\,dN, \tag{5}$$

en la que aparecen las variables intensivas la temperatura T, la presión p y el potencial químico μ como derivadas de la energía interna respecto de las variables extensivas correspondientes (la entropía S, el volumen V y el número de partículas N). En particular, la temperatura se define como

$$\frac{1}{T} = \left(\frac{\partial S}{\partial U}\right)_{V,N}, \tag{6}$$

relación que resultará de importancia en la interpretación posterior, pues establece que la temperatura mide la respuesta de la entropía frente a variaciones de la energía.

El concepto de entropía, introducido por la termodinámica como una magnitud macroscópica de significado abstracto asociada a la irreversibilidad y a la energía, adquiere su interpretación más profunda únicamente al descender a la escala microscópica, tarea que corresponde a la mecánica estadística y que se aborda en la sección siguiente.

3. Mecánica Estadística

La termodinámica describe los sistemas macroscópicos sin hacer referencia a sus constituyentes; sin embargo, todo sistema macroscópico está compuesto por un número inmenso de partículas, del orden del número de Avogadro, cuyo estado individual queda descrito por las leyes de la mecánica. La mecánica estadística constituye el puente entre ambas descripciones: su objeto es deducir las propiedades macroscópicas de un sistema a partir de las propiedades estadísticas del conjunto de sus microestados, entendiendo que el comportamiento observable no depende del detalle de cada configuración microscópica, sino de su distribución estadística.

3.1. Microestados, macroestados y espacio de fases

Un *microestado* de un sistema es una especificación completa del estado de todas sus partículas, en un sistema clásico, el conjunto de todas sus posiciones y momentos. El conjunto de todos los microestados accesibles constituye el espacio de fases del sistema.

Un *macroestado*, en cambio, queda especificado por los valores de unas pocas variables macroscópicas, y es compatible, en general, con un número enorme de microestados distintos.

Un mismo macroestado puede realizarse mediante muchos microestados diferentes, y que el número de tales realizaciones se conoce como la multiplicidad del macroestado, que varía enormemente de un macroestado a otro. Aquellos macroestados que pueden realizarse mediante un mayor número de microestados son más probables de observar, hasta el punto de que, para sistemas de muchas partículas, el macroestado de máxima multiplicidad domina de manera prácticamente absoluta sobre todos los demás.

3.2. El postulado fundamental y la entropía de Boltzmann

La mecánica estadística del equilibrio se rige sobre que, para un sistema aislado en equilibrio, todos los microestados accesibles compatibles con las restricciones macroscópicas,es decir, energía, volumen y número de partículas fijos son igualmente probables. Este postulado, que define el denominado colectivo o ensamble microcanónico, traslada el problema de la descripción del sistema al conteo de sus microestados.

Sea W el número de microestados accesibles compatibles con un macroestado dado. La conexión entre esta cantidad microscópica y la entropía termodinámica macroscópica se establece mediante la célebre relación de Boltzmann,

$$S = k_B \ln W, \tag{7}$$

donde k_B es la constante de Boltzmann, con esta relacion, se puede intuir que la entropía de un macroestado es, salvo la constante k_B, el logaritmo del número de microestados que lo realizan. Un macroestado de elevada entropía es aquel que puede realizarse de innumerables maneras; el estado de equilibrio, caracterizado por la máxima entropía, es simplemente el macroestado de máxima multiplicidad, esto es, el más probable.

La elección del logaritmo en la ecuación (7) responde a como la entropía es una magnitud extensiva, de modo que la entropía de dos sistemas independientes debe ser la suma de sus entropías individuales. Puesto que el número de microestados de un sistema compuesto por dos subsistemas independientes es el producto de sus multiplicidades individuales, $W = W_1 W_2$, y dado que el logaritmo transforma productos en

sumas, la relación logarítmica es la única que garantiza la aditividad de la entropía: $S = k_B \ln(W_1 W_2) = k_B \ln W_1 + k_B \ln W_2 = S_1 + S_2$.

3.3. El ensamble canónico y la distribución de Boltzmann

Cuando el sistema de interés no se encuentra aislado, sino en contacto térmico con un reservorio a temperatura T, la restricción de energía constante se sustituye por la de temperatura constante. El colectivo apropiado en este caso es el ensamble canónico, en el cual la probabilidad de que el sistema se encuentre en un microestado de energía E_r viene dada por la distribución de Boltzmann,

$$p_r = \frac{e^{-\beta E_r}}{Z}, \qquad \beta = \frac{1}{k_B T}, \tag{8}$$

donde β es la temperatura inversa y Z es la función de partición,

$$Z = \sum_r e^{-\beta E_r}, \tag{9}$$

que actúa como constante de normalización y que codifica, además, la totalidad de la información termodinámica del sistema. La distribución de Boltzmann (8) establece que la probabilidad de un microestado decae exponencialmente con su energía, y que la temperatura a través de β regula el peso relativo de los estados de alta y baja energía. Esta distribución, y muy en particular la forma funcional del factor $e^{-\beta E}$, reaparecerá en la formulación del modelo de demanda, constituyendo el fundamento de la analogía central de este trabajo.

La distribución de Boltzmann puede derivarse mediante la maximización de la entropía sujeta a la restricción de energía media fija, procedimiento que anticipa el principio de máxima entropía y que se desarrolla en la sección siguiente.

4. El Principio de Máxima Entropía

4.1. Entropía de la información e inferencia

En 1948, Shannon introdujo en el contexto de la teoría de la comunicación, una medida de la incertidumbre asociada a una distribución de probabilidad. Dada una distribución $\{p_i\}$ sobre un conjunto de posibilidades, la entropía de la información se define como

$$S = -\sum_i p_i \ln p_i. \tag{10}$$

Esta magnitud cuantifica la incertidumbre o la falta de información contenida en la distribución: es nula cuando el resultado es cierto y máxima cuando todos los resultados son equiprobables. La identidad formal entre la expresión (10) y la entropía de la mecánica estadística se puede notar apenas interpretamos el significado de ambas.

Fue Jaynes quien, en 1957, formalizó esto al reinterpretar la mecánica estadística como un caso particular de la inferencia estadística bajo información incompleta [2]. Su planteamiento, conocido como el principio de máxima entropía, resuelve el problema general de asignar una distribución de probabilidad cuando la información disponible es insuficiente para determinarla de manera unívoca. El principio establece que, entre todas las distribuciones compatibles con la información conocida expresada en forma de restricciones, debe seleccionarse aquella que maximiza la entropía de la información.

4.2. Justificación del principio

La legitimidad del principio de máxima entropía descansa sobre una base tanto epistemológica como combinatoria. Desde la perspectiva epistemológica, la distribución de máxima entropía es la que resulta máximamente no comprometida con respecto a la información ausente: cualquier otra distribución compatible con las mismas restricciones incorporaría, de manera implícita, información adicional no justificada por los datos. Elegir la distribución de máxima entropía equivale, por tanto, a abstenerse de introducir suposiciones arbitrarias, y constituye el procedimiento de inferencia más honesto posible ante la incertidumbre.

Desde la perspectiva combinatoria, el principio admite una justificación aún más concreta, que lo conecta directamente con el conteo de microestados de la mecánica estadística. Considérese la asignación de un gran número N de elementos entre un conjunto de categorías, y supóngase que se conocen únicamente ciertos promedios. El número de formas en que puede realizarse una distribución dada de los elementos entre las categorías es un coeficiente multinomial, y la distribución que puede realizarse del mayor número de formas la más probable resulta ser, precisamente, la de máxima entropía. Este argumento, que se desarrollará en detalle en la Sección 6 al derivar el modelo de demanda, demuestra que el principio de máxima entropía no es un postulado *ad hoc*, sino una consecuencia directa de la combinatoria de los sistemas de muchos elementos.

4.3. Formulación matemática mediante multiplicadores de Lagrange

La aplicación del principio de máxima entropía conduce a un problema de optimización con restricciones. Se busca la distribución $\{p_i\}$ que maximiza la entropía (10) sujeta a la

condición de normalización $\sum_i p_i = 1$ y a un conjunto de restricciones que expresan la información disponible, típicamente en forma de valores esperados $\sum_i p_i f_k(i) = \langle f_k \rangle$ para ciertas funciones f_k. La solución de este problema, obtenida mediante el método de los multiplicadores de Lagrange, adopta la forma general

$$p_i = \frac{1}{Z} \exp\left(- \sum_k \lambda_k f_k(i) \right), \tag{11}$$

donde los multiplicadores λ_k quedan determinados por las restricciones y Z es la función de partición que garantiza la normalización. Cuando la única restricción, además de la normalización, es la energía media, la expresión (11) se reduce exactamente a la distribución de Boltzmann (8), con el multiplicador de Lagrange identificándose con la temperatura inversa β. De este modo, la mecánica estadística del equilibrio emerge como un caso particular del principio de máxima entropía, y la temperatura adquiere el significado de un multiplicador de Lagrange asociado a la restricción de energía.

5. La analogía física

La aplicación del formalismo de la mecánica estadística a un sistema que no es de naturaleza física como un sistema de transporte requiere una justificación que precise el alcance y los límites de la analogía, almenos para esta autora . Esta sección aborda dicha justificación, primero en el plano conceptual y a continuación en el plano formal, estableciendo la correspondencia matemática explícita entre ambos sistemas.

5.1. Fundamento conceptual

La clave de la legitimidad reside en el carácter del principio de máxima entropía. Como se ha argumentado en la Sección 4, dicho principio no es una ley específica de la física, sino un principio general de inferencia estadística, aplicable a cualquier sistema del que se disponga de información parcial en forma de restricciones sobre promedios. La mecánica estadística de los sistemas físicos es una de sus aplicaciones aquella en la que las restricciones son la energía, el volumen y el número de partículas, pero de ninguna manera la única. En consecuencia, cuando un sistema de transporte se describe mediante la distribución de sus viajes sujeta a restricciones sobre los totales de origen y destino y sobre el costo total, la aplicación del principio de máxima entropía se ve mediante la aparición de una estructura formal idéntica a la de la mecánica estadística

5.2. Correspondencia formal

La equivalencia entre ambos sistemas puede establecerse de manera explícita comparando las estructuras matemáticas de sus respectivas descripciones. En el ensamble canónico de la mecánica estadística, la probabilidad de un microestado de energía E_r es

$$p_r = \frac{1}{Z}\, e^{-\beta E_r}, \qquad Z = \sum_r e^{-\beta E_r}, \qquad \beta = \frac{1}{k_B T}, \tag{12}$$

donde Z es la función de partición y β la temperatura inversa. En el sistema de transporte, el modelo de máxima entropía asigna a cada par origen-destino una demanda proporcional a

$$T_{ij} \propto e^{-\beta c_{ij}}, \qquad Z_{\mathrm{tr}} = \sum_{ij} O_i D_j\, e^{-\beta c_{ij}}, \tag{13}$$

donde Z_{tr} desempeña el papel de una función de partición del sistema de transporte y β el de una temperatura inversa efectiva. La comparación de las expresiones (12) y (13) revela la correspondencia término a término, que se resume en la Tabla 1.

Del mismo modo que en el sistema físico la energía media se obtiene por derivación de la función de partición, $\langle E\rangle = -\,\partial \ln Z/\partial\beta$, en el sistema de transporte el costo medio de viaje se obtiene de la función de partición análoga mediante la relación idéntica

$$\langle c\rangle = -\,\frac{\partial \ln Z_{\mathrm{tr}}}{\partial \beta}, \tag{14}$$

La temperatura efectiva del sistema puede, en consecuencia, definirse de manera análoga a la temperatura termodinámica [ecuación (6)], como la magnitud que gobierna la respuesta de la entropía de la distribución de viajes frente a variaciones del costo total, estableciendo así una definición operativa.

Cuadro 1: Correspondencia formal entre el sistema físico y el sistema de transporte.

Sistema físico (mecánica estadística)	Sistema de transporte
Partícula	Viaje
Microestado de energía E_r	Viaje de costo c_{ij}
Energía E	Costo generalizado de viaje c
Factor de Boltzmann $e^{-\beta E}$	Factor de disuasión $e^{-\beta c}$
Temperatura inversa $\beta = 1/k_B T$	Sensibilidad al costo β
Función de partición $Z = \sum_r e^{-\beta E_r}$	$Z_{\text{tr}} = \sum_{ij} O_i D_j e^{-\beta c_{ij}}$
Energía total $U = \sum_r p_r E_r$	Costo total $C = \sum_{ij} T_{ij} c_{ij}$
Energía media $\langle E \rangle$	Costo medio de viaje $\langle c \rangle$

5.3. Alcance y límites de la analogía

Se trata de una analogía *formal*, no *material*: lo que comparten el sistema físico y el sistema de transporte es la estructura matemática de la descripción, el mismo problema de optimización, las mismas ecuaciones, las mismas soluciones, y no la naturaleza física de las magnitudes involucradas. El costo de viaje desempeña, en las ecuaciones, el papel formal de la energía, pero no es energía en el sentido físico; la temperatura efectiva del sistema de transporte regula la distribución de los viajes de manera análoga a como la temperatura termodinámica regula la distribución de las energías, pero no corresponde a una agitación térmica ni a un intercambio de calor.

La analogía permite afirmar todo aquello que se deriva de la estructura matemática compartida: la forma exponencial de la distribución de viajes, el papel del parámetro β como regulador de la sensibilidad al costo, la existencia de una función de partición que codifica las magnitudes macroscópicas, y la posibilidad de transiciones entre regímenes

6. Aplicación a la Distribución de Viajes

Establecidos los fundamentos, se procede a la derivación del modelo de demanda, donde el objetivo es determinar la matriz origen-destino T_{ij}, cuyo elemento representa el número de viajes desde la zona i hacia la zona j, a partir de la información parcial disponible sobre los totales de origen y destino.

6.1. Conteo de configuraciones

Considérese la distribución de un total de T viajes entre las celdas de la matriz origen-destino. El número de maneras distintas de asignar los viajes individuales de modo que se obtenga una configuración $\{T_{ij}\}$ es el coeficiente multinomial

$$W = \frac{T!}{\prod_{ij} T_{ij}!}, \tag{15}$$

que se obtiene observando que el número total de ordenamientos de los T viajes es $T!$, y que los viajes asignados a una misma celda son intercambiables sin alterar la configuración, lo que introduce el factor corrector $\prod_{ij} T_{ij}!$. De acuerdo con el postulado de Boltzmann (7), la configuración más probable es la que maximiza esta multiplicidad.

6.2. Aproximación de Stirling y entropía

La maximización de W se facilita considerando su logaritmo y empleando la aproximación de Stirling, $\ln n! \approx n \ln n - n$, válida para valores grandes de n y justificada por la aproximación de la suma $\sum_{k=1}^{n} \ln k$ mediante la integral $\int_1^n \ln x \, dx = n \ln n - n + 1$. Aplicada a la expresión (15), y descartando los términos constantes, conduce a la función objetivo

$$S = -\sum_{ij} T_{ij} \ln T_{ij}, \tag{16}$$

formalmente idéntica a la entropía de Boltzmann-Gibbs-Shannon.

6.3. Restricciones y maximización

La entropía (16) se maximiza sujeta a tres restricciones: la conservación de los viajes originados en cada zona, $\sum_j T_{ij} = O_i$; la conservación de los viajes atraídos por cada zona, $\sum_i T_{ij} = D_j$; y la fijación del costo total de transporte, $\sum_{ij} T_{ij} c_{ij} = C$, donde c_{ij} es el costo generalizado de viaje. Esta última restricción es el análogo formal de la energía total del sistema. Introduciendo multiplicadores de Lagrange λ_i, μ_j y β para cada familia de restricciones, y anulando la derivada del lagrangiano respecto de T_{ij}, se obtiene

$$-\ln T_{ij} - 1 - \lambda_i - \mu_j - \beta c_{ij} = 0, \tag{17}$$

de donde, despejando y reagrupando los factores dependientes del origen y del destino, resulta el modelo gravitacional doblemente restringido

$$\boxed{T_{ij} = A_i \, B_j \, O_i \, D_j \, e^{-\beta c_{ij}}.} \tag{18}$$

El factor $e^{-\beta c_{ij}}$ es el análogo formal del factor de Boltzmann, y el multiplicador β el análogo de la temperatura inversa. Los factores de balance A_i y B_j, mutuamente dependientes, se determinan iterativamente mediante el algoritmo de Furness, garantizando el

cumplimiento de las restricciones de origen y destino. El parámetro β regula la sensibilidad de la demanda al costo: se interpreta como el inverso de una temperatura efectiva del sistema de transporte, cuya calibración se realiza igualando el costo medio de viaje predicho, $\langle c \rangle = \sum_{ij} T_{ij} c_{ij} / \sum_{ij} T_{ij}$, al valor observado.

