

**UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA**



**DIAGNOSTICO ESPACIAL DE LOS ACCIDENTES DE TRÁNSITO
EN EL MUNICIPIO DE SAN SALVADOR 2006 – 2010**

**TRABAJO DE GRADUACIÓN PRESENTADO POR:
WILSON ERNESTO RODRÍGUEZ PACHECO**

**PARA OPTAR AL TITULO DE:
LICENCIADO EN ESTADÍSTICA**

CIUDAD UNIVERSITARIA, DICIEMBRE DE 2013.

AUTORIDADES DE LA UNIVERSIDAD DE EL SALVADOR

RECTOR

Ing. Mario Roberto Nieto Lovo

VICE-RECTOR ACADÉMICO

MSc. Ana María Glower de Alvarado

VICE-RECTOR ADMINISTRATIVO

MSc. Óscar Noé Navarrete

SECRETARIO GENERAL

Dra. Ana Leticia Zavaleta de Amaya

**AUTORIDADES DE LA FACULTAD DE CIENCIAS NATURALES Y
MATEMÁTICA**

DECANO

MSc. Martín Enrique Guerra Cáceres.

VICE-DECANO

Lic. Ramón Arístides Paz Sánchez

SECRETARIO

Lic. Carlos Antonio Quintanilla Aparicio

ESCUELA DE MATEMÁTICA

DIRECTOR

Dr. José Nerys Funes Torres

“DIAGNOSTICO ESPACIAL DE LOS ACCIDENTES DE TRÁNSITO EN EL
MUNICIPIO DE SAN SALVADOR 2006 – 2010”

TRABAJO DE GRADUACIÓN PRESENTADO POR:
WILSON ERNESTO RODRÍGUEZ PACHECO

PARA OPTAR AL TITULO DE:
LICENCIADO EN ESTADÍSTICA

ASESORES

MSC. PEDRO ARMANDO RAMOS ALBERTO: _____

MSC. MARIO GIOVANNI MOLINA MASFERRER: _____

Agradecimientos

Dios, padres, amistades y enemistades

Resumen

Capítulo I: Base teórica y metodológica de la geoestadística.

Conocer fundamentos y procedimientos de la técnica geoestadística como herramienta útil en la modelación de fenómenos de carácter espacial.

Capítulo II: Aplicación y análisis de los métodos geoestadísticos a los accidentes de tránsito en el municipio de San Salvador 2006 – 2010.

Aplicar la técnica geoestadística a los accidentes de tránsito ocurridos en el municipio de San Salvador en el periodo 2006 – 2010 y determinar la distribución que tiene el fenómeno en el área de estudio en base a un test de aleatoriedad y estimar el mejor modelo de interpolación kriging que se ajuste a los datos

Capítulo III: Presentación de resultados.

Presentación de los mapas obtenidos con el estimador kriging de manera adecuada, de forma tal que contribuya a una mejor comprensión y exposición de dichos resultados.

Contenido

Introducción.....	i
Planteamiento del problema.....	ii
Justificación.....	iii
Objetivos.....	iv
Marco teórico.....	v
Metodología.....	vii
1 Base teórica y metodológica de la geoestadística.....	1
1.1 Límites de la estadística clásica.....	3
1.2 Estadística espacial.....	3
1.2.1 Técnicas estadísticas espaciales.....	4
1.2.2 Datos espaciales.....	5
1.2.3 Problemas de la estadística espacial.....	5
1.3 Conceptos básicos de la geoestadística.....	6
1.3.1 Definición de geoestadística.....	6
1.3.2 Variables aleatorias regionalizadas.....	6
1.3.3 Hipótesis básicas de la geoestadística.....	8
1.3.4 Cálculos geométricos básicos.....	10
1.4 Manejo de un estudio espacial.....	11
1.4.1 Análisis exploratorio de datos espaciales.....	12
1.4.1.1 Análisis exploratorio clásico.....	12
1.4.1.2 Análisis exploratorio espacial.....	14
1.4.2 Análisis estructural.....	15
1.4.2.1 Análisis del patrón espacial.....	15
1.4.2.2 Análisis de correlación espacial.....	18
1.4.3 Construcción del semivariograma.....	22
1.4.3.1 Semivariograma experimental.....	23
1.4.3.2 Modelos teóricos del semivariograma.....	26
1.4.4 Modelos de interpolación kriging.....	29
1.4.4.1 Construcción del kriging.....	29
1.4.4.2 Vecindad kriging.....	30
1.4.4.3 Ecuaciones kriging.....	32
1.4.4.4 Validación del modelo kriging.....	38
2. Aplicación y análisis de los métodos geoestadístico a los accidentes de tránsito en el municipio de San Salvador 2006 – 2010.....	40
2.1 Análisis estructural.....	41
2.1.1 Análisis del patrón espacial.....	41
2.1.2 Análisis de correlación espacial.....	43
2.1.3 Estimación de parámetros para el semivariograma.....	45

2.1.4 Validación del modelo kriging.....	48
3. Presentación de resultados.....	51
3.1 Características geográficas, sociales y económicas del municipio de San Salvador.....	52
3.2 Principales características de los accidentes de tránsito en el municipio de San Salvador 2006 – 2010.....	55
3.2.1 Relación: Accidentes de tránsito y edad.....	56
3.2.2 Relación: Accidentes de tránsito y hora.....	57
3.2.3 Relación: Accidentes de Tránsito y día.....	58
3.2.4 Relación: Causa del accidente y tipo de accidente.....	59
3.2.5 Frecuencia y características de los accidentes de tránsito por distrito.....	60
3.3 Distribución espacial de los accidentes de tránsito.....	63
3.3.1 Centro Histórico.....	63
3.3.2 Distrito 1.....	69
3.3.3 Distrito 2.....	75
3.3.4 Distrito 3.....	81
3.3.5 Distrito 4.....	87
3.3.6 Distrito 5.....	93
3.3.7 Distrito 6.....	99
Conclusiones.....	105
Recomendaciones.....	106
Referencias bibliográficas.....	107

Introducción

Hoy en día vivimos en una realidad donde la seguridad vial juega un rol primordial en la calidad de vida de la sociedad. Es evidente que los accidentes de tránsito se han convertido en un problema de salud pública que despierta gran preocupación por tratarse de un problema que constituye una de las principales causas de defunción, discapacidad, daños físicos y psicológicos que sufren las personas en su posterior reincorporación a la sociedad, al tener que vivir en condiciones de salud diferentes a las vividas con anterioridad. Además, los accidentes de tránsito representan pérdidas de producción, daños materiales y una sobrecarga en los servicios de salud, lo que significa una nada despreciable inversión económica para el estado y la ciudadanía en general.

Pese a ser la parte más dramática del problema del transporte, hasta el momento, en el país, no se tiene antecedentes de estudios relacionados al tema, a diferencia con los realizados con otras problemáticas de salud. En ese sentido y con la finalidad de proporcionar un panorama cuantitativo sobre el comportamiento y tendencia de los percances viales y de utilidad para los sectores interesados del país, en el diagnóstico espacial de los accidentes de tránsito en el municipio de San Salvador 2006 – 2010 se analizan las diferentes causas, tipos y lugares donde suceden estos eventos, dentro de los límites de los distritos del municipio de San Salvador. El punto de partida de la investigación es asumir una intuición geográfica - probabilística que tiene como base un concepto espacial, siendo más precisos el estudio de variables distribuidas en el espacio (eventos con ubicación geográfica).

El conocimiento y la información generada es posible básicamente a dos grandes recursos desarrollados en los últimos años en investigaciones de gestión territorial, el primero es la aplicación de la técnica geoestadística, metodología que forma parte de un conjunto de herramientas de la estadística espacial y tradicionalmente aplicada en el estudio de recursos mineros, que posteriormente fue exportada a las ciencias ambientales y al estudio de imágenes satelitales. Y en segundo lugar el uso notable de los sistemas de información geográfica (SIG), que sin duda constituye una de las más innovadoras tecnologías para manejar estudios geográficos con grandes volúmenes de información. Se utiliza el programa estadístico SPSS ampliamente conocido y usado en investigaciones estadísticas, el CrimeStat un programa de estadística espaciales, específicamente para el análisis de ocurrencia de eventos en un área determinada y el programa ArcGIS en la simulación y obtención del modelo que mejor se ajuste a los datos para la predicción espacial.

Esperando que este documento sea de utilidad como una línea base para monitorear los lugares, causas y consecuencias que dejan los percances viales, aspiramos a llamar la atención de las instituciones competentes y que sirva para definir prioridades en acciones y decisiones responsables que lleven a mejores resultados.

Planteamiento del problema

Los accidentes de tránsito ocurren cuando los individuos efectúan erradas maniobras en escenarios de poco control, cuando se viaja en vehículos que no están en condiciones seguras, cuando las calles están en malas condiciones y/o con mala señalización o simplemente por las costumbres intolerantes de la sociedad en general.

