

Descripción del comportamiento de las velocidades de las rutas SITP, a partir de regresiones geográficas ponderadas utilizando variables Socioeconómicas y viales



**UNIVERSIDAD DE MANIZALES
FACULTAD DE CIENCIAS E INGENIERÍA
ESPECIALIZACIÓN EN SISTEMAS DE INFORMACIÓN GEOGRÁFICA
MANIZALES
2018**

**DESCRIPCIÓN DEL COMPORTAMIENTO DE LAS VELOCIDADES DE LAS
RUTAS SITP, A PARTIR DE REGRESIONES GEOGRÁFICAS PONDERADAS
UTILIZANDO VARIABLES SOCIOECONÓMICAS Y VIALES**

Trabajo de Grado presentado como opción parcial para optar
al título de Especialista en Información Geográfica

**UNIVERSIDAD DE MANIZALES
FACULTAD DE CIENCIAS E INGENIERÍA
ESPECIALIZACIÓN EN SISTEMAS DE INFORMACIÓN GEOGRÁFICA
MANIZALES
2018**

CONTENIDO

1. ÁREA PROBLEMÁTICA.....	14
2. OBJETIVOS.....	16
2.1 OBJETIVO GENERAL.....	16
2.2 OBJETIVOS ESPECÍFICOS.....	16
3. JUSTIFICACIÓN.....	17
4. MARCO TEÓRICO.....	18
4.1 IMPORTANCIA DE LOS ESTUDIOS DE VELOCIDAD.....	18
4.2 ANTECEDENTES.....	23
4.2.1 Importancia de estudios de Velocidad.....	24
4.2.2 Definición de Velocidad.....	25
4.2.3 Tipos de Velocidad.....	26
4.2.4 Estudios de Velocidad.....	27
4.2.5 Modelamiento estadístico.....	28
4.2.6 Econometría Espacial.....	30
4.2.7 Modelos de regresión.....	32
4.2.8 Síntesis general.....	34
5. METODOLOGÍA.....	35
5.1 TIPO DE TRABAJO Y ACTIVIDADES.....	35
5.2 DATOS REQUERIDOS.....	35
5.3 SOFTWARE.....	36
5.4 ETAPAS DE TRABAJO.....	36
5.5 ZONA DE ESTUDIO.....	38
5.6 PREPROCESAMIENTO DE LA INFORMACION.....	41
6. RESULTADOS.....	50
6.1 PROCESAMIENTO CAPA PREDIAL.....	50
6.2 VALIDACIÓN DE SUPUESTOS MEDIANTE OLS.....	56
6.3 CONFIGURACION DE PARAMETROS PARA LA REGRESION.....	60

6.4 RESULTADO DE LA REGRESION CON MAYOR AJUSTE.....	62
7. CONCLUSIONES	73
8. RECOMENDACIONES.....	75
9. BIBLIOGRAFÍA	76

LISTA DE FIGURAS

Figura 1. Variables influyentes en un sistema de transporte masivo.....	20
Figura 2. Zona de estudio	39
Figura 3. Vías principales localidad de Engativá.....	40
Figura 4. Zonas de operación del SITP	41
Figura 5. Clasificación Por Usos.....	45
Figura 6. Detalle base datos de Velocidades	49
Figura 7. Geodatabase Base SITP.....	56
Figura 8. Depuración de segmentos viales.....	57
Figura 9. Resultados OLS en ArcGis	58
Figura 10. Geodatabase Resultados_Regresion	61
Figura 11. Resumen regresión del día 12	63
Figura 12. Resumen estadísticos de coeficientes de regresion	64
Figura 13. Detalle del CIV con R2 local de 0,5.....	65
Figura 14. Detalle del CIV con R2 local de 0,6.....	66
Figura 15. Detalle del CIV con R2 local de 0,7.....	67
Figura 16. Detalle del CIV con R2 local de 0,8.....	68
Figura 17. Detalle del CIV con R2 local de 0,9.....	69
Figura 18. Comportamiento del R2 en la mañana	70
Figura 19. Comportamiento del R2 en el medio día	70
Figura 20. Comportamiento del R2 en la tarde	71

LISTA DE TABLAS

Tabla 1. Matriz de Contigüidad de Torre	31
Tabla 2. Etapas del trabajo.....	36
Tabla 3. Puntaje por gravedad de accidentes	41
Tabla 4. Puntaje por gravedad de accidentes	42
Tabla 5. Descripción de información geográfica IDECA.....	43
Tabla 6. Descripción de información geográfica DANE.....	44
Tabla 7. Reclasificación atributo Tipo_Super(tipo de superficie).....	54
Tabla 8. Reclasificación atributo Clasificación vial	54
Tabla 9. Reclasificación atributo Estado_Mal(Estado malla vial).....	54
Tabla 10. Atributos capa resultante	55
Tabla 11. Variables del modelo de regresión	56
Tabla 12. Cálculo del I de Moran.....	60
Tabla 13. Cálculos de GWR con distinto número de vecinos.....	61
Tabla 14. Resultado de los R ² de las GWR	62

LISTA DE CUADROS

Cuadro 1. Topología para la capa Lote.....	46
Cuadro 2. Errores de correspondencia información DANE-IDECA	47
Cuadro 3. Generacion de la capa manzana a partir de la capa Lote	48
Cuadro 4. Muestra de los segmentos viales.....	50
Cuadro 5. Muestra de los Bufer y su influencia.....	51
Cuadro 6. Influencia del segmento vial.....	52
Cuadro 7. Proceso de Intersect y Dissolve	53

LISTA DE ANEXOS

Anexo A. Mapa zona de Estudio	78
Anexo B. Mapa de velocidades hora pico mañana	79
Anexo C. Mapa de velocidades hora pico Medio Día.....	80
Anexo D. Mapa de velocidades hora pico tarde.....	81
Anexo E. Mapa comparativo CIV completo Vs. Velocidad por CIV estimadas por GWR.....	82
Anexo F. Base de datos	83

GLOSARIO

Base datos geográfica: Es una colección de datos organizados de la tal manera que sirvan efectivamente para una o varias aplicaciones SIG, comprende la asociación entre dos componentes principales: datos espaciales y atributos o datos no espaciales.

Cartografía: Es la ciencia que se encarga del trazado y el estudio de mapas geográficos.

CIV: Código de identificación vial, es un numero que identifica cada segmento vial.

Fiabilidad estadística: Permite garantizar la validez y precisión de análisis estadísticos, se relaciona con la capacidad de reproducir los resultados tantas veces como sea necesario. Esto es esencial, ya que genera confianza en el análisis estadístico y en los resultados obtenidos.

Geoestadística: Es un conjunto de técnicas para el análisis y predicción de valores distribuidos en el espacio y/o en el tiempo, dichos valores se asumen correlacionados entre sí.

Geo procesos: Son las funciones de un GIS que mediante operaciones concretas da un información precisa procede de otras preexistentes , se trata de herramientas orientadas a la recopilación y tratamiento de informaciones espaciales, y datos geográficos

Lote: Es un espacio geográfico no separado por otro lote con o sin construcciones y/o edificaciones, perteneciente a personas naturales o jurídicas, es la unidad mínima catastral

Manzana: Código que identifica una subdivisión física existente en los sectores catastrales, separadas entre sí por vías de tránsito vehicular o peatonal y/o limitadas por accidentes naturales como cerros, acequias, ríos, etc., las que deben estar geo-referenciadas para su identificación catastral. Contiene información de loteo, acotamientos, nomenclatura vial y domiciliaria, construcción, número de pisos y sectorización

Patrón espacial: Es un conjunto de rasgos esenciales en un diseño gráfico, mapa o escrito, que ocurren una y otra vez

Prueba estadística: Es una forma de evaluar la evidencia que los datos proporcionan para probar una hipótesis.

Regresión geográfica ponderada: Es una herramienta de geoprocésamiento, que permite analizar la relación entre dos o más variables para explicar un fenómeno dentro de un espacio geográfico.

Variable explicativa: Variable que mediante análisis y, proceso estadístico y regresiones simples dan explicación de un fenómeno o de una variable explicada en concreto.

Variable explicada: Son las variables que se pretenden estudiar y explicar su comportamiento a través de regresiones simples

RESUMEN

Frente de los diferentes problemas que enfrenta el Sistema Integrado de Transporte SITP que no han permitido se convierta en el sistema de transporte ideal para la ciudad de Bogotá por la falta de acogida entre los usuarios y el bajo cumplimiento de garantías, es necesario abordar soluciones que adopten las nuevas tecnologías de la información geográfica.

Utilizando técnicas de análisis multivariado como lo son las regresiones geográficas ponderadas, se busca explicar a partir de variables del entorno de funcionamiento, el comportamiento de las velocidades del SITP, ubicando la zona de estudio en la localidad de Engativá en la ciudad de Bogotá.

La evaluación y explicación de variables demográficas, caracterización catastral y de los segmentos viales por donde circula la flota vehicular, permite identificar patrones que sometidos a diferentes pruebas estadísticas miden la confiabilidad con que explican el fenómeno en estudio.

Finalmente se realiza una serie de análisis espaciales de las variables estudiadas, y de cómo se relacionan entre sí para explicar las velocidades en la flota de las rutas SITP, con ayuda de herramientas de visualización geográfica y cartografía se verifican y contextualizan los resultados obtenidos dentro un marco geográfico.

PALABRAS CLAVES: regresión geográfica ponderada, geoestadística, base datos geográfica, geo procesos.

ABSTRACT

The constant problems faced by the Integrated Public Transport System (SITP), They have not allowed it to become the ideal transportation system for the city of Bogotá due to the lack of acceptance among users and the low compliance with guarantees. Due to this, it is necessary to address solutions that adopted new geographic information technologies.

Using multivariate methods such as weighted geographical regressions, we seek to explain, from variables of the operating environment, the speed behavior of the SITP, locating the study area in the town of Engativa in the city of Bogotá.

The evaluation and explanation of the demographic variables, the cadastral characterization and the road segments through which the vehicle fleet circulates, allow identifying patterns that are sometimes different reliability tests with which they explain the phenomenon under study.

Finally, a series of spatial analyzes of the variables is carried out, the relationship between them for the speeds in the fleet of the SITP routes, with the help of the tools of geographic visualization, cartography and contextualization of the results obtained in the geographical frame.

KEY WORDS: weighted geographical regression (GWR), Geostatistics, Geographical databases, geographic processing

INTRODUCCIÓN

El SITP (Sistema Integrado de Transporte Público de Bogotá) es considerado como un instrumento del Plan Maestro de Movilidad, para mejorar la calidad de vida de la población en la ciudad de Bogotá, la implementación del sistema se ha hecho de forma gradual dentro de la ciudad, uno de los principales propósitos es lograr una cobertura total del servicio en la ciudad a partir de nuevos esquemas de rutas dejando de lado los sistemas de transporte públicos convencionales.

Este sistema de transporte enfrenta problemáticas de efectividad que comprometen el éxito del mismo en Bogotá, ya que existen graves deficiencias en la culturización ciudadana complementado con necesidades insatisfechas por parte de los usuarios.

El adecuado manejo de la información de funcionamiento del SITP dentro de un contexto geográfico, puede contribuir al mejoramiento de las políticas de funcionamiento, ya que parte del éxito de este medio de transporte depende de cómo funciona y como se relaciona con el entorno por el que circula para cumplir las diferentes necesidades de los usuarios, entender e interpretar las velocidades que manejan los buses en la flota de la compañía puede ser el principio para reformular rutas programadas ,cronogramas de cumplimiento o zonas de cobertura.

La implementación de las nuevas tecnologías de la información geográfica permite la caracterización de zonas geográficas específicas para explicar diferentes fenómenos, este proyecto busca ofrecer lineamientos que faciliten el entendimiento de las velocidades la las Rutas SITP a partir de las relaciones espaciales que se generan entre variables explicativas que hacen parte del entono de funcionamiento del sistema, a partir de análisis geoestadísticos.

Se quiere con el desarrollo del presente proyecto mostrar la importancia y la influencia que ejerce sobre cualquier fenómeno la manipulación de variables geográficas y espaciales, además de proponer nuevos modelos explicativos que conjuguen técnicas de estadística espacial y de logística de funcionamiento, con el fin de ofrecer una herramienta practica a la hora de realizar estimaciones sobre las velocidades de SITP.

1. ÁREA PROBLEMÁTICA

La idea de los buses del Sistema Integrado de Transporte Público SITP surge bajo el mandato del alcalde Samuel Moreno en el año 2007, y la primera ruta entro en funcionamiento en el año 2012, como un medio para alcanzar las tendencias en movilidad de los sistemas de transporte masivos.

La falta de sensibilización con los usuarios de transporte público produjo traumatismos y desorientación que no permitieron acoger y dar uso con plenitud del sistema, la implementación del SITP, que conllevó a discontinuar las rutas de transporte público tradicionales, afectando la cobertura mínima de transporte en algunas zonas de Bogotá, y convirtiendo así el SITP como el medio de transporte predominante en la ciudad por lo que es importante el análisis del funcionamiento de la flota de buses.

Actuales estudios y proyecciones dan idea de las muchas falencias con las que cuenta el sistema y las fallas que se pronostican a largo plazo y que afectaran indiscutiblemente la movilidad en la ciudad, ya que al no tener una gran acogida el SITP crea inconformidades en los ciudadanos que se manifiestan en mayor ingreso de automóviles y motos en las calles de la ciudad. Uno de los principales problemas del sistema es la eficiencia de los buses donde la mayoría de los usuarios afirman que las rutas no pasan con la frecuencia esperada por los paraderos, por lo que es imposible ajustarse a los tiempos de ruta o hacer de este medio un sistema predecible, por lo cual el actual sistema de transporte integrado en la ciudad de Bogotá no puede acoger las nuevas tecnologías de transporte inteligente de las grandes ciudades a nivel mundial.

El otro problema que surgió con el pasar del tiempo y el aumento de buses del SITP, fue el incremento en los accidentes de tránsito producto de factores como errores humanos, vetustez de la Flota, excesos de velocidades para cumplir cronogramas de rutas, mal estado de la infraestructura vial de la ciudad, rutas mal diseñadas.

Al ser entonces el SITP el sistema de transporte de mayor cobertura es preocupante la eficiencia de la Flota, el incumplimiento en tiempos de ruta y velocidades promedio de los buses del sistema, factores causales de las problemáticas anteriormente nombradas. La estructuración del proyecto en un principio esperaba una velocidad de 18 kilómetros por hora en la ruta, siendo hoy de 12 kilómetros por horas y ubicándose entre las más bajas de las velocidades de lo diferentes medios de transporte que funcionan en la ciudad.

Para el desarrollo del presente trabajo se analizarán una muestra de velocidades de rutas del SITP de la ciudad de Bogotá, a partir del modelamiento de variables espaciales y su forma de relacionarse dentro de un espacio geográfico, buscando

evaluar la influencia que tiene sobre las velocidades de la flota del SITP , se espera explicar la velocidad de las rutas del SITP en función de variables como características de la infraestructura vial, índices de accidentalidad y variables socioeconómicas entre otras ,con el propósito de generar una herramienta que permita mejorar las política de funcionamiento del sistema.

2. OBJETIVOS

2.1 OBJETIVO GENERAL

Describir el comportamiento de las velocidades de las rutas SITP, a partir de regresiones geográficas ponderadas utilizando variables Socioeconómicas y viales para la localidad de Engativá en Bogotá.

2.2 OBJETIVOS ESPECÍFICOS

- Determinar, evaluar y establecer las variables que intervienen en el comportamiento de las velocidades de las rutas SITP.
- Caracterizar las calzadas viales de la Localidad de Engativá con variables de infraestructura vial y socioeconómicas que se relacionen entre si y sean determinantes en el cálculo de las velocidades promedio de las Rutas SITP.
- Desarrollar regresiones geográficas ponderadas que permita explicar y analizar el comportamiento de las velocidades SITP a partir de comportamientos estadísticos determinados por variables de infraestructura vial y socioeconómicas.
- Evaluar a partir de diferentes análisis espaciales y productos cartográficos el comportamiento de cada variable espacial estudiada y como se relación espacialmente según el modelo de regresión propuesto.

3. JUSTIFICACIÓN

Resolver la problemática de movilidad en la ciudad de Bogotá, es una necesidad de primera mano ya que repercute en muchas instancias de la calidad de vida de la población, los usuarios de los estratos más bajos presentan los tiempos de viaje más altos; a su vez, los estratos bajos destinan cerca del 14% de su ingreso a movilizarse, mientras que en los estratos más altos apenas el 5%”(Secretaria Distrital de planeación, 2012).

La correcta evaluación y determinación de las problemáticas del SITP, permitirá contribuir al entendimiento integral del sistema y poder plantear así estrategias para mejorarlo, contribuyendo a la disminución de viajes en vehículos particulares y motocicletas, lo que tendría un impacto progresivo y positivo en el mejoramiento de la movilidad de la ciudad de Bogotá, atendiendo así las tendencias de crecimiento y densificación urbana de la ciudad que requieren de una dinámica de movilidad fluida.

Acoger el SITP como el transporte público por preferencia en Bogotá necesita de atender a la actuales problemáticas del sistema. Los SIG son una herramienta para plantear posibles soluciones o alternativas de cambio sobre el sistema, ya que ofrece la visión desde un marco espacial a una problemática social, es por eso que cada aporte que se realice para el mejoramiento del sistema desde la conciencia ciudadana que permita un mejor uso y aprovechamiento hasta el diseño e implementación de técnicas de análisis espacial son de gran valor.

La planificación del transporte público en la ciudad de Bogotá, hace parte de las políticas de desarrollo urbano ya que tienen un impacto sobre la movilidad desde una perspectiva, socioeconómica y espacial teniendo en cuenta las diferentes variables del entorno que condiciona la velocidad y flujo vehicular de las rutas SITP, por lo cual la correcta planificación del transporte público SITP repercute positivamente en la toma de decisiones en políticas de desarrollo.

4. MARCO TEÓRICO

4.1 IMPORTANCIA DE LOS ESTUDIOS DE VELOCIDAD

Las técnicas y políticas de mejoramiento del transporte público en la ciudad de Bogotá pueden verse orientadas por estudios realizados en grandes ciudades del mundo, se han realizado varias comparaciones de los diferentes niveles de servicios públicos mediante recolección de información obtenida de terreno para la obtención de indicadores heterogéneos como lo muestran (Muñoz, et al., 2013) en el marco del XVI Congreso de Ingeniería de Transporte en Santiago de Chile donde se pretendió caracterizar las fuentes de variabilidad en los diferentes viajes en transporte público en determinadas zonas geográficas en la ciudad de México, Bogotá, Lima y Santiago.

Al identificar las etapas en las que se realizan los viajes donde se estiman las velocidades promedio de operación en función de: los tiempos de espera, intervalos entre buses, tiempos de bajadas, tiempos de subida, tiempos promedio de caminatas y tiempo de transbordo. La evaluación arrojó que ciudades como Bogotá, Lima y México al ser ciudades con poca regulación del sistema de transporte público, impulsa al sistema a operar con grandes flotas de buses pequeños para aumentar la frecuencia y disminuir los tiempos de espera.

Se encontró también que el servicio de transporte de las ciudades de Bogotá y Santiago son muy similares en cuanto a su demanda diaria de viajes y las velocidades promedio de los mismos aunque Santiago cuenta con la operación de metro y Bogotá tiene un sistema de bus de tránsito rápido (BTR).

Analizando solo los sistemas integrados de transporte público en nuestro país, son relativamente recientes, respondiendo a la necesidad de las grandes ciudades en las cuales la población día a día crece y se requieren soluciones para mover grandes volúmenes de pasajeros en el menor número de vehículos posibles, con el fin de reducir los impactos en la movilidad; la mayor parte de los diagnósticos y evaluaciones han estado orientados a los Buses de Tránsito Rápido (BTR por sus siglas en inglés) debido a que "los primeros pasos hacia un transporte público organizado se dieron en Bogotá. Aunque siempre se pensó que el metro era la opción más efectiva para dar fin a la congestión, la experiencia que ha tenido Bogotá con) ha mostrado que existen opciones de menor costo que también son viables" (Yepes, et al., 2013).

Sin embargo con el pasar del tiempo y para reducir algunos impactos negativos de estos sistemas se inicio hablar de los sistemas integrados de transporte, en los cuales no solo se tienen los conocidos transmilenios, megabuses, Mio, etc. sino

que se brinda la conectividad con las rutas y buses de transporte tradicional organizados y operados en el caso de Bogotá por 13 empresas organizadas. Con esto se busca reducir la congestión accidentalidad y cobertura en zonas marginales.

En la mayoría de estudios realizados con el fin de evaluar los SITP, el punto clave se encuentra en el balance entre el mayor numero de buses y frecuencias para garantizar el bienestar de los usuarios, pero buscando que los costos operativos sean positivos. El otro aspecto relevante son las reducciones en tiempos de viaje "que significan un mejor acceso a los trabajos, el estudio y el comercio, cayeron entre 5,5 y 51 minutos promedio por día por pasajero"(Yepes,et al.2013). Sin embargo en Bogotá este aspecto no es tan favorable debido a que los desplazamientos son bastante largos y la población se queja de la falta de rutas y buses. Ya que "es posible encontrar altos niveles de congestión en todo tipo de vías, incluso en zonas residenciales. De esta forma, sería posible plantear que parte del problema generalizado en la congestión de Bogotá es la poca tendencia a permitir la expansión de la ciudad" (Behrentz, Carrizosa y Acevedo, 2009, p. 205).

Los estudios también muestran que al no satisfacer los temas tratados anteriormente " los SITM colombianos no son sostenibles. Se encuentran atados a un círculo vicioso de desequilibrio financiero de los operadores, reducción de frecuencias del servicio y reducción de la demanda"(Yepes,et al.2013).