7. La Congestión como Fenómeno Crítico

La descripción del sistema de transporte en el marco de la física estadística permite interpretar la congestión como un fenómeno crítico análogo a una transición de fase. Esta sección desarrolla dicho concepto con el aparato de la teoría de los fenómenos críticos, precisando los conceptos de parámetro de orden, transición y comportamiento singular, y estableciendo la correspondencia con el sistema de tráfico.

7.1. Transiciones de fase y parámetro de orden

Una transición de fase es un cambio cualitativo y abrupto del estado macroscópico de un sistema, que ocurre cuando un parámetro de control,la temperatura, la presión o la densidad, atraviesa un valor crítico. Las transiciones se clasifican según el comportamiento de las derivadas de la energía libre. En una transición de primer orden, las primeras derivadas de la energía libre como la entropía o el volumen, presentan una discontinuidad, asociada a un calor latente; ejemplos característicos son la fusión y la ebullición. En una transición de segundo orden, o continua, las primeras derivadas son continuas pero las segundas, como la capacidad calorífica o la susceptibilidad, divergen en el punto crítico; el ejemplo paradigmático es la transición ferromagnética en el punto de Curie.

La caracterización de una transición de fase se realiza mediante un *parámetro de orden* ϕ, definido como una magnitud que distingue las dos fases: se anula en la fase desordenada y adopta un valor no nulo en la fase ordenada. En la transición ferromagnética, el parámetro de orden es la magnetización espontánea, que es nula por encima de la temperatura crítica T_c y crece de forma continua por debajo de ella. En las proximidades del punto crítico, el parámetro de orden frecuentemente exhibe un comportamiento en forma de ley de potencia,

$$\phi \sim |\,\epsilon\,|^{\beta_c}, \qquad \epsilon = \frac{g - g_c}{g_c}, \tag{19}$$

donde g es el parámetro de control, g_c su valor crítico, ϵ la distancia relativa al punto crítico y β_c un exponente crítico. Análogamente, la susceptibilidad χ, que es la respuesta del parámetro de orden ante una perturbación externa, diverge en el punto crítico según

$$\chi \sim |\,\epsilon\,|^{-\gamma}, \tag{20}$$

con un exponente crítico γ. Una característica notable de los fenómenos críticos es la *universalidad*: sistemas físicamente muy diversos comparten los mismos exponentes críticos, que dependen únicamente de propiedades generales como la dimensionalidad del sistema y la simetría del parámetro de orden, y no de los detalles microscópicos de las interacciones.

Esta universalidad es la que justifica que modelos idealizados como el modelo de Ising o los autómatas celulares de tráfico, capturen el comportamiento crítico de sistemas reales mucho más complejos.

7.2. La transición de fase en el flujo de tráfico

El tráfico vehicular exhibe un comportamiento estructuralmente análogo al de los sistemas que experimentan transiciones de fase. Al aumentar la densidad de vehículos, el sistema transita de un régimen de flujo libre en el que los vehículos circulan a velocidad próxima a la de diseño y la interacción entre ellos es despreciable a un régimen congestionado caracterizado por la formación de colas y la propagación de ondas de parada y

arranque. Esta transición se manifiesta ya en el diagrama fundamental [ecuación (23)]: la densidad crítica k_c separa la rama de flujo libre, en la que el flujo crece con la densidad, de la rama congestionada, en la que decrece, marcando el punto en el que el sistema cambia cualitativamente de comportamiento.

Este carácter crítico fue establecido de manera rigurosa por el modelo de autómata celular de Nagel y Schreckenberg [3]. En dicho modelo, la vía se discretiza en celdas que pueden estar ocupadas o vacías, y los vehículos avanzan según reglas estocásticas de aceleración, frenado ante el vehículo precedente y desaceleración aleatoria. A pesar de su simplicidad, el modelo reproduce la formación espontánea de embotellamientos y exhibe una transición entre una fase de flujo libre, para densidades bajas, y una fase congestionada, para densidades altas, con propiedades críticas análogas a las de las transiciones de fase de la materia [4, 5], con esto se puede predecir que, la transición de tráfico está emparentada con las transiciones que se observan en procesos de exclusión, como el proceso de exclusión simple totalmente asimétrico, para los que existen soluciones exactas y diagramas de fase caracterizados.

7.3. El parámetro de orden del sistema de transporte

En el presente trabajo, la transición se caracteriza tomando como parámetro de control el factor de escala de la demanda λ, que regula la densidad efectiva de vehículos sobre la red, y como parámetro de orden la proporción de viajes completados ϕ, definida como el cociente entre el número de viajes que alcanzan su destino dentro del horizonte de simulación y el número total de viajes generados,

$$\phi(\lambda) = \frac{N_{\text{completados}}(\lambda)}{N_{\text{total}}(\lambda)}. \tag{21}$$

Esta elección satisface los requisitos de un parámetro de orden: en la fase de flujo libre, para $\lambda < \lambda_c$, la práctica totalidad de los viajes se completa y $\phi \approx 1$; al superar la demanda crítica λ_c, el sistema ingresa en la fase congestionada y ϕ decae abruptamente hacia valores reducidos, reflejando que una fracción creciente de los viajes no logra completarse. La demanda crítica λ_c desempeña, en esta analogía, el papel de la temperatura crítica T_c de la transición ferromagnética, y su localización constituye el objetivo central de la caracterización.

La correspondencia con la densidad crítica del diagrama fundamental es directa, el factor de demanda λ determina la densidad de vehículos que se inyecta sobre los tramos; cuando esta densidad, en los tramos que constituyen el cuello de botella del sistema, alcanza la densidad crítica k_c, dichos tramos ingresan en la rama congestionada del diagrama fundamental, y la congestión se propaga hacia atrás degradando el desempeño global. La demanda crítica λ_c es, por tanto, la manifestación macroscópica a escala de la red completa de la densidad crítica que gobierna la transición a nivel de cada tramo, estableciendo un vínculo cuantitativo entre la descripción local del flujo y la descripción global del sistema.

Bajo esta perspectiva, la evaluación de las estrategias de optimización desarrollada en el Capítulo 13 adquiere una interpretación física, y reside en que una intervención es más efectiva cuanto mayor es el desplazamiento que produce en la demanda crítica λ_c, esto es, cuanto más amplía el dominio de la fase de flujo libre

8. El Simulador UXsim

La componente de simulación de este trabajo se implementa mediante UXsim, un simulador de flujo de tráfico de red de código abierto desarrollado íntegramente en el lenguaje Python [26, 27]. UXsim computa la evolución dinámica del tráfico sobre una red a partir de una demanda origen-destino dependiente del tiempo, en un esquema de asignación dinámica de tráfico. Esta sección describe el modelo matemático que implementa y las funcionalidades que ofrece, con el detalle necesario para comprender su papel en la metodología.

8.1. El modelo de ondas cinemáticas

UXsim se fundamenta en la teoría de ondas cinemáticas de Lighthill, Whitham y Richards [6], que describe el tráfico como un flujo continuo gobernado por la ecuación de continuidad en una dimensión,

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0, \tag{22}$$

donde $k(x,t)$ es la densidad de vehículos y $q(x,t)$ el flujo. Esta ecuación expresa la conservación del número de vehículos: toda variación temporal de la densidad en un punto se compensa con la divergencia del flujo. La ecuación se cierra mediante una relación empírica entre el flujo y la densidad, $q = Q(k)$, denominada diagrama fundamental. UXsim adopta un diagrama fundamental triangular, definido por la velocidad de flujo libre u, la velocidad de retroceso de las ondas de congestión w y la densidad de atasco k_j, según

$$Q(k) = \begin{cases} u\,k & \text{si } k \le k_c, \\ w\,(k_j - k) & \text{si } k > k_c, \end{cases} \tag{23}$$

donde la densidad crítica k_c separa la rama de flujo libre en la que la velocidad es constante e igual a u de la rama congestionada en la que un aumento de la densidad reduce el flujo. Esta densidad crítica constituye, a nivel de un tramo individual, el análogo del punto de transición de fase.

8.2. El modelo de Newell y la resolución mesoscópica

Para la resolución del flujo a nivel de cada tramo, UXsim emplea la versión mesoscópica del modelo de seguimiento vehicular simplificado de Newell [7], equivalente al modelo de ondas cinemáticas con diagrama fundamental triangular. El modelo de Newell se basa en el principio de que la trayectoria espacio-temporal de un vehículo reproduce la de su predecesor, desplazada en el espacio y en el tiempo. La posición $x_i(t)$ del vehículo i se determina como

$$x_i(t) = \min\left\{\, x_i(t-\tau) + u\,\tau,\;\; x_{i-1}(t-\tau) - \delta \,\right\}, \tag{24}$$

donde $\tau = 1/(w\,k_j)$ es el desfase temporal asociado a la propagación de las ondas de congestión y $\delta = 1/k_j$ es el espaciamiento mínimo en condiciones de atasco. El primer término representa el avance en régimen libre a la velocidad u; el segundo, la restricción impuesta por el vehículo precedente. El operador mínimo selecciona automáticamente el régimen aplicable, reproduciendo la formación y propagación de colas. El carácter mesoscópico de la implementación en el que los vehículos se agrupan en paquetes para reducir el costo computacional permite simular redes de escala urbana preservando la fidelidad de la dinámica.

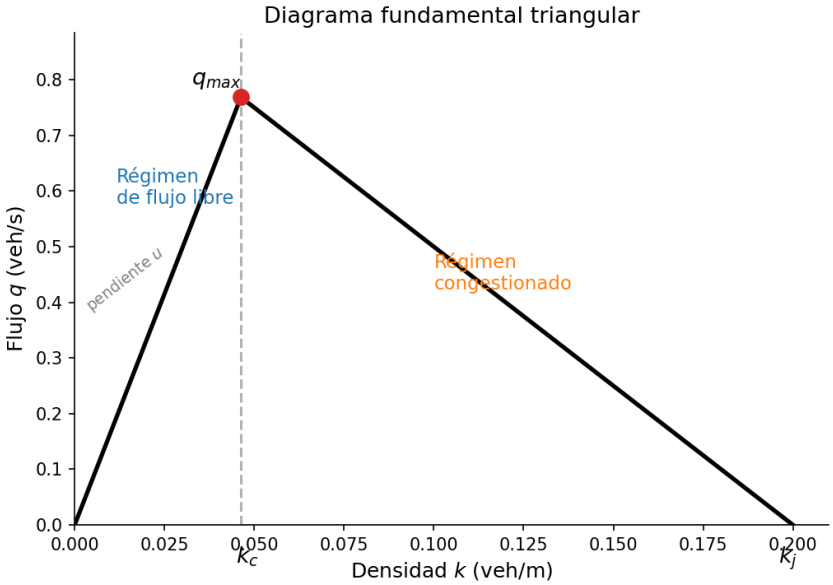


Figura 1: Diagrama fundamental triangular del flujo de tráfico. La rama ascendente corresponde al régimen de flujo libre, en el que el flujo crece con la densidad a velocidad constante u; la rama descendente corresponde al régimen congestionado. La densidad crítica k_c, en la que el flujo alcanza su valor máximo q_{max}, separa ambos regímenes y constituye el análogo, a nivel de un tramo, del punto de transición de fase.

8.3. Modelo de nodos, semáforos y elección de ruta

La transferencia de vehículos entre tramos, esto es, la resolución de las intersecciones, se realiza mediante un modelo de nodos que determina, en cada paso de tiempo, el número de vehículos que pueden transferirse en función de la oferta y la demanda disponibles y de las prioridades establecidas. Este modelo permite incorporar de manera natural la operación de los semáforos, mediante la asignación de intervalos de verde y de rojo a los distintos grupos de enlaces que confluyen en un nodo; un enlace únicamente transfiere vehículos durante la fase de verde de su grupo. La elección de ruta de los vehículos se rige por un principio de óptimo dinámico de usuario, según el cual cada vehículo tiende a seleccionar la ruta de menor costo dadas las condiciones de la red, si bien en el presente trabajo las rutas pueden restringirse explícitamente para representar los recorridos fijos de los servicios troncales.

8.4. Herramientas de análisis y visualización

UXsim proporciona un extenso conjunto de herramientas para el análisis de los resultados, esenciales para la interpretación física perseguida en este trabajo. A escala global, permite calcular el tiempo total de viaje, el retraso medio y la proporción de viajes completados, magnitudes que actúan como métricas de desempeño y como parámetros de orden. A escala de cada par origen-destino y de cada tramo, proporciona el volumen de tráfico, la velocidad media y los estados dinámicos del flujo. Entre las herramientas de visualización se cuentan los diagramas espacio-tiempo de densidad y de trayectorias que revelan la formación y propagación de las ondas de congestión en el interior de un tramo, las curvas de conteo acumulado de llegadas y salidas que cuantifican el número de vehículos en cola y el retraso, el diagrama fundamental macroscópico de la red, los ma-

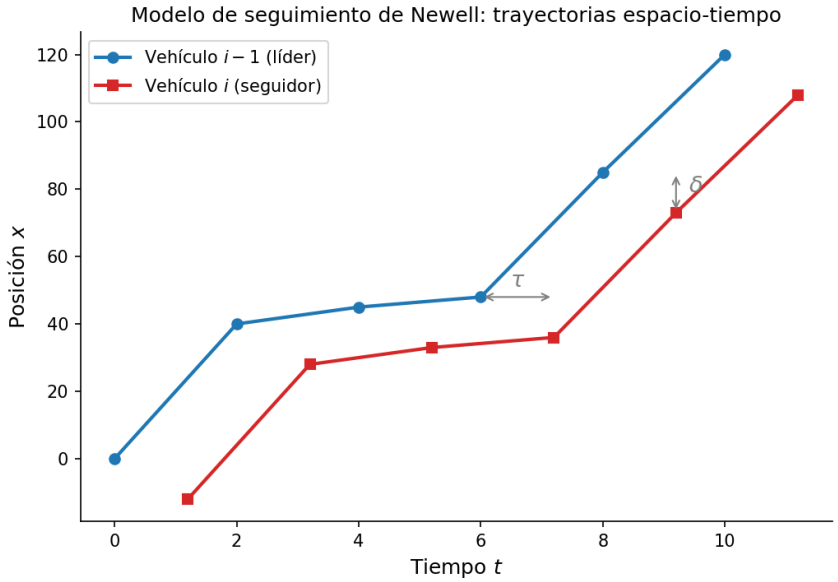


Figura 2: Representación espacio-tiempo del modelo de seguimiento de Newell. La trayectoria del vehículo seguidor reproduce la del líder, desplazada un tiempo τ y un espacio δ. El tramo de menor pendiente en el centro corresponde a la reducción de velocidad ante la congestión, y el desfase constante entre trayectorias refleja el espaciamiento mínimo en dichas condiciones.

pas de la red coloreados por velocidad o congestión, y la animación de la dinámica del tráfico sobre la red completa. Los resultados pueden exportarse a estructuras de datos para su tratamiento posterior. Este conjunto de funcionalidades permite caracterizar el sistema simultáneamente a escala local y global, en plena correspondencia con el marco físico-estadístico adoptado en este trabajo.

9. Objetivos

9.1. Objetivo General

Desarrollar un modelo físico-estadístico del sistema de transporte masivo TransMilenio de Bogotá, fundamentado en los principios de la termodinámica y la mecánica estadística, que permita caracterizar la congestión como un fenómeno crítico y formular estrategias de optimización cuantitativamente verificadas mediante simulación computacional.

9.2. Objetivos Específicos

- Reconstruir la red troncal del sistema TransMilenio a partir de datos geográficos y operacionales reales y simular su dinámica de flujo mediante el paquete de código abierto UXsim.
- Modelar las 105 rutas del sistema TransMilenio utilizando principios de la termodinámica y la estadística, interpretando el flujo de usuarios como partículas térmicas en un sistema dinámico.
- Formular y evaluar cuantitativamente propuestas de optimización del sistema, midiendo su efecto a través del desplazamiento que producen sobre la demanda crítica.