Según la Organización Panamericana de la Salud en conjunto con la Organización Internacional de la Salud, en los últimos años en El Salvador las cuatro principales causas de muerte en el país fueron las enfermedades cardiovasculares con una tasa de 81.6%, el cáncer 48.5%, los accidentes de tránsito y la delincuencia con una tasa de 38.7% y 26% respectivamente. No es extraño que El Salvador siendo el país más pequeño de la región, con graves problemas económicos, sociales y con mayor densidad de población se vea perjudicado con este tipo problemática.

Continuando con el informe de la Organización Panamericana de la Salud OPS, El Salvador ocupa la quinta posición como sistema de tránsito más inseguro y su tasa de accidentalidad de 15.8 muertes por accidente de tránsito por cada 100,000 habitantes se encuentra por arriba de la tasa media de la región de las Américas. Un 70% de la sociedad salvadoreña interactúa y se enfrenta todos los días con un tráfico vehicular muy complejo y peligroso, debido al aumento de automotores el cual puede notarse en el incremento de los embotellamientos que a diario se generan en las horas de mayor congestión.

Además de acuerdo a datos proporcionados por la Subdirección de Tránsito Terrestre de la PNC y Viceministerio de Transporte, que son las únicas instituciones que recolectan estadísticas básicas sobre accidentes de tránsito en el país; entre el período 2006 y 2010 han ocurrido 31,356 accidentes de tránsito y 5,499 muertes por accidentalidad vial registrados, lo que equivaldría a decir un promedio de tres muertes diarias en el mismo intervalo de años.

Un elemento clave para reducir los accidentes de tránsito sería el conocer las zonas donde usualmente ocurren estos, en el entendido que al conocer dichas zonas, están disponibles de mayor atención por parte de las autoridades involucradas en el tema, lo que consecuentemente podría significar evitar accidentes. Por esta razón es necesario investigar nuevas metodologías que permitan realizar pronósticos adecuados y nos planteamos el construir un mapa de frecuencias de accidentes tránsito ocurridos entre los años 2006 y 2010 en el municipio de San Salvador utilizando la técnica geoespacial como eje fundamental en la investigación.

Justificación

Como se ha mencionado anteriormente las condiciones culturales, demográficas y económicas de nuestro país han provocado un aumentado en los accidentes de tránsito, llegando a cifras alarmantes. Empíricamente aumenta el costo de la atención médica especializada, en un país donde los seguros de vida y de accidentes son inalcanzables para la gran mayoría, haciendo que los gastos casi siempre recaigan sobre las familias de los accidentados. Por lo anteriormente descrito, este proyecto es de suma importancia y la razón principal de realizarlo es el de colaborar a las instituciones que velan por la seguridad vial, a la búsqueda de posibles soluciones que conlleven a reducir o prevenir los accidentes de tránsito.

Consecuentemente las medidas que se adopten a partir de la información que se obtenga de este estudio serán de beneficio para la comunidad en general, lo que podríamos explicarlo por el hecho siguiente: los accidentes de tránsito tienen un costo en cuidados médicos, intervenciones de los servicios de policía, reparaciones de vehículos, pérdidas de producción económica, personas heridas o fallecidas, etc. Evitar un accidente equivaldría a ahorrar dinero al estado y a la sociedad lo que vendría a mejorar la calidad de vida de todos.

El reto futuro es transformar este primer esfuerzo de su forma impresa a un sistema de información activo, flexible y permanente, que sea capaz de generar información con periodicidad regular. La creación de un sistema de información geográfica para la prevención de accidentes de tránsito en el municipio de San Salvador sería de gran utilidad para toda la ciudadanía y serviría a su vez como modelo para poder ser aplicado posteriormente en otras ciudades del país o con otros fenómenos de estudio.

Objetivos

General:

- Desarrollar un modelo geoestadístico para el diagnóstico espacial de los accidentes de tránsito.

Específicos:

- Estudiar mediante un análisis estructural el fenómeno de los accidentes de tránsito.
- Detectar las zonas de mayor intensidad.
- Determinar zonas de alto riesgo vial.
- Identificar las variables que más inciden en los accidentes de tránsito.
- Efectuar comparaciones numéricas y dinámicas de los accidentes de tránsito entre los diferentes años.

Marco teórico

1 - Antecedentes históricos: La constante búsqueda de técnicas para realizar estudios más profundos y que nos dejen resultados fiables ha venido evolucionando y perfeccionándose en la medida que se han desarrollado estudios en mayor cantidad y mejor calidad tecnológica. Las tecnologías de información geográfica, especialmente los sistemas de información geográfica (SIG) constituye una herramienta poderosa y muy útil para los estudios de análisis y gestión territoriales, dado sus capacidades para capturar, almacenar, manipular, analizar, modelar y representar datos espaciales georeferenciados¹.

El estudio de fenómenos de tipo espacial, por medio de métodos geoestadísticos, surgió a partir de los años sesenta, especialmente con el propósito de predecir valores de las variables en sitios no muestreados. Como antecedentes suelen citarse trabajos de Sichel (1947; 1949) y Krige (1951). El primero observó la naturaleza asimétrica de la distribución del contenido de oro en las minas surafricanas, la equiparó a una distribución de probabilidad lognormal y desarrolló las fórmulas básicas para esta distribución. Ello permitió una primera estimación de las reservas, pero bajo el supuesto de que las mediciones eran independientes, en clara contradicción con la experiencia de que existen "zonas" más ricas que otras. Una primera aproximación a la solución de este problema fue dada por geólogo G. Krige que propuso una variante del método de medias móviles, el cual puede considerarse como el equivalente al krigeado simple, que es uno de los métodos de estimación lineal en el espacio con mayores cualidades teóricas². La formulación rigurosa y la solución al problema de predicción vinieron de la mano de Matheron (1962) en la escuela de minas de París. En los años sucesivos la teoría se fue depurando, ampliando su campo de validez y reduciendo las hipótesis necesarias. De la minería la técnica geoestadística, se ha "exportado" a muchos otros campos como hidrología, física del suelo, ciencias de la tierra y más recientemente al monitoreo ambiental.

Aunque la aplicación de la herramienta geoestadística es bastante reciente, son innumerables los ejemplos en los que se ha utilizado esta técnica en estudios ambientales con el ánimo de predecir fenómenos espaciales. La columna vertebral del análisis geoestadístico es la determinación de la estructura de autocorrelación entre los datos y su uso en la predicción a través de las técnicas conocidas como kriging.

2 –Definición clásica de la geoestadística: La geoestadística se define como la aplicación de los métodos probabilísticos a fenómenos que varían en el espacio, su objetivo principal es la estimación, predicción y simulación de dichos fenómeno³.

¹Instituto de Andaluz del patrimonio histórico, M.C. Pizarro Moreno. 2004

²Geoestadística teoría y aplicación, Ramón Giraldo Henao. 2002.

³Isaaks&Srivistava. 1989.

Una definición formal sería que la geoestadística interpreta cada valor $Z(x)$ como una realización particular de una variable aleatoria y el conjunto de éstas dentro del dominio D constituye una función aleatoria $\{Z(x); x \in D\}$ ⁴.

Un concepto clave en la comprensión y el análisis de un fenómeno espacial es que el conjunto de observaciones presente dependencia espacial, es decir una observación está relacionada con todas las demás, pero las observaciones cercanas están más relacionadas de las observaciones lejanas⁵. Generalmente la variación espacial de los fenómenos es muy compleja para ser descrita por medio de modelos matemáticos sencillos y aunque pareciera que es completamente aleatoria, estos fenómenos presentan un grado de correlación espacial. Una forma de describir la continuidad espacial es mediante la herramienta del semivariograma, que es una función que mide el grado de similitud de una variable en función de la distancia de separación⁶.

3 - Utilidad de la geoestadística y los SIG: El desarrollo de los sistemas de información geográfica (SIG) y la aplicación de la geoestadística ha supuesto un nuevo impulso al análisis espacial aplicado a la ecología y a las ciencias ambientales⁷. Analizar los patrones de distribución espacial a variables medibles en el espacio a partir de muestreos realizados en el área de interés ha permitido la realización de las "mejores interpolaciones" en aquellos lugares donde no se conoce la magnitud del atributo investigado. Son muchas las disciplinas en las que interesa estudiar algún tipo de propiedad que tiene cierta característica espacial. En estas disciplinas la metodología espacial puede ser de ayuda en alguna o en muchas etapas del estudio, desde el diseño inicial del muestreo hasta la representación final de los resultados obtenidos.

⁴Metodologías para la estimación del semivariograma de un proceso espacial, Ángela M. Diblasi. 2002.