Existen diversos indicadores en los estudios con el evaluar el estado de los SITP, pero todos se centran en tres temas: Reducción de tiempos de viaje, accidentalidad y contaminación. La mayoría de estos desarrollados por el Estado, debido a que este es quien ha financiado en buena parte el desarrollo de estos SITP, en cuanto a la infraestructura y el mobiliario urbano necesario para satisfacer las necesidades del sistema. algunos de los indicadores son: "kilómetros recorridos por la flota, volumen de pasajeros transportados, ingresos y gastos obtenidos por la operación, tiempos de viaje, accidentalidad, contaminación producida, accesibilidad entre otros"(ZIMA,2014).

Como resultado de estos estudios se propusieron algunas recomendaciones con el fin de hacer eficiente el sistema, entre las cuales están los carriles preferenciales y la adquisición de flota vehicular nueva " La simple adaptación de infraestructura para buses mediante carriles exclusivos y compra de vehículos nuevos genera ahorros en operación y beneficios de tiempo de viaje"(Hidalgo,2005), sin embargo también se proponen algunos indicadores con el fin de evaluar la rentabilidad del sistema, "el Índice de Pasajeros por kilómetro (IPK) que mide la productividad operacional, la velocidad promedio de recorrido y la productividad del capital (promedio de pasajeros por bus por día).

Sin embargo no debemos dejar de lado los problemas de movilidad que existen en Bogotá, está el alto crecimiento de las tasas de motorización, la constante expansión geográfica de la ciudad, lo que lleva a altos niveles de congestión por la intensificación del uso del automóvil y la mala infraestructura vial , por lo que se considera pertinente cuestionar el diseño de las rutas del actual SITP y de la matriz de origen-destino usada para establecer el cubrimiento de las rutas SITP y colocación de los puntos de conexión entre dichas rutas.

La evaluación, análisis y detección en problemática de asignación de rutas según (Galindo, et al.2015), parte de la recopilación de datos en campo como: tiempo de recorrido y tiempos de paradas a través de los métodos de Point Check y Ride Check , métodos que consisten en calcular los tiempos que demoran las rutas en hacer un recorrido previamente pre-diseñado y el tiempo en que tiene que pasar un bus de cada ruta en cada paradero, algunos de los resultado que arrojo esta metodología es que la frecuencia en las rutas es muy deficiente a pesar de que las distancia entre paraderos son cortas, los tiempos de espera entre cada servicio son aproximadamente 20 minutos y que existen paraderos con muy poca demanda de pasajeros los cuales deben ser reprogramados en otras rutas o evaluar su real necesidad.

En el estudio de (Yepes et al 2014) se proponen algunos indicadores o impactos a evaluar de forma directa e indirecta sobre los usuarios tal como se pueden ver en la Figura 1.

Figura 1. Variables influentes en un sistema de transporte masivo



Fuente: (Yepes et al 2014)

Los factores indirectos como el precio y uso del suelo están enfocados en el impacto positivo de las obras desarrolladas para garantizar el funcionamiento del sistema las cuales a su vez aumentan la percepción de seguridad. Además de la conectividad de los distintos barrios de la ciudad. El objetivo de las evaluaciones a los SITP también va enfocado a que "en el corto plazo las medidas deben dirigirse a revisar la sostenibilidad financiera, la optimización de rutas y la eficiencia técnica y del capital del sistema. En el mediano y largo plazo, deben evaluarse los impactos sobre la estructura urbana". (Yepes et al. 2013).

Por la necesidad latente de investigar e indagar sobre las deficiencias y posibles soluciones a las problemáticas del Sistema Integrado de Transporte Público, el uso de indicadores hace que las mediaciones se realicen bajo lineamientos oficiales, para esto el Ministerio de Transporte creo el Manual de Indicadores De Monitoreo Y Evaluación de Proyectos de Transporte Urbano en Colombia. Caben nombrar algunos como los indicadores de desempeño del transporte local que ofrecen una visión del comportamiento de los sistemas de transporte, Indicadores del ambiente urbano que informan sobre aspectos de movilidad , medio ambiente y salud publica impactados por el transporte público , por último se encuentran los indicadores de eficiencia y efectividad que suministran la información sobre eficiencia y efectividad en la operación factores que dependen estrechamente de variables como los insumos (cantidad de vehículos) , los productos (número de kilómetros recorridos en un día) y los resultados del transporte (usuarios transportados).

Dentro de las temáticas abordadas en la investigación se quiere dar especial importancia al Sistema Integrado de Transporte desde la cobertura del servicio y la accesibilidad que tiene los usuarios con discapacidades físicas. (Martínez,2012) dirigió su investigación a encontrar los aciertos y faltas del SITP enfocándose en lo que se ha tratado del tema, que se ha hecho y que está por hacer frente a esta población.

A través de estadísticas basadas en entrevistas a una muestra de población en Bogotá con Discapacidad Física y que son usuarios activos del sistema ,se encontró que un gran porcentaje requiere sillas de ruedas para su movilidad aspecto a tener en cuenta para calcular el espacio necesario en los buses y que son personas que deben rodar para acceder a los servicios de transportes aproximadamente 40 minutos, y que 31% de los viajes que realizan se debe a trabajo y 29% a terapias médicas , se identificó que el principal problema al momento de usar el transporte se da por la congestión y saturación de gente en los buses, seguido de la falta de adaptaciones a los vehículos del sistema de transporte público y en tercer lugar la intolerancia de otros usuarios.

En un estudio realizado por un grupo de la Escuela Colombiana de Ingeniería, se analizaron todas las variables descritas anteriormente de una forma más

detallada, en cuanto a los tiempos de viajes propusieron analizar los tiempos de recorrido entre paraderos, tiempos de subida y bajada de pasajeros y los tiempos de recorrido por pasajeros, todo esto buscando reducir los tiempos de desplazamientos.

El último punto importante estudiado es la cultura en el sistema, puesto que se evidencia que gran parte de los bogotanos desconocen cómo funciona el SITP, las rutas, paraderos y puntos de recargas de las tarjetas, razón por la cual los usuarios siguen usando el transporte tradicional. Con todo lo anterior se busca que el sistema pueda seguir adelante a pesar de los problemas económicos que lo aquejan y buscando la integración con los sistemas de transporte que vendrán a futuro como el Metro y los trenes de cercanías para conexión con los municipios de la sabana.

Por otro lado en la más reciente encuesta de percepción del SITP del 2016 realizada por la Cámara de Comercio de Bogotá se obtuvieron como resultado que la calificación del SITP por parte de los usuarios fue de 2.6 y los tiempos de viajes promedio aumentaron a 102 minutos. Esto a causa de la falta de buses y la demora de frecuencias

Una de las preguntas interesantes de este estudio es la razón por la cual utiliza el SITP, la mayoría respondió que es la única forma de transporte, dejando de lado otras opciones como la comodidad, seguridad, reducción de tiempo de viajes, entre otras. Por lo cual la inconformidad de los usuarios es notoria tanto así que el 68% de los usuarios piensan que el SITP empeora.

Se debe entender la falta de acogimiento del servicio SITP y por ello su mala imagen, como un problema cultural por la falta de conocimiento y difícil acceso a la compra de los pasajes, acompañado de la desinformación y desactualización sobre rutas nuevas o que entran a reemplazar las rutas tradicionales de buses a las que se encontraban acostumbrados los usuarios, desviando el sentido primordial del sistema que es mejorar los niveles en la calidad del transporte público como estrategia para desestimular el uso del vehículo particular e incrementar la calidad de vida de los habitantes de Bogotá y se está convirtiendo en un sistema que produce grandes pérdidas de dinero en la ciudad al no alcanzar las metas planeadas de cubrimiento y de implementación de rutas de transporte bajo estándares de calidad .

Por lo cual a manera de recomendaciones de los usuarios, ellos desean que las recargas de las tarjetas se pudieran realizar en línea, tener información más clara de las rutas y recorridos y tener la información en tiempo de real sobre la frecuencia de los recorridos, con esto saben cuánto tiempo esperar.

4.2 ANTECEDENTES

Realizando una revisión a estudios previos que se hayan realizado para determinar el comportamiento de las velocidades de los sistemas de transporte público con respecto a variables de su entorno, estos son muy escasos y valiosos. ya que un análisis de este tipo es de vital importancia para determinar tiempos de operación y buscar reducir al máximo las bajas velocidades a través de la implementación de estrategias que contribuyan a aumentar la velocidad de operación.

Una primera aproximación la encontramos en el trabajo de (Alessandrini,2015) el cual aplica algoritmos de aprendizaje estadísticos para la predicción de velocidades en el sistema de Transantiago, en su trabajo el compara varios métodos como lo son las redes neuronales, maquinas de soporte vectorial y redes bayesianas pero a diferencia de trabajos previos el autor busca hacerlo en tiempo real, sin embargo en su modelo el utiliza los reporte de velocidad del GPS de los vehículos para pronostica por cuadrantes cual sería la velocidad en un día posterior, como resultado de su investigación obtuvo que el mejor método fue el de redes neuronales con el cual se obtuvo un error del 20% en las predicciones, las cuales mejoran en las zonas u horas donde se generan congestiones.

Algo de resaltar del trabajo de (Alessandrini, 2015) y a lo que apunta el trabajo que desarrollamos es trabajar por franjas, es decir segmentar los datos en horas pico y valle ya que las condiciones varían y por tanto los resultados pueden ser mejores y la inclusión de variables externas como por ejemplo la accidentalidad.

Otro trabajo que aporta al desarrollo de nuestro proyecto es el de (Valencia,2008), la autora desarrollo un modelo de planificación para los corredores de transporte público, en este se deja constancia de algo que es vital para que cualquier sistema de transporte masivo funcione, el tema de la percepción de los usuarios acerca del servicio, y el tema clave aquí es la movilidad del sistema, es decir los usuario buscan que sus desplazamientos sean rápidos para poder estar a tiempo en sus actividades y no perder el tiempo en trancones.

En el estudio de valencia ,se aplican varios modelos como lo son las tradicionales redes optimas de transporte y estudios microeconómicos los cuales son valiosos para los inversionistas en estos sistemas, mas no para los usuarios, en su estudio también revela como la ubicación de los paraderos y semáforos causan la reducción en las velocidades de operación; esto es algo que es muy notable para el caso del SITP debido a que es muy común ver una fila de buses en un paradero, lo cual genera congestiones y retrasos en la operación.

Ya aterrizando a estudios de modelos predictivos en el país, podemos analizar el articulo de (Sabogal ,et al. 2015), en el cual plantean un modelo de regresión lineal

para estimar los tiempos de viaje, este artículo es interesante debido a que los autores lo desarrollan mediante la captura de información por medio de sensores los cuales miden la velocidad de los buses y el tiempo de recorrido.

Con estas dos variables formulan modelos de regresión lineal, uno aplicando los tiempos de viajes explicado mediante la distancia del recorrido, y el segundo explica tiempos de llegada en función de los tiempos de viajes y las velocidades. Sin embargo los resultados no son positivos debido a que los resultados hacen que no se cumplan algunos supuestos estadísticos, lo cual desencadena en que el modelo no realice predicciones con una certeza alta.

Sin embargo llegan a una conclusión significativa y es que aunque los modelos no sean buenos para la predicción, en un caso particular si fu bueno y se ajusto a la realidad , y fue en tramos donde no habían estaciones ni intercesiones , con lo cual se confirma nuevamente que este tipo de variables se deben tener en cuenta en los estudios posteriores a realizar.

Debido a que el tema de los paraderos y los semáforos han sido factores relevantes de acuerdo al análisis de estudios previos, decidimos revisar bibliografía acerca de este tema y su razón de ser en cuanto a su influencia en el tema de las velocidades, en un artículo de (Valencia et al.2008) ellas muestran como su ubicación y características de entorno hacen que los tiempo de viaje aumenten o disminuyan de acuerdo a su correcta configuración.

En el artículo se detallan tiempos de aproximación, tiempos de cargue y descargue de pasajeros y tiempo de salida, a su vez que relacionan la ubicación de estos con matrices de origen y destino, con el fin de minimizar el número de paradas, aunque estas deben estar equilibradas ya que para los usuarios tener mayor número de paraderos garantiza un mayor nivel de aceptación por el sistema, sin embargo esto genera un aumento en los tiempos de detención de los buses y con ello disminución en las velocidades de operación de las rutas.

4.2.1 Importancia de estudios de Velocidad

Como ya lo hemos visto a lo largo del presente estudio, el análisis de velocidad en los sistemas integrados de transporte es algo fundamental, puesto que las repercusiones tanto para las empresas como para los usuarios son muy notables, ya que influyen directamente en los ingresos de las empresas, debido a que al tener un vehículo más tiempo del normal en un recorrido por causa de trancones o mala planificación de paradas, genera por ejemplo un mayor consumo de combustibles, y la falta de disponibilidad del vehículo en un momento determinado lleva en muchos casos a tener que poner en operación más vehículos de los necesarios para cubrir la operación.

En cuanto a las repercusiones para los usuarios podemos ver como lo que buscan ellos es poder movilizarse de forma rápida entre sus sitios de trabajo, casas, etc.. Debido a que tiempos en trancones, y retrasos en la espera de los buses, genera que los usuarios busquen otras alternativas de transporte como el transporte informal o los medios alternativos como bicicletas, lo cual genera que el nivel de usuarios disminuya y con ello los ingresos del sistema.

Por tanto los estudios de velocidad son fundamentales como una solución a la reducción de tiempo de transporte en las distintas rutas, esto teniendo en cuenta la configuración del entorno de la operación, es decir para poder mantener un nivel de velocidad adecuado a parte de evitar las congestiones propias de cualquier sistema de movilidad, se requiere de la infraestructura adecuada, como lo veremos más adelante la correcta ubicación de paraderos la coordinación de la semaforización y el estado de la malla vial, hace que la velocidad pueda variar de forma significativa y con ello los tiempos de operación.

Por tanto y como lo dice (Mendez,2009) "la velocidad debe ser estudiada, regulada y controlada con el fin de que origine un perfecto equilibrio entre el usuario, el vehículo y la vía, de tal manera que se garantice la seguridad".

4.2.2 Definición de Velocidad

Este concepto va asociado a la percepción del movimiento de los objetos, con respecto un patrón de tiempo, a esto se agrega el término de dirección y se formula como tal el concepto de velocidad. Como bien se conoce la velocidad se expresa como una relación entre la distancia y el tiempo expresada mediante la siguiente fórmula:

$$V = \frac{\Delta d}{\Delta t}$$

Donde

V representa la velocidad

Δd Representa la distancia recorrida

Δt Representa el tiempo que tarda el recorrido

De acuerdo análisis que se requiera se pueden trabajar distintos tipos de velocidad, y dependiendo del tipo también se ejecuta un estudio particular para analizarlas; de forma general **tenemos los siguientes tipos de velocidades:**

4.2.3 Tipos de Velocidad

Velocidad de punto

Este tipo de velocidad es la más común, debido a que su cálculo se obtiene cuando un vehículo pasa por un punto determinado sobre el cual se realiza la medición, también es denominada velocidad instantánea ya que se calcula en un lugar y tiempo específico.

Velocidad media temporal.

Es calculada a partir de los datos de la velocidad de punto, es decir se toma un rango de tiempo determinado y se realiza el promedio de los datos de las velocidades de punto de un grupo de vehículos objeto de estudio.

Velocidad media espacial

Al igual que la media temporal, está dada por el promedio de un conjunto de datos de velocidades de punto, esta se diferencia de la media temporal en cuanto a que se toman las velocidades en un mismo instante para un tramo vial, es decir se toma la muestra de vehículos que se encuentren en un mismo instante sobre un lugar determinado.

Velocidad de recorrido

Este tipo de velocidad es la más trabajada en el análisis de recorridos de una flota, en esta se analiza la distancia recorrida entre un punto de origen y un punto de fin y el tiempo total que tarda este recorrido. En ella van incluidas los factores a tener en cuenta en el análisis operativo de tiempos de recorridos, como lo son los tiempos de detención en paraderos, congestiones viales, semáforos, velocidad aplicada al vehículo por el conductor, entre otros factores.

En la velocidad de recorrido también se aplican las anteriormente descritas es decir se puede calcular la velocidad media temporal de recorridos dividiendo el tiempo total de recorrido de n vehículos, entre las distancias recorridas y la media espacial del recorrido estimando las velocidades de punto para un grupo de buses en un determinado tramo vial.

Este tipo de estudio va en la búsqueda de comparar que tan bueno o malos son los tiempos de recorrido entre distintas rutas o incluso poder llegar a comparar los recorridos de una misma ruta a lo largo del día con el fin de determinar patrones y así poder hacer más eficiente la operación.

Velocidad de marcha

En ella se analiza el tiempo y la velocidad en los cuales el vehículo estuvo efectivamente en movimiento, es decir al tiempo total del recorrido entre un punto de origen y destino, se le descuentan cualquier tiempo de detención del vehículo como por ejemplo tiempo de detención en semáforos y paraderos. Este tipo de análisis es muy útil para diferenciar entre los tiempos efectivos del recorrido y tiempos detención.

Velocidad de proyecto

Este tipo de velocidad es calculada mediante datos de variables relativas al diseño vial, es decir no se determina como las anteriores a partir de los datos que proporciona el vehículo, sino de las características de su entorno, como por ejemplo el diseño y estado de la vía; debido a que las velocidades varían si la vía es principal o en un barrio, además del estado ya que si hay presencia de huecos se deberá de reducir con el fin de no dañar el vehículo.

También se incluyen en este estudio, datos de movilidad ya que por lo general se determinan zonas y horarios en los cuales se puede presentar congestión vehicular, lo ideal en este tipo de estudio es lograr que la velocidad llegue a ser constante, aunque para las grandes ciudades es complejo de llevar a la realidad esta meta, a menos que se utilicen ciertas estrategias como por ejemplo los carriles exclusivos.

4.2.4 Estudios de Velocidad

Ya descritos los principales tipos de velocidad que son relevantes para un estudio de velocidades en sistemas de transporte masivo, ahora nos enfocamos a explicar los estudios que se pueden llegar hacer con estos datos y sus aplicaciones.

Estudio de velocidad de punto

Este tipo de estudio es de los más utilizado, a causa de que por lo general se desea determinar la velocidad de un grupo determinado de vehículos en ciertas zonas con el fin determinar algunos patrones de tráfico o de entorno que hacen aumentar o disminuir la velocidad. Su aplicación tiene relación con:

- Tendencias de velocidad: análisis de velocidades por punto de muestreo para un periodo de tiempo determinado, buscando determinar patrones.
- Determinación de sitios con problemas de velocidades: al tener puntos de muestreo se pueden ubicar las zonas en las que se presentan problemas de velocidad, con el fin de aplicar las correspondientes estrategias para mejorar los niveles de fluidez vehicular.

- Planeación de la operación de tránsito: al tener datos puntuales por vehículo, se puede determinar cuáles van más rápidos y cuáles más lento y con esto realizar modelamientos que permitan reducir niveles de accidentes y poder fijar límites de velocidad por zonas.
- Estudios de antes y después: se puede aplicar estos estudios para analizar por ejemplo el impacto de cierres viales por obras, o el resultado de aplicar algún tipo de política de mejoramiento de tráfico.

Estudio de velocidad de recorrido

Con este tipo de estudios se busca determinar las causas de las demoras de tráfico, poder detectar su ubicación, hora en la que se registra y otras características y con esto poder mejorar el nivel de velocidad en estas zonas, o de no ser posible poder replantear las rutas con un recorrido alternativo. con este tipo de estudios se puede llegar a determinar:

- Eficiencia de rutas: Al poder determinar el tiempo real de operación y compararlo con el tiempo en el cual los vehículos se encuentran detenidos, con esto se pueden visualizar aquellas zonas en las cuales se presentan fallas de movilidad.
- Cálculo de costos usuario-vía: Con este tipo de análisis se puede estudiar los costos que afrontarían los usuarios de las vías, por un plan de mejoramiento.
- Demoras en intersecciones: Con la identificación de las zonas donde demora más el cruce vehicular, se puede llegar a determinar las causas asociadas, como por ejemplo los cruces existentes en la intersección o la falta de los mismos, tiempos de semaforización asociados, etc.; este tipo de estudios son apropiados para la mejora de elementos viales o para un diseño geométrico de las vías adecuados.
- Estudios y demoras del transporte público: este estudio es el más relevante al trabajo que desarrollaremos, a causa de que en ellos se mide la calidad del servicio prestado, buscando que el tiempo de recorrido origen destino sea el indicado reduciendo las demoras a través de la ingeniería de rutas.

4.2.5 Modelamiento estadístico

Con el pasar de los años y el desarrollo de teorías matemáticas y estadísticas se han venido elaborando modelos que ayudan a la toma de decisiones buscando analizar el comportamiento, determinando tendencias; y con esto generar predicciones de lo que podría pasar en un futuro de acuerdo a una serie de parámetros o condiciones preestablecidas. Para conocer en detalle el Modelamiento estadístico y geoestadístico definiremos algunos conceptos básicos.

Modelo

Se define como una representación de un elemento de la realidad, según (Vera, 2003) "los modelos pueden ser físicos (descritos por variables medibles), análogos (diagrama de flujo) y simbólicos (matemáticos, lingüísticos, esquemáticos). Los modelos matemáticos o cuantitativos son descritos por un conjunto de símbolos y relaciones lógico–matemáticas".

Estos deben estar soportados en leyes físicas o matemáticas que garanticen su veracidad y con ello se puedan realizar mediciones que garanticen un nivel probabilístico de certeza de la predicción.

Estimaciones y estimadores

Las estimaciones buscan usar los datos de una muestra en un modelo con el fin de explicarlo y predecir un valor, las estimaciones pueden ser de tipo puntual o por intervalos; la puntual hace referencia al reemplazo en un modelo de un valor puntual, mientras que en intervalo se aplica una fórmula determinada para un rango de números, esto debido a que un modelo puede llegar a ser explicado por distintas expresiones matemáticas.

Los estimadores son los datos que se pueden obtener a partir de datos muestrales, de acuerdo a (Vera, 2003) "Las propiedades más deseables de un estimador son: Que la distribución de muestreo este concentrada alrededor del valor del parámetro y que la varianza del estimador sea lo menor posible". Con lo anterior se busca dar una precisión al modelo y con ello una buena predicción.