10. Plan de trabajo

Tema	Actividades	Semanas	Horas
Fase 1: Recolección y estructura de datos	Identificación de nodos principales (portales y estaciones); recolección de rutas, velocidades, frecuencias y demanda; obtención de datos institucionales; limpieza, normalización y organización de la base de datos.	1	74
Fase 2: Formulación del modelo físico–estadístico	Traducción del sistema a un modelo termodinámico; formulación del modelo de máxima entropía para matrices OD; derivación del modelo gravitacional doblemente restringido; ajuste preliminar del parámetro de sensibilidad al costo b.	2	74
Fase 3: Implementación computacional y simulaciones	Configuración de nodos y rutas en UXsim; integración de datos empíricos; programación de escenarios base y horas pico; validación interna del modelo; ejecución de simulaciones; generación de mapas de calor, histogramas y análisis preliminar.	3	96
Fase 4: Integración de resultados y optimización	Cálculo y análisis entrópico; identificación de cuellos de botella y transiciones de fase; análisis cuantitativo; formulación de propuestas de optimización; evaluación del impacto operativo; redacción de resultados y discusión.	4	74
Fase 5: Documento final y socialización	Integración del marco teórico, metodología y simulaciones; revisión académica; elaboración de conclusiones; preparación de presentación oral; entrega final del trabajo.	5	72
TOTAL		**5 semanas**	**400 horas**

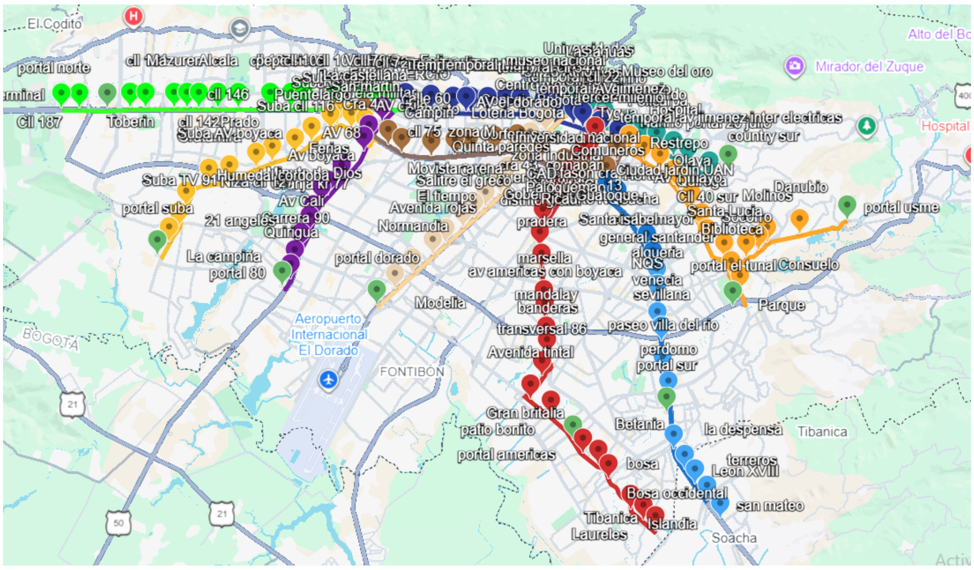


Figura 3: Red troncal de transmilenio construida con base en la red real

11. Metodología

La metodología desarrollada en este trabajo articula dos componentes, la construcción de una representación fiel de la red física del sistema TransMilenio a partir de datos geográficos y operacionales reales, y la formulación de un modelo de demanda de base física que alimenta la simulación dinámica del flujo. El presente capítulo describe en detalle ambos componentes, en primer lugar se presentan las fuentes de datos empleadas; a continuación se detalla la proyección geográfica y la construcción de la malla física de la red, incluyendo el tratamiento de los recorridos de las rutas y la incorporación de los semáforos; posteriormente se formula el modelo de demanda por máxima entropía y su acoplamiento con la simulación; y finalmente se describe la configuración de los escenarios simulados. Con el fin de ilustrar la implementación, se incluyen los fragmentos de código representativos de cada etapa; la implementación completa se encuentra en el Anexo A.

11.1. Fuentes de datos

La construcción del modelo se sustenta en tres conjuntos de datos de acceso público, provenientes de fuentes oficiales del Distrito de Bogotá y de la entidad TRANSMILENIO S.A.

El primer conjunto corresponde a la localización geográfica de las estaciones y al trazado de los corredores troncales, obtenido en formato KML (*Keyhole Markup Language*). Este formato codifica, por una parte, la posición de cada estación como un punto georreferenciado en coordenadas de longitud y latitud, y por otra, el trazado de cada corredor como una polilínea (*LineString*) compuesta por una sucesión ordenada de puntos que describe la geometría real de la vía, incluidas sus curvaturas.

El segundo conjunto corresponde a la definición de las rutas del sistema troncal, estructurada en un archivo dataset construido con base en la pagina oficial de rutas en cada tramo, cada ruta se describe mediante su estación de origen, la secuencia ordenada de estaciones intermedias que conforman su recorrido, su estación de destino y su franja horaria de operación. Este conjunto permite reconstruir la secuencia de estaciones servidas por cada servicio, distinguiendo ambos sentidos de circulación.

El tercer conjunto corresponde a la red semafórica de la ciudad, obtenida del portal de datos abiertos de Bogotá. Este conjunto contiene la localización georreferenciada de cada intersección semaforizada, junto con atributos operativos como el número de fases del ciclo y la identificación de los grupos de señales asociados específicamente al sistema BRT, atributo que permite aislar los semáforos que regulan el paso de los buses troncales del resto de la red semafórica urbana.

11.2. Proyección geográfica

Dado que los datos de origen se expresan en coordenadas geográficas angulares (longitud y latitud), y que la simulación requiere trabajar con distancias métricas sobre un plano, se aplica una proyección de las coordenadas al plano tangente local. Tomando como origen del sistema de referencia la posición del Portal El Tunal, de coordenadas (ϕ_0, λ_0), cada punto de coordenadas (ϕ, λ) se transforma en coordenadas planas (x, y) expresadas en metros mediante las relaciones

$$x = R\,(\lambda - \lambda_0)\,\frac{\pi}{180}\cos\phi_0, \qquad y = R\,(\phi - \phi_0)\,\frac{\pi}{180}, \tag{25}$$

donde $R = 6\,378\,137$ m es el radio ecuatorial de la Tierra y los ángulos se expresan en grados. Esta proyección de plano tangente resulta adecuada para el dominio de estudio, dado que la extensión de la ciudad es reducida frente al radio terrestre y el error de distorsión introducido es despreciable a la escala de análisis. La distancia euclidiana entre dos puntos proyectados se calcula de la forma habitual,

$$d(p_1, p_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \tag{26}$$

La implementación de esta transformación se presenta a continuación:

```
def proyectar_a_metros(lon, lat):
    R = 6378137.0                       # radio de la Tierra en metros
    d_lat = math.radians(lat - LAT_ORIGEN)
    d_lon = math.radians(lon - LON_ORIGEN)
    x = d_lon * R * math.cos(math.radians(LAT_ORIGEN))
    y = d_lat * R
    return x, y
```

La función recibe las coordenadas de longitud y latitud de un punto y devuelve su posición en metros relativa al origen; el factor $\cos\phi_0$ corrige la convergencia de los meridianos a la latitud de la ciudad.

11.3. Construcción de la malla física de la red

La representación de la red en el simulador exige definir un conjunto de nodos y un conjunto de enlaces dirigidos que los conectan, cada uno caracterizado por una longitud y unos parámetros de capacidad. Un aspecto central de este trabajo consistio en garantizar que dichos enlaces representen tramos físicamente existentes de la red, evitando conexiones entre estaciones que, aun siendo consecutivas en el recorrido de una ruta, se encuentran geográficamente distantes. Este problema se manifiesta particularmente en las rutas expresas, que enlazan estaciones separadas por varios kilómetros omitiendo las estaciones intermedias.

11.3.1. Reconstrucción de los corredores

Para construir una malla físicamente consistente, se reconstruye en primer lugar la secuencia real de estaciones que compone cada corredor troncal, aprovechando la geometría de las polilíneas del archivo KML. Para cada polilínea, se proyecta cada estación sobre su trazado y se determina tanto la distancia perpendicular de la estación a la polilínea como su posición de arco, es decir, la longitud recorrida a lo largo del trazado hasta el punto de proyección. La proyección de un punto p sobre un segmento definido por los extremos a y b se calcula mediante el parámetro

$$t = \frac{(p - a) \cdot (b - a)}{\|b - a\|^2}, \tag{27}$$

acotado al intervalo $[0, 1]$, y se extiende a la polilínea completa buscando el segmento que minimiza la distancia perpendicular. La implementación de esta proyección se muestra a continuación:

```
def proyectar_punto_en_segmento(p, a, b):
    ax, ay = a; bx, by = b; px, py = p
    dx, dy = bx - ax, by - ay
    seglen2 = dx*dx + dy*dy
    if seglen2 == 0:
        return 0.0, a, distancia(p, a)
    t = ((px - ax)*dx + (py - ay)*dy) / seglen2
    t = max(0.0, min(1.0, t))          # acotar al segmento
    proy = (ax + t*dx, ay + t*dy)
    return t, proy, distancia(p, proy)
```

Se considera que una estación pertenece a un corredor si su distancia perpendicular al trazado es inferior a un umbral de 120 m. Las estaciones asociadas a un mismo corredor se ordenan entonces según su posición de arco, lo que reconstruye la secuencia física real de estaciones a lo largo de ese corredor. La longitud de cada tramo entre estaciones consecutivas se calcula como la diferencia de sus posiciones de arco, lo que proporciona la distancia real siguiendo la curvatura de la vía, en lugar de la distancia en línea recta.

11.3.2. Composición de la malla y conectividad

La malla física se compone a partir de dos familias de enlaces. La primera corresponde a los tramos entre estaciones consecutivas de cada corredor, obtenidos según el procedimiento anterior. La segunda corresponde a las adyacencias cortas presentes en la definición de las rutas: para cada par de estaciones consecutivas en el recorrido de una ruta cuya distancia sea inferior a un umbral de 1 800 m y que no haya sido ya incorporada por un corredor, se añade el enlace correspondiente. Esta segunda familia rescata conexiones reales que la geometría de las polilíneas no cubría, y garantiza la conectividad de la malla. Por el contrario, los pares de estaciones separados por distancias superiores al umbral característicos de los recorridos expresos, no se incorporan como enlaces directos, evitando así la aparición de tramos espurios de gran longitud. La rutina de inserción de enlaces, que conserva la menor longitud cuando un tramo aparece en más de un corredor, se muestra a continuación:

```
def agregar_arista(a, b, length):
    if a == b or length <= 1:
        return
    if b not in adyacencia[a] or length < adyacencia[a][b]:
        adyacencia[a][b] = length      # enlace en ambos sentidos
        adyacencia[b][a] = length
```

La consistencia de la malla resultante se verifica mediante un análisis de componentes conexas, que confirma que la totalidad de las estaciones queda integrada en una única componente conectada. Esta verificación es un requisito indispensable, pues garantiza que existe un camino físico entre cualquier par de estaciones, condición necesaria tanto para la expansión de los recorridos como para el cálculo de la matriz de costos.

11.3.3. Expansión de los recorridos

Una vez construida la malla física, cada ruta del sistema definida en los datos de origen como una secuencia de estaciones servidas se expande a su recorrido físico completo. Para ello, se determina, entre cada par de estaciones servidas consecutivas, el camino de menor costo sobre la malla mediante el algoritmo de Dijkstra. De este modo, un servicio expreso que enlaza dos estaciones distantes recorre, en la representación del modelo, la totalidad de los tramos físicos intermedios que atraviesa, compartiendo dichos tramos con los servicios que sí se detienen en las estaciones intermedias. Este tratamiento resulta esencial para reproducir de manera realista la congestión, que emerge precisamente de la superposición de múltiples servicios sobre los tramos compartidos de las troncales. La implementación del algoritmo de Dijkstra empleado es la siguiente:

```
def camino_mas_corto(origen, destino):
    if origen == destino:
        return [origen]
    dist = {origen: 0}; prev = {}; pq = [(0, origen)]; visto = set()
    while pq:
        d, u = heapq.heappop(pq)
        if u in visto:
            continue
        visto.add(u)
        if u == destino:
            break
        for v, w in adyacencia[u].items():
            nd = d + w
            if v not in dist or nd < dist[v]:
                dist[v] = nd; prev[v] = u
                heapq.heappush(pq, (nd, v))
    camino = [destino]
    while camino[-1] != origen:
        camino.append(prev[camino[-1]])
    camino.reverse()
    return camino
```

11.4. Incorporación de los semáforos

La red semafórica se filtra para conservar únicamente las intersecciones que regulan el paso de los buses troncales, identificadas mediante el atributo de grupos de señales BRT. De este subconjunto se descartan los registros con coordenadas inconsistentes, detectados por su distancia anómala al origen del sistema de referencia. Cada semáforo restante se asocia a la estación más cercana siempre que la distancia entre ambos sea inferior a un umbral de 200 m, criterio que asigna a las estaciones los semáforos que efectivamente regulan su acceso.

La operación de cada intersección semaforizada se modela mediante un ciclo de dos fases: una fase de verde, durante la cual se permite el paso de los buses de la troncal, y una fase de rojo, durante la cual el paso queda interrumpido. Puesto que el modelo representa exclusivamente el flujo de los buses troncales, la fase de rojo representa el intervalo del ciclo asignado a los movimientos transversales del tráfico mixto, no incluido explícitamente en la simulación. La duración del verde se determina a partir del número real de fases del semáforo, de modo que un semáforo de n fases asigna a la troncal una fracción $1/n$ de un ciclo de duración total $\mathcal{C}$; la duración del verde es entonces $\mathcal{C}/n$ y la del rojo el complemento. La rutina correspondiente se muestra a continuación:

```
def signal_de(estacion):
    if estacion not in senal_por_estacion:
        return [0]                    # sin semáforo: siempre verde
    fases = senal_por_estacion[estacion]
    verde = max(10, round(CICLO_S / fases))
    rojo  = max(10, CICLO_S - verde)
    return [verde, rojo]
```

Dado que los tiempos precisos de semaforización no se encuentran disponibles públicamente, la duración total del ciclo se adopta como un parámetro del modelo, fijado en un valor representativo de la operación urbana.

11.5. Modelo de demanda por máxima entropía

La demanda de viajes se estima mediante el modelo gravitacional doblemente restringido derivado en el marco teórico [ecuación (18)], resuelto por el algoritmo de Furness. Los insumos del modelo son la matriz de costos, los totales de origen y destino, y el parámetro de sensibilidad al costo.

11.5.1. Matriz de costos

La matriz de costos c_{ij} se define como la distancia del camino de menor costo entre cada par de estaciones sobre la malla física, calculada mediante el algoritmo de Dijkstra y expresada en kilómetros. Esta elección vincula el costo generalizado de viaje con la distancia real recorrida a través de la red, de manera coherente con la geometría del sistema.

11.5.2. Totales de origen y destino

Ante la ausencia de datos desagregados de demanda por estación, los totales de origen O_i y de destino D_j se estiman distribuyendo el total diario de viajes del sistema de manera proporcional a la conectividad de cada estación, medida como el número de rutas que la sirven. Esta hipótesis se fundamenta en la observación de que las estaciones de mayor conectividad —portales y estaciones de intercambio— concentran una fracción mayor de la demanda. El total diario de viajes se fija en un valor representativo de la operación del sistema, del orden de un millón y medio de viajes en el componente troncal. Por simetría, se adopta $O_i = D_i$, hipótesis razonable para un día completo de operación en el que los desplazamientos de ida y regreso tienden a equilibrarse.