⁵Tobler Waldo, 1999.

⁶Carr Clark, 1979.

⁷Taylor, 1984.

Metodología

El objetivo del Diagnostico Espacial de los Accidentes de Tránsito en el municipio de San Salvador es conocer dónde y bajo que patrones se presentan los accidentes de tránsito, y de esta manera renovar las políticas orientadas a mejorar la seguridad y bienestar de la población. Para poder alcanzar los objetivos, tomaremos en cuenta los aspectos que a continuación especificamos:

a) Revisión bibliográfica: Buscar la más completa información en libros, documentos, revistas, internet, etc., sobre la geoestadística y aplicaciones. Así como también para el desarrollo del estudio es necesaria toda la información referente al tema, para poder tener los fundamentos teóricos de nuestra investigación.

b) Información disponible: Identificar instituciones que generan información estadística y cartográfica relacionada con los accidentes de tránsito del municipio de San Salvador. Esta etapa es clave para el desarrollo exitoso de nuestro trabajo de investigación y consiste en investigar sobre qué datos están disponibles y de esa manera ver hasta qué punto es factible el estudio (delimitarlo en espacio y tiempo). La recopilación de la información que se necesita juega un papel primordial, pues el contar con información adecuada es lo que nos permitirá obtener buenos resultados. Es importante el contacto directo con las instituciones proveedoras de información para la investigación, en nuestro caso: la Sub Dirección de Tránsito Terrestre de la PNC institución que nos proporcionara en conjunto un total de 18 variables registradas en el periodo 2006 - 2010.

Las 18 variables listadas a continuación:

1	Edad	10	Departamento
2	Tipo daño	11	Municipio
3	Resultado	12	Causa
4	Sexo	13	Naturaleza
5	Tipo	14	Dirección
6	Fecha	15	Distrito
7	Mes	16	Región
8	Hora	17	Fecha fallecido
9	Día	18	Fallecido lugar

c) Cobertura geográfica y temporal: La investigación se realizara a nivel del municipio de San Salvador en los 7 distritos y procesando los accidentes de tránsito registrados en el periodo comprendido entre enero de 2006 a diciembre de 2010.

d) Aplicación y análisis de datos: Generalmente los trabajos geoestadísticos se llevan a cabo en tres etapas:

1. Análisis exploratorio de los datos: Se estudian los datos, sin tener en cuenta su distribución geográfica. Sería una etapa de aplicación de la estadística clásica. Se comprueba la consistencia de los datos, eliminándose los erróneos e identificándose las distribuciones de las que provienen, indicadores de tendencia, frecuencia, tasas de la accidentalidad vial y realizar una depuración de variables, escogiendo las más pertinentes a la descripción del fenómeno.
 2. Análisis estructural: Construcción de modelos para identificar el tipo de distribución espacial que caracteriza a los accidentes de tránsito. Se realizan pruebas de aleatoriedad y se calcula la función semivariograma explique la variabilidad espacial del fenómeno.
 3. Estimación espacial: En base a la información que proporcione el semivariograma se aplica el krigeado que es un método geoestadístico de estimación de puntos. Esta técnica de interpolación se basa en la premisa que la estructura de correlación espacial seleccionada es continua en toda el área de estudio. La importancia en la aplicación de esta técnica es el obtener estimaciones de la variable en los puntos no muestrales, considerando el margen de error entre los puntos muestrales y los estimados.
- e) Publicación de resultados:** Integración y redacción del reporte, que permite conocer la distribución espacial y caracterización de los accidentes de tránsito.

1**Base teórica y metodológica de la
geoestadística.**

Introducción:

La estadística aplicada a la geografía ha experimentado un avance espectacular en las últimas dos décadas, introduciéndose el concepto de “espacio” como eje fundamental del análisis. Este avance se ha visto acompañado por un rápido desarrollo de aplicaciones estadísticas integradas en los sistemas de información geográfica, constituyéndose de esta forma en un conjunto de herramientas imprescindibles en la planificación territorial.

En un estudio espacial son utilizados diversos procedimientos geoestadísticos de estimación y/o simulación. Esto es, a partir de un conjunto de muestras tomadas en localizaciones de la región donde se manifiesta el fenómeno a estudiar y consideradas representativas de su realidad, estos procedimientos permiten la descripción o caracterización de las variables con dos fines ligados entre sí, primero, proporcionar valores estimados en localizaciones de interés donde no se tiene información alguna y segundo generar valores que en conjunto presenten iguales características de dispersión que los datos originales. La geología y la minería es el campo típico para la aplicación de estos modelos, campo en el que surge y se desarrolla la geoestadística como ciencia aplicada y que posteriormente es exportada a todo tipo de estudio ambiental, económico, social, salud y monitoreo de imágenes satelitales.

Este capítulo es un primer acercamiento a los procedimientos que ofrece la geoestadística lineal para realizar estimaciones en estudio de fenómeno de contexto espacial. Contiene tres partes bien diferenciadas: en la primera se muestra la definición y conceptualización de la estadística espacial como un área de la estadística general que utiliza procedimientos inferenciales con observaciones georreferenciadas; la segunda se dedica a la revisión de la teoría geoestadística, como una técnica alternativa para la inferencia de fenómenos de tipo espacial.

La tercera parte explica en cómo se debe abordar un estudio espacial, que como cualquier estudio estadístico parte del análisis exploratorio de los datos utilizando técnicas estadísticas clásicas como las medidas de tendencia central y dispersión. Una siguiente fase sería la captación de la estructura de dependencia espacial en el área donde se presenta el

fenómeno, a través de los estadísticos vecino más próximo, K de Ripley, test I de Moran, test C de Geary y la construcción de los semivariogramas adecuados a la variabilidad natural de los datos. Por último aplicar los procedimientos de estimación kriging aptos de utilización en función de las características particulares del problema objeto de estudio.

En resumen, este capítulo pretende ser una guía tanto teórica como práctica, para todos aquellos que pretendan introducirse al nuevo mundo de la geoestadística.

La gran relevancia que tiene actualmente a nivel mundial el tema ambiental ha hecho que los profesionales en estadística encaminen esfuerzos en el desarrollo de nuevas técnicas apropiadas para el análisis de información enmarcada dentro de este contexto. Como consecuencia de este impulso surgió una nueva rama de la estadística denominada “estadística espacial”.

Son muchas las disciplinas en las que interesa estudiar algún fenómeno que tenga una propiedad espacial. En estas disciplinas la metodología puede ser de ayuda en alguna o en muchas etapas del estudio, desde el diseño inicial del muestreo hasta la representación final de los resultados obtenidos. Aunque la geoestadística ha supuesto una mejora considerable en la representación gráfica aplicada a las ciencias, para conseguir buenos resultados es preciso conocer sus fundamentos sobre las que se basa.

1.1 Límites de la estadística clásica

La razón principal por lo que no se aplica la estadística clásica a estudios de tipo espacial, es que esta considera los datos como realizaciones independientes de una misma variable aleatoria, es decir, se supone que no tienen relaciones entre sí y que siguen la misma distribución de probabilidad. Sin embargo, cuando los datos están ubicados en el espacio geográfico, las hipótesis de la estadística clásica son rara vez aceptables en particular, si bien simplifica los cálculos estadísticos, la hipótesis de independencia de las observaciones resulta poco realista en el marco espacial. Intuitivamente, observaciones próximas tienen valores semejantes, mientras que aquellas que están más alejadas entre sí tienen una menor relación entre ellas.

Así, en general, no puede considerarse modelar las variables aleatorias en el espacio por medio de funciones determinísticas, debido a su extrema complejidad y tampoco se puede asimilar que los datos medidos pertenecen a variables aleatorias independientes. La estadística espacial ofrece técnicas que entregan una representación intermedia (a la vez aleatoria y funcional), al basarse en el concepto de función aleatoria regionalizada.

1.2 Estadística espacial

Una definición formal de la estadística espacial es que trata con el análisis de una serie de realizaciones u observaciones de un proceso estocástico $\{Z(x); x \in D\}$, en el que $x \in R^d$ y representa una ubicación en el espacio d-dimensional, $Z(x)$ es una variable aleatoria en la ubicación x y x varía sobre un conjunto de índices $D \subset R^d$.⁸

Tomada en su sentido metodológico más amplio, este término engloba a todo análisis que utiliza la herramienta estadística a partir de una serie de mediciones con dimensión espacial, ya sea que esta dimensión se refiera a la naturaleza del fenómeno objeto de estudio o solo como objeto espacial para la interpretación del análisis de las variables.

⁸Metodologías para la estimación del semivariograma de un proceso espacial, Ángela M. Diblasi. 2002.