Geoestadística

A partir de los años 70 se inicio a involucrar la estadística aplicada al componente geográfico, con el fin de poder predecir, simular y estimar el comportamiento en zonas geográficas en las cuales no hubieran datos muestrales, por lo general sus aplicaciones van orientadas a temas ambientales y de minería. Es una de las tres ramas de la estadística espacial, las otras dos son los lattices y los patrones espaciales.

Conceptualmente podemos decir que la geoestadística busca modelar comportamientos en un espacio continuo, esto quiere decir que si se carece de este fundamento, las predicciones no tendrían sentido alguno o simplemente se estarían dando perspectivas erradas de un fenómeno.

De acuerdo a (Giraldo,2002) la geoestadística opera básicamente en dos etapas. La primera es el análisis estructural, en la cual se describe la correlación entre puntos en el espacio. En la segunda fase se hace predicción en sitios de la región no muestreados por medio de la técnica kriging, esto de acuerdo al autor debido a

que existe otro método con el cual se puede trabajar llamado IDW(inverso de la distancia).

4.2.6 Econometría Espacial

La econometría surgió como una herramienta de la economía , mediante la cual matemáticos y estadísticos a través de modelos buscaban formular modelos econométricos con el fin de explicar el comportamiento de variables y su relación con otras, por ejemplo explicar el precio del suelo, en función de características del entorno como el estrato, el valor de las viviendas, etc.

A medida que paso el tiempo, y estos modelos fueron tomando mayor importancia y sus resultados eran de gran relevancia en diversos estudios económicos se empezó a analizar la manera de involucrar el componente espacial a estos modelos, es por esto que en 1956 Isard da uno de los primeros pasos acuñando el termino de economía espacial, con esto busco diferenciar la parte espacial, de los modelos tradicionales dando un nuevo enfoque y buscando distinguir los modelos econométricos espaciales de aquellos que no manejan este componente. Básicamente el componente espacial cobra gran importancia al poder analizar variables que se encuentran sobre un territorio, las cuales pueden llegar a tener un comportamiento determinado de acuerdo a su ubicación o proximidad entre variables.

Uno de los primeros términos en manejarse en este tipo de modelos es la autocorrelacion espacial, esto debido a las características propias de los datos geográficos, debido a que por lo general un determinado componente espacial suele ser parecido mientras más cercano se encuentre de otro, después de esto los dos fenómenos analizados en la econometría espacial son la dependencia espacial y la heterogeneidad de los datos.

Dependencia Espacial

En la econometría clásica, se plantea que los datos no pueden tener autocorrelacion ya que esto genera perturbaciones en los modelos lineales planteados, sin embargo en los modelos econométricos espaciales, este término tiene gran importancia de acuerdo a (Gujarati,1995) la autocorrelacion espacial muestra la falta de independencia entre los errores de distintas observaciones espaciales, en términos más claros esta dependencia se refleja como la relación que existe entre un punto X_i y otro punto X_j ; de acuerdo al fenómeno presentado la autocorrelacion puede ser positiva o negativa; positiva cuando un fenómeno o tendencia en un área determinada ,puede expandirse a las áreas vecinas y el comportamiento se mantiene constante, mientras que es negativa cuando el fenómeno se localiza en un área pero no se deja pasar a los vecinos.

Heterogeneidad de los datos

Los datos espaciales tienden a tener un comportamiento espacial distinto en cuanto a su ubicación, estos datos con comportamiento heterogéneo son con los cuales se busca explicar fenómenos, de estas variaciones se destacan dos fenómenos, según (pineda,2006).

- **La inestabilidad estructural:** está vinculada a la estructura espacial de las observaciones, es decir cuando no existe homogeneidad en éstas, se tiene inestabilidad espacial, que repercute en las formas funcionales y en la variación de los parámetros de regresión.
- **La heteroscedasticidad:** vinculada a los procesos espaciales, surge ante la omisión de variables o formas de especificación erróneas, lo cual provoca errores de medición.

Estos fenómenos pueden ser analizados con las pruebas estadísticas de la econometría básica como el test de White o la prueba de cambio estructural de Chow. para estimar estos modelos existen cuatro métodos

- ✓ Switching regressions.
- ✓ Variación de coeficientes aleatorios.
- ✓ Expansión espacial.
- ✓ Regresiones ponderadas geográficas.

Matriz de contigüidad

Se define la matriz de contigüidad W , como aquella que representa una región en el espacio a través de sus filas y columnas, en ella se expresan las relaciones existentes en el espacio, la matriz básica es representadas por 0 y 1 donde 1 es igual a contigüidad mientras que 0 representa lo contrario. Existen muchas formas de crear la matriz de contigüidad, sin embargo la más usadas es la matriz de contigüidad de torre descrita a continuación.

Tabla 1. Matriz de Contigüidad de Torre

	B	
B	A	B
	B	

Fuente: Acevedo et al,2008

En esta $W_{ij}=1$ para unidades que comparten un lado en común con la región de interés, así como esta existen varios modelos de acuerdo al análisis espacial que se desea desarrollar y la caracterización de la región de estudio. Es muy común

realizar transformaciones a la matriz de contigüidad , de forma que los elementos de cada fila sumen 1, esta matriz se denomina de primer orden estandarizada y se describe así:

$$C = W_z$$

Esto se hace buscando que al multiplicar esta matriz por un vector de observaciones de una variable y, la nueva matriz del producto

$$y^* = Cy$$

representa una nueva variable igual a la media de las observaciones de las regiones contiguas(Acevedo et al,2008).

El índice I de Moran

Es una de las distintas técnicas usadas para determinar si existe o no autocorrelacion espacial, este método es un ajuste matemático de la medición de autocorrelacion a elementos espaciales, se define así:

$$I = \frac{n}{\sum_{i=1}^n (y_i - \bar{y})^2} \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n \sum_{j=1}^n W_{ij}}$$

Donde

$\sum_{i=1}^n \sum_{j=1}^n W_{ij} (y_i - \bar{y})(y_j - \bar{y})$ es la covarianza para la unidad de estudio en la región i ,j, al realizar la multiplicación entre $(y_i - \bar{y})(y_j - \bar{y})$ se determina hasta que punto varían las observaciones.

W_{ij} son términos de la matriz de contigüidad definidos así **1** si i, j son adyacente , **0** en caso contrario.

$\sum_{i=1}^n \sum_{j=1}^n W_{ij}$ calcula en número de relación existentes en la región de estudio, y finalmente

$\frac{n}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{1}{V(y)}$ es una división por la varianza de los elementos.

Este índice busca garantizar la eficiencia de los estimadores a partir de la estimación de la correlación y los errores que esto pueda llegar a generar.

4.2.7 Modelos de regresión

Modelo de regresión lineal

Una vez identificadas las características de los datos, se busca linealizarlas en un modelo de regresión, para esto partimos del modelo de regresión lineal clásico.

$$Y = xB + e$$

Este modelo básico la variable dependiente Y, está en función de las variables independientes X, además de una perturbación e., estos modelos deben cumplir los siguientes supuestos.

$E\left(\frac{e}{X}\right) = 0$ el valor esperado de las perturbaciones debe ser 0

$$Var\left(\frac{e_i}{X}\right) = \sigma^2 \quad i=1, \dots, n$$

$$Cov\left(\frac{e_i}{X}, \frac{e_j}{X}\right) = 0 \quad i=1, \dots, n$$

las perturbaciones tienen una varianza uniforme y no están correlacionadas

Modelo de regresión espacial

con base en lo anterior se establecen los modelos de regresión espacial el cual se formula a continuación

$$\begin{aligned} Y &= xB + U \\ U &= \lambda WU + e \end{aligned}$$

Donde U es el termino que explica las relaciones espaciales, basados en la matriz de contigüidad W la cual explica los elementos a analizar en función de su contigüidad. Este método se establece cuando el modelo de regresión clásico no es capaz de explicar un fenómeno espacial.

Modelo de regresión geográfico ponderado

Un modelo de regresión tradicional se define así:

$$Y_i = \beta_1 + \beta_2 \beta_2 + \dots + \beta_k \beta_k$$

Este modelo como se explicaba anteriormente no considera las variaciones que se pueden dar con respecto al espacio, sin embargo para corregir esto y poder realizar un análisis espacial, se introducen las coordenadas de cada dato $u_i v_j$ con esto podemos reescribir la ecuación como:

$$Y_i = \beta_1(u_i v_j) + \beta_2 \beta_2(u_i v_j) + \dots + \beta_k \beta_k(u_i v_j)$$

Los modelos de regresión geográficos ponderados buscan analizar la no estacionalidad de los datos, por esta razón se define también la matriz de parámetros por medio de la cual se analiza la vecindad de cada punto.

$$\beta = \begin{bmatrix} \beta_1(u_1v_1) & \beta_2(u_1v_1) & \cdots & \beta_k(u_1v_1) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_1(u_nv_n) & \beta_2(u_nv_n) & \cdots & \beta_k(u_nv_n) \end{bmatrix}$$

4.2.8 Síntesis general

El modelamiento estadístico es una herramienta diseñada para realizar predicciones a partir de tendencias y estimadores; Existen ciencias como la geoestadística que aborda: primero, mediante el componente geográfico la predicción de comportamientos en espacio geográficos determinados a partir de la descripción de la correlación existente entre puntos en el espacio y segundo, predicciones en zonas no muestreadas por diferentes métodos de interpolación; La ciencia de la Econometría Espacial surgió como una ciencia alterna y consisten en la explicación de comportamientos sobre un territorios de una variable y su relación o dependencia de otras variables con un componente espacial como su ubicación o proximidad entre variables, la econometría espacial aborda así los temas de dependencia espacial y la heterogeneidad de los datos .

La dependencia espacial se explica a través de la autocorrelación espacial existente en una muestra y de la dependencia que existe en la relación de dos puntos, el comportamiento heterogéneo de los datos se da debido a su ubicación. Una de las técnicas usadas para determinar si existe o no autocorrelación espacial es el Índice de Moran que mide la variabilidad o covarianza en la región.

El modelo de Regresión espacial se emplea para explicar las relaciones espaciales a partir de la proximidad de los elementos y los modelos de regresión geográfica ponderados incluyen las coordenadas geográficas de los puntos muestrales tomados, se busca así explicar la no estacionalidad de los datos mediante diferentes análisis de vecindad.

5. METODOLOGÍA

5.1 TIPO DE TRABAJO Y ACTIVIDADES

Para el desarrollo del trabajo se plantean las siguientes actividades, las cuales hacen parte de una investigación aplicada en la cual se busca explicar a partir de variables geográficas, el comportamiento de las velocidades del SITP

- Recopilar la información geográfica de una muestra de rutas del SITP, para un periodo de 10 días, con el fin de procesarlas en ArcGis.
- Determinar cuáles son las problemáticas más comunes del SITP, en temas de velocidad, de acuerdo a la revisión y análisis de estudios previos realizados.
- Analizar y depurar la información objeto de estudio con el fin de garantizar la calidad de los datos.
- Estructurar y elaborar una base de datos espacial donde se aloje la información de las rutas del caso de estudio en cuanto a variables de velocidad de los vehículos, históricos de accidentalidad y variables socioeconómicas. Entre otras.
- Elaborar las respectivas regresiones geográficamente ponderadas, para cada día objeto de estudio, con el fin de establecer para que día y periodo de tiempo las variables explicativas nos dan un modelo ajustado.
- Generar la cartografía necesaria para determinar zonas representativas del sistema en cuanto a las variables estudiadas

5.2 DATOS REQUERIDOS

- Información geográfica actualizada sobre las velocidades de las rutas del SITP a estudiar
- Estudios estadísticos acerca de la accidentalidad en la que se han visto involucrados buses del SITP
- Información Socioeconómica de las zonas a estudiar, obtenida a partir de capas Predio catastral y manzana catastral.
- Información geográfica acerca del estado de la malla vial.
- Información geográfica de las condiciones o características de la infraestructura vial

Lo anterior con el fin de realizar el análisis de las condiciones actuales de la operación, mirar sus fallas y limitantes en cuanto a los niveles de velocidad de los buses y la relación de estas con las variables socioeconómicas y espaciales.

Las bases de datos en formato Excel con la información de las estadísticas y demás estudios se especializara en ArcGis y se trabajaran en el sistema World Geodetic System WGS84, además de coordenadas planas origen Bogotá con el fin de realizar los respectivos análisis espaciales en cuanto a tiempos y distancias, entre otros factores a estudiar.

5.3 SOFTWARE

- ✓ ArcGis 10.X,
- ✓ Excel
- ✓ Qgis

5.4 ETAPAS DE TRABAJO

Tabla 2. Etapas del trabajo

Etapa	Objetivo	Actividades	Resultados
Evaluación de los estudios previos realizados al SITP	Analizar los estudios previos en cuanto a las problemáticas que afectan al SITP	Elaboración del estado del arte	Estado del arte.
Análisis de la información no espacial insumo para el trabajo	Comprender y analizar la información base para el desarrollo del presente trabajo	Estudios de la operación del SITP.	Metodología actual de operación del SITP.
		Estudios de accidentalidad donde el SITP esté involucrado.	Base de datos con la información ordenada y depurada.

		Variables socioeconómicas relativas al SITP.	Base de datos con la información ordenada y depurada
		Información relativa al estado de la malla Vial y su tipología	Base de datos con la información ordenada y depurada
Análisis de información geográfica relativa al proyecto	Reconocer la información geográfica base del proyecto.	Revisión de los datos geográfico, realizando la respectiva depuración de la información.	Shapefile con datos de las rutas. Y las correspondientes variables socioeconómicas, estado de la malla vial, Etc.
Elaboración del estudio de acerca de las velocidades del SITP y su relación con variables del entorno	Desarrollar el estudio correspondiente con la información recolectada en los pasos anteriores.	Caracterizar las zonas donde se presentan las variaciones en la velocidad de los buses.	Shapefile con ubicación de los datos estadísticos.
		Estudio de la accidentalidad en el sistema.	Espacialización de los lugares donde se presentaron los accidentes.
		Consolidación de una base de datos con todas las variables necesarias, con las cuales se quiere llegar a explicar las velocidades del SITP.	ShapeFile con su base de datos con los atributos requeridos.
		Determinación mediante la utilización de mínimos cuadrados ordinario (OLS), de las variables	Variables explicativas para realizar las regresiones

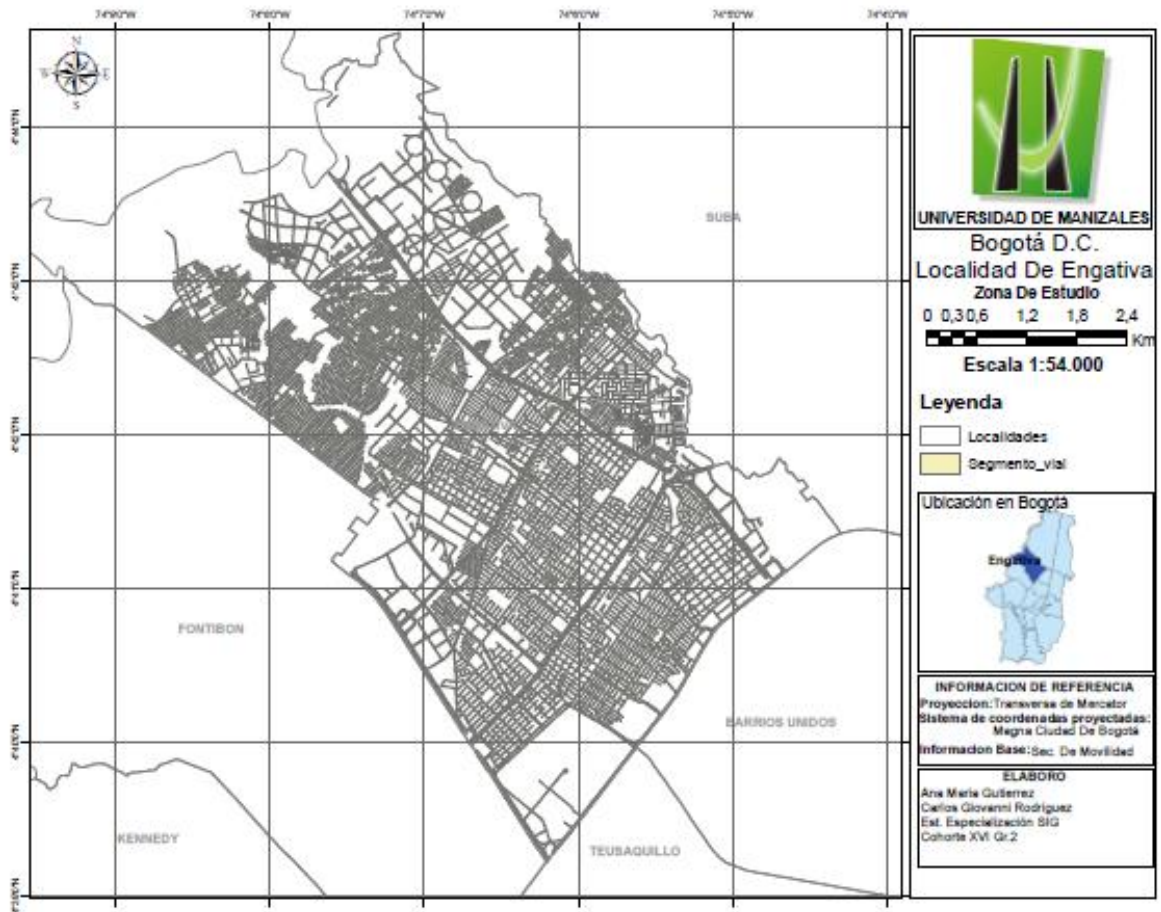
		explicativas requeridas para el cálculo de las regresiones, además de la aplicación de las pruebas estadísticas requeridas.	
		Calculo de regresiones geograficas ponderadas, a partir de las variables socioeconómicas, estado de la malla vial, Etc. determinadas con anterioridad	Shapefile y database con la información resultante de las regresiones geográficas ponderadas(GWR)
Resultados del estudio general realizado	Salidas gráficas y estructuradas de la información obtenida	Ubicación de la información espacial obtenida en el estudio	Cartografía con las normas técnicas correspondientes
		Organización y estructuración de la información espacial obtenida	Base de datos espacial con la información organizada

Fuente: Elaboración Propia

5.5 ZONA DE ESTUDIO

De acuerdo al análisis realizados, se eligió trabajar sobre la localidad de Engativá debido a que en esta encontramos los factores característicos propios del análisis a desarrollar, es decir las variables prediales (uso del suelo, densidad poblacional, estratificación), Además de acuerdo a (Cogua,2014) Engativá es la decima localidad más grande en superficie en Bogotá. Por lo cual es representativa en cuanto al estudio a desarrollar.

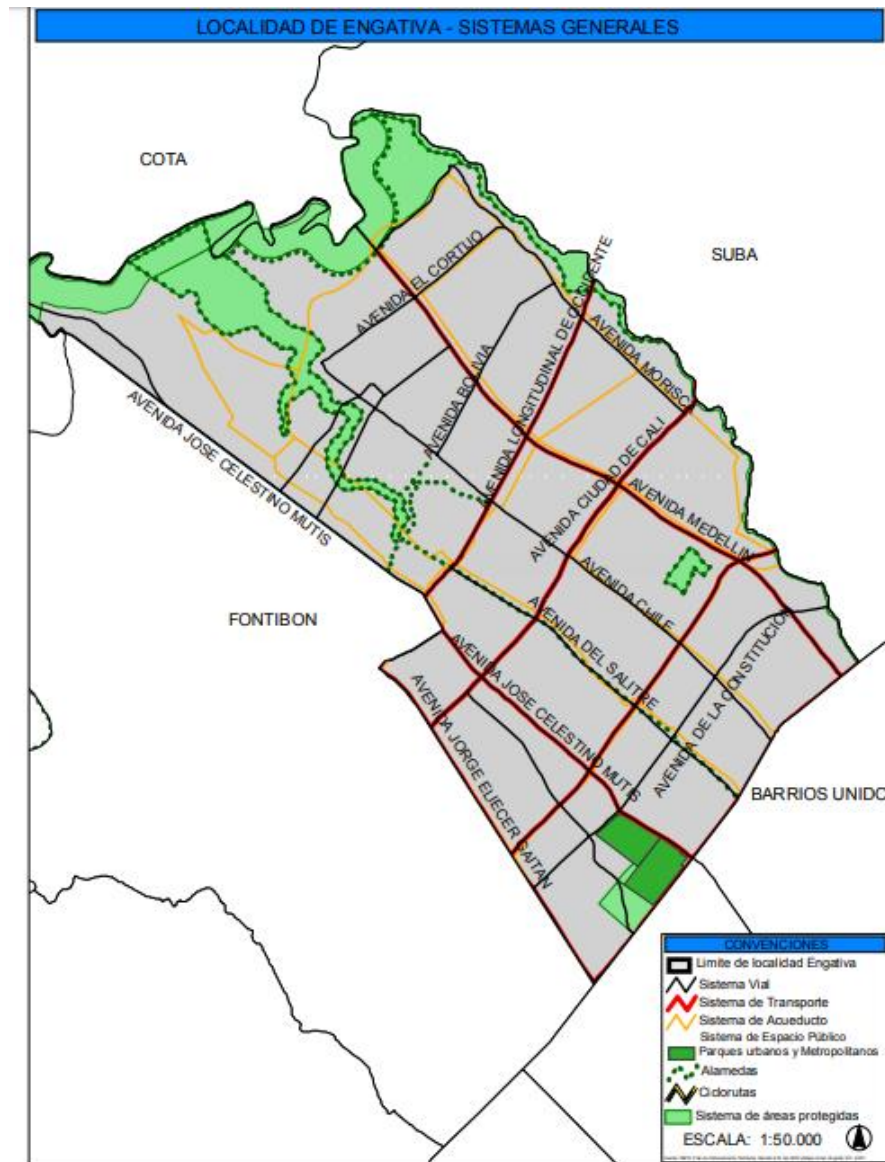
Figura 2. Zona de estudio



Fuente : Elaboración propia

Las variables viales (Estado malla vial, accidentalidad, huecos) también de gran importancia debido a que en la localidad se encuentran vías importantes que atraviesan toda la ciudad como la Av. Ciudad de Cali, Av. Boyacá y la Autopista Bogotá Medellín (CII80), las cuales se describen en la Figura 3

Figura 3. Vías principales localidad de Engativá



Fuente: <https://www.institutodeestudiosurbanos.info/endatos/0200/02-050-transporte/docs/ENGATIVA.pdf>

Además de lo anteriormente descrito, la importancia de esta localidad también radica en la cantidad de rutas del SITP que recorren la localidad, de las 13 zonas en las que se dividió la ciudad (figura 4) para la operación del SITP, Engativá ocupa el 5 Puesto con cantidad de rutas, con un total de 35 rutas.