11.5.3. Calibración del parámetro de sensibilidad

El parámetro β se calibra de modo que el costo medio de viaje predicho por el modelo [ecuación (**??**)] corresponda a un valor realista para el sistema. Un barrido sistemático de β permite observar la transición del sistema entre un régimen de alta temperatura efectiva en el que la demanda se distribuye sobre desplazamientos largos y un régimen de baja temperatura efectiva en el que la demanda se concentra en desplazamientos cortos, seleccionando el valor que reproduce la longitud media de viaje característica del sistema real. La solución del modelo se obtiene mediante el algoritmo de Furness, implementado como sigue:

```
def furness(O, D, C, b, max_iter=200, tol=1e-3):
    T = np.outer(O, D) * np.exp(-b * C)    # semilla: O_i D_j exp(-b c_ij)
    for _ in range(max_iter):
        T_old = T.copy()
        rs = T.sum(axis=1); rs[rs == 0] = 1
        T *= (O / rs)[:, None]             # ajustar filas   -> O_i
        cs = T.sum(axis=0); cs[cs == 0] = 1
```

```
        T *= (D / cs)[None, :]              # ajustar columnas -> D_j
        if np.abs(T - T_old).max() < tol:
            break
    return T
```

El término `np.exp(-b * C)` corresponde al factor de Boltzmann $e^{-\beta c_{ij}}$, mientras que los ajustes sucesivos de filas y columnas implementan el cálculo iterativo de los factores de balance A_i y B_j.

11.6. Acoplamiento demanda-simulación y escenarios

La matriz origen-destino obtenida constituye la demanda que se inyecta en el simulador. Cada elemento T_{ij} de la matriz se traduce en una demanda entre la estación i y la estación j, distribuida a lo largo del día de acuerdo con un perfil horario que reproduce la estructura característica de la demanda de transporte urbano, con dos periodos de máximo flujo de personas, uno en la mañana y otro en la tarde,separados por periodos de menor demanda. Con el fin de mantener manejable el número de pares de demanda, se descartan los pares cuyo volumen de viajes es inferior a un umbral mínimo, que representan una fracción despreciable de la demanda total. La inyección de la demanda, escalada por franjas horarias, se implementa como sigue:

```
for h_ini, h_fin, frac in perfil_horario:
    t0, t1 = hora_a_seg(h_ini), hora_a_seg(h_fin)
    dur = t1 - t0
    for i in range(n):
        for j in range(n):
            if i == j:
                continue
            viajes = T[i, j] * frac    # fracción de la demanda en esta franja
            if viajes >= UMBRAL_VIAJES:
                W.adddemand(orig=lista_estaciones[i], dest=lista_estaciones[j],
                            t_start=t0, t_end=t1, flow=viajes/dur)
```

La simulación se configura con los parámetros propios del sistema troncal: una velocidad de flujo libre correspondiente a la velocidad promedio de operación de los buses, y un número de carriles por sentido acorde con la infraestructura de las troncales principales. El paso de tiempo de la simulación y el tamaño de agrupación de vehículos se ajustan para equilibrar la precisión del cálculo y el costo computacional.

Sobre esta configuración base se definen distintos escenarios que combinan dos ejes de variación. El primer eje corresponde al tipo de día, representado mediante un factor de escala de la demanda: un día laboral típico, un día de fin de semana y un día festivo, con niveles de demanda decrecientes. El segundo eje corresponde a la condición ambiental, distinguiendo entre condiciones secas y condiciones de lluvia; estas últimas se modelan mediante una reducción de la velocidad de flujo libre y de la capacidad efectiva de los tramos, que representan el efecto del clima adverso sobre la operación. Adicionalmente, y con el propósito de caracterizar la transición de fase del sistema, se realiza un barrido del factor de escala de la demanda, registrando para cada nivel las métricas de desempeño como proporción de viajes completados, velocidad media y retraso, que actúan como

parámetros de orden del sistema. Este conjunto de escenarios permite localizar la demanda crítica que separa el régimen de flujo libre del régimen congestionado, y evaluar su desplazamiento en función de las condiciones operativas y ambientales.

12. Resultados y Discusión

En este capítulo se presentan y analizan los resultados obtenidos. Se inicia con la calibración del modelo de demanda y la caracterización de la matriz origen-destino resultante; a continuación se examina en detalle el escenario de día laboral en condiciones secas, empleado como caso de referencia, a partir del conjunto completo de herramientas de visualización; posteriormente se comparan los seis escenarios definidos por la combinación del tipo de día y la condición ambiental; y finalmente se presenta la caracterización de la transición de fase del sistema y el efecto de la lluvia sobre la demanda crítica.

12.1. Calibración del modelo de demanda

La aplicación del modelo de máxima entropía sobre la red reconstruida, compuesta por las estaciones del sistema y los tramos físicos de las troncales, produce la matriz origen-destino de la cual se deriva la demanda de la simulación. El parámetro de sensibilidad al costo β se calibró mediante un barrido sistemático, registrando para cada valor la entropía relativa del sistema y el costo medio de viaje resultante. La Tabla 2 resume este barrido.

Cuadro 2: Barrido de calibración del parámetro β: entropía relativa y costo medio de viaje.

β	$S/S_{\text{máx}}$	$\langle c \rangle$ (km)
0.001	0.976	11.75
0.010	0.976	11.56
0.050	0.973	10.57
0.100	0.962	9.06
0.150	0.943	7.50
0.200	0.919	6.13
0.500	0.805	2.45
1.000	0.708	1.05

El barrido evidencia el papel de β como temperatura inversa del sistema, anticipado en el marco teórico. Para valores pequeños de β es el régimen de alta temperatura efectiva, la entropía relativa es próxima a la unidad y el costo medio de viaje es elevado, lo que indica una demanda ampliamente distribuida sobre desplazamientos largos. A medida que β aumenta y la temperatura efectiva disminuye, tanto la entropía como el costo medio descienden de forma monótona, reflejando una concentración progresiva de la demanda en desplazamientos cortos. Se seleccionó el valor $\beta = 0{,}15$, que produce un costo medio de viaje de 7,5 km y una entropía relativa de 0,94, por corresponder a una longitud media de viaje realista para el sistema troncal. Con este valor, los pares origen-destino de mayor demanda resultan ser aquellos que conectan estaciones cercanas dentro de un mismo corredor, en concordancia con el comportamiento esperado del sistema real.

12.2. Escenario de referencia: día laboral en condiciones secas

El escenario de día laboral en condiciones secas constituye el caso de referencia del estudio, al representar la operación del sistema bajo la demanda máxima y la infraestructura en condiciones normales. Este escenario se analiza a continuación con el conjunto completo de herramientas de visualización, que en los escenarios posteriores se emplearán de forma selectiva.

12.2.1. Evolución espacial de la congestión

Las Figuras 4, 5 y 6 presentan el estado de la red en tres instantes sucesivos del periodo pico ($t = 3600$, 7200 y 9000 s), con los tramos coloreados según su velocidad. En el primer instante la red se encuentra mayoritariamente en régimen de flujo libre, con congestión incipiente en unos pocos nodos. Conforme avanza el periodo pico, la congestión representada por los tramos de baja velocidad se concentra y se intensifica en la región central de la red, donde convergen múltiples corredores, mientras que los corredores periféricos permanecen en flujo libre. Este patrón reproduce el comportamiento característico del sistema real, en el que los puntos de intercambio del centro de la ciudad constituyen los principales cuellos de botella.

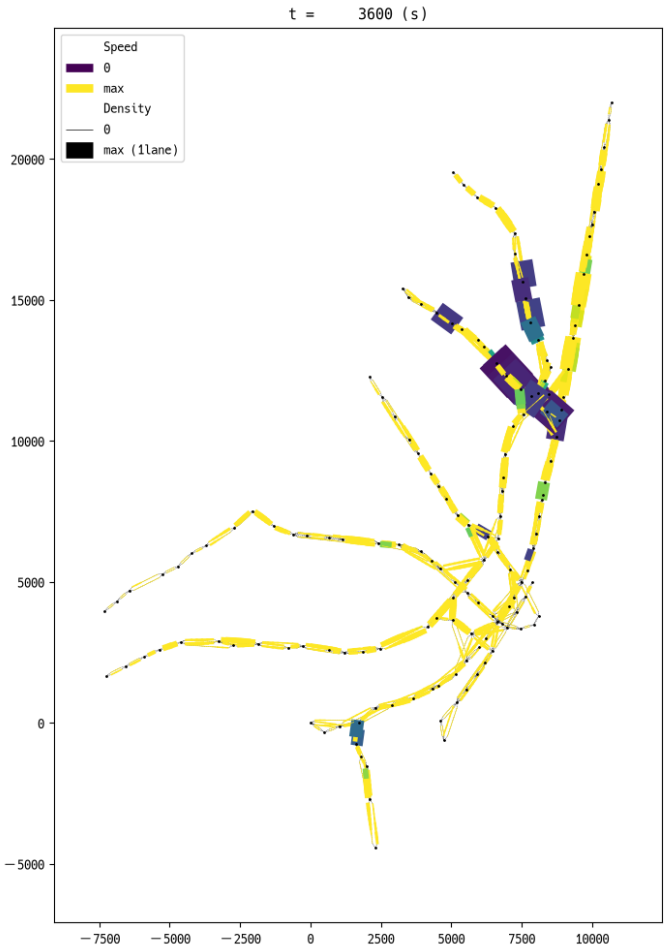


Figura 4: Estado de la red en $t = 3600$ s (día laboral seco). Los colores representan la velocidad de los tramos.

12.2.2. Análisis de un tramo congestionado

Para caracterizar la dinámica de la congestión a nivel local se selecciona el tramo de mayor retraso de la red. La Figura 7 muestra las curvas de conteo acumulado de dicho tramo: la curva de llegadas (superior) y la curva de salidas (inferior), junto con el tiempo de viaje instantáneo. La separación vertical entre ambas curvas cuantifica el número de vehículos acumulados en el tramo en cada instante, mientras que su separación horizontal representa el tiempo de demora experimentado. Se observa que, tras un periodo inicial

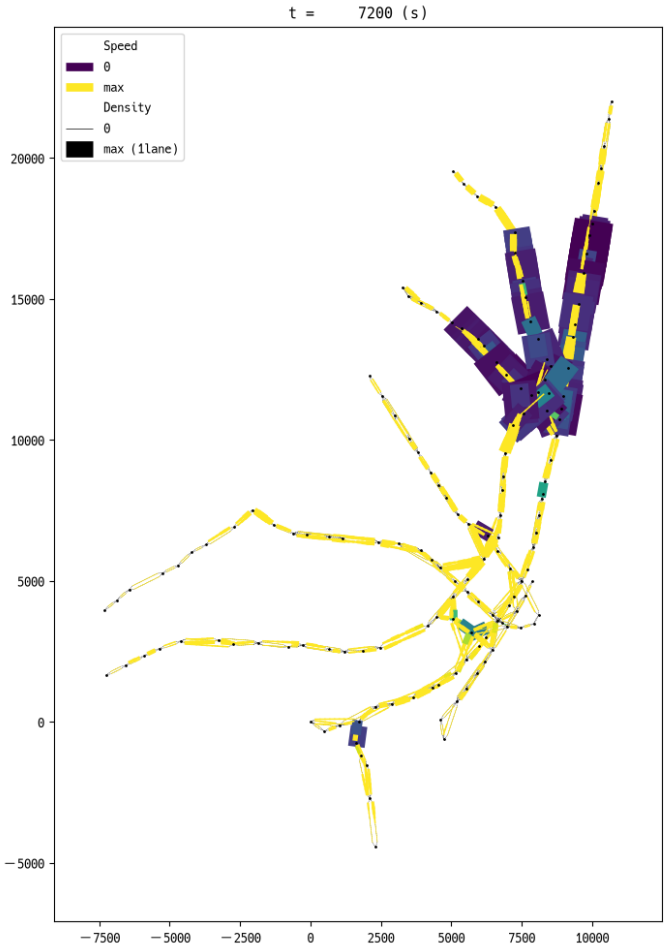


Figura 5: Estado de la red en $t = 7200$ s (día laboral seco).

de flujo libre en el que ambas curvas permanecen próximas, se produce una separación creciente a partir del inicio del periodo pico, indicativa de la acumulación de vehículos y del incremento del tiempo de viaje.

La Figura 8 presenta la visualización de las condiciones promedio de la red, en la que cada tramo se colorea según su nivel de congestión y su grosor es proporcional al volumen de tráfico. Esta representación permite identificar de un vistazo los corredores más solicitados y los que sufren mayor degradación de su desempeño a lo largo de toda la simulación.

Finalmente, las Figuras 9 y 10 presentan los diagramas espacio-tiempo de densidad y de trayectorias del tramo seleccionado. El diagrama de densidad revela la formación de bandas de alta densidad que se propagan a lo largo del tramo, correspondientes a ondas de congestión, mientras que el diagrama de trayectorias muestra el avance de los vehículos individuales, cuyas líneas exhiben el patrón escalonado característico de la detención y el arranque sucesivos en condiciones de congestión y ante los semáforos.

12.3. Comparación de escenarios

La Tabla 3 resume las métricas de desempeño de los seis escenarios, definidos por la combinación del tipo de día (laboral, sábado y festivo) y la condición ambiental (seca y de lluvia).

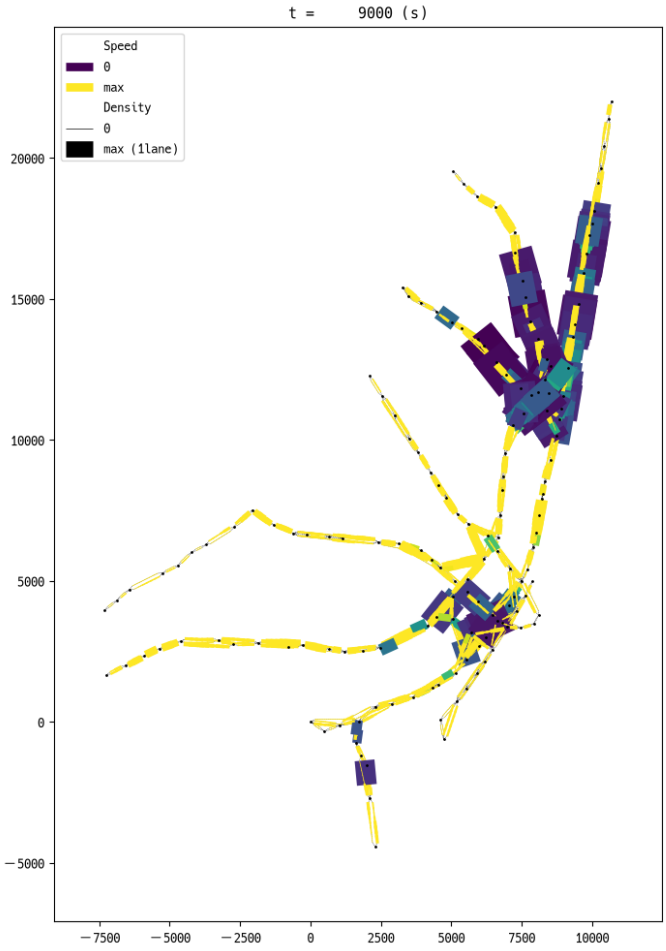


Figura 6: Estado de la red en $t = 9000$ s (día laboral seco), en el momento de máxima congestión del periodo pico.

Los resultados revelan tres comportamientos de interés. En primer lugar, el día laboral es el único tipo de día en el que el sistema se aproxima al colapso: en condiciones secas completa apenas el 87,6 % de los viajes, con una razón de retraso de 0,561, mientras que el sábado y el festivo completan entre el 93 % y el 97 % con razones de retraso considerablemente menores. Esto indica que únicamente la demanda del día laboral supera el umbral crítico del sistema, en tanto que los días de menor demanda operan cómodamente en el régimen de flujo libre. Este hallazgo es consistente con la interpretación de la congestión como un fenómeno de umbral, y no como una degradación proporcional a la demanda.

En segundo lugar, el efecto de la lluvia depende fuertemente del régimen de demanda en el que se encuentre el sistema. En el día laboral, la lluvia produce un deterioro drástico: la velocidad media desciende de 16,6 a 2,9 m/s y la proporción de viajes completados cae del 87,6 % al 61,4 %. En cambio, en el día festivo la lluvia apenas altera el desempeño global. Esta dependencia se explica por la proximidad del sistema al punto crítico: cuando la demanda es elevada y el sistema opera cerca de la saturación, la reducción de capacidad inducida por la lluvia basta para precipitar el colapso; cuando la demanda es baja y existe un amplio margen de capacidad, la misma reducción no alcanza a saturar el sistema. La vulnerabilidad del sistema frente a las condiciones ambientales es, por tanto, una función del régimen de demanda.