La geoestadística considera:

- a) **Objetos localizados por naturaleza:** Se refiere al estudio de hechos y fenómenos físicos, biológicos y humanos de la superficie terrestre, es decir trata con estudios puramente en el contexto geográfico. Por ejemplo variables que permiten caracterizar la organización espacial de un entorno: la forma y la densidad de un semillero de puntos, (hábitats, explotaciones agrícolas, temperatura) o el grado de heterogeneidad de la utilización del suelo.
- b) **Sólo como objeto espacial:** Se estudian entidades espaciales localizadas (regiones, ciudades, comunas, parcelas), pero esta localización no desempeña un papel “activo” en el análisis estadístico. Estas entidades son en general descritas por variables relativas a la demografía, la actividad de la sociedad, la utilización comercial, etc., que no son específicamente espaciales. Estas entidades y esas variables cualitativas y cuantitativas pueden ser eficazmente analizadas por métodos estadísticos clásicos como la correlación, la regresión, el análisis de datos descriptivos, etc. El carácter espacial de las entidades interviene entonces solamente en la interpretación de los resultados, al final del análisis estadístico. El espacio juega un simple rol de “soporte”.

Para analizar objetos localizados existen técnicas estadísticas específicas, una de las más utilizada es la geoestadística, que da cuenta globalmente de la tendencia de los lugares próximos a reunirse (autocorrelación) o, por el contrario que vinculan la dispersión de una variable y la distancia a un centro dado, esto permite poner en evidencia las discontinuidades en la distribución espacial del fenómeno estudiado. Otros métodos no toman en cuenta la estructura espacial del fenómeno en estudio y al no tener en cuenta la autocorrelación en los análisis estadísticos se puede cometer el error de introducir sesgos en la estimación de los parámetros.

1.2.1 Técnicas estadísticas espaciales

Para analizar objetos localizados existen técnicas estadísticas específicas, la aplicación de estas depende de las características que presenten el conjunto de datos asociadas a la región de estudio del proceso estocástico de interés. A continuación se mencionan dichas técnicas y se describen las propiedades del dominio (campo de estudio) para cada una de éstas.

- a) **Geoestadística:** Las ubicaciones provienen de un dominio continuo (*D continuo*), estas observaciones son seleccionadas mediante una muestra. Es importante resaltar que en la geoestadística es primordial la continuidad espacial, si no hay continuidad espacial pueden hacerse predicciones carentes de sentido y no cumplir con el propósito esencial de la interpolación.
- b) **Lattices: (enmallados):** Las ubicaciones pertenecen a un dominio discreto y son seleccionadas por el investigador (*D fijo*). Estas pueden estar regular o irregularmente

espaciadas. Básicamente las ubicaciones de interés deben de corresponder a agrupaciones, más que a un conjunto de puntos del espacio.

- c) Patrones Espaciales: Las ubicaciones pertenecen a un dominio que puede ser discreto o continuo y su selección no depende del investigador (*D aleatorio*). Una vez se ha hecho la selección de sitios (muestra) es posible hacer medidas de variables aleatorias en cada uno de ellos, en general el propósito de análisis en estos casos es el de determinar si la distribución de los puntos dentro de la región es aleatoria, agregada o uniforme.

1.2.2 Datos espaciales

Los datos espaciales o geográficos se definen como cualquier información o característica (posición, atributo o forma) de objetos o fenómenos que tengan una ubicación bien definida sobre un sistema de coordenadas (x, y) con respecto a la superficie de la tierra. Las localizaciones pueden ser puntos o regiones referenciadas, normalmente este tipo de información puede ser expresada mediante tres tipos de objetos espaciales o figuras geométrica:

- a) Punto: Cuando representan objetos o eventos cuya dimensión es insignificante o nula para determinada escala. Se encuentran determinados por las coordenadas terrestres medidas por latitud y longitud. Por ejemplo, casas, cajero de un banco, estación de autobuses, etc.
- b) Líneas: Objetos abiertos que cubren una distancia dada y comunican varios puntos o nodos, aunque debido a la forma esférica de la tierra también se le consideran como arcos. Por ejemplo carreteras, ríos, vías de trenes, etc.
- c) Polígonos: Son datos que describen objetos o sucesos que existen como superficies o áreas. Suelen representarse como figuras planas conectadas por distintas líneas u objetos cerrados que cubren un área determinada. Ejemplo de este tipo de datos son países, departamentos, parcelas o lagos, etc.

1.2.3 Problemas de la estadística espacial

La mayor parte de las técnicas estadísticas espaciales suponen varios problemas cuando intentamos aplicarlas a variables distribuidas en el espacio, generalmente estas son:

- a) Uno de los criterios básicos de la estadística convencional es la independencia de los elementos que componen una muestra unos de otros. Al trabajar con datos espaciales este precepto no se cumple debido a la autocorrelación espacial. El resultado puede ser que los estadísticos resultantes aparezcan sesgados hacia los valores predominantes en los puntos muestreados.
- b) Efectos de escala, los resultados pueden variar en función de la escala espacial con que se midan las variables. Por ejemplo la densidad de población varía si la medimos a escala 1:200000 (todo el poblamiento aparece agregado) o a escala 1:25000 (se tiene en cuenta la dispersión del poblamiento).

- c) Efectos de borde, muchas técnicas de análisis espacial fallan cuando aparecen fronteras al otro lado de las cuales los fenómenos son diferentes o no aparecen.

1.3 Conceptos básicos de la geoestadística

Para poder “describir” y “entender” un fenómeno regionalizado, es necesario elaborar una representación matemática, una primera solución consiste en utilizar un modelo determinístico. En general, este enfoque conduce a una evaluación precisa de la variable regionalizada a partir de un número limitado de observaciones, pero requiere conocer la raíz del fenómeno y las leyes físicas o ecuaciones matemáticas que rigen la distribución de la variable regionalizada.

No obstante en general, los fenómenos regionalizados en estudio son extremadamente complejos y su comprensión puede ser tan parcial que un modelamiento determinístico es imposible o ilusorio. Estamos entonces obligados a renunciar a una descripción determinística del fenómeno y recurrir a un modelo probabilístico. Este proceder resulta operatorio, pues permite formalizar tanto los conocimientos como las incertidumbres que se tiene del fenómeno regionalizado.

1.3.1 Definición de geoestadística

La Geoestadística es una técnica estadística que se especializa en el análisis y la modelación de la variabilidad espacial en ciencias de la tierra. Su objeto de estudio es la predicción de fenómenos en espacio y/o tiempo⁹. Dichos fenómenos están caracterizados por la distribución espacial de una o más variables, que denominamos como variables aleatorias regionalizadas.

Si $Z(x)$ es el valor de una variable en un punto x , intentamos representar el comportamiento y variabilidad de esa variable en el espacio de interés o dominio, partiendo del hecho de que la realidad nos muestra que existe una similitud estadística, una correlación entre pares de valores $\{Z(x_i), Z(x_i) + h\}$ al menos cuando la distancia h entre ellos no supera un cierto valor. El interés en representar dicha variabilidad surge porque nos permite hacer estimaciones del valor de la variable en los puntos que nos interesen dentro del dominio.

La geoestadística interpreta cada valor $Z(x)$ como una realización particular de una variable aleatoria y el conjunto de éstas dentro del dominio D constituye una función aleatoria $\{Z(x); x \in D\}$.

1.3.2 Variables aleatorias regionalizadas

El objeto sobre el cual trabajaremos será una descripción matemática del fenómeno regionalizado, por medio de una o varias funciones numéricas llamadas “variables

⁹Variables regionalizadas y su estimación, George Matheron. 1962.

aleatorias regionalizadas”, que describen el comportamiento de ciertas propiedades o atributos relacionados al fenómeno de estudio.

Una definición más rigurosa matemáticamente equivalente consistiría en decir que una variable regionalizada es una variable aleatoria definida en un punto x del espacio R^d $\{Z(x); x \in R^d\}$.¹⁰

Si a cada punto x que pertenece a un dominio en el espacio le hacemos corresponder una variable aleatoria $Z(x)$, que en sentido general pueden ser dependientes, entonces el conjunto de variables aleatorias espacialmente distribuidas $\{Z(x_i); x_i \in \Omega, i = 1, 2, \dots, n\}$ será una función aleatoria. Las variables regionalizadas que se presentan a observación, poseen características cualitativas, ligadas estrechamente a la estructura natural del fenómeno. Entre estas características importantes que la estadística clásica es incapaz de expresar y que deben obligatoriamente ser tomadas en cuenta por la teoría de las variables regionalizadas, están:

- a) Localización: Una variable regionalizada no toma sus valores en cualquier lugar, sino más bien, en una región bien determinada del espacio, que se llama campo geométrico o dominio.
- b) Continuidad: Una segunda característica esencial es el grado de mayor o menor relación espacial en una región. El modelo de relación espacial más frecuente es la continuidad, que es una forma sencilla y clara de identificar la dependencia entre los espacios y que éstos respondan de forma idónea a sus exigencias.
- c) Anisotropía: En tercer lugar, una regionalización puede ser anisótropa, es decir, cuando el fenómeno en estudio carece de una estructura interna bien definida. Por ejemplo puede existir una dirección privilegiada a lo largo de la cual los valores se modifican lentamente, mientras que éstos varían mucho más rápido en la dirección opuesta o perpendicular. Esta característica en general está asociada a la naturaleza de fenómenos en estudio.
- d) Fenómeno de transición: Otros tipos de estructuras se pueden manifestar ligados a la aparición en el campo geométrico de la variable de una red de discontinuidades, se habla entonces de una manera general de fenómenos de transición. Puede suceder que la variable la cual sea constante o casi constante al interior de los estratos tenga cambios bruscos al pasar de una localización a otra.
- e) Momentos de primer orden: Si la función de distribución de $Z(x_i)$ tiene una media definida, será una función de la localización x_i .

$$m(x_i) = E[Z(x_i)]$$

¹⁰Metodologías para la estimación del semivariograma de un proceso espacial, Ángela M. Diblasi. 2002.