Figura 4. Zonas de operación del SITP



Fuente: <http://www.sitp.gov.co/>

5.6 PREPROCESAMIENTO DE LA INFORMACION

Actividad 1

Recopilar la información geográfica de una muestra de rutas del SITP, para un periodo de 10 días. con el fin de procesarlas en ArcGis.

Capa de Segmento vial

Esta capa es fundamental debido a que en ella se va agrupar toda la información vial, debido a que en ella está el elemento más importante con el cual se pueden agrupar datos y son los CIV, estos son una serie de polígonos en los cuales se dividen todas las vías de la ciudad para llevar el inventario de las mismas, cada CIV está identificado con un número único de identificación al cual se le relaciona distinto tipo de información de entidades como la secretaria de movilidad el IDU, entre otras. La capa trae los siguientes atributos descritos en la tabla 3. trabajados bajo el sistema de coordenadas MAGNA_Ciudad_Bogota

Tabla 3. Puntaje por gravedad de accidentes

Nombre del Atributo	Descripción
PK_ID_CALZ	Código de identificación por CIV, para integración con otras bases de datos
CIV	Código de identificación vial
Superficie	Tipo de recubrimiento del segmento vial
ClasificaM	Clasificación vial del segmento
Km_Carril	Longitud del segmento vial

PCI_2017_1	Indicador del estado del pavimento
Estado2017	Estado del pavimento

Fuente: Elaboración propia

A esta capa no se le realizó ningún geoproceso inicial, sin embargo a esta se le adicionaron datos de accidentalidad y huecos, con el fin de poder pasar esta información a la capa de información predial.

Capa Accidente _vial

Esta capa contiene el registro georeferenciado de los accidentes ocurridos en la localidad de Engativá, de esta capa el atributo fundamental con el cual trabajamos, es la gravedad de accidente. La secretaria de movilidad categoriza la gravedad de accidentes en tres clases:

- **Solo daños:** Cuando se generan afectaciones a vehículos o infraestructura.
- **Con Heridos:** Cuando en el accidente se ve perjudicada alguna persona
- **Con Muerto:** Cuando en el accidente ocurre el deceso de alguna persona.

De acuerdo a los estudios que realiza la secretaria de movilidad para poder cuantificar las tres categorías anteriormente descritas, ellos le dan el siguiente puntaje de acuerdo a la gravedad, razón por la cual nosotros aplicamos este mismo puntaje. Tal como se muestra en la tabla 4

Tabla 4. Puntaje por gravedad de accidentes

Gravedad	Puntaje
Solo daños	1
Con heridos	2
Con muertos	13

Fuente: Secretaria de Movilidad

En ArcGis se agregó a la capa una columna llamada Puntaje en el cual se le calculó el puntaje a cada accidente, una vez con esta información se procedió a realizar un Join espacial entre esta capa y la de segmento vial, con el fin de crear una nueva capa llamada Accidentes, en la cual el puntaje de cada accidente estuviera amarrado al CIV donde se presentó.

Finalmente se realizó un summarize, con el fin de obtener la totalidad de la puntuación de la gravedad de los accidentes presentados en cada segmento vial, con esta información se realizó un join de la tabla resultante del Summarize "Suma_Accidentes" y la capa de "segmento vial" con el fin de que en esta quedara alojada la suma de la gravedad de accidentes presentada en cada CIV.

Capa Huecos_Engativa

En esta capa se encuentra almacenado el inventario de los huecos en la localidad, cada uno georeferenciado, por lo cual con esta capa se realizó un join espacial con la capa de segmento vial, con el fin de que los huecos quedaran asociados a un CIV. Una vez relacionados, se procedió a realizar un summarize con el fin de determinar la cantidad de huecos en cada CIV.

El resultado del summarize se alojó en la tabla sum_hueco y una vez con este resultado, se procedió a realizar un join con la capa de segmento vial con el fin de que en ella, quedara alojada la cantidad de huecos por CIV.

Actividad 2

Estandarizar, generalizar y depurar la información socioeconómica de acuerdo a los alcances y necesidades del proyecto

La información socioeconómica empleada para la descripción del entorno dentro del que se desenvuelve el funcionamiento del SITP se obtiene a partir de los predios catastrales, manzanas catastrales y de cómo se relacionan, las bases de datos empeladas fueron obtenidas de IDECA y DANE.

IDECA: La Infraestructura de Datos Espaciales para el Distrito Capital – IDECA, es el conjunto de datos, estándares, políticas, tecnologías y acuerdos institucionales, que, de forma integrada y sostenida, facilitan la producción, disponibilidad y acceso a la información geográfica del Distrito Capital, con el fin de apoyar su desarrollo social, económico y ambiental, la información suministrada es pública y de libre uso a los diferentes usuarios a través de un portal web de fácil manejo.

Se obtuvo información libre de IDECA en formato digital .shp, formato compatible con los software seleccionados para trabajar, corresponde a la siguiente información:

Tabla 5. Descripción de información geográfica IDECA

NOMBRE DE LA CAPA GEOGRÁFICA	DESCRIPCIÓN
Lote	Suministra información geográfica relacionada con los lotes catastrales, cuenta con atributos tales como el estrato socioeconómico, uso y categorías de los lotes y extensión entre otras.
Localidad	Suministra información geográfica relacionada con las localidades que hacen parte de la ciudad de Bogotá

Fuente: Elaboración propia

DANE: El Departamento Administrativo Nacional de Estadística (DANE), es la entidad responsable de la planeación, levantamiento, procesamiento, análisis y difusión de las estadísticas oficiales de Colombia, ofrece al país y al mundo más de 70 investigaciones de todos los sectores de la economía, industria, población, sector agropecuario y calidad de vida, entre otras.

La información obtenida del DANE corresponde a datos de nivel de densidad poblacional, y número de casas por unidades de manzanas catastrales, obtenidas de investigaciones como los registros del SISBEN, censo Poblacional , y viviendas registradas en catastro.

Tabla 6. Descripción de información geográfica DANE

NOMBRE DE LA CAPA GEOGRÁFICA	DESCRIPCIÓN
Manzana	Suministra información alfanumérica que relaciona la cantidad de personas por vivienda y la cantidad de unidades familiares o viviendas por manzana catastral, identificándolas con ID manzana generado por el Dane.

Fuente: Elaboración propia

Estandarización, Y estructuración Capa Lote

Según catastro Distrital el uso de una construcción corresponde a la destinación que se le da a los elementos materiales de la estructura urbana en las distintas actividades ciudadanas, corresponde a la actividad económica que se le está dando a la construcción en un predio al momento de su reconocimiento catastral.

La capa Lote obtenida del portal web de IDECA, cuenta con los usos y estratos socioeconómicos oficiales asignados por catastro Distrital, donde existen más de 60 usos formales y reconocidos por esta entidad en donde hay categorías y subcategorías para cada uso, para la presente investigación se generalizo los usos catastrales en 5 grandes grupos según sus característica

Figura 5. Clasificación Por Usos

USO CATASTRAL	RECLASIFICACION USO
HABITACIONAL (Menor o igual a tres pisos)	RESIDENCIAL
HABITACIONAL (Mayor o igual a cuatro pisos)	
HABITACIONAL (MENOR O IGUAL A 3 PISOS)	
ENRAMADAS COBERTIZOS CANEYES	

USO CATASTRAL	RECLASIFICACION USO
ACTIVIDAD ARTESANAL (antes Industria Artesanal)	INDUSTRIAL
INDUSTRIA MEDIANA	
ACTIVIDAD ARTESANAL (antes Industria Artesanal)	
INDUSTRIA GRANDE	

INSTITUCIONAL PUNTUAL	RECLASIFICACION USO
COLEGIOS Y UNIVERSIDADES DE 1 A 3 PISOS	DOTACIONAL
IGLESIAS	
COLEGIOS Y UNIVERSIDADES DE 4 O MAS PISOS	
CLINICAS, HOSPITALES, CENTROS MEDICOS GRANDES	
INSTALACIONES MILITARES	
COLISEOS	
COLEGIOS Y UNIVERSIDADES DE UNO A TRES PISOS	
CLINICAS, HOSPITALES, CENTROS MEDICOS GRANDES	
INSTITUCIONAL PUNTUAL	
CEMENTERIOS	
MUSEOS	

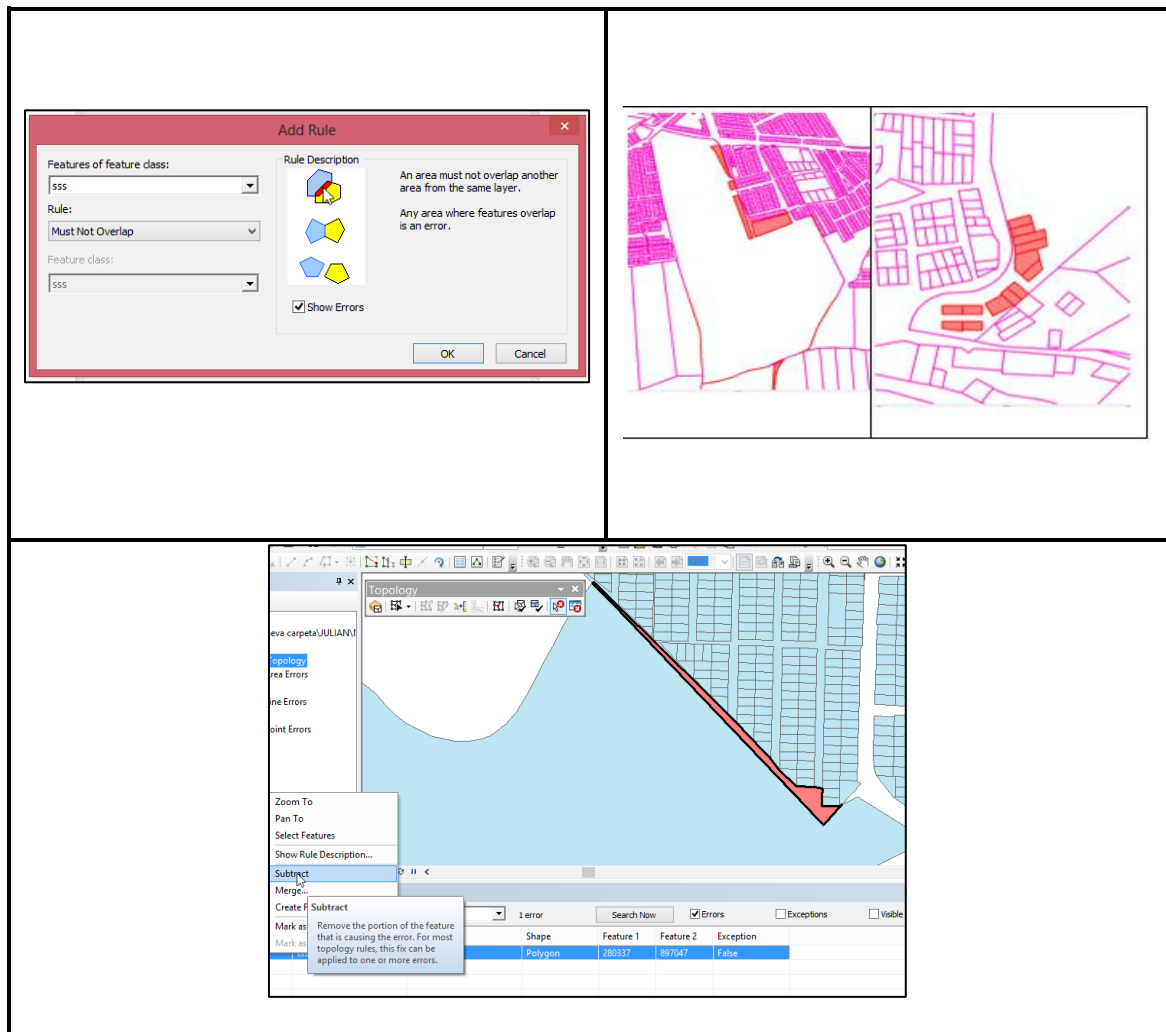
USO CATASTRAL	RECLASIFICACION USO
PARQUEO LIBRE	OTROS
URBANIZABLE NO URBANIZADO	
SECADEROS	
LOTES EN	

USO CATASTRAL	RECLASIFICACION USO
COMERCIO PUNTUAL	COMERCIAL
COMERCIO EN CORREDOR COMERCIAL	
OFICINAS OPERATIVAS	
CENTRO COMERCIAL GRANDE	
COMERCIO PUNTUAL (Local comercial)	
RESTAURANTES	
GRANDES ALMACENES (antes Bodega Comercial)	
HOTELES	
DEPOSITO DE ALMACENAMIENTO	
TEATROS Y CINEMAS	
EDIFICIOS DE PARQUEO	
BODEGAS DE ALMACENAMIENTO	
MOTELAS, AMOBLADOS, RESIDENCIAS	
CLUBES DE MAYOR EXTENSIÓN	
PISCINAS	
BODEGA ECONOMICA	
PLAZAS DE MERCADO	
GALPONES GALLINEROS	
ESTABLOS PESEBRERAS CABALLERIZAS	
COCHERAS MARRANERAS PORQUERIZAS	
KIOSCOS	
SILOS	

Fuente: Elaboración propia, basado en Documento técnico de usos de la construcción y destinos económicos de los predios (Unidad Administrativa Especial de Catastro Distrital,2013)

Se realizaron edición, sobre la capa Lote, ya que presentaba debilidades de integridad, de consistencia espacial y estructura en los polígonos, los cuales son corregidos al validar las reglas topológicas a consideración, en búsqueda de garantizar proximidad, conectividad, y adyacencia entre los lotes representados por polígonos.

Cuadro 1. Topología para la capa Lote



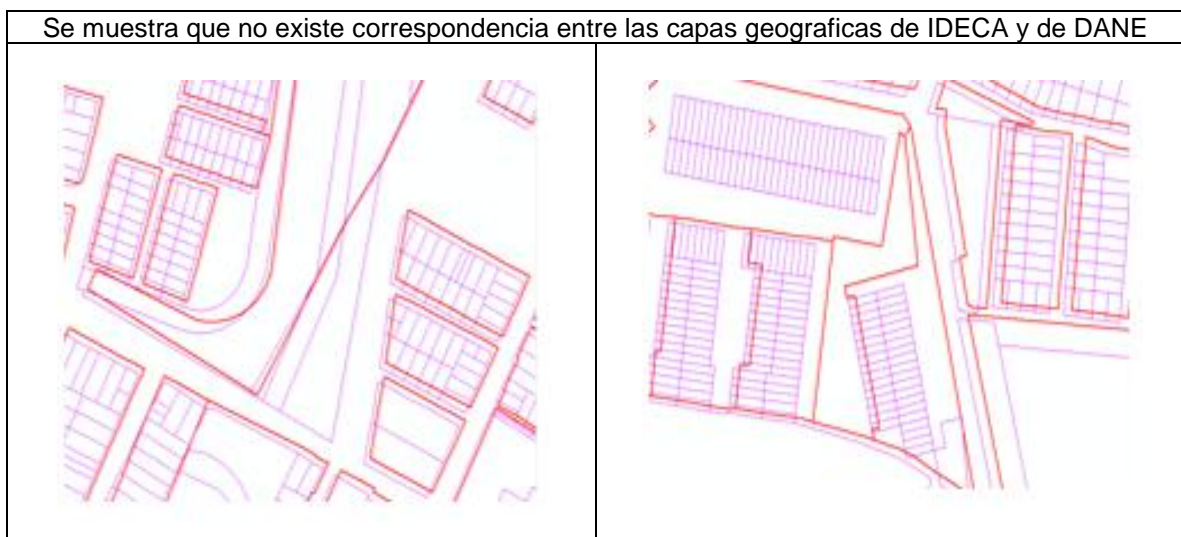
Fuente: Elaboración propia en ArcGis, 2018

Edición y generación de capa manzana a partir de la capa Lote

La información geográfica de los lotes de la base de datos de IDECA y manzanas de la base de datos del DANE no coincide espacialmente, existe desplazamiento entre las dos capas y problemas de temporalidad, aspecto que dificulta realizar análisis de tipo espacial entre ellas y compromete la exactitud de la información obtenida.

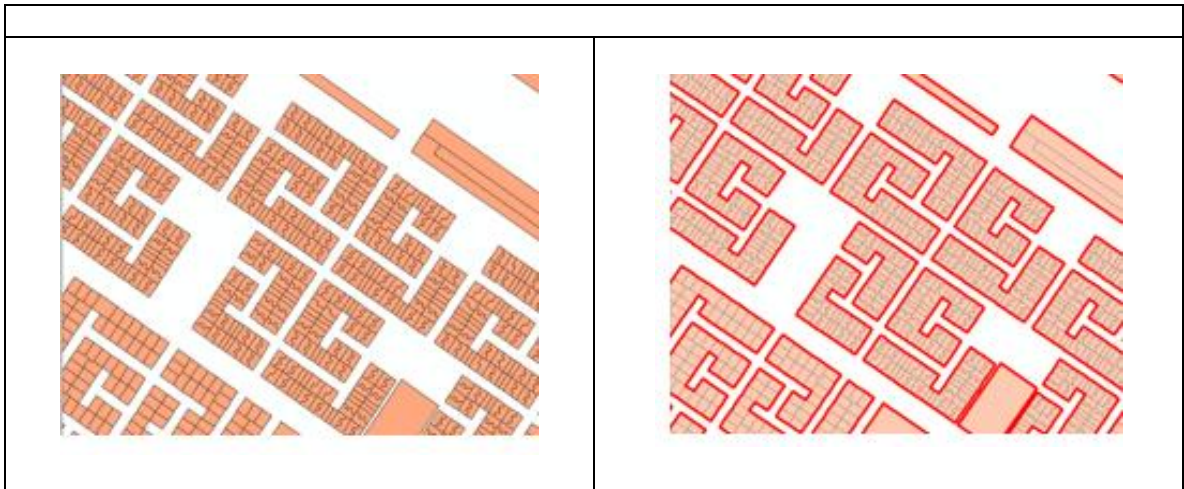
Se decide así generar la capa manzana a partir de la integración de la capa lote obtenida de IDECA, conservando la información atributiva de la capa manzana obtenida del DANE, utilizando la herramienta de geoprocésamiento Dissolve por el identificador único ManzCodigo, donde todos los predios que corresponden a una misma manzana cuentan con el mismo identificador.

Cuadro 2. Errores de correspondencia información DANE-IDECA



Fuente: Elaboración propia en ArcGis, 2018

Cuadro 3. Generacion de la capa manzana a partir de la capa Lote



Fuente: Elaboración propia en ArcGis, 2018

Actividad 3

Analizar y depurar la información objeto de estudio con el fin de garantizar la calidad de los datos.

Conformación de capa base para regresiones

Una vez depuradas las capas base de información(segmento Vial, predial, Accidentes y cantidad de huecos), se procedió a realizar la unificación de todas en un solo shape. Para esto la información se trajo mediante un join con el elemento clave de todas las capas que es el CIV así:

De la capa de segmento vial se tomaron los siguientes atributos:

- Estado de la malla vial
- Tipo de superficie
- Clasificación de la malla vial

De la capa de Accidente Vial se trajo el total del puntaje registrado en cada CIV.

De la capa Huecos_Engativa se trajo el total de huecos por cada CIV.

De la capa Lote_Engativa_Segmento se trajo:

- Uso(residencial, comercial, etc.)
- Personas_X_manzana
- Estrato.

Toda esta información se incluyo en el Feature_class CIV_completo, ya que con este, mas la información de velocidades se procederá al cálculo de las regresiones

Conformación de capa de velocidades

Para el desarrollo del trabajo se utilizó información de las velocidades del SITP en los CIV de la localidad de Engativá, para esto se tomó una muestra de dos semanas del mes de Julio de 2017 de lunes a viernes, considerando que en este periodo de tiempo las velocidades son más estables, esto se debe a varios factores que diferencian los fines de semana como por ejemplo, el día sábado no hay pico y placa y las actividades laborales y escolares se reducen por lo cual el volumen de gente que se moviliza en la ciudad es mayor, por lo cual aumenta el volumen de carros en las calles lo que genera reducción en niveles de velocidad, y los días domingo sucede todo lo contrario la gente se concentra en sus casas o salen de la ciudad lo que genera menor número de vehículos en las calles lo cual aumenta los niveles de velocidad.

Por tanto se trabajó con 10 días y se tomaron 3 horas en las cuales la velocidad en la ciudad se reduce, estas son la 7Am, hora en la cual la mayoría de los bogotanos realizan sus desplazamientos hacia el trabajo o estudio, las 12 Pm hora en la cual el pico y placa se levanta y la mayoría de gente saca sus carros para desplazarse a almorzar y las 7 Pm hora en la cual la gente retorna a sus casas.

Este filtro se hizo debido a que el volumen de información es considerable por varios factores como lo son: existen 3082 CVI en la localidad de Engativá en los cuales se registra algún dato de velocidad del SITP, y por cada hora hay varios reportes de velocidad en cada CIV lo que lleva a una base de datos robusta, después de aplicar el filtro obtuvimos una base con cerca de 1.040.000 filas de datos.