En tercer lugar, se observa que en el día festivo la razón de retraso en condiciones de

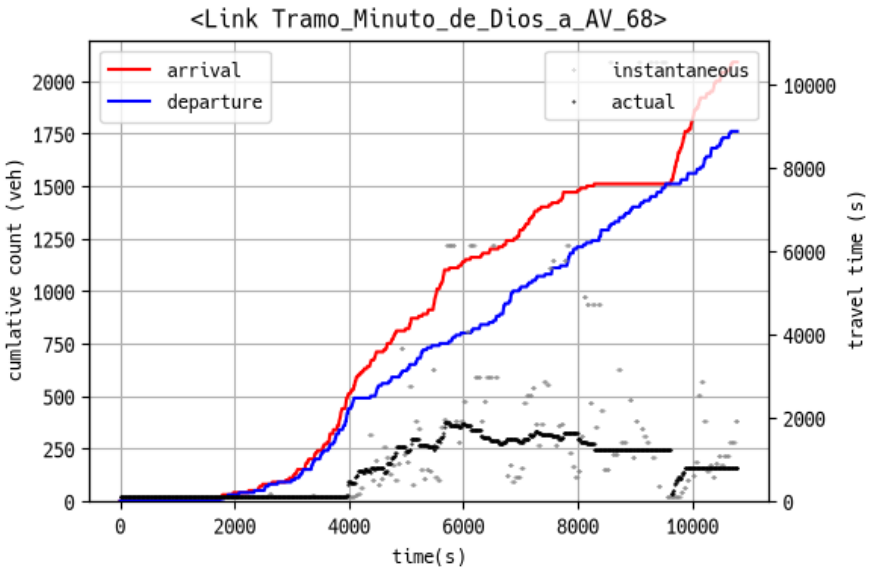


Figura 7: Curvas de conteo acumulado del tramo de mayor retraso (día laboral seco). La curva superior corresponde a las llegadas y la inferior a las salidas; su separación cuantifica los vehículos acumulados y la demora.

Cuadro 3: Métricas de desempeño de los seis escenarios simulados.

Escenario	Completados (%)	$\langle t \rangle$ (s)	Retraso (s)	Razón de retraso	Vel. (m/s)
Laboral seco	87.6	614.4	344.8	0.561	16.6
Laboral lluvia	61.4	930.2	605.0	0.650	2.9
Sábado seco	94.8	306.8	78.3	0.255	13.6
Sábado lluvia	93.0	451.5	140.2	0.311	9.0
Festivo seco	96.9	199.8	28.9	0.144	14.7
Festivo lluvia	96.0	268.1	33.2	0.124	10.8

lluvia (0,124) resulta ligeramente inferior a la de condiciones secas (0,144). Este resultado, aparentemente contraintuitivo, se explica por la naturaleza de la métrica: la razón de retraso se calcula respecto del tiempo de viaje en condiciones ideales, que a su vez depende de la velocidad de flujo libre; al reducirse dicha velocidad en condiciones de lluvia, la referencia se modifica, de modo que la razón de retraso deja de ser directamente comparable entre condiciones de distinta velocidad. En términos absolutos, el tiempo medio de viaje en festivo lluvia (268,1 s) es mayor que en festivo seco (199,8 s), lo que confirma que la lluvia sí incrementa el tiempo de viaje, aun cuando la razón de retraso no lo refleje directamente.

Las Figuras correspondientes a los mapas de red y a las condiciones promedio de los escenarios restantes confirman visualmente estos hallazgos: los escenarios de fin de semana y festivo presentan una red mayoritariamente en flujo libre, mientras que el escenario de día laboral en lluvia exhibe la mayor extensión de tramos congestionados.

12.4. Transición de fase y demanda crítica

Con el fin de caracterizar de manera sistemática la transición entre el régimen de flujo libre y el régimen congestionado, se realizó un barrido del factor de escala de la demanda

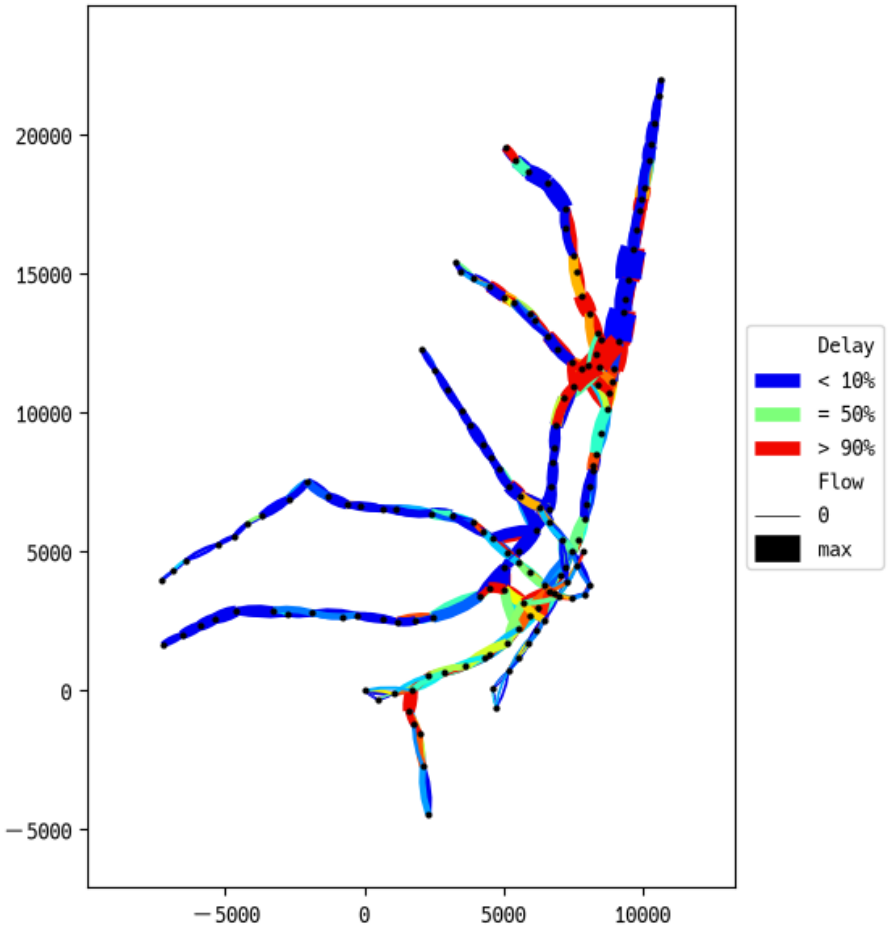


Figura 8: Condiciones promedio de la red (día laboral seco). El color indica el nivel de congestión y el grosor el volumen de tráfico.

λ, manteniendo constante la estructura de la matriz origen-destino y registrando, para cada nivel de demanda, la proporción de viajes completados y el retraso medio. Este procedimiento se llevó a cabo tanto en condiciones secas como en condiciones de lluvia. La Figura 11 presenta los resultados.

El panel izquierdo de la Figura 11, que muestra la proporción de viajes completados en función de λ, exhibe la firma característica de una transición de fase. Para valores bajos de la demanda, el sistema completa la práctica totalidad de los viajes y se mantiene en el régimen de flujo libre. Al superar un valor crítico λ_c, la proporción de viajes completados decae de forma abrupta, marcando el paso al régimen congestionado. La proporción de viajes completados actúa así como parámetro de orden de la transición, análogo a la magnetización en una transición ferromagnética o a la densidad en una transición líquido-vapor. La localización de la caída permite estimar la demanda crítica del sistema, que se sitúa en el entorno de $\lambda_c \approx 0{,}5$ en condiciones secas.

La comparación de las curvas seca y de lluvia constituye uno de los resultados centrales de este trabajo. La curva correspondiente a condiciones de lluvia se sitúa sistemáticamente por debajo de la curva seca en la región de transición, lo que indica que, para un mismo nivel de demanda, el sistema en condiciones de lluvia completa una menor proporción de viajes. Equivalentemente, la transición al régimen congestionado ocurre a un valor de demanda menor: la lluvia desplaza el punto crítico λ_c hacia la izquierda. Este resultado

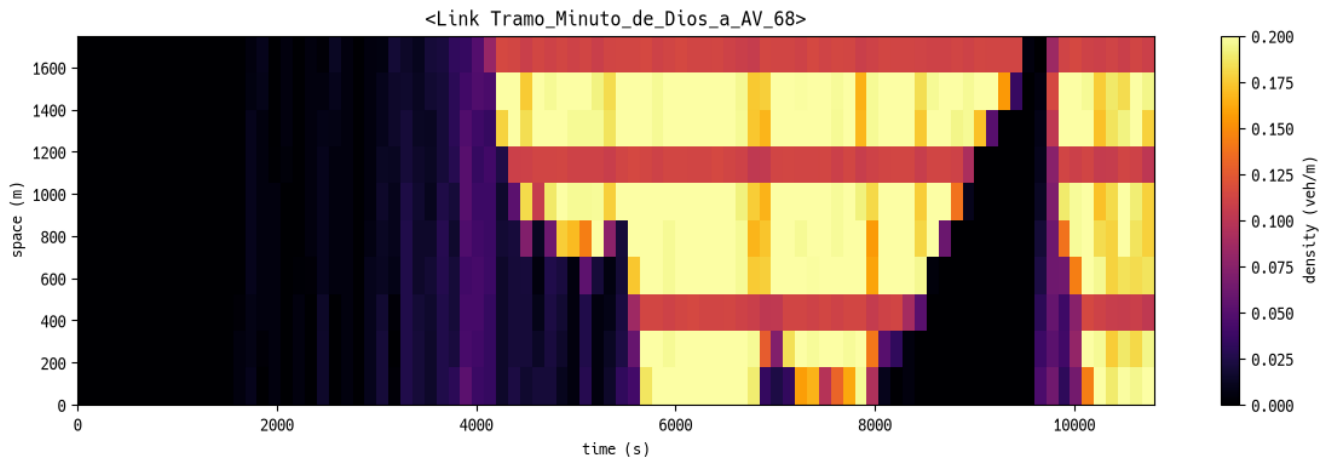


Figura 9: Diagrama espacio-tiempo de densidad del tramo de mayor retraso (día laboral seco). Las bandas claras corresponden a regiones de alta densidad (congestión).

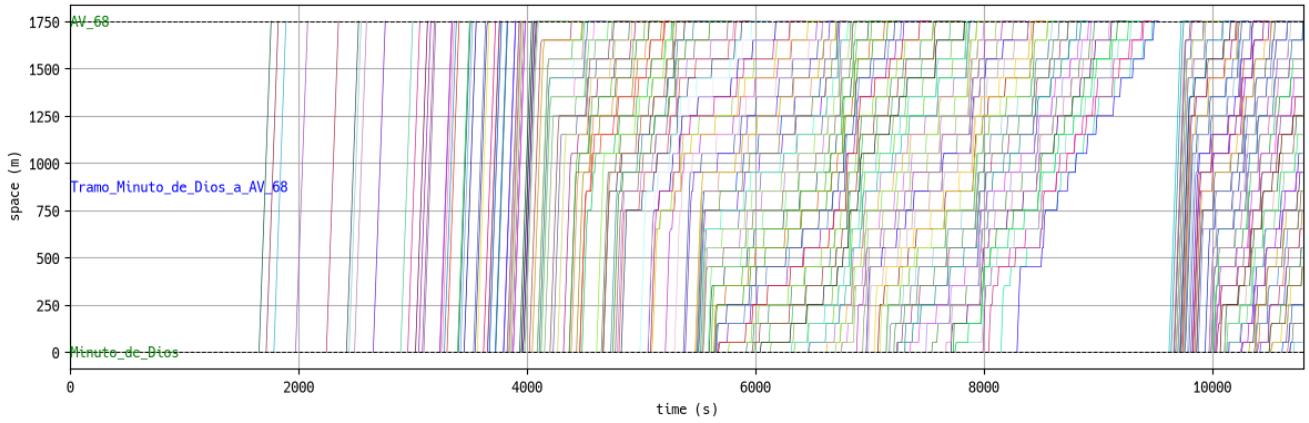


Figura 10: Diagrama espacio-tiempo de trayectorias del tramo de mayor retraso (día laboral seco). El patrón escalonado evidencia la detención y arranque de los vehículos.

formaliza y cuantifica la observación empírica de que las condiciones climáticas adversas reducen la resiliencia del sistema, y proporciona una interpretación física precisa: al disminuir la capacidad efectiva de la red, la lluvia reduce la demanda que el sistema puede absorber antes de colapsar.

El panel derecho de la Figura 11, que muestra el retraso medio en función de λ, complementa esta interpretación y revela un fenómeno adicional de interés físico. El retraso medio aumenta con la demanda hasta alcanzar un máximo en las proximidades del punto crítico, y desciende posteriormente para valores mayores de λ. Este comportamiento no lineal es consistente con la teoría del diagrama fundamental del tráfico: más allá de la densidad crítica, el sistema ingresa en la rama congestionada en la que el flujo efectivo disminuye, de modo que un número progresivamente menor de vehículos logra completar su viaje, y el retraso medio, calculado únicamente sobre los viajes completados deja de crecer. El máximo del retraso medio señala, por tanto, la frontera entre los dos regímenes, en concordancia con la demanda crítica estimada a partir del parámetro de orden.

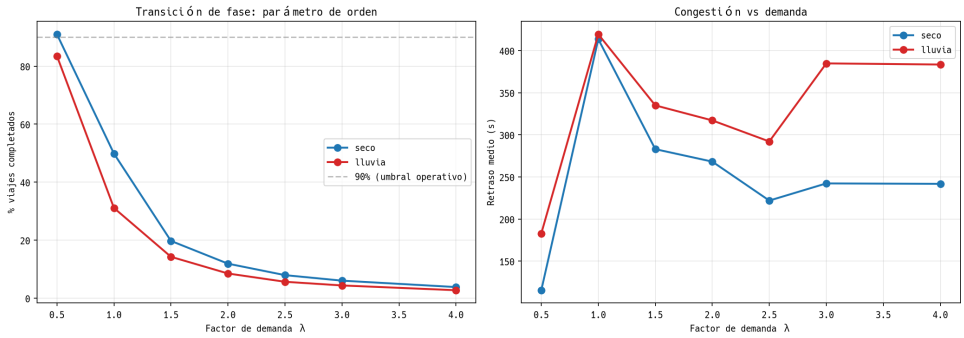


Figura 11: Transición de fase del sistema. Izquierda: proporción de viajes completados (parámetro de orden) en función del factor de demanda λ. Derecha: retraso medio en función de λ. Las curvas azul y roja corresponden a condiciones secas y de lluvia, respectivamente.

13. Optimización del Sistema

El tercer objetivo de este trabajo consiste en formular y evaluar propuestas de optimización del sistema TransMilenio a partir de los patrones de congestión identificados. A diferencia de las aproximaciones puramente cualitativas, en las que las propuestas de mejora se enuncian sin una verificación cuantitativa, en este trabajo cada intervención se implementa explícitamente en el modelo y su efecto se mide mediante el desplazamiento que produce sobre la demanda crítica del sistema, es decir, sobre el valor del factor de demanda λ_c que marca la transición del régimen de flujo libre al régimen congestionado. De este modo, la optimización se plantea como un problema físico bien definido: una intervención es beneficiosa en la medida en que desplaza el punto crítico hacia valores mayores de la demanda, ampliando el rango en el que el sistema opera de manera funcional.

El presente capítulo se organiza en varias etapas. En primer lugar se describe la metodología de evaluación y la identificación automática de los tramos críticos. A continuación se evalúan de manera individual tres propuestas de intervención, con el fin de aislar el efecto de cada una. Posteriormente se construye un escenario de optimización integral que combina la totalidad de las intervenciones, representando la cota superior de mejora alcanzable. Seguidamente se desarrolla la interpretación termodinámica de los resultados, y finalmente se plantean las líneas de investigación futura.

13.1. Metodología de la evaluación

La evaluación de las propuestas de optimización se fundamenta en el barrido del factor de escala de la demanda λ introducido en el Capítulo 12. Para cada configuración del sistema, la situación actual y cada una de las propuestas de mejora se ejecuta una serie de simulaciones a distintos niveles de demanda, registrando la proporción de viajes completados como parámetro de orden. La comparación de las curvas resultantes permite determinar el desplazamiento de la demanda crítica atribuible a cada intervención.