- f) Momento de segundo orden: Si la varianza de $Z(x_i)$ existe, entonces se define como el momento de segundo orden y será también una función de la localización x_i .

$$Var[Z(x_i)] = E[Z(x_i) - m(x_i)]^2$$

Si la varianza de las variables $Z(x_i)$ y $Z(x_j)$ existe entonces la covarianza de éstas también existe y es función de las localizaciones x_i y x_j .

$$Cov[Z(x_i), Z(x_j)] = E[Z(x_i) - m(x_i)][Z(x_j) - m(x_j)]$$

Si $x_i = x_j$ entonces:

$$Cov[Z(x_i), Z(x_j)] = VarZ(x_i)$$

1.3.3 Hipótesis básicas de la geoestadística

Como la forma en que se presenta la información es muy diversa, la geoestadística se construye asumiendo condiciones de estacionariedad. Por lo que es necesario aceptar el cumplimiento de ciertas hipótesis sobre el carácter de la función aleatoria o procesos estocásticos estudiados, estas hipótesis son la estacionariedad estricta, la estacionariedad de segundo orden, la hipótesis intrínseca y los procesos cuasiestacionarios¹¹.

- a) Estacionariedad estricta: En la práctica, un caso particular importante se presenta cuando una distribución de dimensiones finitas permanece invariante para cualquier translación arbitraria de los puntos por medio de un vector h o lo que es lo mismo, si la función de distribución del vector aleatorio

$$Z = [Z(x_1), Z(x_2), \dots, Z(x_n)]$$

Es idéntica a la del vector:

$$Z = [Z(x_1 + h), Z(x_2 + h), \dots, Z(x_n + h)]$$

Tal función aleatoria se llama estacionariedad estricta. Físicamente, esto significa que el fenómeno es homogéneo en el espacio y por decirlo así se repite a sí mismo en todo el espacio.

- b) Estacionariedad de segundo orden: Cuando una función aleatoria es estacionaria, sus momentos, si existen son invariantes ante cualquier translación. Si consideramos sólo los momentos de primer orden para los puntos x y $x + h$ de R^n tenemos

$$E[Z(x)] = m$$

$$E[(Z(x) - m)(Z(x + h) - m)] = C(h)$$

¹¹Teoría geoestadística, Journel y Huijbregts 1978

En la primera ecuación tenemos la expresión de la media que es constante y en la segunda, la de la covarianza que depende solo de h . Por definición, una función que satisfaga las condiciones anteriores, es una función débilmente estacionaria o estacionaria de segundo orden en el sentido amplio. En general no se cuenta con los elementos como para hacer de está una afirmación más fuerte.

- c) La hipótesis intrínseca: Una hipótesis más suave consiste en suponer que para cada vector h el incremento $\gamma(x) = Z(x+h) - Z(x)$, es una función aleatoria estacionaria en x . En ese caso, una función aleatoria $Z(x)$, se dice intrínseca cuando su esperanza matemática existe y no depende de la localización x .

$$E[Z(x)] = m, \forall x$$

Para todo vector h el incremento $[Z(x+h) - Z(x)]$ tiene varianza finita y no depende de la localización x :

$$Var[Z(x+h) - Z(x)] = E[Z(x+h) - Z(x)]^2 = 2\gamma(h); \forall x.$$

Cuando se cumple esta condición se dice que la función aleatoria $Z(x)$ es homogénea. Esta condición, se encuentra con bastante frecuencia en la naturaleza, pues existen muchos procesos que no tiene varianza finita y sin embargo, poseen una función de semivariograma finita. La estacionariedad de segundo orden, siempre implica la condición intrínseca (homogeneidad), sin embargo la relación inversa no siempre se cumple.

- d) Procesos cuasiestacionarios: En la práctica la función estructural, covarianza o semivariograma, es sólo usada por límites $h = b$. El límite b , representa la extensión de la región en la que el fenómeno estudiado conserva cierta homogeneidad del comportamiento de $Z(x)$. En otros casos, b pudiera ser la magnitud de una zona homogénea donde dos variables $Z(x)$ y $Z(x+h)$ puedan ser consideradas dependientes en la misma zona de homogenización si $h > b$.

En tales casos, podemos, y verdaderamente debemos, estar satisfecho con una función estructural $\gamma(x, x+h)$, lo que no es más que estacionaridad local (para distancias h menores que el límite b).

En la práctica son dos las hipótesis que más se presentan: La estacionariedad de segundo orden y la hipótesis intrínseca¹². Estas condiciones de estacionariedad se asumen en el desarrollo teórico, en la práctica deben ser verificadas en los datos antes de comenzar un estudio geoestadístico, para lo que se puede realizar un análisis exploratorio de la información, de modo que se refleje así, el grado de confiabilidad en la aplicación de estos métodos.

¹²Elementos de la geoestadística, Armstrong y Carignan. 1997.

1.3.4 Cálculos geométricos básicos

Prácticamente todos los cálculos que se llevan a cabo en un sistema de georeferenciación se basan en la posición, distancia y las relaciones topológicas entre objetos, se definen a continuación conceptos básicos a la distancia entre puntos.

- a) Distancia entre dos puntos, en casi todos los procedimientos de análisis espacial se incluye este concepto. Aunque se han definido varios tipos de distancia, se trabaja fundamentalmente con la distancia euclidiana por ser la más apropiada para la realidad espacial. La distancia entre los puntos i y j es también la longitud del segmento recto entre los puntos i y j .

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

- b) Área del trapecio situado bajo un segmento, se resuelve cómo la suma de las áreas del triángulo y el rectángulo bajo el segmento.

$$At_{i,i+1} = (x_{i+1} - x_i)y_i + \frac{(x_{i+1} - x_i)(y_{i+1} - y_i)}{2} = (x_{i+1} - x_i) \frac{(y_{i+1} - y_i)}{2}$$

Hay que tener en cuenta que el resultado será positivo si $x_{i+1} > x_i$ y negativo en caso contrario. En este último caso el valor correcto es el valor absoluto de $At_{i,i+1}$.

- c) Punto de corte de dos segmentos, es aquel que resulta de resolver el sistema de que forman las ecuaciones de la recta de ambos segmentos.

$$Y_a = AX_a + B$$

$$Y_b = CX_b + D$$

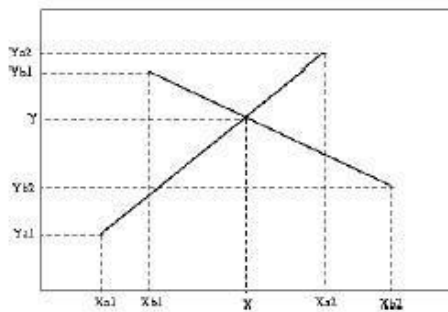


Figura 1.1 Punto de cruce entre dos segmentos

Dónde:

$$A = \frac{Y_{a2} - Y_{a1}}{X_{a2} - X_{a1}}, B = Y_{a1}$$

$$C = \frac{Y_{b_2} - Y_{b_1}}{X_{b_2} - X_{b_1}}, D = Y_{b_1}$$

Si el sistema tiene solución y el punto resultado forma parte de ambos segmentos, entonces la solución nos da el punto de corte (figura 1.1). En realidad el sistema tiene solución siempre, salvo que los segmentos sean paralelos, por lo que es necesario comprobar que la solución forme parte de ambos segmentos a la vez y sea, por tanto, su punto de corte. Es decir, hay que comprobar que "X" esté entre (X_{a_1}, X_{a_2}) y entre (X_{b_1}, X_{b_2}) y que "Y" esté entre (Y_{a_1}, Y_{a_2}) y entre (Y_{b_1}, Y_{b_2}) .