Ya teniendo una base de datos manejable, se empezó a trabajar por medio de una tabla dinámica en Excel con el fin de tener los datos necesarios que requerimos para desarrollar el modelo geoestadístico, tal como se observa en la figura 6.

Figura 6. Detalle base datos de Velocidades

CVI	14			15		
	7:00:00 a. m.	12:00:00 p. m.	5:00:00 p. m.	7:00:00 a. m.	12:00:00 p. m.	5:00:00 p. m.
5007254	16,11	12,34	11,70	15,79	11,06	13,36
9000930	11,40	30,17		12,89	15,86	12,69
9000931	18,02	32,60	11,55	22,35	31,51	22,08
9001036	16,87	24,85	14,06	26,68	29,95	24,40
9001379	20,47	26,17	18,37	25,23	32,83	26,28
9001736	12,89	15,33	12,05	16,76	17,27	15,54
9001904	15,75	16,89	14,52	22,55	22,37	22,64

Fuente: Elaboración propia

En la tabla básicamente tenemos un cuadro resumen con la siguiente información: la lista de todos los CIV de la localidad de Engativá, por cada día de la muestra una columna por cada hora a trabajar, de cada una de estas horas tomamos el promedio con los datos de velocidad que tuviera cada CIV

6. RESULTADOS

6.1 PROCESAMIENTO CAPA PREDIAL

El diseño del proyecto se hizo para explicar las velocidades de las rutas SITP en la localidad de Engativá, el estudio se realiza exclusivamente sobre los segmentos viales por las que transita las diferentes rutas del SITP, para lo cual se realiza la selección de los mismos y se genera la nueva capa de segmentos viales a estudiar.

Cuadro 4. Muestra de los segmentos viales



Fuente: Elaboracion propia en ArcGis, 2018

El proposito de la investigacion es modelar el comportamiento de las velocidades de las rutas del SITP en funcion de las diferentes variables que se relacionan con el funcionamiento de los buses de suervicio.

Fue necesario la selección exclusivamente de los Lotes que se relacionan con las segmentos viales depurados y seleccionados, ya que no es conveniente incluir en la investigacion características de lotes que se encuentren distanciados de los

segmentos viales seleccionados ya que no tienen relación directa no ejercer influencia directa y no se consideran explicativos en el modelo geoestadístico.

Ya que los segmentos viales no se relacionan de forma espacial con los Lotes, es necesario generar zonas de influencia respecto a los segmentos viales, lo cual se realiza con la herramienta Buffer, permitiendo seleccionar exclusivamente los lotes con una ubicación aproximada a los segmentos en evaluación.

Cuadro 5. Muestra de los Buffer y su influencia

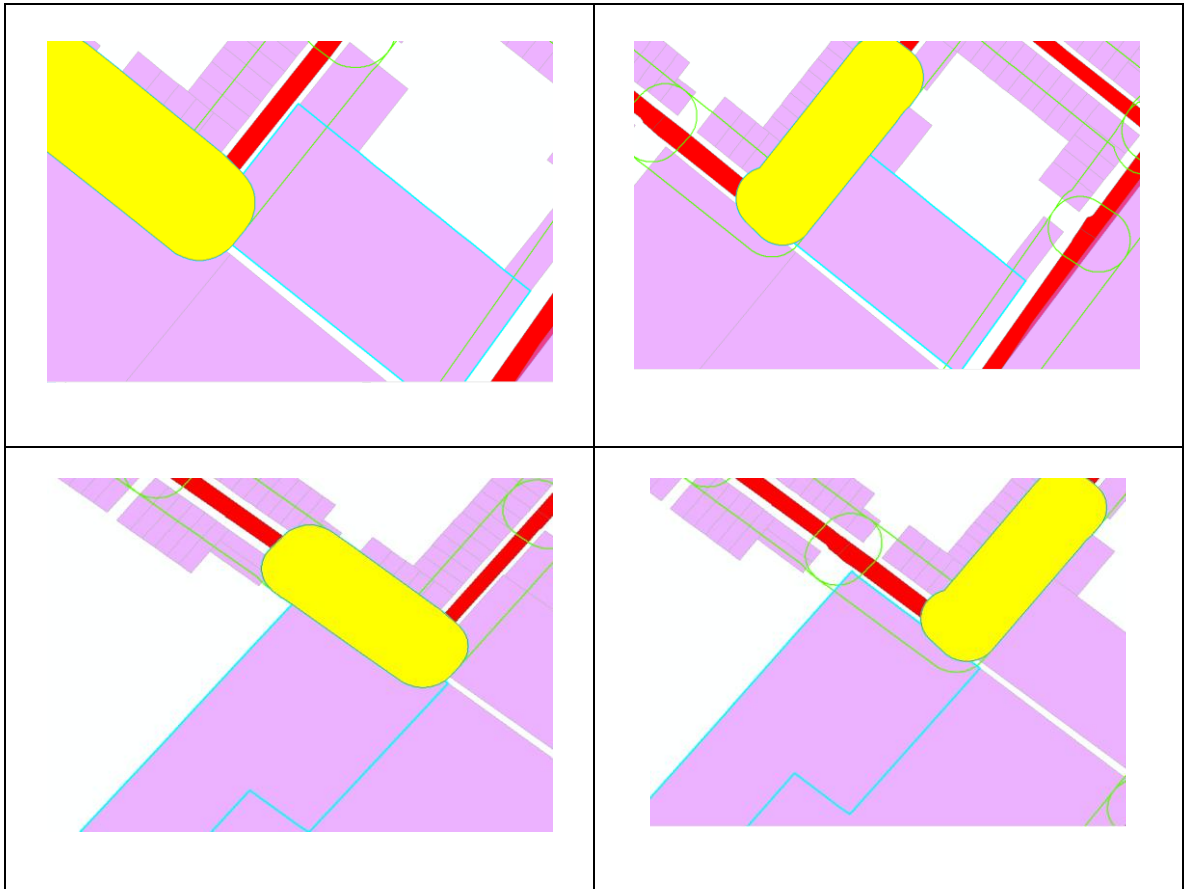


Fuente: Elaboración propia en ArcGis, 2018

La capa resultante del geo proceso de generación de la zona de influencia a partir del Buffer, hereda toda la información atributiva de la capa segmento vial (tipo de superficie, clasificación, estado, puntaje de accidentalidad, cantidad de huecos).

Ya que es necesario que cada elemento de la capa lote herede la información correspondiente al segmento vial que ejerza influencia sobre él, en donde cada lote se puede ver influenciado exclusivamente por un segmento vial, en muchos casos un lote tiene influencia de más de un buffer por lo cual se debe realizar proceso de edición apoyados en geoprocésamientos para corregir esta inconsistencia.

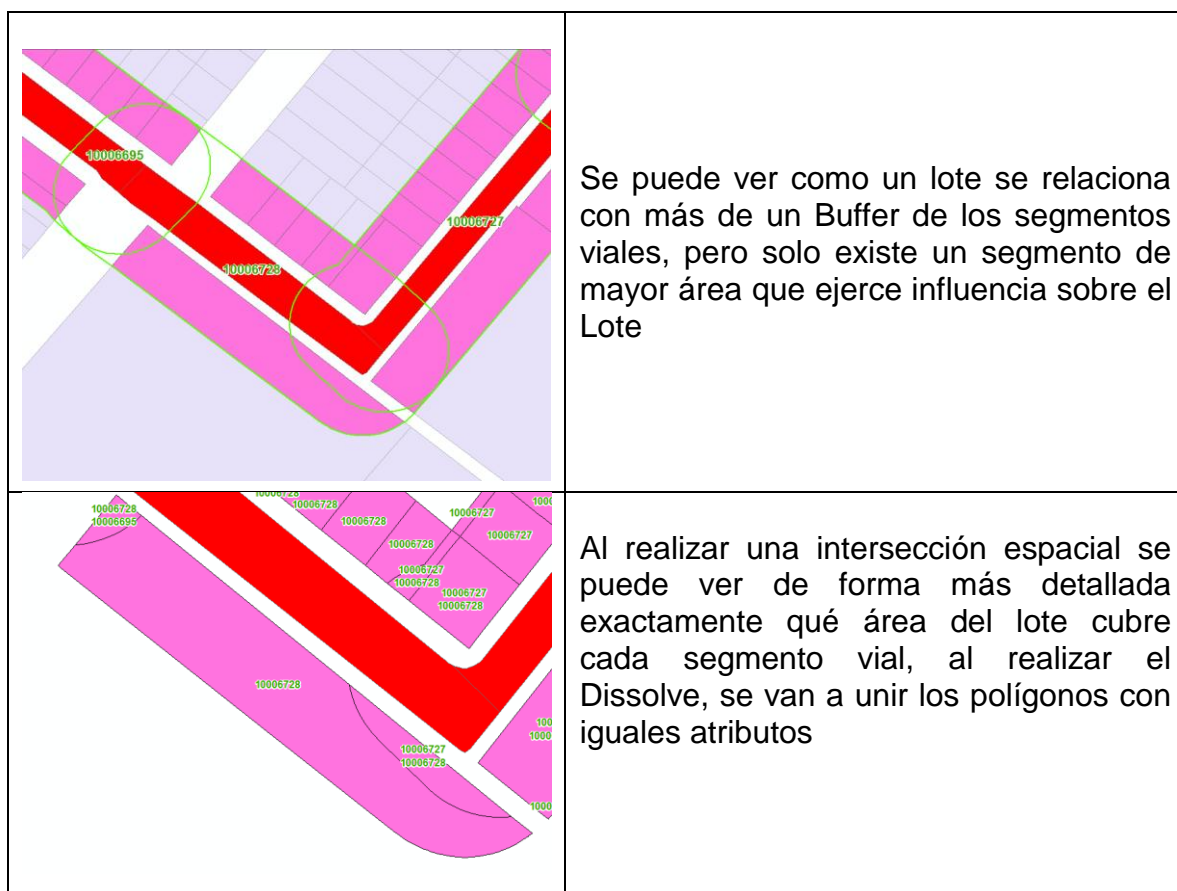
Cuadro 6. Influencia del segmento vial



Fuente: Elaboracion propia en ArcGis, 2018

Se procedió a realizar un Intersect entre la capa lote y la capa Buffer obteniendo así un polígono nuevo por cada lote que intercepte con más de un segmento vial, paso siguiente se genero un Dissolve en donde se especificó que por cada ID lote se unieran los polígonos que cuentan con el mismo CIV (id único de la capa segmento vial)

Cuadro 7. Proceso de Intersect y Dissolve



Fuente: Elaboración propia en ArcGis, 2018

Posteriormente se trabaja sobre la capa producto del Dissolve y se seleccionan los polígonos de mayor área por cada ID lote garantizando así que solo exista un segmento vial relacionado a un lote y que este sea el de mayor área.

Después de establecer la relación uno a uno para cada predio y cada segmento vial en evaluación se hereda la información de los segmentos viales a cada Lote con el uso de la herramienta Join.

Conformación de la capa de regresiones

Con base en la capa creada *CVI_completo*, se procedió a reclasificar, aquellos atributos cualitativo, ya que para poder realizar los cálculos de regresiones se requiere que los atributos sean cuantitativos, estos se clasificaron de menor a mayor según el tipo. A continuación presentamos las clasificaciones realizadas:

- Tipo_Super(tipo de superficie)

Tabla 7. Reclasificación atributo Tipo_Super(tipo de superficie)

Atributo	Clasificador
Adoquín Arcilla	1
Adoquín Concreto	2
Afirmado	3
Flexible	4
Mixtos	5
Rígidos	6

Fuente: Elaboración propia

- Clasificación

Tabla 8. Reclasificación atributo Clasificación vial

Atributo	Clasificador
Local	1
Intermedia	2
Arterial	3
Troncal	4

Fuente: Elaboración propia

- Estado_Mal(Estado malla vial)

Tabla 9. Reclasificación atributo Estado_Mal(Estado malla vial)

Atributo	Clasificador
Sin PCI	1
Malo	2
Regular	3
Bueno	4

Fuente: Elaboración propia

- First_Recla(Usos predial)

Con este atributo se decidió tomar el total de predios por uso, ya que si los clasificamos como los atributos anteriores las regresiones pueden perder sentido, razón por la cual se crearon 5 nuevos atributos

- Comercial
- Industrial
- Dotacional
- Residencial
- Otros

A cada uno se le asigno la suma de predios con estos usos, los cuales están influenciados por cada CIV. Finalmente la capa resultante quedo conformada así:

Tabla 10. Atributos capa resultante

Atributo	Descripción
CIV	Código identificación vial(ID único)
COMERCIAL	Predios con uso comercial
DOTACIONAL	Predios con uso dotacional
INDUSTRIAL	Predios con uso industrial
OTROS	Predios con uso otros(detallados en la capa predial)
RESIDENCIAL	Predios con uso residencial
ESTRATO	Estrato
PERSONAS_X	Personas por manzana
TIPO_SUPER	Tipo de superficie de la vía
CLASIFIC	Clasificación de la vía
ESTADO_MAL	Estado de la malla vial
PUNT_ACCID	Puntaje de accidentes
CANT_HUECO	Cantidad de huecos

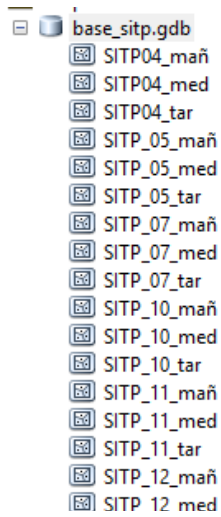
Fuente: Elaboración propia

Conformación de la Geodatabase con la capa de regresiones

Ya con la capa de las variables necesarias para poder realizar las regresiones geográficas ponderadas(GWR), se procedió a crear la Geodatabase para almacenar cada día objeto de estudio y cada periodo de tiempo(mañana, Medio día, tarde).

La Geodatabase se denominó *Base_sitp*, en ellas se creó un feature Class por cada día y periodo de tiempo denominados por ejemplo así *SITP04_Mañ*, para describir el día 04 en la mañana., su estructura esta descrita en la figura 7

Figura 7. Geodatabase Base SITP



Fuente: Elaboración propia en ArcGis, 2018

Cada uno de estos Feature Class traen todas las variables descritas en la actividad anterior, con las cuales se busca explicar las velocidades en cada CIV, además mediante un join se les añadió la velocidad promedio para cada CIV. Ya con esto se procede a la determinación de las variables necesarias para explicar el modelo y al cálculo de las regresiones.

6.2 VALIDACIÓN DE SUPUESTOS MEDIANTE OLS

Con el fin de poder determinar las variables significativas en el modelo se utilizo primero la herramienta OLS(mínimos cuadrados ordinarios) de ArcGis para encontrar las variables explicativas y con estas poder correr las regresiones para cada día y periodo de tiempo objeto de estudio, con el fin de encontrar el día y la hora en la cual el modelo se ajusta mejor a las variables.

Tabla 11. Variables del modelo de regresión

Variable	Variable	Descripción
Dependiente	Velocidad	Velocidad del SITP
Independientes	COMERCIAL	Predios con uso comercial
	DOTACIONAL	Predios con uso dotacional
	INDUSTRIAL	Predios con uso industrial
	OTROS	Predios con uso otros(detallados en la capa predial)
	RESIDENCIAL	Predios con uso residencial
	ESTRATO	Estrato

	PERSONAS_X	Personas por manzana
	TIPO_SUPER	Tipo de superficie de la vía
	CLASIFIC	Clasificación de la vía
	ESTADO_MAL	Estado de la malla vial
	PUNT_ACCID	Puntaje de accidentes
	CANT_HUECO	Cantidad de huecos

Fuente: Elaboración propia

Al realizar unos primeros modelos con OLS se encontraron dos factores, los cuales causaban que el R2 fuera muy bajo (menor al 0,1) y/o que la mayoría de variables explicativas del modelos no fueran significativas, por lo cual no podíamos continuar a calcular las regresiones geográficas ponderadas.

Figura 8. Depuración de segmentos viales



Fuente: Elaboración propia

Con el fin de poder mejorar el modelo, se decidió reducir el tamaño de la muestra, para esto habían dos posibilidades: trabajar una parte de la localidad o trabajar con las vías principales; por lo cual optamos por la segunda, para esto eliminamos aquellos segmentos viales pequeño los cuales hacen referencia a vías de barrio, o segmentos que no estaban comunicados con otros, tal como lo muestra la Figura 8 se eliminaron los segmentos en rojo, y se modelo nuevamente con los segmentos en negro

Antes de ejecutar las regresiones geográficas ponderadas, es pertinente verificar la confiabilidad de las variables a través del análisis de 6 supuestos, la verificación hecha se realiza sobre los datos del miércoles 12 de Julio al medio día

Figura 9. Resultados OLS en ArcGis

Summary of OLS Results								
Variable	Coefficient	StdError	t-Statistic	Probability	Robust_SE	Robust_t	Robust_Pr	VIF [1]
Intercept	15,855423	1,432281	11,070051	0,000000*	1,269308	12,491389	0,000000*	-----
COMERCIAL	-0,310122	0,056023	-5,535604	0,000000*	0,054563	-5,683703	0,000000*	1,120399
RESIDENCIA	0,036747	0,026995	1,361264	0,173691	0,023072	1,592724	0,111492	1,086967
PERSONA_X	0,002396	0,000772	3,103604	0,001969*	0,000865	2,769141	0,005706*	1,014312
TIPO_SUPER	-0,258246	0,282884	-0,912906	0,361456	0,240517	-1,073714	0,283153	1,055405
CLASIFICAC	1,267030	0,320326	3,955444	0,000089*	0,337909	3,749615	0,000197*	1,214121
ESTADO_MAL	0,462827	0,184993	2,501861	0,012472*	0,185280	2,497986	0,012608*	1,022625
PUNT_ACCID	0,012175	0,005261	2,314006	0,020816*	0,005155	2,361795	0,018326*	1,212277

OLS Diagnostics			
Number of Observations:	1240	Number of Variables:	8
Degrees of Freedom:	1232	Akaike's Information Criterion (AIC) [2]:	8255,378011
Multiple R-Squared [2]:	0,053934	Adjusted R-Squared [2]:	0,048558
Joint F-Statistic [3]:	10,033468	Prob(>F), (7,1232) degrees of freedom:	0,000000*
Joint Wald Statistic [4]:	69,683050	Prob(>chi-squared), (7) degrees of freedom:	0,000000*
Koenker (BP) Statistic [5]:	19,988213	Prob(>chi-squared), (7) degrees of freedom:	0,005595*
Jarque-Bera Statistic [6]:	90,945955	Prob(>chi-squared), (2) degrees of freedom:	0,000000*

Fuente: Elaboración propia en ArcGis, 2018

1. Supuesto: los coeficientes deben tener el signo esperado.

Se puede inferir que la velocidad a evaluar de las rutas SITP, es proporcional y aumenta cuando existe presencia de lotes con uso residencial, el análisis de coeficientes indica que cuando el estado de la malla vial es bueno también hay un aumento proporcional con la velocidad, otras variables menos significativas que tiene una relación proporcional positivan a la velocidad es el número de personas por manzana, y la presencia de puntajes por accidentes, según la interpretación de coeficientes.

Variable	Coefficient
Intercept	15,855423
COMERCIAL	-0,310122
RESIDENCIA	0,036747
PERSONA_X	0,002396
TIPO_SUPER	-0,258246
CLASIFICAC	1,267030
ESTADO_MAL	0,462827
PUNT_ACCID	0,012175

El análisis de los coeficientes obtenidos permite inferir a partir de los coeficientes con signo negativo que la velocidad explicada disminuye cuando existen lotes con uso comercial y que la velocidad es sensible al tipo de superficie Flexible donde la malla vial tiende a

flexionarse con el peso.

2. Supuesto: no debe haber redundancia entre las variables explicativas.

VIF [1]
1,120399
1,086967
1,014312
1,055405
1,214121
1,022625
1,212277

la revisión y análisis del valor VIF (factor de inflación de varianza) se encuentran entre el rango de 1,01 a 1,21 siendo valores bajos lo que indica que no existen variables que estén correlacionadas, y permite confiar en las predicciones que se realicen a partir del mismo.

3. Supuesto: Los coeficientes deben ser estadísticamente significativos.

Koenker (BP) Statistic [5]: 19,988213 Prob(>chi-squared), (7) degrees of freedom: 0,005595*

Probability	Robust_SE	Robust_t	Robust_Pr
0,000000*	1,269308	12,491389	0,000000*
0,000000*	0,054563	-5,683703	0,000000*
0,173691	0,023072	1,592724	0,111492
0,001969*	0,000865	2,769141	0,005706*
0,361456	0,240517	-1,073714	0,283153
0,000089*	0,337909	3,749615	0,000197*
0,012472*	0,185280	2,497986	0,012608*
0,020816*	0,005155	2,361795	0,018326*

Se puede ver que el coeficiente de Koenker (BP) es significativo estadísticamente por lo cual se hace caso exclusivamente a la probabilidad robusta que permite inferir que todos los coeficientes son significativos, el coeficiente de uso residencial en los lotes de 0,111492 y según el supuesto no se considera significativo pero por la

naturaleza de la investigación se considera incluirlo ya que se evalúa la velocidad de rutas SITP de usuarios que principalmente se desplazan de zonas residenciales a otras zonas, el coeficiente de tipo de superficie también se considera para la investigación.

4. Supuesto: Los residuos están distribuidos normalmente.

Este supuesto se verifica a través de la prueba Jarque-Bera, lo que permite detectar normalidad en los modelos de predicción.

Jarque-Bera Statistic [6]: 90,945955 Prob(>chi-squared), (2) degrees of freedom: 0,000000*

No cumple el supuesto de normalidad y existe la posibilidad del que modelo este sesgado, ya que el valor es muy cercano a 0 existe evidencia que los coeficientes de las diferentes variables de modelo se ajustan a una normal.

Es recomendable investigar nuevas variables de importancia que puedan ser de gran significancia para explicar y dar mayor confianza a las predicciones del modelo.