Un elemento central de esta metodología es la identificación automática de los tramos que constituyen el principal cuello de botella del sistema, el denominado nudo central, sobre los cuales se concentran varias de las intervenciones. Dicha identificación se realiza ejecutando una simulación del sistema en su configuración base bajo una demanda elevada, y seleccionando los tramos de mayor tiempo medio de viaje. El Fragmento **??** presenta la rutina correspondiente.

```
def tramos_criticos(top_n=25):
    W = World(name="tmp", deltan=10, tmax=3*3600,
              save_mode=0, print_mode=0, show_mode=0)
    for est, (x, y) in estaciones.items():
        W.addNode(name=est, x=x, y=y, signal=signal_de(est))
    for a in list(adyacencia):
        for b_, length in adyacencia[a].items():
            nom = f"Tramo_{a}_a_{b_}"
            if nom not in W.LINKS_NAME_DICT:
                W.addLink(name=nom, start_node=a, end_node=b_, length=length,
                          free_flow_speed=16.6, number_of_lanes=2, signal_group=0)
    for i in range(n):
        for j in range(n):
            if i == j: continue
            v = T[i, j] * 0.24
            if v >= UMBRAL_VIAJES:
                W.adddemand(orig=lista_estaciones[i], dest=lista_estaciones[j],
                            t_start=0, t_end=3*3600, flow=v/(3*3600))
    W.exec_simulation()
    df = W.analyzer.link_to_pandas()
    top = df.sort_values("average_travel_time",
                         ascending=False)["link"].head(top_n).tolist()
    return set(top)
```

La rutina construye la red en su configuración base, inyecta la demanda correspondiente al periodo pico, ejecuta la simulación y extrae, a partir de la tabla de resultados por tramo, el conjunto de los `top_n` tramos con mayor tiempo medio de viaje. Este conjunto, que se almacena para su uso posterior, corresponde geográficamente a la región de convergencia de los corredores en el centro de la red, en concordancia con lo observado en los mapas del Capítulo 12. La identificación automática garantiza que las intervenciones se dirijan a los tramos que efectivamente limitan el desempeño del sistema, y no a tramos seleccionados de manera arbitraria.

13.2. Evaluación individual de las propuestas

Se evalúan tres propuestas de intervención, seleccionadas por su viabilidad y por su correspondencia con estrategias reales de gestión de sistemas de transporte masivo. La primera propuesta (P1), *refuerzo del nudo*, incrementa la capacidad de los tramos del cuello de botella central de dos a tres carriles, representando la ampliación física de la infraestructura. La segunda propuesta (P2), *semaforización optimizada*, reduce la duración del ciclo semafórico de 120 a 80 segundos, representando la mejora de la gestión del tráfico sin ampliación de la infraestructura. La tercera propuesta (P3) corresponde a la combinación de las dos anteriores.

El Fragmento **??** presenta la función que ejecuta una simulación para un nivel de demanda y una configuración dados, parametrizada de modo que permite representar tanto la situación base como cada una de las propuestas.

```
def correr(lam, config):
    W = World(name="opt", deltan=10, tmax=3*3600,
```

```
            save_mode=0, print_mode=0, show_mode=0)
    for est, (x, y) in estaciones.items():
        if est in senal_por_estacion:
            fases = senal_por_estacion[est]
            verde = max(10, round(config["ciclo"] / fases))
            sig = [verde, max(10, config["ciclo"] - verde)]
        else:
            sig = [0]
        W.addNode(name=est, x=x, y=y, signal=sig)
    for a in list(adyacencia):
        for b_, length in adyacencia[a].items():
            nom = f"Tramo_{a}_a_{b_}"
            if nom in W.LINKS_NAME_DICT: continue
            lanes = config["lanes_nudo"] if nom in TRAMOS_NUDO else config["lanes_base
            W.addLink(name=nom, start_node=a, end_node=b_, length=length,
                      free_flow_speed=16.6, number_of_lanes=lanes, signal_group=0)
    for i in range(n):
        for j in range(n):
            if i == j: continue
            v = T[i, j] * 0.24 * lam
            if v >= UMBRAL_VIAJES:
                W.adddemand(orig=lista_estaciones[i], dest=lista_estaciones[j],
                            t_start=0, t_end=3*3600, flow=v/(3*3600))
    W.exec_simulation()
    df = W.analyzer.basic_to_pandas()
    frac = df["completed_trips"].iloc[0] / df["total_trips"].iloc[0]
    delay = df["average_delay"].iloc[0]
    return frac, delay
```

La función recibe el factor de demanda `lam` y un diccionario `config` que especifica los parámetros de la configuración: el número de carriles del nudo (`lanes_nudo`), el número de carriles del resto de la red (`lanes_base`) y la duración del ciclo semafórico (`ciclo`). La condición `nom in TRAMOS_NUDO` determina, para cada enlace, si pertenece al conjunto de tramos críticos identificados previamente, aplicándole en tal caso el número de carriles reforzado. Tras ejecutar la simulación, la función devuelve la proporción de viajes completados y el retraso medio, magnitudes que constituyen las métricas de la evaluación.

El barrido de demanda para las cuatro configuraciones produce los resultados resumidos en la Tabla 4, representados gráficamente en la Figura 12.

Cuadro 4: Proporción de viajes completados (%) en función del factor de demanda λ, para la situación base y las tres propuestas de intervención individuales.

λ	0.5	1.0	1.5	2.0	2.5	3.0
Base	90.5	47.8	20.4	12.3	9.6	7.9
P1: Refuerzo nudo	90.4	51.2	20.6	12.5	9.4	7.8
P2: Semáforo optim.	90.7	43.4	18.9	11.8	9.9	8.4
P3: Combinada	88.9	46.5	19.3	12.2	9.3	7.7

El análisis de la Figura 12 y de la Tabla 4 revela varios aspectos de interés. En primer lugar, las cuatro configuraciones exhiben la misma estructura cualitativa: una proporción

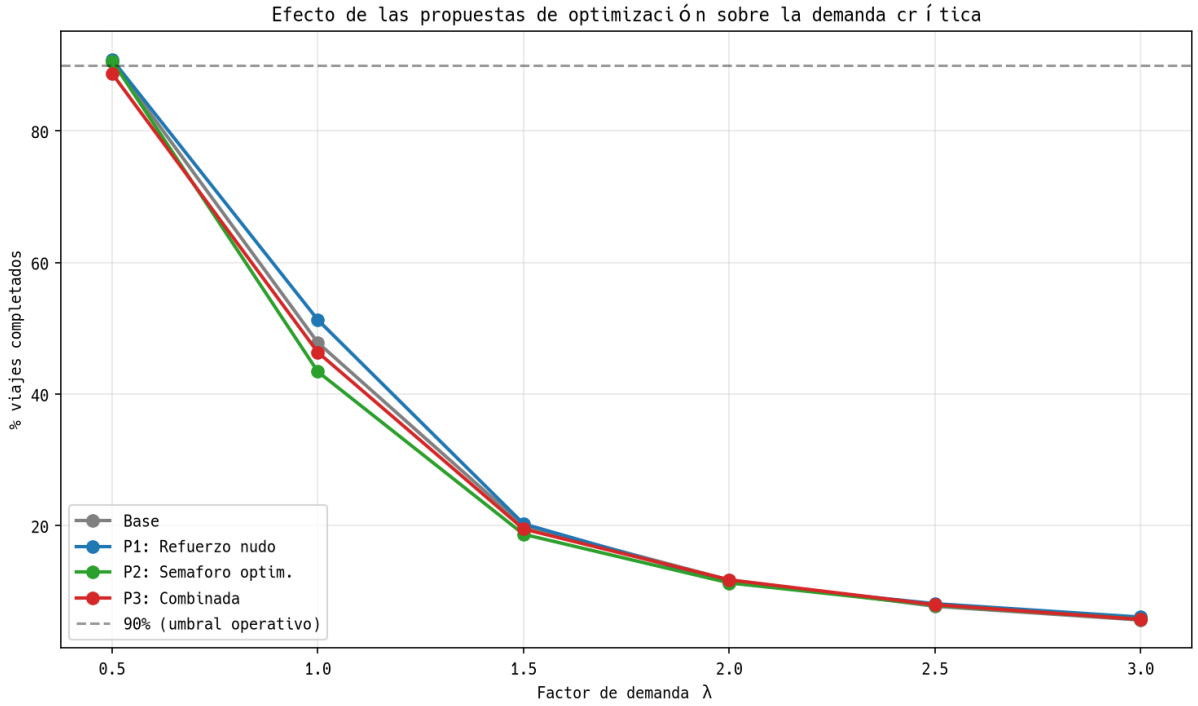


Figura 12: Proporción de viajes completados en función del factor de demanda λ para la situación base y las tres propuestas de intervención individuales. La línea discontinua indica el umbral operativo del 90 %.

de viajes completados próxima al umbral operativo para $\lambda = 0{,}5$, seguida de una caída pronunciada al aumentar la demanda, que conduce al colapso del sistema. La demanda crítica, el valor de λ en el que se cruza el umbral del 90 %, se sitúa en el entorno de $\lambda_c \approx 1{,}0$ para todas las configuraciones, lo que indica que ninguna intervención individual desplaza de manera apreciable el punto crítico.

En segundo lugar, y de manera más reveladora, las diferencias entre las configuraciones se concentran en el régimen de transición, en el entorno de $\lambda = 1{,}0$, donde la pendiente de la caída es máxima. En este punto, el refuerzo del nudo (P1) alcanza la mayor proporción de viajes completados (51,2 %), superando en más de tres puntos porcentuales a la situación base (47,8 %). Este comportamiento admite una interpretación física precisa: en las proximidades del punto crítico, la limitación dominante del sistema es la capacidad de flujo de los tramos del cuello de botella, de modo que su ampliación produce el efecto más notorio. La región de transición es, en el lenguaje de los fenómenos críticos, aquella en la que el sistema exhibe la máxima sensibilidad a las variaciones de sus parámetros, y es allí donde las intervenciones sobre la capacidad manifiestan su mayor influencia.

En tercer lugar, se observa que la optimización semafórica (P2), si bien no mejora la proporción de viajes completados en la zona de transición e incluso la reduce ligeramente respecto de la base, incide de manera preferente sobre el retraso medio de los viajes que sí se completan, magnitud sobre la que actúa directamente al reducir el tiempo de detención en las intersecciones. Esta distinción entre el efecto sobre la completitud de los viajes y el efecto sobre el retraso pone de manifiesto que las distintas intervenciones actúan sobre dimensiones diferentes del desempeño del sistema.

Finalmente, para demandas elevadas ($\lambda \geq 1{,}5$), las cuatro curvas convergen hacia valores prácticamente idénticos, muy por debajo del umbral operativo. Esta convergencia

indica que, una vez que la demanda supera suficientemente la capacidad estructural de la red, las diferencias entre las configuraciones se desvanecen: el sistema colapsa con independencia de la intervención aplicada. Este resultado, lejos de constituir una limitación del análisis, refleja fielmente la naturaleza física del fenómeno y anticipa la necesidad de un enfoque integral, que se aborda en la sección siguiente.

13.3. Escenario de optimización integral

Con el fin de determinar la cota superior de mejora alcanzable, se construye un escenario de optimización integral que combina simultáneamente cuatro intervenciones: el refuerzo de capacidad del nudo central; la optimización del ciclo semafórico, llevado a 60 segundos con una fracción de verde de al menos el 60 % asignada a la troncal, en representación de una priorización semafórica del sistema BRT; el incremento de la velocidad de operación en los tramos reforzados, de 16,6 a 18 m/s; y una estrategia de gestión de la demanda, modelada mediante la redistribución de una fracción del 15 % de la demanda del periodo pico hacia otras franjas horarias, en representación de medidas como el escalonamiento de jornadas laborales o el incentivo de los desplazamientos fuera de las horas de máxima afluencia. El Fragmento **??** presenta la función que implementa este escenario.

```
def correr_integral(lam, integral):
    if integral:
        ciclo = 60                  # semáforo optimizado (ciclo corto)
        verde_frac_min = 0.6        # al menos 60% de verde a la troncal
        lanes_nudo = 3              # refuerzo del nudo
        vel_nudo = 18.0             # mayor velocidad en tramos mejorados
        gestion_demanda = 0.85      # redistribución del 15% del pico
    else:
        ciclo = 120; verde_frac_min = None
        lanes_nudo = 2; vel_nudo = 16.6; gestion_demanda = 1.0
    W = World(name="int", deltan=10, tmax=3*3600,
              save_mode=0, print_mode=0, show_mode=0)
    for est, (x, y) in estaciones.items():
        if est in senal_por_estacion:
            fases = senal_por_estacion[est]
            verde = max(10, round(ciclo / fases))
            if verde_frac_min is not None:
                verde = max(verde, int(ciclo * verde_frac_min))
            sig = [verde, max(10, ciclo - verde)]
        else:
            sig = [0]
        W.addNode(name=est, x=x, y=y, signal=sig)
    for a in list(adyacencia):
        for b_, length in adyacencia[a].items():
            nom = f"Tramo_{a}_a_{b_}"
            if nom in W.LINKS_NAME_DICT: continue
            if nom in NUDO:
                lanes, vel = lanes_nudo, vel_nudo
            else:
                lanes, vel = 2, 16.6
```

```
            W.addLink(name=nom, start_node=a, end_node=b_, length=length,
                      free_flow_speed=vel, number_of_lanes=lanes, signal_group=0)
    for i in range(n):
        for j in range(n):
            if i == j: continue
            v = T[i, j] * 0.24 * lam * gestion_demanda
            if v >= UMBRAL_VIAJES:
                W.adddemand(orig=lista_estaciones[i], dest=lista_estaciones[j],
                            t_start=0, t_end=3*3600, flow=v/(3*3600))
    W.exec_simulation()
    df = W.analyzer.basic_to_pandas()
    frac = df["completed_trips"].iloc[0] / df["total_trips"].iloc[0]
    return frac
```

La estructura condicional inicial establece los parámetros de la configuración integral cuando el argumento `integral` es verdadero, y los de la situación base en caso contrario. Es de destacar el tratamiento de la priorización semafórica, implementado mediante la instrucción `verde = max(verde, int(ciclo * verde_frac_min))`, que garantiza que la fracción de verde asignada a la troncal no sea inferior al umbral establecido, independientemente del número de fases del semáforo. Asimismo, la gestión de la demanda se implementa mediante el factor multiplicativo `gestion_demanda`, que reduce proporcionalmente la demanda inyectada en el periodo pico.

La Figura 13 presenta la comparación entre la situación base y el escenario de optimización integral, y la Tabla 5 resume los valores correspondientes.

Cuadro 5: Proporción de viajes completados (%) en función del factor de demanda λ, para la situación base y el escenario de optimización integral.

λ	0.5	0.8	1.1	1.4	1.7	2.0
Base	97.0	93.0	49.8	28.4	17.2	12.6
Optimización integral	97.8	96.8	96.3	64.1	34.5	24.0

El contraste entre la Figura 13 y la evaluación individual es notable. Mientras que las intervenciones aisladas apenas modificaban la posición del punto crítico, el escenario integral desplaza la demanda crítica desde $\lambda_c \approx 1{,}1$ en la situación base hasta $\lambda_c \approx 1{,}4$ en la configuración optimizada, lo que representa un incremento cercano al 27 % en la demanda que el sistema es capaz de absorber antes de ingresar en el régimen congestionado. El área sombreada de la figura, comprendida entre ambas curvas, representa el conjunto de viajes que el sistema optimizado logra completar y que el sistema actual no; su magnitud, particularmente en el intervalo $1{,}1 \leq \lambda \leq 1{,}7$, ilustra de manera directa el alcance de la mejora.

Los valores de la Tabla 5 cuantifican esta diferencia. En $\lambda = 1{,}1$, un nivel de demanda en el que la situación base ya ha colapsado , completando apenas el 49,8 % de los viajes, el sistema optimizado mantiene un desempeño prácticamente óptimo, completando el 96,3 %. Incluso en $\lambda = 1{,}4$, donde la base completa solo el 28,4 %, la optimización integral sostiene un 64,1 %. La diferencia máxima entre ambas configuraciones, superior a cuarenta puntos porcentuales, se alcanza precisamente en la región de transición del sistema base, confirmando que el mayor beneficio de la optimización se obtiene en el entorno del punto crítico.