- d) Longitud (L) de una línea, es la suma de las distancias entre cada par de vértices consecutivos de la misma, asumiendo que la línea tiene n vértices:

$$\sum_{i=1}^{n-1} d_{i,i+1}$$

- e) Distancia (D) en línea recta entre el inicio y el final de una línea, es decir entre su nodo inicial y su nodo final e índice de sinuosidad (I_s), viene dado por:

$$D = d_{1,N}$$

$$I_s = \frac{L}{D}$$

- f) Longitud máxima del polígono puede definirse como la distancia máxima entre cualquier par de vértices en la línea que define el contorno del perímetro:

$$L_m = \max(d_{i,j})$$

- g) Área de un polígono es el valor absoluto de la suma de todas las áreas de los trapecios bajo los segmentos que forman el polígono:

$$A = \sum_{i=1}^{n-1} At_{i,i+1}$$

Dónde:

$At_{i,i+1}$: Es el área bajo el segmento formado por los vértices i e $i + 1$.

1.4 Manejo de un estudio espacial

Para realizar un estudio espacial es importante considerar tres fases importantes como lo son: el análisis exploratorio de los datos, análisis estructural y la estimación para el análisis geostatístico; el primero, se refiere al análisis de los datos muestrales sin tener en cuenta su distribución geográfica. Sería una etapa de aplicación de la estadística clásica, se

comprueba la consistencia y calidad de los datos, eliminándose los erróneos e identificándose las distribuciones de las que provienen.

El segundo análisis es el estructural, el cual consiste determinar la distribución que tiene el fenómeno en el área de estudio en base a pruebas de ocurrencia de los sucesos entre áreas vecinas y a la construcción de modelos autorizados para el semivariograma, herramienta encargada de explicar la variabilidad del fenómeno con respecto a la distancia.

Por último, la estimación o predicción del fenómeno en zonas donde no se cuenta con información y que se obtienen la técnica de interpolación denominado kriging.

1.4.1 Análisis exploratorio de datos espaciales

El análisis exploratorio de datos sugiere que previo a cualquier análisis estadístico es necesario realizar un examen cualitativo de los datos, hay que comprender y reflexionar sobre la información que ellos contienen.

Previo al uso de métodos geoestadísticos propiamente como tal, conviene realizar un estudio exploratorio de los datos disponibles, el cual tiene tres objetivos:

- a) Analizar mediante herramientas estadísticas simples la cantidad, la calidad y la ubicación de los datos disponibles.
- b) Definir la(s) zona(s) de estudio. Una división del campo en varias sub-zonas puede ser relevante si uno observa cambios abruptos en la distribución espacial de valores, o si el fenómeno lo indica.
- c) Anticipar dificultades o problemas que puedan surgir en la fase de estimación local (por ejemplo, presencia de valores atípicos que se destacan de aquellos de los datos vecinos).

La estadística descriptiva clásica se ocupa de describir los datos a través de gráficos y de algunas medidas de tendencia central y de dispersión, además debe considerarse el análisis exploratorio espacial, que trata sobre el análisis de la nube de puntos en el espacio por tratarse de estudios de tipo geográfico.

1.4.1.1 Análisis exploratorio clásico

Con el objetivo de conocer la información disponible, se puede hacer un análisis de la estadística descriptiva, para ello, existen varias herramientas, siendo la más conocida:

El Histograma: El histograma es una herramienta útil para detectar valores atípicos, representa gráficamente las frecuencias de ocurrencia en función del valor. Para ello es necesario dividir el rango de los valores en intervalos (generalmente, con el mismo ancho) y visualizar la proporción de datos que caben dentro de cada intervalo.

Ahora, cabe destacar que un dato atípico no es forzosamente falso y nunca debe ser eliminado sin razón, por ello es necesario investigar la presencia de este. Junto con el

histograma, es conveniente calcular algunas estadísticas básicas sobre la distribución de valores, entre ellas, podemos distinguir:

Medidas de posición: Estas medidas nos facilitan información sobre la serie de datos que estamos analizando. La descripción de un conjunto de datos, incluye como un elemento de importancia la ubicación de éstos dentro de un contexto de valores posible.

- a) Media aritmética: La media se define como la suma de todos los valores observados, dividido por el número total de observaciones.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- b) Mediana: Es el valor que se encuentra en el centro de los datos, es decir, nos permite conocer el valor que se encuentra exactamente en la mitad del conjunto de datos después que las observaciones se han ubicado en serie ordenada. Esta medida nos indica que la mitad de los datos (50%) se encuentran por debajo de este valor y la otra mitad (50%) por encima del mismo. Designando la mediana como M_c , y distinguimos dos casos:

- 1- Si n es impar la mediana es el valor que ocupa la posición $(n + 1)/2$ una vez que los datos han sido ordenados (en orden creciente o decreciente), es el valor central:

$$M_c = X_{(n+1)/2}$$

- 2- Si n es par, la mediana es la media aritmética de los valores centrales:

$$M_c = \frac{(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})}{2}$$

- c) Moda: Es el valor que más veces se repite dentro de los datos.
- d) Cuartiles: Grupos que dividen la distribución en partes iguales; entendidas éstas como intervalos que comprenden la misma proporción de valores.

$$Q_i = L_i + \frac{\sum f_i/4 - f_{aa}}{f_i} * I_c$$

- e) Máximo y mínimos: Indican el valor máximo y el valor mínimo de la serie de datos, además de establecer el rango en el cual se distribuyen los valores.

Medidas de dispersión: también llamadas medidas de variabilidad, muestran la variabilidad de una distribución, indicando por medio de un número, si las diferentes puntuaciones de una variable están muy alejadas de la mediana y/o media. Cuanto mayor sea ese valor,

mayor será la variabilidad, cuanto menor sea, más homogénea será a la mediana y/o media. Así se sabe si todos los casos son parecidos o varían mucho entre ellos.

- a) Varianza: Promedio aritmético de la desviación cuadrática entre cada valor y la media. Esta medida cuantifica la dispersión del histograma y se expresa en el cuadrado de la unidad de la variable en estudio.

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{x} - x_i)^2$$

- b) Desviación estándar: Se define como la raíz cuadrada de la varianza. Este valor es una medida (cuadrática) que informa de la media de distancias que tienen los datos respecto de su media aritmética, expresada en las mismas unidades que la variable.

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\bar{x} - x_i)^2}$$

- c) Coeficiente de variación: Razón entre la desviación estándar y la media; mostrando una mejor interpretación porcentual del grado de variabilidad que la desviación típica o estándar.

$$C_v = \frac{\sigma}{|\bar{x}|} * 100$$

1.4.1.2 Análisis exploratorio espacial

Cuando se trabaja con datos espaciales deben considerarse si las observaciones, efectivamente han sido georeferenciadas en el espacio geográfico de interés. El análisis exploratorio espacial puede definirse como el conjunto de técnicas que describen y visualizan las distribuciones espaciales, identifican localizaciones atípicas, descubren esquemas de asociación espacial, agrupamientos o puntos calientes y sugieren estructuras espaciales u otras formas de heterogeneidad espacial.¹³

A continuación, se da una breve descripción de las principales herramientas para el análisis exploratorio espacial.

- a) El despliegue de los datos: Permite visualizar las ubicaciones en el espacio. Cuando se trabaja en el espacio tridimensional, el despliegue puede hacerse con proyecciones (plantas y secciones) o vistas en perspectiva. Los mapas de datos pueden incluir información sobre los valores medidos y así dar una idea preliminar de la distribución y continuidad espacial de la variable en estudio, por ejemplo, utilizando símbolos de forma o tamaño variable o escalas de colores.

¹³Calculo de autocorrelación espacial, LucAnselin. 1995

- b) Determinación de los clúster: Lo más sencillo suele ser, ajustar agrupaciones entre observaciones vecinas. Esto puede darnos un primer acercamiento sobre las posibles zonas con mayor influencia en la región. Suele usarse escala de colores para distinguir los clúster con mayor número de observaciones con los de menor.
- c) Nube de Correlación: Consiste en visualizar los valores de una variable en función de otra, lo cual requiere que ambas variables hayan sido medidas en los mismos sitios. Su utilización nos permite ver la relación par a par o “correlación” de ambas variables, así como detectar datos atípicos (puntos que se alejan del resto de la nube).

1.4.2 Análisis estructural

Es una de las partes más importantes de la geoestadística puesto que se encarga de caracterizar la organización espacial del fenómeno regionalizado en una localidad en función de la variable aleatoria que se estudia.¹⁴

Para analizar esta estructura espacial es importante contrastar diversos supuestos que presentan tales procesos; es decir, dicho análisis consiste en estimar y modelar una función que refleje la correlación espacial de la variable regionalizada a partir de la adopción razonada de la hipótesis más adecuada acerca de su variabilidad.