5. Supuesto: El valor R^2 fuertemente ajustado.

El valor R^2 indica cuanta variación existe en el modelo que explica la variable dependiente, el valor obtenido por el modelo corresponde a 0.5, tiende a 1 y es significativo, lo que permite inferir que el modelo es explicativo frente al fenómeno de las velocidades del SITP que se quiere explicar, aunque el R^2 es bajo al calcular las regresiones geográficamente ponderadas el ajuste del modelo será mayor

6. Supuesto: Los residuos no deben estar espacialmente autocorrelacionados.

Este supuesto se verifica a partir de la prueba de Moran, que se basa en las ubicaciones de las variables, evalúa si el patrón de las variables se encuentra agrupado, disperso o aleatorio.

Con la ayuda de la herramienta para determinar la autocorrelación espacial por medio del índice de Moran en ArcGIS, logramos determinar el grado que existe de correlación en las variables con las cuales queremos desarrollar las regresiones geográficas ponderadas.

Tabla 12. Cálculo del I de Moran

Variable	Índice De Moran	P Valué	Z-Score
Comercial	0,4514	0,0000	42,97
Residencial	0,4073	0,0000	38,83
Personas X Manzana	0,4373	0,0000	41,61
Tipo De Superficie	0,5836	0,0000	55,46
Clasificación Vial	0,7493	0,0000	71,10
Estado de la Vía	0,3796	0,0000	36,05
Puntaje Accidente	0,6616	0,0000	63,08

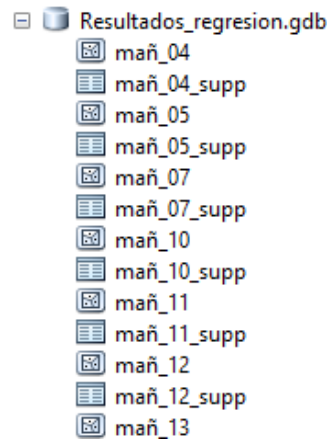
Fuente: Elaboración propia en ArcGis, 2018

El valor P es estadísticamente significativo y la puntuación z es positiva, indicando un patrón de dispersión espacial aleatorio se rechaza la hipótesis nula, la distribución espacial de los valores altos y los valores bajos se agrupan espacialmente.

6.3 CONFIGURACION DE PARAMETROS PARA LA REGRESION

Con las variables explicativas del modelo se procedió a calcular las respectivas regresiones para cada día, estos resultados se guardaron en una Geodatabase denominada resultados_regresion, en esta se agrupa cada resultado obtenido por día y hora.

Figura 10.Geodatabase Resultados_Regresion



Fuente: Elaboración propia en ArcGis, 2018

El cálculo de las regresiones se realizó empleando las siguientes características para la función GWR:

El Tipo de Kernel utilizado fue adaptativo, este se ajusta a las características propias de nuestro modelo, debido a que en las zonas densas utiliza un rango de vecinos pequeño, mientras que en las zonas con baja densidad él se extiende buscando un número mayor de vecinos.

El método de ancho de banda trabajado fue mediante por un número de vecinos especificados ya que los métodos de AICC y CV por su formulación matemática, nos mostraron como resultados modelos con un ajuste bajo. Mientras que al colocar el número de vecinos de forma manual los resultados son mejores.

Sin embargo antes de establecer el número de vecinos ideal para nuestro modelo, realizamos varias pruebas para analizar los resultados obtenidos.

Tabla 13. Cálculos de GWR con distinto número de vecinos

# de vecinos	R2	R2 ajustado	Velocidades para explicar	Velocidades explicadas	% de explicación
30	0,82	0,61	1240	534	43%
40	0,71	0,50		609	49%
50	0,65	0,45		697	56%

Fuente: Elaboración propia

Con los resultados obtenidos se optó por trabajar con un número de 50 vecinos, ya que con este garantizamos que el modelo explique más del 50% de los datos de velocidad, y también se garantiza una confiabilidad de por lo menos el 50% en las estimaciones.

Como resultados del cálculo de las regresiones geográficas ponderadas obtuvimos que las variables explicativas del modelo se ajustan mejor para los días miércoles en la hora pico del medio día

Tabla 14. Resultado de los R2 de las GWR

	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES
DIA	3	4	5	6	7
R2 Mañana	N/A	0,4357	0,4655	N/A	0,1986
R2 Medio Día	N/A	0,5360	0,5837	N/A	0,4578
R2 Tarde	N/A	0,5830	0,3553	N/A	0,4111
DIA	10	11	12	13	14
R2 Mañana	0,5548	0,5562	0,5100	0,2580	0,4503
R2 Medio Día	0,5666	0,3901	0,6546	0,5777	0,4856
R2 Tarde	0,4099	0,3252	0,4250	0,4939	0,3588
DIA	17				
R2 Mañana	0,4262				
R2 Medio Día	0,4471				
R2 Tarde	0,5326				

Fuente: Elaboración propia

6.4 RESULTADO DE LA REGRESION CON MAYOR AJUSTE

De acuerdo al análisis realizado los modelos de regresión con el mayor ajuste fueron los del día miércoles en el periodo de medio día, por lo cual a continuación analizaremos los resultados de las regresiones:

En el informe resumen de la GWR en ArcGis para el día 12 se muestra lo siguiente

Figura 11. Resumen regresión del día 12

OBJECTID *	VARNAME	VARIABLE	DEFINITION
1	Neighbors	50	
2	ResidualSquares	465,037647	
3	EffectiveNumber	19,815561	
4	Sigma	3,444984	
5	AICc	339,958137	
6	R2	0,766675	
7	R2Adjusted	0,654637	
8	Dependent Field	0	medio
9	Explanatory Field	1	COMERCIAL
10	Explanatory Field	2	RESIDENCIA
11	Explanatory Field	3	Persona_x
12	Explanatory Field	4	Tipo_Super
13	Explanatory Field	5	Clasificac
14	Explanatory Field	6	Estado_mal
15	Explanatory Field	7	Punt_accid

Fuente: Elaboración propia en ArcGis, 2018

1 Neighbors (Vecinos):[50], este valor se definió de acuerdo al análisis inicial con el fin de obtener un buen ajuste del modelo, básicamente son los vecinos con los cuales se realiza la estimación.

2 Residual Squares:[465,03] este valor es la suma de los residuales del modelo, para obtener un modelo con un buen ajuste este valor debe ser pequeño, es decir la variación entre el valor verdadero y el valor predicho por el modelo debe ser mínima, en nuestro modelo aunque el valor es alto es el mas bajo de todos los modelos calculados

3 Effective Number:[19,81] este valor es una relación entre la varianza de los valores ajustados y el ancho de banda, su análisis varia de acuerdo al ancho de banda, para el caso de nuestro modelo como el ancho es pequeño el numero efectivo tiende a ser cercano al número de vecinos es decir 50.

4 Sigma:[3,44] este valor es el cociente de la raíz cuadrada de los residuales, sobre los grados de libertad del modelo, al igual que los residuales, se busca que este valor tienda a 0 para tener un modelo ajustado, en nuestro modelo el valor es bajo por lo cual el ajuste es significativo.

5 AICc:[339] con este valor se analizar el rendimiento del modelo, al igual que los anteriores indicadores, se espera que el valor tienda a 0 y es un buen indicador para comparar el modelo GWR con el OLS, ya que el de menor AICc presenta el mejor ajuste, como en nuestro caso donde en OLS el valor era de 8299(Figura 9) frente a 339 del GWR.

6R2: [0,76] nos muestra que tanto se explica el modelo con las variables de predicción, se busca que el valor sea cercano a 1, es decir que se explique el

modelo en su totalidad, en nuestro modelo este valor es muy cercano a , por lo cual la explicacion del modelo es aceptable

7R2 Adjusted:[0,65] al igual que el R2 se busca que este valor sea cercano a 1 , por lo general el R2 ajustado es menor que el R2 debido a que normaliza las variables, es decir compensa el efecto del ingreso de nuevas variables en el modelo; al igual que el R2 este valor tiende a 1 y esta por encima del 0,5 es decir que se esta explicando gran parte del modelo.

Con respecto a las variables utilizadas para la regresión se usaron aquellas que fueron significativas en la prueba realizada mediante OLS las cuales aparecen en el siguiente grafico

Figura 12. Resumen estadísticos de coeficientes de regresion

Summary of OLS Results							
Variable	Coefficient	StdError	t-Statistic	Probability	Robust_SE	Robust_t	Robust_Pr VIF [1]
Intercept	15,855423	1,432281	11,070051	0,000000*	1,269308	12,491389	0,000000* -----
COMERCIAL	-0,310122	0,056023	-5,535604	0,000000*	0,054563	-5,683703	0,000000* 1,120399
RESIDENCIA	0,036747	0,026995	1,361264	0,173691	0,023072	1,592724	0,111492 1,086967
PERSONA_X_	0,002396	0,000772	3,103604	0,001969*	0,000865	2,769141	0,005706* 1,014312
TIPO_SUPER	-0,258246	0,282884	-0,912906	0,361456	0,240517	-1,073714	0,283153 1,055405
CLASIFICAC	1,267030	0,320326	3,955444	0,000089*	0,337909	3,749615	0,000197* 1,214121
ESTADO_MAL	0,462827	0,184993	2,501861	0,012472*	0,185280	2,497986	0,012608* 1,022625
PUNT_ACCID	0,012175	0,005261	2,314006	0,020816*	0,005155	2,361795	0,018326* 1,212277

Fuente: Elaboración propia en ArcGis, 2018

Como se puede ver los P values(Probability) de la mayoría de las variables tienden a 0, por lo cual de acuerdo a la estadística clásica se rechaza la hipótesis nula, es decir la variable es significativa ya que la variable predictora tiene una relacion importante en el valor a predecir.

Al ejecutar las regresiones mediante GWR se decidio trabajar con las 6 variables significativas, y ademas incorporar la variable residencial, la cual hace referencia al numero de viviendas residenciales; ya que en la localidad y de acuerdo al analisis realizado a la base de datos e Inumero de predios residenciales es alto y por esto debe tener alguna relacion con las velocidades del SITP en cuanto a volumen de pasajeros y paradas realizadas; Ademas se incluyo el tipo de superficie puesto que la velocidad del SITP, se debe ver influenciada por la superficie de rodura, es decir debe haber mayor velocidad en una via pavimentada,a comparacion de una via en adoquin .

Analizando los valores de los coeficientes resultante de la regresión mediante GWR tenemos que de 1935 CIV, el GWR predijo valores para 817, es decir el 42% de estos, como bien sabemos la regresion GWR es calculada una a una por cada valor a predecir, es decir que por cada velocidad hay una regresion que la explica,

y esta a su vez es diferente de las otras regresiones, razón por la cual de manera general analizaremos los resultados para aquellas con un R2 local del 0.5, 0.6, 0.7, 0.8 y 0.9

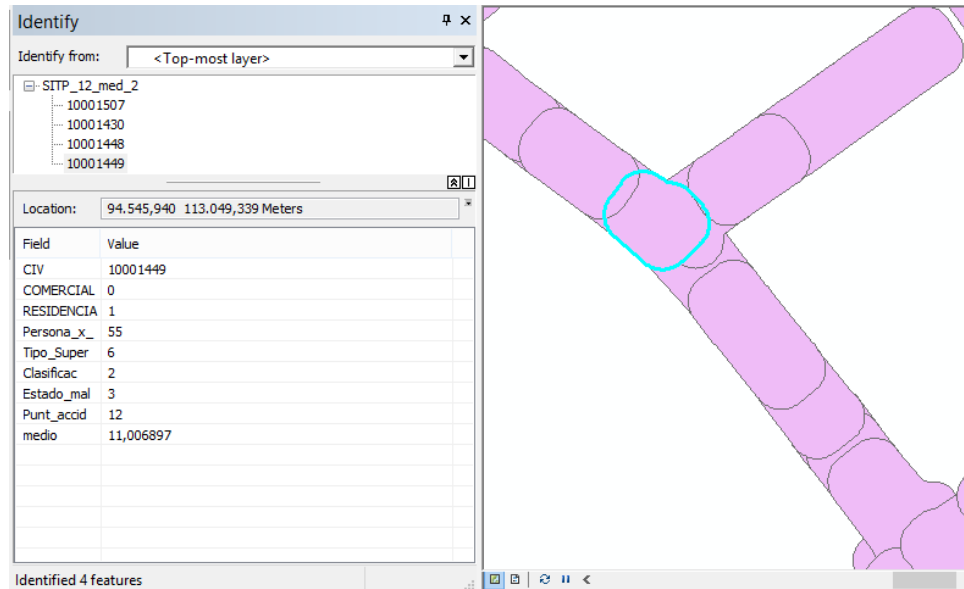
Tabla 15. Resumen de resultados de regresión GWR

StdError	Observed	LocalR2	Predicted	Intercept	C1_COMERCI	C2_RESIDEN	C3_Persona	C4_Tipo_Su	C5_Clasifi	C6_Estado_	C7_Punt_ac
2,97	11,01	0,50	9,48	12,75	-0,66	0,18	0,02	-0,88	-0,23	1,09	-0,18
3,01	25,13	0,60	29,51	104,15	0,60	-0,32	-0,02	-13,70	-6,74	0,32	-0,03
3,16	24,70	0,70	24,97	38,26	-0,80	0,53	-0,03	-5,43	9,29	-4,78	0,17
2,97	21,96	0,80	23,39	-15,64	-0,85	0,58	0,03	4,84	-2,84	3,38	0,03
3,03	27,57	0,92	26,44	-13,62	-0,59	-0,03	0,02	4,46	2,64	-0,37	0,03

Fuente: Elaboracion propia ArcGis, 2018

Como se puede apreciar en la tabla 15, no se puede realizar un análisis general de cada variable con la cual queremos explicar la velocidad, debido a que el comportamiento de los coeficientes es diferente para cada CIV a explicar, por lo cual a continuación presentamos un resumen de las características de cada CIV por R2 Local

Figura 13. Detalle del CIV con R2 local de 0,5



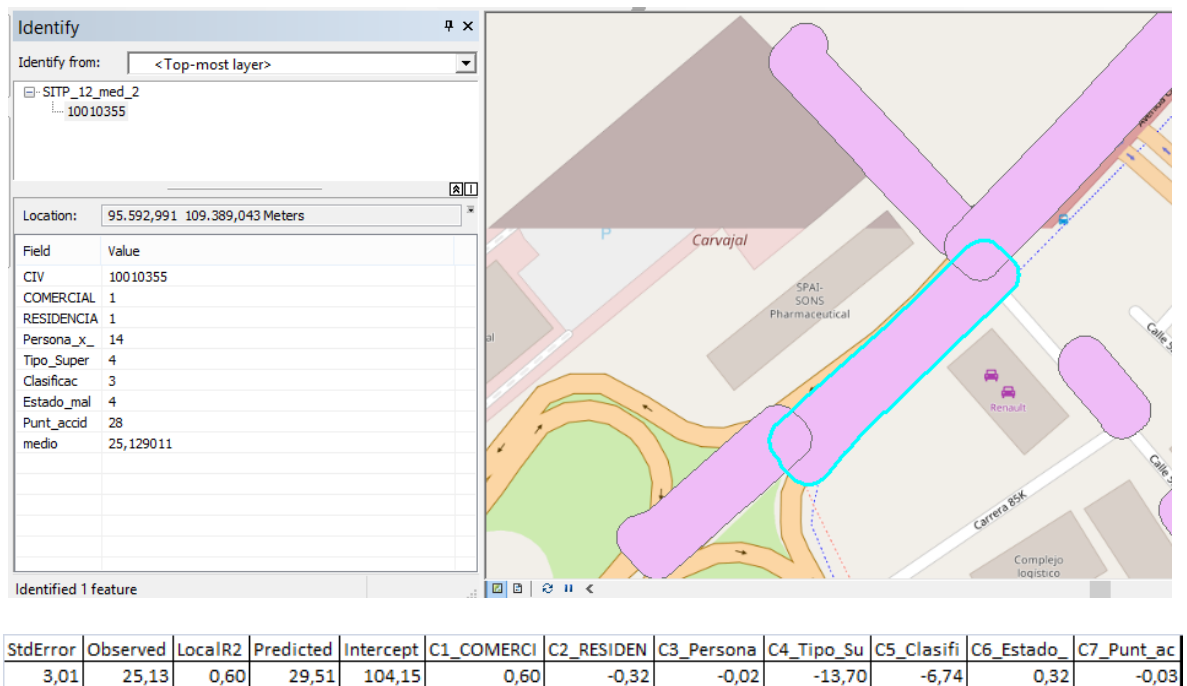
Observed	StdError	LocalR2	Predicted	C1_COMERCI	C2_RESIDEN	C3_Persona	C4_Tipo_Su	C5_Clasifi	C6_Estado_	C7_Punt_ac
11,01	2,97	0,50	9,48	-0,66	0,18	0,02	-0,88	-0,23	1,09	-0,18

Fuente: Elaboración propia ArcGis, 2018

De forma general tenemos que los predios comerciales no influyen ya que no hay, los predios residenciales influyen de forma positiva aunque muy poco debido a que solo hay 1 y la constante es de 0.18, la personas influyen de forma positiva al modelo pero de forma mínima ya que si bien hay 55 personas en este CIV la

constante es de 0,02 la menor de todas; el tipo de superficie 6 es rígido y hay una contradicción en el modelo, debido a que el signo del coeficiente debería ser positivo ya que este tipo de pavimento hace que la velocidad aumente; en cuanto a la clasificación el tipo 2 es intermedia, es decir una vía principal que conecta varios barrios por lo cual la movilidad puede ser densa y por esta razón el coeficiente es negativo, es decir la velocidad se reduce; en cuanto al estado el tipo 3 es regular lo cual no afecta en la velocidad y por tanto el coeficiente es positivo es decir no se reduce la velocidad en estas condiciones; finalmente el puntaje de accidente con 12, la constante es negativa, debido a que con esta cantidad de accidentes la velocidad es probable que se reduzca.

Figura 14. Detalle del CIV con R2 local de 0,6

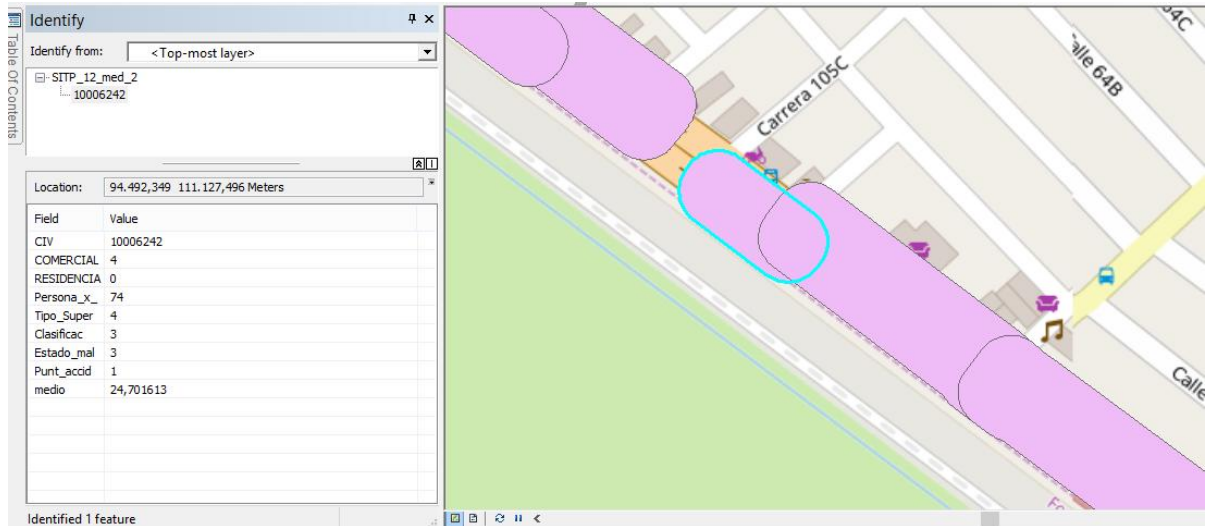


Fuente: Elaboración propia ArcGis, 2018

De forma general tenemos que los predios comerciales influyen poco ya que solo hay 1, los predios residenciales influyen de forma negativa aunque muy poco debido a que solo hay 1 y la constante es de -0.32, la personas influyen de forma negativa al modelo pero de forma mínima ya hay 14 personas en este CIV la constante es de -0.02 la menor de todas; el tipo de superficie 4 es flexible y sigue habiendo una contradicción en el modelo, debido a que el signo del coeficiente debería ser positivo ya que este tipo de pavimento hace que la velocidad aumente; en cuanto a la clasificación el tipo 3 es local, es decir una vía que conecta gran parte de la ciudad por lo cual la movilidad puede ser densa y por esta razón el coeficiente es negativo, es decir la velocidad se reduce; en cuanto al estado el tipo 3 es regular lo cual no afecta en la velocidad y por tanto el coeficiente es positivo

es decir no se reduce la velocidad en estas condiciones; finalmente el puntaje de accidente con 28, la constante es negativa, debido a que con esta cantidad de accidentes la velocidad es probable que se reduzca, sin embargo el coeficiente es bajo por lo cual no influye tanto en la regresión resultante.