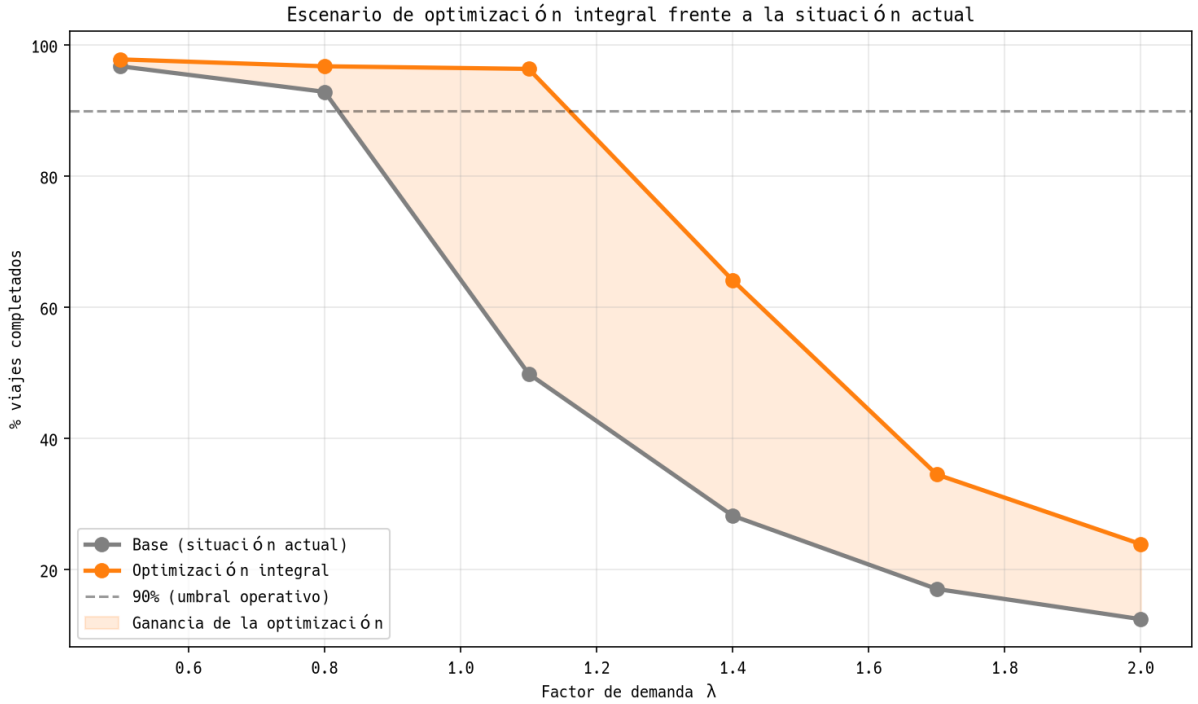


Figura 13: Proporción de viajes completados en función del factor de demanda λ para la situación base (gris) y el escenario de optimización integral (naranja). El área sombreada representa la ganancia de desempeño atribuible a la optimización.

Es igualmente significativo el comportamiento del sistema bajo demandas extremas. En $\lambda = 2{,}0$, ambas curvas se aproximan nuevamente 12,6 % para la base frente a 24,0 % para la configuración optimizada, y aunque la optimización mantiene una ventaja, el sistema optimizado también ha ingresado en el régimen congestionado. Este resultado revela la existencia de un límite físico irreductible: la optimización operativa desplaza el punto crítico, pero no lo elimina, pues la capacidad estructural de la red impone una cota máxima a la demanda absorbible. Más allá de dicha cota, la única vía para preservar la funcionalidad del sistema es la actuación directa sobre la demanda.

13.4. Interpretación de la optimización

La transición entre el régimen de flujo libre y el régimen congestionado, observada al variar el factor de demanda λ, es formalmente análoga a una transición de fase en un sistema termodinámico, en la que el factor de demanda desempeña el papel de un parámetro de control análogo a la densidad de partículas de un fluido.

En efecto, en la analogía desarrollada en el marco teórico, los viajes o los buses que los materializan constituyen las partículas de un sistema de muchos cuerpos que se desplaza sobre la red. El factor de demanda λ regula el número de tales partículas presentes en el sistema y, por tanto, su densidad efectiva. Al incrementar λ, la densidad de vehículos sobre los tramos aumenta hasta alcanzar la densidad crítica k_c del diagrama fundamental [ecuación (23)], a partir de la cual el sistema abandona la rama de flujo libre e ingresa en la rama congestionada. La demanda crítica λ_c identificada en este trabajo es, en consecuencia, la manifestación macroscópica a escala de la red completa de la densidad crítica que gobierna la transición a nivel de cada tramo. El parámetro de orden de esta transición, la

proporción de viajes completados, es análogo a las magnitudes que distinguen las fases en los sistemas físicos, como la fracción de fase condensada en una transición líquido-vapor: próximo a la unidad en la fase de flujo libre, decae abruptamente al atravesar el punto crítico.

Bajo esta perspectiva, las intervenciones de optimización pueden interpretarse como modificaciones de los parámetros que determinan la ubicación del punto crítico del sistema. El refuerzo de la capacidad de los tramos incrementa la densidad crítica k_c a la que ocurre la transición, mientras que la optimización de la semaforización reduce las pérdidas de capacidad asociadas a las detenciones, desplazando de igual modo el umbral de saturación. La gestión de la demanda, por su parte, no modifica el punto crítico, sino que reduce la densidad efectiva de partículas para un mismo nivel de demanda nominal, alejando al sistema de la región crítica. El efecto conjunto de estas intervenciones el desplazamiento del punto crítico hacia valores mayores de λ equivale, en el lenguaje termodinámico, a un aumento de la capacidad del sistema para permanecer en la fase ordenada de flujo libre bajo condiciones de creciente solicitación.

Esta interpretación permite comprender por qué las intervenciones individuales resultan insuficientes y por qué el enfoque integral es cualitativamente superior. En las proximidades de una transición de fase, el sistema exhibe una elevada susceptibilidad: pequeñas variaciones de los parámetros de control producen grandes variaciones del estado macroscópico. Una intervención individual actúa sobre un único parámetro, produciendo un desplazamiento limitado del punto crítico que, dada la abrupta pendiente de la transición, se traduce en una mejora modesta. El enfoque integral, en cambio, actúa simultáneamente sobre varios parámetros la capacidad, las pérdidas por detención y la densidad efectiva, cuyos efectos se acumulan para producir un desplazamiento sustancial del punto crítico. La existencia de un límite irreductible bajo demandas extremas encuentra asimismo una interpretación natural: así como ningún ajuste de las condiciones externas puede impedir la transición de fase de un sistema físico una vez superado ampliamente su punto crítico, ninguna optimización operativa puede evitar el colapso de la red cuando la demanda excede de manera suficiente su capacidad estructural.

13.5. Líneas de investigación futura

Los resultados obtenidos en este trabajo abren diversas líneas de investigación que podrían profundizar y extender el modelo desarrollado.

Una primera línea corresponde al refinamiento del modelo de demanda mediante la incorporación de datos reales desagregados de afluencia por estación. En el presente trabajo, los totales de origen y destino se estimaron a partir de la conectividad de las estaciones, ante la ausencia de datos públicos desagregados. La disponibilidad de registros de validación de acceso permitiría calibrar la matriz origen-destino con mayor precisión y contrastar las predicciones del modelo de máxima entropía con la demanda observada, fortaleciendo la validez empírica de los resultados.

Una segunda línea consiste en la extensión del modelo hacia una representación más detallada del sistema de transporte público, que incorpore explícitamente el comportamiento de los pasajeros en las estaciones, los procesos de abordaje y descenso, los tiempos de permanencia y la capacidad de los andenes. Dado que el simulador empleado modela el flujo de los vehículos pero no la dinámica de los pasajeros en las paradas, la integración de un modelo de transporte público específico permitiría capturar fenómenos de congestión que se originan en las estaciones, complementando la congestión de los tramos analizada

en este trabajo.

Una tercera línea de investigación, de carácter metodológico, corresponde a la aplicación de técnicas de aprendizaje automático a la optimización del sistema. La metodología desarrollada en este trabajo, que evalúa cada intervención mediante simulación, podría integrarse como entorno de evaluación de un agente de aprendizaje por refuerzo, capaz de explorar de manera autónoma el espacio de posibles intervenciones, distribución de carriles, tiempos semafóricos, estrategias de gestión de la demanda y de proponer configuraciones óptimas que maximicen la demanda crítica del sistema. Un enfoque de esta naturaleza requeriría la generación de un extenso conjunto de simulaciones de entrenamiento, pero permitiría abordar la optimización como un problema de búsqueda en un espacio de alta dimensión, superando las limitaciones de la evaluación de un número reducido de configuraciones predefinidas.

Una cuarta línea corresponde a la incorporación de la variabilidad estocástica y de escenarios de perturbación. El análisis desarrollado consideró condiciones promedio para cada tipo de día y de clima; la introducción de fluctuaciones aleatorias en la demanda y en la capacidad, así como la simulación de eventos de perturbación, fallas operativas, cierres de estaciones, picos súbitos de demanda, permitiría estudiar la resiliencia del sistema desde la perspectiva de la respuesta a perturbaciones, un tema de creciente relevancia en el estudio de los sistemas complejos.

Finalmente, una quinta línea consiste en la extensión del análisis termodinámico mediante la formalización de magnitudes adicionales. La caracterización del sistema en términos de una temperatura efectiva podría complementarse con la definición y el cálculo de análogos de otras magnitudes termodinámicas, capacidades, susceptibilidades, exponentes críticos, lo que profundizaría la analogía entre el sistema de transporte y los sistemas físicos y permitiría una descripción más completa de la naturaleza de la transición de fase observada.

14. Conclusiones

El presente trabajo desarrolló un modelo físico-estadístico del sistema de transporte masivo TransMilenio, integrando la estimación de la demanda de viajes mediante el principio de máxima entropía con la simulación mesoscópica del flujo sobre la red real de la ciudad. A partir de este modelo se caracterizó la transición entre el régimen de flujo libre y el régimen congestionado, se evaluó el efecto de las condiciones operativas y ambientales sobre dicha transición, y se formularon y verificaron propuestas de optimización. Las principales conclusiones que se derivan de este trabajo se exponen a continuación.

Se construyó una representación fiel de la red física del sistema a partir de datos geográficos y operacionales reales. El procedimiento desarrollado basado en la reconstrucción de los corredores a partir de la geometría de los trazados, la composición de una malla físicamente consistente y la expansión de los recorridos mediante el cálculo de caminos de menor costo permitió representar la totalidad de la red troncal como una única componente conectada, sobre la que los distintos servicios comparten los tramos de las troncales. La incorporación de los semáforos reales que regulan el paso de los buses constituyó un elemento determinante para reproducir la congestión, al introducir el principal mecanismo de pérdida de capacidad del sistema.

La aplicación del modelo de máxima entropía permitió estimar la matriz origen-destino como la distribución más probable de los viajes, y confirmó la interpretación del parámetro de sensibilidad al costo como una temperatura inversa del sistema. El barrido de dicho parámetro reprodujo la transición del sistema entre un régimen de alta temperatura efectiva con desplazamientos largos y elevada entropía y un régimen de baja temperatura efectiva con desplazamientos cortos y menor entropía, permitiendo su calibración a un valor que reproduce una longitud media de viaje realista.

El resultado central del trabajo es la caracterización de la congestión como una transición de fase. El barrido del factor de demanda reveló que la proporción de viajes completados, tomada como parámetro de orden, se mantiene próxima a la unidad en el régimen de flujo libre y decae de manera abrupta al superar una demanda crítica, en plena analogía con las transiciones de fase de los sistemas físicos. Este hallazgo demuestra que la congestión no es una degradación proporcional al aumento de la demanda, sino un fenómeno de umbral: el sistema permanece funcional hasta un punto crítico a partir del cual colapsa rápidamente.

La evaluación cuantitativa de las propuestas de optimización demostró que las intervenciones individuales, el refuerzo de la capacidad del nudo central y la optimización de la semaforización, producen mejoras localizadas en el régimen de transición, pero resultan insuficientes para desplazar de manera significativa la demanda crítica. Por el contrario, el escenario de optimización integral, que combina simultáneamente el refuerzo de la capacidad, la priorización semafórica, la mejora de la velocidad operativa y la gestión de la demanda, desplaza la demanda crítica en un orden del veintisiete por ciento, ampliando de manera sustancial el rango de operación funcional del sistema. La interpretación termodinámica de este resultado reveló que la superioridad del enfoque integral es una consecuencia directa de la naturaleza crítica de la transición: sólo la acción simultánea sobre múltiples parámetros de control permite trasladar de forma efectiva la frontera entre las fases del sistema.

Un hallazgo con implicaciones para la planificación del transporte, es la existencia de un límite físico irreductible a la optimización operativa. Bajo demandas suficientemente extremas, ni siquiera la combinación de todas las intervenciones evita el colapso del

sistema, pues la capacidad estructural de la red impone una cota máxima a la demanda absorbible. Este resultado subraya que las mejoras de infraestructura y de gestión operativa, aunque necesarias y efectivas dentro de un rango, deben complementarse necesariamente con estrategias de gestión de la demanda para garantizar la resiliencia del sistema.

Referencias

[1] Wilson, A. G. (1970). *Entropy in Urban and Regional Modelling*. Pion, Londres.

[2] Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *Physical Review*, 106(4), 620–630.

[3] Nagel, K., & Schreckenberg, M. (1992). A Cellular Automaton Model for Freeway Traffic. *Journal de Physique I*, 2(12), 2221–2229.

[4] Schadschneider, A. (2000). Statistical Physics of Traffic Flow. *Physica A*, 285(1–2), 101–120.

[5] Helbing, D. (2001). Traffic and Related Self-Driven Many-Particle Systems. *Reviews of Modern Physics*, 73(4), 1067–1141.

[6] Lighthill, M. J., & Whitham, G. B. (1955). On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society of London A*, 229(1178), 317–345.

[7] Newell, G. F. (2002). A Simplified Car-Following Theory: A Lower Order Model. *Transportation Research Part B*, 36(3), 195–205.

[8] Schweitzer, F. (2018). Sociophysics: To the extent that individuals interact with each other in prescribed ways, their collective social behavior can be modeled and analyzed. *Physics Today*, 71(2), 40–46.

[9] Caldarelli, G., Wolf, S., & Moreno, Y. (2018). Physics of humans, physics for society. *Nature Physics*, 14, 870.

[10] Helbing, D. (2012). Pluralistic modeling of complex systems. *PLoS ONE*, 7(9), e50775.

[11] Du, W.-B., Zhou, X.-L., Jusup, M., & Wang, Z. (2016). Physics of transportation: Towards optimal capacity using the multilayer network framework. *Scientific Reports*, 6, 19059.

[12] Zachary, D. S. (2022). Modelling shifts in social opinion through an application of classical physics. *Scientific Reports*, 12, 5485.

[13] Mavromatidis, L. (2022). *Constructal thermodynamics and its semantic ontology in autopoietic, digital, and computational architectural and urban space open systems*. Elsevier.

[14] Gürdal, M., Gürsoy, E., Gedik, E., & Mavromatidis, L. (2023). The first and second law analyses of thermodynamics for a human living at Kastamonu and Karabük cities in Türkiye. *Thermal Science and Engineering Progress*.

[15] Wang, X., Jia, Y., Shao, M., Chang, Z., & Zhang, P. (2024). Optimization Algorithms for Urban Energy Balance: The Application of Thermodynamic Methods in Low-Carbon Ecological City Design. IIETA.

[16] Cederlof, G., Ramírez, F. A., Correal, N. A., Chala, M. C., Hoyos, M. D., Ochoa, A. F., & Rivera, A. L. (2024). Energy as object and relation: Thermodynamics, space, and emergent energy geographies. *Geoforum*.

[17] Sugar, L., & Kennedy, C. (2023). Thermodynamics of urban growth revealed by city scaling. *Physica A*. Department of Civil Engineering, University of Toronto and University of Victoria, Canada.

[18] Wilson, C., & Shonk, J. K. P. (2023). Microscale to neighbourhood scale: Impact of shading on urban climate. *Building and Environment*.