Por su importancia, es necesario antes de llegar a la conclusión de que el proceso presenta una autocorrelación espacial, analizar el proceso con técnicas estadísticas de distancias.

1.4.2.1 Análisis del patrón espacial

En muchas ocasiones dependiendo de la escala de estudio tales elementos pueden describirse mediante sus coordenadas espaciales (x, y) , generándose así un conjunto de datos que recibe el nombre de patrón espacial de puntos¹⁵. La metodología habitual en el estudio de estas estructuras consiste en realizar un test de aleatoriedad y describir si los eventos dentro de una región de estudio comparten la misma estructura (aleatorio, uniforme o agrupado). En la figura 1.2, se describe dicho patrón:

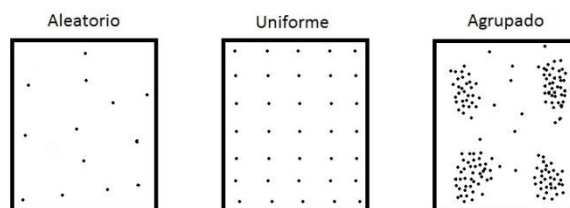


Figura 1.2: Patrón espacial

Entre las técnicas de análisis para inferir sobre la estructura espacial que presenta la nube de puntos del fenómeno regionalizado, están el análisis del vecino más próximo y el K de

¹⁴Geoestadística aplicada: Modelación estadística de fenómenos en espacio y tiempo, Martín Díaz. 2002.

¹⁵Geoestadística aplicada: Modelación estadística de fenómenos en espacio y tiempo, Martín Díaz. 2002.

Ripley. Su función se basa en la ocurrencia de eventos entre espacios próximos y como estos varían entre vecindades. A continuación se definen dichas técnicas:

- a) **Análisis del Vecino Más Próximo (NNA):** Esta es una herramienta esencial para el análisis de distancia. Su objetivo es inferir, por medio de qué patrón espacial tendrá un proceso, es decir, a partir de la hipótesis nula de aleatoriedad contrastar si el proceso presenta una estructura aleatoria, uniforme o agrupada.

El índice del vecino más próximo se define como un cociente de dos medidas que compara la distancia entre los puntos más cercanos y la media distancia aleatoria. La primera medida se expresa de la siguiente manera:

$$\text{Distancia del vecino más próximo} = d(n, n) = \sum_{i=1}^n \frac{\text{Min}(d_{i,j})}{N}$$

Donde $\text{Min}(d_{i,j})$ es la distancia entre cada punto y su vecino más cercano y N es el número de puntos en la distribución. Así, se calcula la distancia de un solo punto a cada otro punto y la distancia más pequeña es seleccionada.

Entonces el proceso a seguir es que se toma el próximo punto y se calcula la distancia a todo el resto de los puntos (incluyendo el primer punto medido) con el vecino más próximo seleccionado y agrega la primera distancia mínima.

Este proceso se repite hasta que todos los puntos hayan tenido su vecino más próximo seleccionado. La suma total de la distancia mínima es entonces dividida por N , el tamaño de la muestra de puntos para producir así una media distancia mínima.

La segunda medida resumen es la media distancia aleatoria (o la media aleatoria de la distancia más próxima). Está definida por:

$$\text{Media distancia aleatoria} = d(\text{ran}) = \frac{1}{2} \sqrt{\frac{A}{N}}$$

Donde A es el área de la región y N es el número de eventos o incidentes. Puesto que A esta definida por el cuadrado de la unidad de medida (ejemplo, miles de cuadrados, miles de kilómetros, etc.), ésta produce una media aleatoria de la distancia en la misma unidad (es decir, miles, metros, etc.).

Así tenemos que el índice del vecino más próximo es el cociente de estas dos medidas, de la distancia vecina más cercana observada y la media distancia aleatoria, decir:

$$\text{Índice del vecino más próximo} = NNI = \frac{d(n, n)}{d(\text{ran})}$$

El resultado de este test determina el patrón espacial de la distribución en base al valor del índice resultante; su interpretación se resume en la tabla 1.1:

<i>Test</i>	<i>Hipótesis</i>	<i>Valor del Índice</i>
<i>Vecino más próximo</i>	<i>Ho: Patrón aleatorio</i>	$NNI = 0$
	<i>Hi: Patrón uniforme</i>	$NNI > 1$
	<i>Hi: Patrón agrupado</i>	$NNI < 1$

Tabla 1.1: Índice del vecino más próximo

El test de significancia del índice del vecino más próximo se hace mediante la prueba de una normal Z como el cociente de la diferencia de la distancia vecina más próxima observada y la media distancia aleatoria entre el error estándar de la media distancia aleatoria, así:

$$SE_{d(ran)} = \sqrt{\frac{(4 - \pi)A}{4\pi N^2}} \cong \frac{0.26136}{\sqrt{\frac{N^2}{A}}}$$

Con A siendo el área de la región y N el número de puntos.

- b) **K de Ripley:** Otro estadístico de análisis de distancia es el llamado K de Ripley. Se trata de un operador que permite establecer el tipo, la intensidad y el rango del patrón espacial a través del análisis de las distancias existentes entre todos los puntos.

La función K es el número esperado de puntos en un círculo de radio r centrado en un punto arbitrario (el cual no es contado) dividido por la intensidad del patrón. Esta función describe características de los procesos de puntos a varias escalas.

Este test, de igual manera que el estadístico del vecino más próximo, busca probar si el proceso presenta una estructura aleatoria, uniforme o agrupada. Para ello, se construye una K teórica de tipo *CSR* (Complete Spatial Randomness) y se compara con la función $K(r)$ de la función observada. El valor teórico esperado de la función K bajo el supuesto de *CSR* es igual a πr^2 . De tal manera que:

<i>Test</i>	<i>Hipótesis</i>	<i>Valor del Índice</i>
<i>K de Ripley</i>	<i>Ho: Patrón aleatorio</i>	$K(r) = \pi r^2$.
	<i>Hi: Patrón uniforme</i>	$K(r) < \pi r^2$
	<i>Hi: Patrón agrupado</i>	$K(r) > \pi r^2$

Tabla 1.2: Índice K de Ripley

El estimador de su función observada $K(r)$ es el promedio, dado por:

$$K(r) = \frac{1}{\lambda N} \sum_{i=1}^n \sum_{i \neq j}^n I(d_{i,j} \leq r)$$

Dónde:

d_{ij} : Es igual a la distancia entre el i-esimo y j-esimo punto.

$$I(d_{i,j} \leq r) = \begin{cases} 1; & \text{si } I(d_{i,j} \leq r) \\ 0; & \text{en otro caso.} \end{cases}$$

$\lambda = \frac{N}{A}$: La intensidad estimada.

1.4.2.2 Análisis de correlación espacial

Los estadísticos definidos anteriormente son esenciales para efectuar el análisis estructural especialmente, en lo que respecta al análisis de distancia. El análisis de correlación tiene que ver también con este tipo de análisis, con la ventaja que por medio de este análisis se conocerá si el proceso espacial presenta autocorrelación o dependencia espacial.

Existen varias técnicas de análisis de correlación espacial, dependiendo del tipo de variable y la dimensión de la variabilidad objeto de investigación. Las más utilizadas para el análisis de variables cuantitativas son aquellas que aplican diferentes test como: el test I de Moran, el test C de Geary, y una especialmente a través del desarrollo de la geoestadística como la técnica del semivariograma. Para analizar la correlación espacial es importante conocer la matriz de pesos espacial (figura 1.3), la cual es una generalización de la matriz de interacciones o contigüidades (conocida habitualmente como matriz de pesos, ponderaciones, retardos o contactos espaciales). La forma más común de incorporación de elementos de la configuración espacial consiste en la utilización de matrices de pesos espaciales.

Dado que en el área de las geociencias se dispone de información de las variables de estudio para cada uno de las divisiones del territorio, tiene sentido incorporar las matrices de pesos. Desde un punto de vista estadístico se incumple la hipótesis de independencia de las observaciones y se debe recoger de algún modo la interdependencia existente. La matriz de pesos espaciales W tiene la siguiente forma:

$$W = \begin{pmatrix} 0 & w_{1,2} & \cdots & w_{1,n} \\ w_{2,1} & 0 & \cdots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,1} & w_{n,2} & \cdots & 0 \end{pmatrix}$$

Figura 1.3: Matriz de pesos.

Donde $w_{i,j}$ refleja la intensidad de la interdependencia entre cada par de observaciones i y j . La matriz es cuadrada, simétrica y debe tener todos sus elementos finitos y no negativos. La manera más sencilla de definir la intensidad de la interdependencia es atribuir el valor $w_{i,j} = 1$, si las unidades son físicamente adyacentes y $w_{i,j} = 0$, si no lo son o son elementos de la diagonal principal. Si definimos así los pesos hablamos de matriz de contactos binarios.