Figura 15. Detalle del CIV con R2 local de 0,7

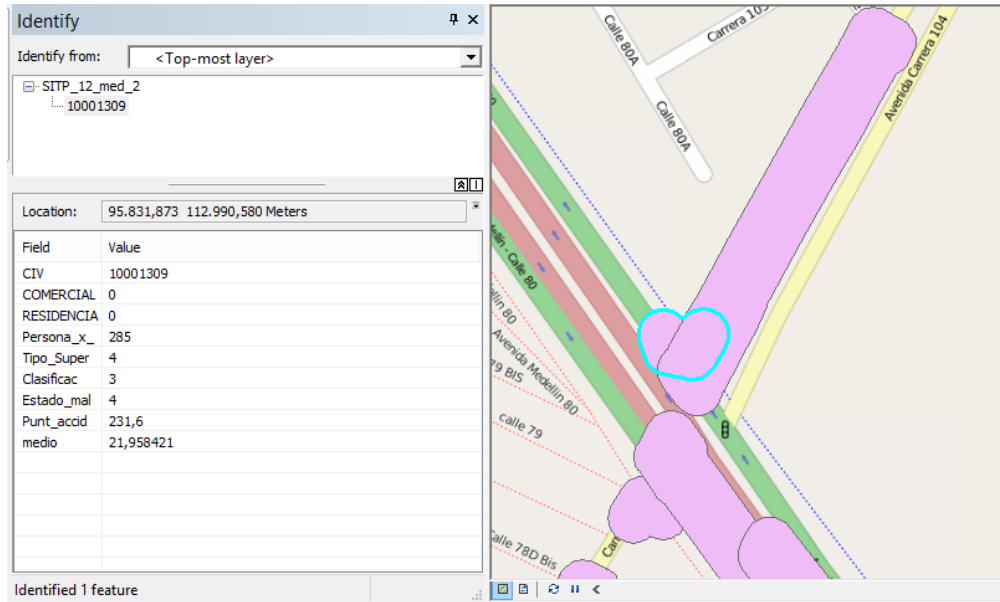


StdError	Observed	LocalR2	Predicted	Intercept	C1_COMERCI	C2_RESIDEN	C3_Persona	C4_Tipo_Su	C5_Clasifi	C6_Estado_	C7_Punt_ac
3,16	24,70	0,70	24,97	38,26	-0,80	0,53	-0,03	-5,43	9,29	-4,78	0,17

Fuente: Elaboración propia ArcGis, 2018

De forma general tenemos que los predios comerciales influyen negativamente hay 4 pero hacen que se reduzca la velocidad lo cual es usual y por tanto la constante es negativa, los predios residenciales influyen de forma positiva sin embargo como no hay de este tipo, no se afecta la regresión; las personas influyen de forma negativa hay 74 personas por lo cual por esta densidad que puede ser alta la velocidad se puede reducir; el tipo de superficie 4 es flexible y sigue habiendo una contradicción en el modelo, debido a que el signo del coeficiente debería ser positivo ya que este tipo de pavimento hace que la velocidad aumente; en cuanto a la clasificación el tipo 3 es local, es decir una vía que conecta gran parte de la ciudad por lo cual la movilidad puede ser densa y por esta razón el coeficiente debería ser negativo y no como se muestra en el modelo; en cuanto al estado el tipo 3 es regular lo cual no afecta la velocidad en cuanto a reducirla, por lo cual el coeficiente debería ser positivo y no como lo calculo el modelo; finalmente el puntaje de accidente como solo hay 1 su influencia en el modelo es mínima y por tanto el coeficiente es positivo.

Figura 16.Detalle del CIV con R2 local de 0,8

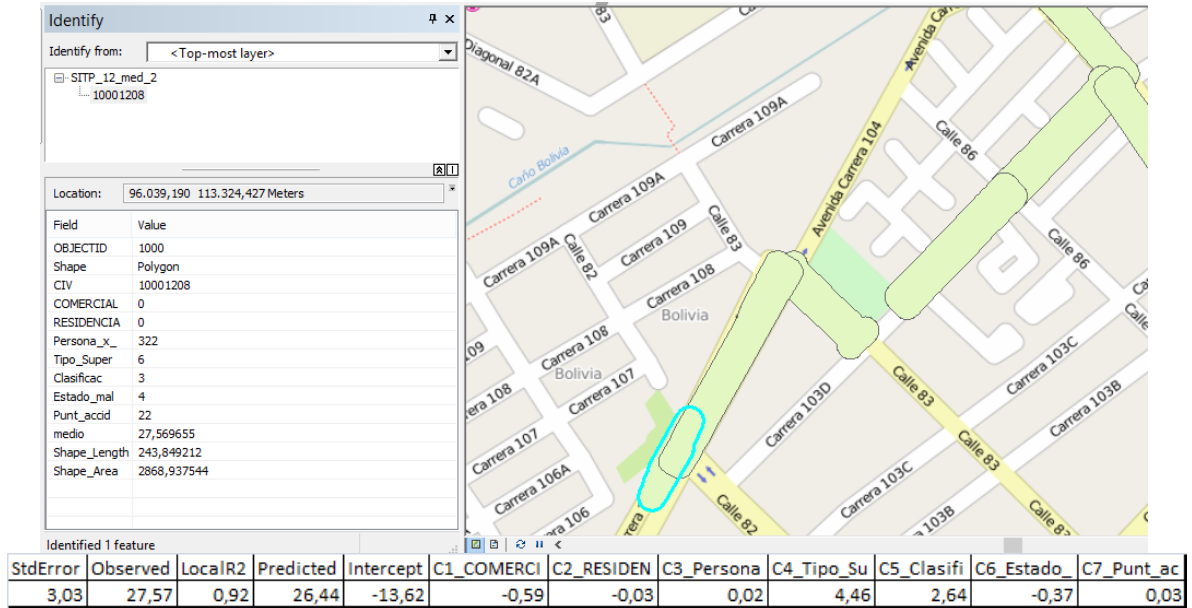


StdError	Observed	LocalR2	Predicted	Intercept	C1_COMERCI	C2_RESIDEN	C3_Persona	C4_Tipo_Su	C5_Clasifi	C6_Estado_	C7_Punt_ac
2,97	21,96	0,80	23,39	-15,64	-0,85	0,58	0,03	4,84	-2,84	3,38	0,03

Fuente: Elaboración propia ArcGis, 2018

De forma general tenemos que tanto los predios comerciales como residenciales no influyen en este modelo debido a que no hay presencia de estos en el CIV analizado, sin embargo podemos observar como el coeficiente de la variable comercial es negativo es decir si hubiera presencia de este tipo de predio en la estimación de la velocidad esta se reduciría en presencia de predios comerciales, caso contrario a los predios residenciales; las personas influyen de forma positiva hay 285 personas por lo cual por esta densidad que puede ser alta, la velocidad se puede reducir; el tipo de superficie 4 es flexible aquí el signo del coeficiente es el correcto, ya que este tipo de pavimento hace que la velocidad aumente; en cuanto a la clasificación el tipo 3 es local, es decir una vía que conecta gran parte de la ciudad por lo cual la movilidad puede ser densa y por esta razón el coeficiente es negativo; en cuanto al estado el tipo 4 es Bueno, con lo cual el signo positivo del coeficiente es el correcto puesto que si no hay huecos las velocidades son buenas, es decir en promedio de 20 a 30 Km/h; finalmente el puntaje de accidente es el más alto de todos sin embargo el valor del coeficiente es el más bajo por influencia en el modelo es mínima.

Figura 17.Detalle del CIV con R2 local de 0,9

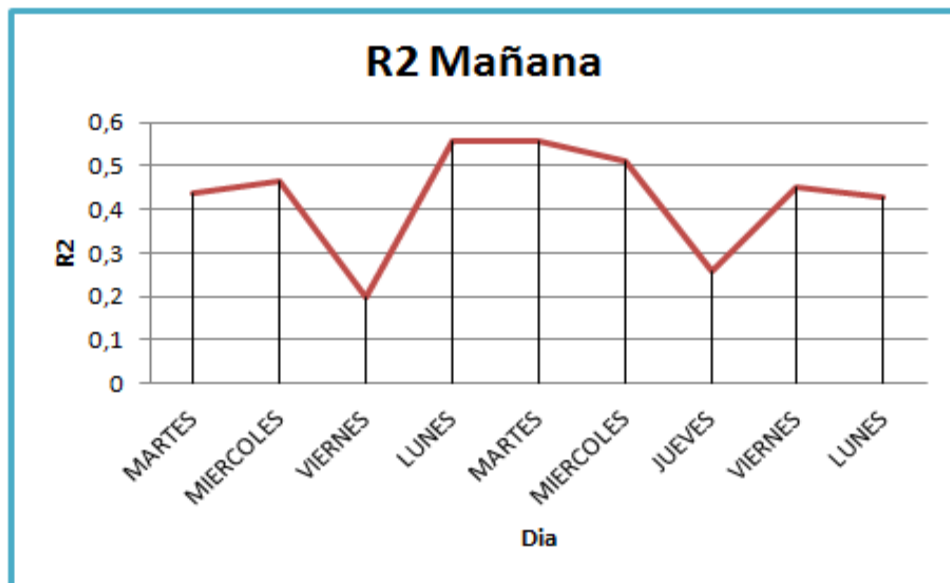


Fuente: Elaboración propia ArcGis, 2018

De forma general tenemos que tanto los predios comerciales como residenciales no influyen en este modelo debido a que no hay presencia de estos en el CIV analizado, sin embargo podemos observar como el coeficiente de la variable comercial y residencial es negativo, quiere decir que si hubieran predios de estas categorías la velocidad del SITP se reduciría; las personas influyen de forma positiva hay 322 personas por lo cual por esta densidad la cual es la más alta, la velocidad aumenta; el tipo de superficie 6 es Rígido aquí el signo del coeficiente es el correcto, ya que este tipo de pavimento es el mejor lo cual hace que la velocidad aumente; en cuanto a la clasificación el tipo 3 es local, es decir una vía que conecta gran parte de la ciudad por lo cual la movilidad puede ser densa y por esta razón el coeficiente debería ser negativo, en el modelo es positivo; en cuanto al estado el tipo 4 es Bueno, el signo debería ser positivo; finalmente el puntaje de accidente 27 no influye tanto en el modelo puesto que el valor del coeficiente es el más bajo.

Analizando gráficamente el comportamiento del R2 a través de los días tenemos lo siguiente:

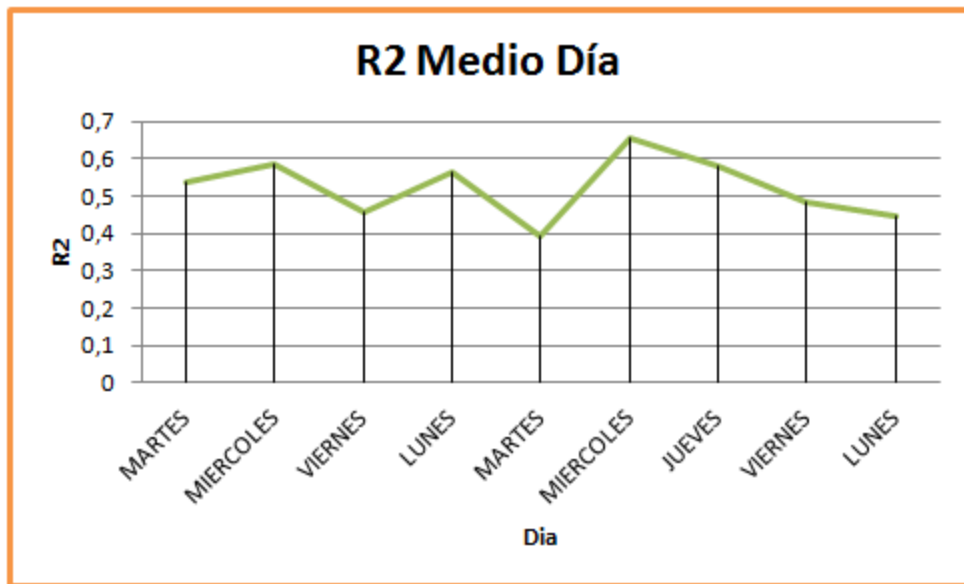
Figura 18. Comportamiento del R2 en la mañana



Fuente: Elaboración propia

Para el cálculo en las mañanas se obtienen modelos con R2 muy variables con un mínimo el día viernes de 0,19 y un máximo de 0,55 para el día lunes y martes, sin embargo no se encuentra ningún patrón o tendencia en el comportamiento de un día a otro, por ejemplo de viernes a lunes la primera semana el comportamiento es al aumento, mientras que en la segunda el comportamiento es decreciente por lo cual es poco confiable generar alguna premisa.

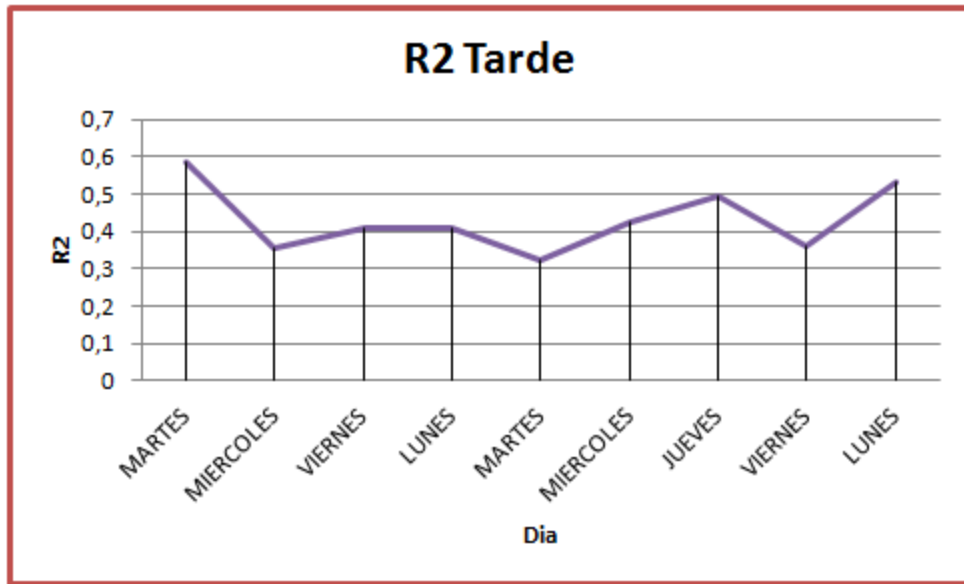
Figura 19. Comportamiento del R2 en el medio día



Fuente: Elaboración propia

Para los modelos calculados a medio día obtenemos un valor mínimo de R2 de 0,39, el cual es el más alto de los tres periodos de tiempo de análisis (mañana, medio día y tarde), lo mismo sucede con la tarde con un valor máximo de R2 de 0,65 para el día miércoles; analizando de forma general los R2 son los más altos por lo cual los modelos son más ajustados, aparte de esto podemos analizar que hay una periodicidad en el comportamiento de los R2, por ejemplo de martes a miércoles el comportamiento es creciente y el miércoles es donde se obtiene el mayor R2 para las dos semanas; el comportamiento de los R2 tiende a ser decreciente de miércoles a viernes.

Figura 20. Comportamiento del R2 en la tarde



Fuente: Elaboración propia

Analizando los datos de los R2 obtenidos tenemos que el R2 mínimo es de 0,32 para el día martes, mientras que el R2 más alto lo tenemos también el martes, razón por la cual no hay consistencia en los resultados obtenidos, también analizando el comportamiento entre días este es contradictorio ya que para la primer semana de martes a miércoles decrece mientras que en la segunda es creciente, lo mismo pasa de miércoles a viernes, el comportamiento es creciente para la primer semana mientras que para la segunda el comportamiento es decreciente.

7. CONCLUSIONES

- Los modelos de regresión geográficamente ponderados GWR al incluir el componente espacial en su análisis presentan una notable ventaja sobre los modelos de mínimos cuadrados ordinarios OLS, con lo cual el resultado de estos procesos presenta un grado de confiabilidad mayor, sin embargo hay que tener en cuenta que los modelos de OLS sirven como un primer filtro para la elección de las variables explicativas en las regresiones geográficas ponderadas.
- La naturaleza del problema estudiado, que busca explicar las velocidades en un sistema de transporte masivo no es posible abordarlo desde la aplicación de modelos de Regresión Lineal convencionales, por lo cual en el presente trabajo se busco mostrar la relevancia y diferentes aplicaciones de los modelos de regresión espacial ponderados y la importancia de incluir el componente geográfico que enmarca diferentes fenómenos.
- El modelo planteado pretende ser la base para futuros estudios que busquen explicar el comportamiento de velocidades y la dinámica de los diferentes medios de transporte que existen en las grandes ciudades, mediante la inclusión de más variables explicativas referentes a aspectos que no se hayan tenido en cuenta en el desarrollo del presente proyecto.
- Con base en los resultados obtenidos del modelo que presenta el mayor ajuste, podemos decir que el ajuste es bueno para aquellos CIV donde la velocidad se encuentra dentro de un promedio, ya que para aquellas zonas donde la velocidad es alta, el valor predicho presenta un mayor desfase el cual puede estar entre los 10 a 15 Km entre el valor real y el de la predicción.
- Para la selección de las variables explicativas del modelo, se debe tener en cuenta no incluir en lo posible variables categorizadas ya que pueden generar una multicolinealidad local, este mismo problema sucede en el caso de incluir variables binarias; como en nuestro caso el estrato y predios de tipo dotacional, Industrial y Otros las cuales con el primer muestreo mediante OLS fueron no significativas ,por lo cual se retiraron del calculo de las regresiones
- Con el método de regresión geográficamente ponderado, logramos encontrar un buen ajuste para el día miercoles a medio día con un R^2 de 0,65 y por tanto una buena explicación a las velocidades en partes de la localidad con características especiales, razón por la cual podemos llegar a concluir que este

método es bueno para aplicarlos a una pequeña zona con una muestra considerable.

8. RECOMENDACIONES

- Para trabajos posteriores es recomendable tratar de sistematizar procesos, con el fin de hacer más eficientes los cálculos y así poder ampliar las muestras de datos al analizar zonas geográficamente más grandes como la ciudad de Bogotá.
- Se deben Incluir el mayor número de variables explicativas de la velocidad como la semaforización, y medidas de flujo vehicular por cada CIV, entre otras; que se encuentren sujetas a toda la operación logística del sistema y que se ajusten al modelo de regresión geográfica ponderada, permitiendo así realizar más análisis de fiabilidad estadística y despreciar variables poco significativas.
- Ya que el propósito es estudiar el comportamiento de las velocidades de la flota del SITP, es posible derivar un segundo análisis relacionado con el cumplimiento en cronogramas de rutas, se sugiere incluir en los análisis variables de semaforización y estudiar la influencia de estas sobre los modelos.
- La investigación puede ser complementada con técnicas de analítica de datos a través de visores web como el software Tableau, donde se pueda modelar los resultados y comportamiento espaciales al incluir o excluir variables del modelo y visualizar de forma más sencilla los patrones.
- Los resultados de las regresiones geográficas ponderadas en el software ArcGIS, se ven afectados por la inclusión de variables cualitativas las cuales se convierten en variables categorizadas las cuales pueden llegar a presentar correlación y o multicolinealidad por lo tanto es preciso, probar diferentes software y aplicaciones de modelamiento geoestadístico , para realizar una comparación en cuento a los resultados obtenidos.

9. BIBLIOGRAFÍA

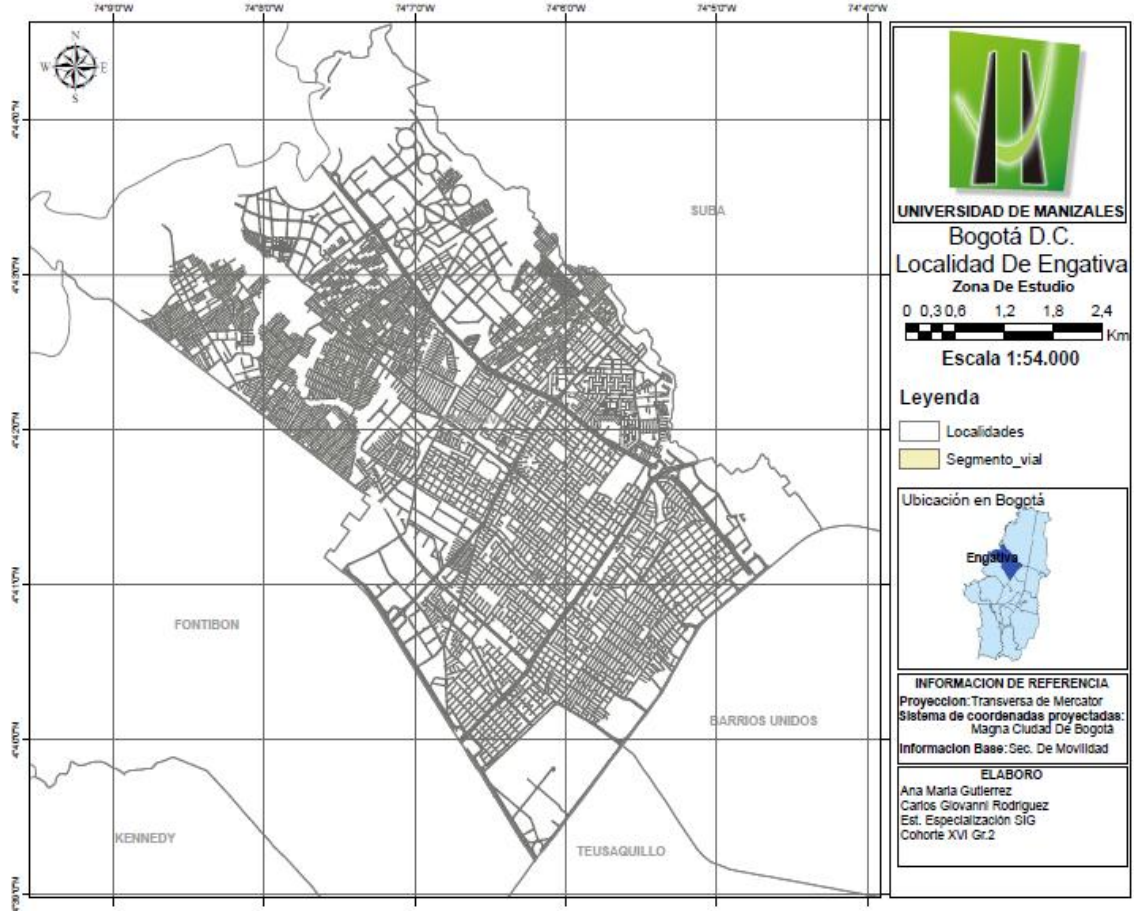
- Acevedo, I., Velásquez, E. (2008). Algunos conceptos de la econometría espacial y el análisis exploratorio de datos espaciales. *Revista Ecos de Economía* No. 27. Medellín, octubre de 2008, pp. 9- 34
- Alcaldía Mayor de Bogotá(2012). *Diseño y Evaluación de Alternativas de Accesibilidad para los Usuarios del SITP en Condición de Discapacidad*". Bogotá D.C.
- Alessandrini, N(2015). *Aplicación de algoritmos de aprendizaje estadísticos para predecir velocidades de buses con información en tiempo real*. Pontificia universidad Católica de Chile. Santiago De Chile.
- Behrentz, E., Carrizosa, J., Acevedo, J. (2009). "Lograr la sostenibilidad: un debate entre Bogotá y la Región". En: *Revista de Estudios Sociales*, 32, pp. 204 210.
- Cámara de comercio de Bogotá(2016) *Encuesta de percepción sobre las condiciones, calidad y servicio a los usuarios de Transmilenio,SITP, y TPC*.Bogotá D.C.
- Cogua, M (2014), *Dinámica de la construcción por usos localidad Engativá*, UAECD, Bogotá D.C.
- Departamento nacional de planeación(2014). *Síntesis de Evaluación de los Sistemas Integrados de Transporte Masivo de Colombia*. Bogotá, D. C.
- Duque, J., Velásquez, H., Agudelo J(2011). *Infraestructura pública y precios de vivienda: una aplicación de regresión geográficamente ponderada en el contexto de precios hedónicos*. *Revista Ecos de Economía* No. 33. Medellín, Diciembre de 2011, pp. 95- 122
- Giraldo(2002). *Introducción a la geoestadística, Teoría y aplicación*.Universidad Nacional De Colombia. Bogotá D.C.
- Gujarati, DN (1995). *Econometría Básica (3ª Ed.)*. McGraw-Hill International Editions.New York
- Hidalgo Darío(2005).*Comparación de alternativas de transporte público masivo-Una aproximación conceptual*. revista de ingeniería 21 facultad de ingeniería Universidad de los Andes, p.p 94-104.
- Jolonch Palau, Javier (2013). *Análisis del transporte masivo y la movilidad en Bogotá*. Universidad & Empresa, No. 24, p.p. 15-23.
- Méndez, D (2009).Curso: Ingeniería de tránsito, recuperado de [http://www.academia.edu/4852383/Abril_2009_Abril_2009_MAESTRIA_EN_VIAS_TERRESTRES_Proped%C3%A9utico_MAESTRIA_EN_VIAS_TERRESTRES_MAESTRIA_EN_VIAS_TERRESTRES_Proped_Proped_%C3%](http://www.academia.edu/4852383/Abril_2009_Abril_2009_MAESTRIA_EN_VIAS_TERRESTRES_Proped%C3%A9utico_MAESTRIA_EN_VIAS_TERRESTRES_MAESTRIA_EN_VIAS_TERRESTRES_Proped_Proped_%C3%99)

A9_%C3%A9_utico_utico_Curso_Ingenier_Curso_Ingenier%C3%AD_%C3%ADa_de_Transito_a_de_Transito.