[19] Tyagi, P. K., Kumar, R., Paschalis, A., Bohnenstengel, S. I., & van Reeuwijk, M. (2023). Thermodynamic modeling and performance optimization of nanofluid-based photovoltaic/thermal system using central composite design scheme of response surface methodology. Imperial College London, University of Reading, University of Cyprus.

[20] Li, E., Huang, J., Yu, H., Liu, S., He, W., Zhang, W., Pang, H., & Zhang, C. (2024). Photoaged tire wear particles hinder the transport of Pb(II) in urban soils under acid rain: Experimental and numerical investigations. *Water Research*.

[21] Gu, Y., Shi, X., Jin, P., & Wang, X. C. (2025). Assessing entropy changes in urban sewer systems during pollutant degradation. *Water Science & Technology*, 91(4), 424.

[22] Moreno Triviño, J., & Castro García, W. (2020). *General guidelines for the design of BRT routes in the Public Transport Integrated System of Bogotá*. National University of Colombia.

[23] Ramírez, F. A., Correal, N. A., Chala, M. C., Hoyos, M. D., Ochoa, A. F., & Rivera, A. L. (2021). Improving BRT route design through code: the case of Bogotá's. En *14th Conference on Transport Engineering* (6–8 de julio).

[24] TransMilenio S.A. (2025). *Mapa interactivo de TransMilenio*. Recuperado de `https://www.transmilenio.gov.co/publicaciones/150402/publicaciones-mapa-interactivo-de-transmilenio/`

[25] Cámara de Comercio de Bogotá. (2025). *En Bogotá - Región viven más de 10 millones de habitantes*. Recuperado de `https://www.ccb.org.co`

[26] Seo, T. (2025). *UXsim: Urban traffic simulator* [Repositorio GitHub]. Recuperado de `https://github.com/toruseo/UXsim`

[27] Seo, T. (2023). UXsim: An Open Source Macroscopic and Mesoscopic Traffic Simulator in Python — a Technical Overview. *arXiv preprint arXiv:2309.17114*.

A. Anexo

En este anexo se presenta la implementación completa del modelo desarrollado en este trabajo, escrita en lenguaje Python y organizada en las mismas etapas descritas en la metodología. El código se apoya en las bibliotecas `numpy` para el cálculo numérico, `xml.etree.ElementTree` y `csv` para la lectura de los datos, `heapq` para el algoritmo de Dijkstra, `matplotlib` para la visualización, y `uxsim` para la simulación del flujo de tráfico.

A.1. Configuración, geometría y lectura de datos

El primer bloque establece los parámetros de configuración, define las funciones de proyección geográfica y de geometría sobre las polilíneas, y lee los tres conjuntos de datos: las estaciones y los trazados del archivo KML, las rutas del sistema troncal y la red semafórica.

```
import xml.etree.ElementTree as ET
import math, csv, heapq
from collections import defaultdict
import numpy as np
import matplotlib.pyplot as plt
from uxsim import World

# --- Configuración ---
KML_PATH = "transmilenio.kml"
CSV_PATH = "transmilenio_rutas.csv"
SEM_PATH = "Red_Semaforica.csv"
LAT_ORIGEN = 4.571180071494235
LON_ORIGEN = -74.13999213342498
UMBRAL_DIST_A_RUTA = 120
UMBRAL_ADYACENCIA_M = 1800
MIN_ESTACIONES_CORREDOR = 5
UMBRAL_SEM_M = 200
MAX_BOGOTA_M = 30000
CICLO_S = 120
TOTAL_VIAJES_DIA = 1_500_000
b = 0.15

# --- Geometría ---
def proyectar_a_metros(lon, lat):
    R = 6378137.0
    d_lat = math.radians(lat - LAT_ORIGEN)
    d_lon = math.radians(lon - LON_ORIGEN)
    return d_lon * R * math.cos(math.radians(LAT_ORIGEN)), d_lat * R

def distancia(p1, p2):
    return math.hypot(p1[0]-p2[0], p1[1]-p2[1])

def proyectar_punto_en_segmento(p, a, b_):
    ax, ay = a; bx, by = b_; px, py = p
```

```
    dx, dy = bx-ax, by-ay
    seglen2 = dx*dx + dy*dy
    if seglen2 == 0:
        return 0.0, a, distancia(p, a)
    t = max(0.0, min(1.0, ((px-ax)*dx + (py-ay)*dy) / seglen2))
    return t, (ax+t*dx, ay+t*dy), distancia(p, (ax+t*dx, ay+t*dy))

def proyectar_punto_en_polilinea(p, poli):
    acum = [0.0]
    for i in range(1, len(poli)):
        acum.append(acum[-1] + distancia(poli[i-1], poli[i]))
    mejor_d, mejor_arco = float("inf"), None
    for i in range(len(poli)-1):
        t, _, d = proyectar_punto_en_segmento(p, poli[i], poli[i+1])
        if d < mejor_d:
            mejor_d, mejor_arco = d, acum[i] + t*distancia(poli[i], poli[i+1])
    return mejor_arco, mejor_d

def distancia_real(pa, pb):
    mejor_score, mejor_long = float("inf"), None
    for poli in rutas_geom:
        aa, da = proyectar_punto_en_polilinea(pa, poli)
        ab, db = proyectar_punto_en_polilinea(pb, poli)
        if da < UMBRAL_DIST_A_RUTA and db < UMBRAL_DIST_A_RUTA and
        da+db < mejor_score:
            mejor_score, mejor_long = da+db, abs(ab-aa)
    return (mejor_long, True) if mejor_long is not None else
    (distancia(pa, pb), False)

# --- Lectura del KML ---
tree = ET.parse(KML_PATH); root = tree.getroot()
ns = {'kml': 'http://www.opengis.net/kml/2.2'}
estaciones = {}
for pm in root.findall('.//kml:Placemark', ns):
    nn = pm.find('kml:name', ns); pt = pm.find('.//kml:Point/kml:coordinates', ns)
    if nn is not None and pt is not None:
        nombre = nn.text.strip().replace(" ", "_")
        lon, lat = map(float, pt.text.strip().split(',')[:2])
        estaciones[nombre] = proyectar_a_metros(lon, lat)
rutas_geom = []
for pm in root.findall('.//kml:Placemark', ns):
    ls = pm.find('.//kml:LineString/kml:coordinates', ns)
    if ls is not None:
        pts = [proyectar_a_metros(*map(float, tok.split(',')[:2]))
               for tok in ls.text.strip().split()]
        rutas_geom.append(pts)

# --- Lectura de las rutas ---
```

```
def parse_est(t):
    return [e.strip() for e in t.split(";") if e.strip()]
rutas_bus = {}
with open(CSV_PATH, encoding="utf-8") as f:
    cont = {}
    for row in csv.DictReader(f):
        r = row["ruta"].strip()
        seq = [row["origen"].strip()] + parse_est(row["recorrido_intermedio"]) \
              + [row["final"].strip()]
        i = cont.get(r, 0); rutas_bus[f"{r}_dir{i}"] = seq; cont[r] = i+1

# --- Lectura de los semáforos ---
senal_por_estacion = {}
for fila in csv.DictReader(open(SEM_PATH, encoding="utf-8-sig")):
    if fila["GRUPOS_BRT"].strip() in ("", "N/A"):
        continue
    try:
        lon, lat = float(fila["X"]), float(fila["Y"]); fases
        = int(fila["NUMEROFASES"])
    except (ValueError, KeyError):
        continue
    p = proyectar_a_metros(lon, lat)
    mejor_est, mejor_d = None, float("inf")
    for e, pe in estaciones.items():
        d = distancia(p, pe)
        if d < mejor_d:
            mejor_d, mejor_est = d, e
    if mejor_d > MAX_BOGOTA_M or mejor_d > UMBRAL_SEM_M:
        continue
    if mejor_est not in senal_por_estacion or fases
    > senal_por_estacion[mejor_est]:
        senal_por_estacion[mejor_est] = fases

def signal_de(est):
    if est not in senal_por_estacion:
        return [0]
    fases = senal_por_estacion[est]
    verde = max(10, round(CICLO_S / fases))
    return [verde, max(10, CICLO_S - verde)]
```

A.2. Construcción de la malla física

El segundo bloque construye la malla física de la red: reconstruye los corredores a partir de las polilíneas, compone la malla añadiendo las adyacencias cortas de las rutas, e implementa el algoritmo de Dijkstra empleado para la expansión de los recorridos y para el cálculo de la matriz de costos.

```
adyacencia = defaultdict(dict)
def agregar_arista(a, b_, length):
```

```
    if a == b_ or length <= 1: return
    if b_ not in adyacencia[a] or length < adyacencia[a][b_]:
        adyacencia[a][b_] = length; adyacencia[b_][a] = length

# Sub-tramos de cada corredor (ordenados por arco)
for poli in rutas_geom:
    sobre = []
    for est, p in estaciones.items():
        arco, d = proyectar_punto_en_polilinea(p, poli)
        if d < UMBRAL_DIST_A_RUTA:
            sobre.append((arco, est))
    if len(sobre) < MIN_ESTACIONES_CORREDOR: continue
    sobre.sort()
    for (aa, a), (ab, b_) in zip(sobre, sobre[1:]):
        agregar_arista(a, b_, abs(ab-aa))

# Adyacencias cortas de las rutas
for seq in rutas_bus.values():
    for a, b_ in zip(seq, seq[1:]):
        if a == b_ or a not in estaciones or b_ not in estaciones
        or b_ in adyacencia[a]:
            continue
        length, _ = distancia_real(estaciones[a], estaciones[b_])
        if length <= UMBRAL_ADYACENCIA_M:
            agregar_arista(a, b_, length)

# Dijkstra
_cache = {}
def camino_mas_corto(origen, destino):
    if origen == destino: return [origen]
    if (origen, destino) in _cache: return _cache[(origen, destino)]
    dist = {origen: 0}; prev = {}; pq = [(0, origen)]; visto = set()
    while pq:
        d, u = heapq.heappop(pq)
        if u in visto: continue
        visto.add(u)
        if u == destino: break
        for v, w in adyacencia[u].items():
            nd = d + w
            if v not in dist or nd < dist[v]:
                dist[v] = nd; prev[v] = u; heapq.heappush(pq, (nd, v))
    if destino not in prev:
        _cache[(origen, destino)] = None; return None
    cam = [destino]
    while cam[-1] != origen: cam.append(prev[cam[-1]])
    cam.reverse(); _cache[(origen, destino)] = cam
    return cam
```

A.3. Modelo de demanda por máxima entropía

El tercer bloque implementa el modelo de demanda: calcula la matriz de costos mediante Dijkstra, estima los totales de origen y destino a partir de la conectividad de las estaciones, y resuelve el modelo gravitacional doblemente restringido mediante el algoritmo de Furness.

```
lista_estaciones = sorted(estaciones.keys())
n = len(lista_estaciones)

# Matriz de costos (distancia por la red, en km)
C = np.zeros((n, n))
for i, a in enumerate(lista_estaciones):
    for j, b_est in enumerate(lista_estaciones):
        if i == j: continue
        cam = camino_mas_corto(a, b_est)
        C[i, j] = sum(adyacencia[x][y] for x, y in zip(cam, cam[1:]))
        if cam else 0
C[C == 0] = C[C > 0].max() * 2
np.fill_diagonal(C, 0)
C_km = C / 1000.0

# Totales de origen y destino (proporcionales a la conectividad)
rutas_por_est = {e: 0 for e in lista_estaciones}
for seq in rutas_bus.values():
    for e in set(seq):
        if e in rutas_por_est: rutas_por_est[e] += 1
peso = np.array([rutas_por_est[e] for e in lista_estaciones], dtype=float)
peso[peso == 0] = 1; peso /= peso.sum()
O = peso * TOTAL_VIAJES_DIA
D = peso * TOTAL_VIAJES_DIA

# Algoritmo de Furness (máxima entropía)
def furness(O, D, C, b, max_iter=200, tol=1e-3):
    T = np.outer(O, D) * np.exp(-b * C)
    for _ in range(max_iter):
        T_old = T.copy()
        rs = T.sum(axis=1); rs[rs == 0] = 1; T *= (O/rs)[:, None]
        cs = T.sum(axis=0); cs[cs == 0] = 1; T *= (D/cs)[None, :]
        if np.abs(T - T_old).max() < tol: break
    return T

T = furness(O, D, C_km, b)
```

A.4. Construcción del mundo, demanda y simulación

El cuarto bloque construye el mundo de simulación en UXsim —nodos con sus semáforos y enlaces de la malla física—, inyecta la matriz origen-destino como demanda distribuida según el perfil horario, y ejecuta la simulación.

```
DELTAN = 5
TMAX = 68400
UMBRAL_VIAJES = 20

def hora_a_seg(h, mm=0):
    return ((h - 4) * 60 + mm) * 60

perfil = [(4,6,0.05),(6,9,0.24),(9,12,0.12),(12,14,0.10),
          (14,16,0.09),(16,19,0.26),(19,21,0.09),(21,23,0.05)]

W = World(name="TransMilenio_OD", deltan=DELTAN, tmax=TMAX,
          print_mode=1, save_mode=1, show_mode=1)

# nodos con semáforo
for nombre, (x, y) in estaciones.items():
    W.addNode(name=nombre, x=x, y=y, signal=signal_de(nombre))

# enlaces de la malla física
for a in list(adyacencia):
    for b_, length in adyacencia[a].items():
        nom = f"Tramo_{a}_a_{b_}"
        if nom not in W.LINKS_NAME_DICT:
            W.addLink(name=nom, start_node=a, end_node=b_, length=length,
                      free_flow_speed=16.6, number_of_lanes=2, signal_group=0)

# demanda de la matriz OD por franjas horarias
for h_ini, h_fin, frac in perfil:
    t0, t1 = hora_a_seg(h_ini), hora_a_seg(h_fin)
    dur = t1 - t0
    for i in range(n):
        for j in range(n):
            if i == j: continue
            viajes = T[i, j] * frac
            if viajes >= UMBRAL_VIAJES:
                W.adddemand(orig=lista_estaciones[i], dest=lista_estaciones[j],
                            t_start=t0, t_end=t1, flow=viajes/dur)

# ejecutar simulación
W.exec_simulation()
W.analyzer.print_simple_stats()
```

A.5. Barrido de la transición de fase y optimización

El último bloque implementa el barrido del factor de demanda para caracterizar la transición de fase, y las funciones de evaluación de las propuestas de optimización. Ambas rutinas construyen el mundo de simulación para cada configuración y nivel de demanda, y registran la proporción de viajes completados como parámetro de orden. Estas funciones se presentaron parcialmente en el Capítulo 13; se reproduce aquí la función de barrido

genérica.

```
def barrer_demanda(lambdas, vel=16.6, lanes=2, ciclo=120):
    resultados = []
    for lam in lambdas:
        W = World(name="barrido", deltan=10, tmax=3*3600,
                  save_mode=0, print_mode=0, show_mode=0)
        for est, (x, y) in estaciones.items():
            if est in senal_por_estacion:
                fases = senal_por_estacion[est]
                verde = max(10, round(ciclo / fases))
                sig = [verde, max(10, ciclo - verde)]
            else:
                sig = [0]
            W.addNode(name=est, x=x, y=y, signal=sig)
        for a in list(adyacencia):
            for b_, length in adyacencia[a].items():
                nom = f"Tramo_{a}_a_{b_}"
                if nom not in W.LINKS_NAME_DICT:
                    W.addLink(name=nom, start_node=a, end_node=b_, length=length,
                              free_flow_speed=vel, number_of_lanes=lanes,
                              signal_group=0)
        for i in range(n):
            for j in range(n):
                if i == j: continue
                v = T[i, j] * 0.24 * lam
                if v >= 20:
                    W.adddemand(orig=lista_estaciones[i], dest=lista_estaciones[j],
                                t_start=0, t_end=3*3600, flow=v/(3*3600))
        W.exec_simulation()
        df = W.analyzer.basic_to_pandas()
        frac = df["completed_trips"].iloc[0] / df["total_trips"].iloc[0]
        resultados.append((lam, frac))
    return resultados
```

El código completo, incluyendo las rutinas de generación de las figuras y de los escenarios comparativos, se encuentra disponible en el repositorio del proyecto.