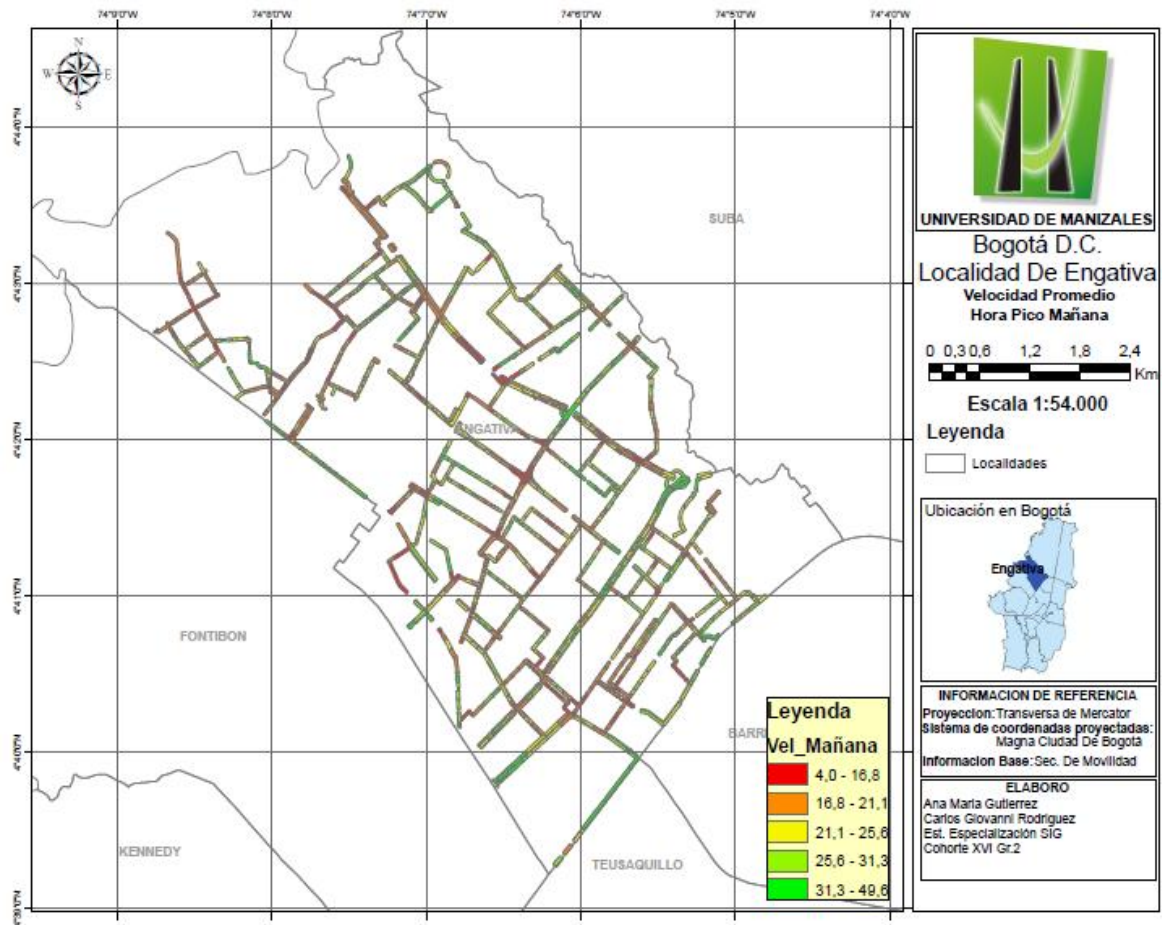
- Pérez, J.(2006). Econometría espacial y ciencia regional, Investigación Económica, vol. LXV, núm. 258, octubre-diciembre, 2006, pp. 129-160, Distrito Federal, México.
- Puentes, S., Hernández, V., Rivera J.(2015). Aplicación de ITIL al SITP. Escuela Colombiana De Ingeniería Julio Garavito Bogotá D.C.
- Sabogal, O.A, Hincapié, J.D., Santa Chávez, J,J. , Escobar J.W.. (2015). Modelos de regresión lineal para estimación de tiempos de viaje en sistemas de transporte masivo. Ciencia e Ingeniería Neogranadina, 25(1), pp.77 - 89
- Secretaria Distrital de Planeación (2011) Documento Técnico de Soporte Modificación al Plan de Ordenamiento Territorial de Bogotá, Bogotá, D. C.
- Universidad de Oviedo. Conceptos de construcción de un MDE, Parte 16 interpolación. Recuperado de <http://www6.uniovi.es/~feli/CursoMDT/Tema1/Page16.html>
- Valencia, A(2008). Modelo para planificación, operación y diseño físico de corredores de transporte público de superficie. Universidad de Chile. Santiago de Chile.
- Valencia, A(2008).Influencia del espaciamiento de paraderos en la velocidad comercial en corredores de transporte público de superficies
- Yepes, T., Ramírez, J., Targa, F.(2013). La integración de los sistemas de transporte urbano en Colombia. Una reforma en transicion.Fedesarrollo.

ANEXOS

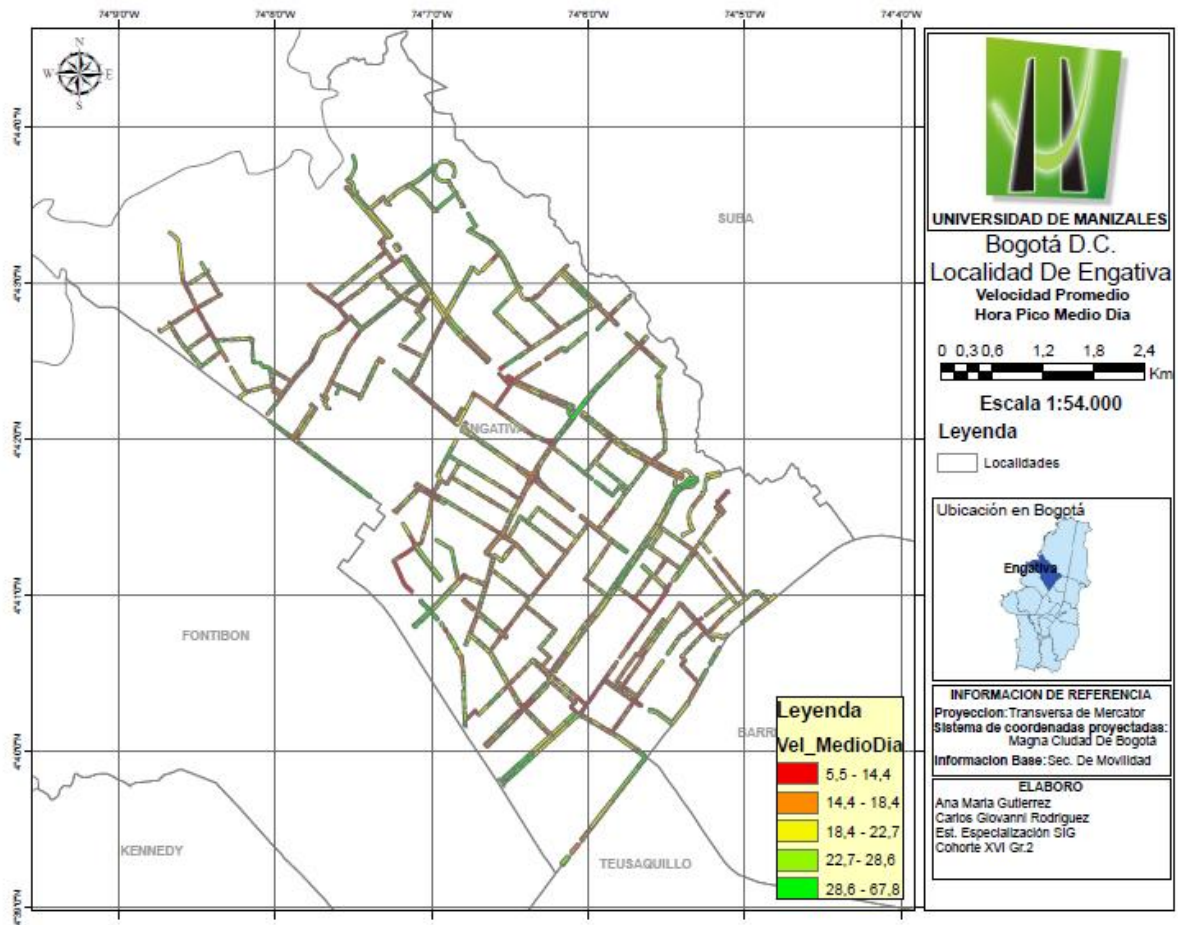
Anexo A. Mapa zona de Estudio



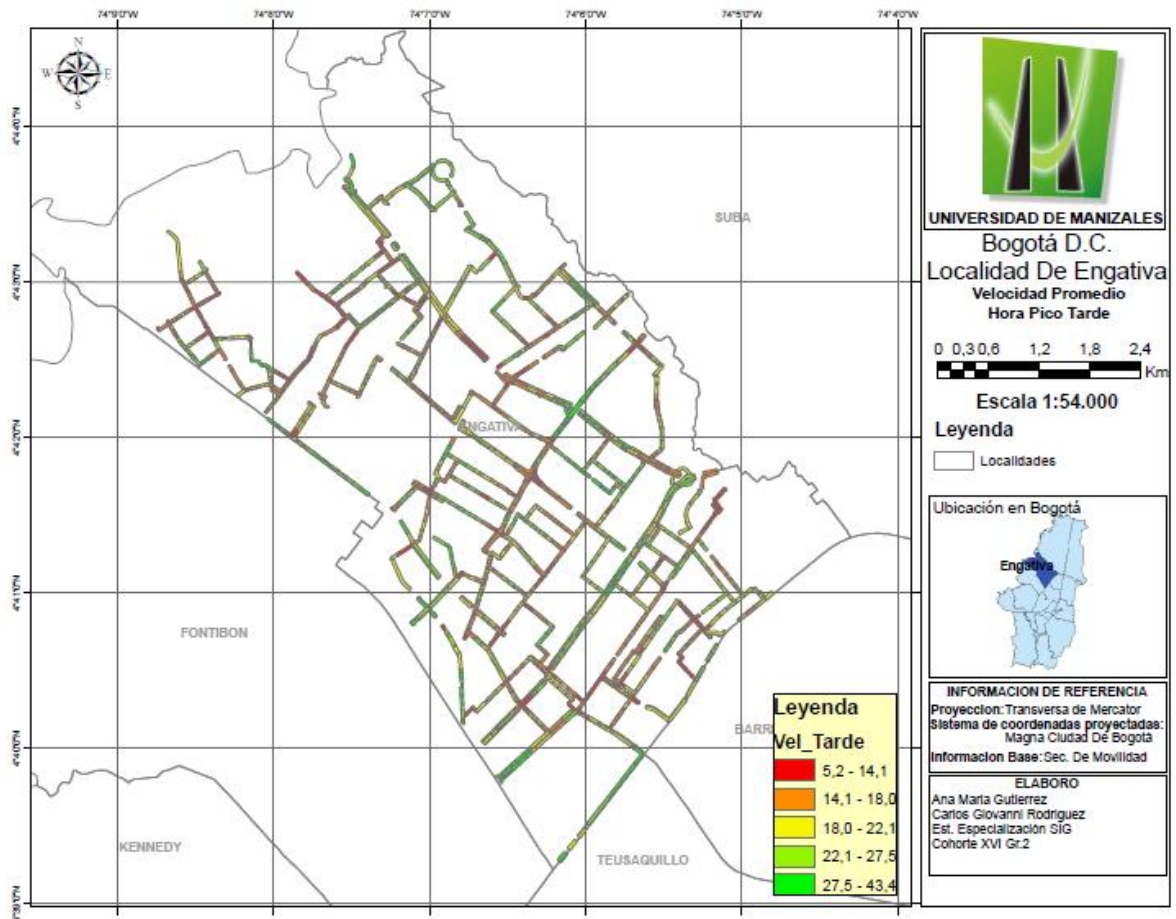
Anexo B. Mapa de velocidades hora pico mañana



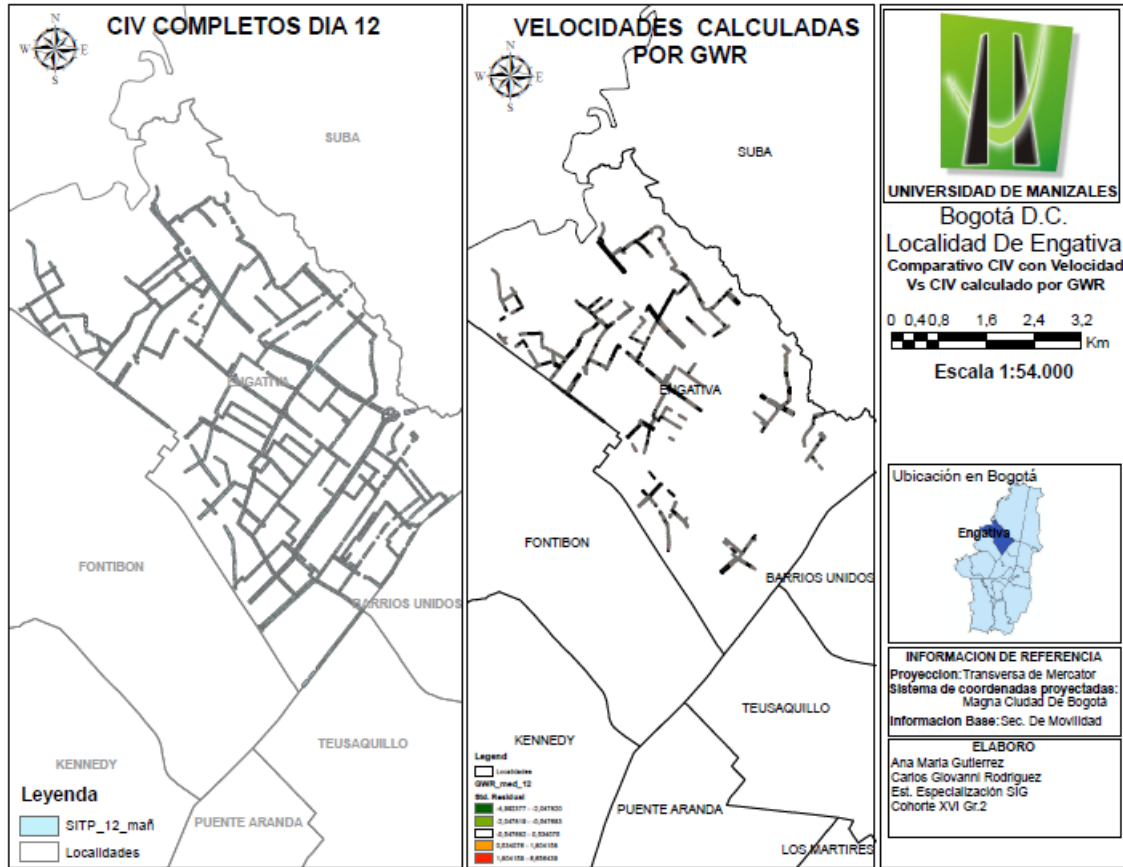
Anexo C. Mapa de velocidades hora pico Medio Día



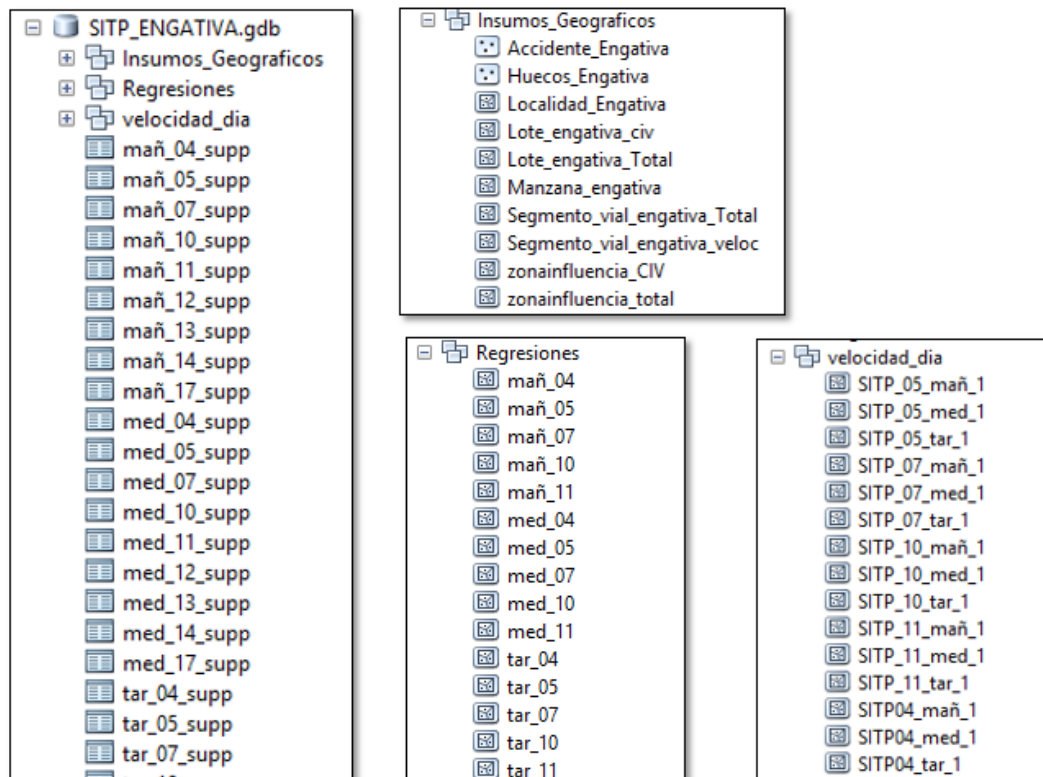
Anexo D. Mapa de velocidades hora pico tarde



Anexo E. Mapa comparativo CIV completo Vs. Velocidad por CIV estimadas por GWR



Anexo F. Base de datos



Para el almacenamiento de la información necesaria y para el desarrollo del proyecto se generó una File Geodatabase con nombre SITP_ENGATIVA, al tratarse de estudio centralizado en la ciudad de Bogotá se trabajó bajo el sistema de coordenadas de Magna Ciudad Bogotá.

La estructura de la base de datos se compone de las tablas resultantes de cada regresión realizada para cada día de estudio en los diferentes horarios, que permiten determinar los, diferentes coeficientes, valores de errores estándar, los valores residuales de cada variable estudiada y 3 Data Frame así:

- **Insumos Geográficos:** donde se almacena las capas geográficas de la localidad de Engativá necesarias para el desarrollo y localización de la investigación, incluye las capas intermedias generadas en los geoprocесamientos empleados en el proceso.
- **Regresiones:** en este Data Frame se almacenan las capas espaciales resultantes de las Regresiones realizadas en el programa Arcgis para cada día estudiado y cada diferente horario evaluado.
- **Velocidad_dia:** incluye una capa espacial por cada día de estudio, donde se almacena la información de las variables geográficas consolidadas por cada segmento vial, complementada con la variable de velocidad medida en tres

diferentes horarios: mañana , medio día y tarde , donde para cada horario se genera una capa espacial contenida a su vez en este Data Frame .