Ahora bien, en el contexto espacial, las cosas no son tan sencillas, dadas las muchas direcciones en que el desfase puede tener lugar. Los siguientes criterios de simple contigüidad binaria, sirven para definir los pesos para las unidades adyacentes a la unidad i .

- a) Criterio de contigüidad torre, los desplazamientos espaciales serían:

$$y_{i-1,j}; y_{i,j-1}; y_{i+1,j}; y_{i,j+1}$$

- b) Criterio de contigüidad del alfil, los desplazamientos serían:

$$y_{i-1,j-1}; y_{i+1,j-1}; y_{i+1,j+1}; y_{i-1,j+1}$$

- c) Criterio de contigüidad reina, que es una combinación de los dos anteriores.

La figura 1.4 describe los criterios de desplazamientos:

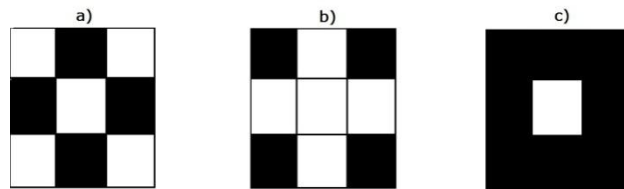


Figura 1.4: Diagrama de retardos espaciales

El propósito es contrastar la hipótesis de que una variable se encuentre distribuida de forma totalmente aleatoria en un sistema espacial o si por el contrario, existe algún tipo de asociación significativa de valores similares o distintos entre regiones vecinas. Los estadísticos que realizan este tipo de contraste son: el test I de Moran y test C de Geary.

- a) Test I de Moran: Este test fue formulado como una función de una variable, considerada en los puntos del espacio (i, j) en desviaciones a la media y los elementos de la matriz binaria de interacciones o peso espaciales w_{ij} . Esta expresión inicial de Moran podría ser generalizada de la siguiente manera:

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

Dónde:

w_{ij} : Es el elemento de la matriz de pesos correspondiente al par (i, j) .

$S_o = \sum_{i=1}^n \sum_{i \neq j}^n w_{ij}$: Es la suma de los pesos espaciales.

\bar{y} : Es el valor medio o esperado de la variable.

N : Número de observaciones o tamaño muestral.

Interpretación: La herramienta de autocorrelación espacial (I de Moran global) es una estadística deductiva, lo que significa que los resultados del análisis siempre se interpretan dentro del contexto de la hipótesis nula. Para la estadística I de Moran global, la hipótesis nula establece que el atributo que se analiza está distribuido en forma aleatoria entre las entidades del área de estudio; es decir, los procesos espaciales que promueven el patrón de valores observado constituyen una opción aleatoria.

Cuando el valor P que devuelve esta herramienta es estadísticamente significativo, puede rechazar la hipótesis nula. La siguiente tabla resume la interpretación de los resultados:

El valor P no es estadísticamente significativo.	No se puede rechazar la hipótesis nula. Es posible que la distribución espacial de los valores de entidades sea el resultado de procesos espaciales aleatorios.
El valor P es estadísticamente significativo y la puntuación z es positiva.	Puede rechazarse la hipótesis nula. La distribución espacial del fenómeno en estudio está agrupada espacialmente.
El valor P es estadísticamente significativo y la puntuación z es negativa.	Puede rechazarse la hipótesis nula. La distribución espacial del fenómeno en estudio está más dispersa espacialmente de lo que se esperaría si los procesos espaciales subyacentes fueran aleatorios. Un patrón espacial disperso suele reflejar algún tipo de proceso competitivo.

Respecto a la distribución del contraste I de Moran, cuando el tamaño de la muestra es suficientemente amplio, la expresión estandarizada del test I de Moran se distribuye como una normal tipificada $N(0,1)$. Es así como el proceso inferencial utiliza valores estandarizados Z de cada uno de ellos obtenidos, a través del cociente entre la diferencia del valor inicial y la media teórica, y la desviación típica teórica, como:

$$Z_I = \frac{I - E[I]}{SD[I]}$$

Dónde:

$E[I]$: Es la media teórica del estadístico I .

$SD[I]$: Desviación típica del estadístico I .

- b) Test de C de Geary: Este test es parecido al test I de Moran, ya que es función de la variable "y" y los elementos de la matriz de interacciones espaciales w_{ij} , puede ser también generalizado de la siguiente manera:

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

Dónde:

w_{ij} : Es el elemento de la matriz de pesos correspondiente al par (i, j) .

$S_o = \sum_{i=1}^n \sum_{i \neq j}^n w_{ij}$: Es la suma de los pesos espaciales.

\bar{y} : Es el valor medio o esperado de la variable.

N : Número de observaciones o tamaño de la muestra.

Interpretación: El valor de C de Geary se encuentra entre los valores 0 y 2. El test C de Geary es inversamente proporcional a I de Moran, pero no es idéntica. I de Moran es una medida de autocorrelación espacial global, mientras que C Geary es más sensible a la autocorrelación espacial local.

Valores entre 0.95 y 1.05 Significa que no hay autocorrelación espacial y que la distribución espacial de los valores de entidades sea el resultado de procesos espaciales aleatorios.

Valores inferiores a 0.95 Demuestran el aumento de autocorrelación espacial positiva y presencia de un patrón agrupado.

Valores mayores a 1.05 Muestran el aumento de autocorrelación espacial negativa, con tendencia al patrón agrupado.

También, en este caso, puede afirmarse que el valor estandarizado de C se distribuye como una normal tipificada, $N(0,1)$ asintótica, lo cual se obtiene a partir de la expresión de los momentos de C de la manera siguiente:

$$Z_c = \frac{C - E[C]}{SD[C]}$$

Dónde:

$E[C]$: Es la media teórica del estadístico C .

$SD[C]$: Desviación típica del estadístico C .

1.4.3 Construcción del semivariograma

Los valores de una variable regionalizada no son independientes, en el sentido que un valor observado en un sitio proporciona información sobre los valores de los sitios vecinos. En la interpretación probabilística de la variable regionalizada, esta noción intuitiva de dependencia está descrita por la distribución espacial de la función aleatoria, que modela la manera como se relacionan los valores observados en distintos sitios por una distribución de probabilidad multivariable.

En muchos problemas la descripción de la distribución espacial se limita a los primeros momentos. El momento de orden 1 (esperanza) hace intervenir un solo sitio a la vez y no entrega realmente información sobre dependencia espacial. En cambio, los momentos de orden 2 (covarianza) están definidos con la ayuda de dos sitios, es decir del más pequeño conjunto que se puede considerar para describir la “interacción” entre valores. Son estos momentos los que entregan una descripción elemental y operatoria de la continuidad espacial de la variable regionalizada.

Con los indicadores descritos anteriormente se ha llegado a un análisis de distancia para conocer la estructura y correlación espacial del proceso. Ahora lo que resta es estudiar la forma en que puede ser ajustado el proceso, es decir, luego de haber inferido en el rechazo de la hipótesis nula de aleatoriedad en el proceso, queda por conocer qué modelo teórico se ajusta a la distribución de los datos espaciales. En ese sentido el estudio se abordará mediante la herramienta denominada semivariograma, una forma de cuantificar la variabilidad que existe en una distribución, midiendo el grado de correlación existente entre los valores de la variable en cada punto y la distancia entre ellos.

El propósito primordial del semivariograma, integrado en las técnicas geoestadísticas, va más allá de la determinación de la correlación, esto porque al cuantificar la relación de una variable medida en una serie de puntos, puede conocer/predecir esa misma variable medida en puntos situados a distancias conocidas, pero que no han sido muestreados. El estudio variográfico está compuesto por el cálculo del semivariograma experimental y el ajuste a este de un modelo teórico conocido¹⁶.

¹⁶Armstrong y Carignan, 1997.

1.4.3.1 Semivariograma experimental

El cálculo del semivariograma experimental es la herramienta geoestadística más importante en la determinación de las características de variabilidad y correlación espacial del fenómeno estudiado, es decir, tener conocimiento de cómo la variable cambia de una localización a otra, el semivariograma representa la herramienta más importante de que dispone el análisis geoestadístico para el estudio de fenómenos en el espacio. Este análisis tiene como condicionantes: la distribución estadística, la existencia de valores aberrantes o anómalos, la presencia de zonas homogéneas o posibles zonaciones en la distribución de las leyes. El semivariograma se define como la media aritmética de todos los cuadrados de las diferencias entre pares de valores experimentales separados una distancia h . La función $\gamma(h)$ se denomina semivariograma y su expresión es la siguiente:

$$\gamma(h) = \frac{1}{2N_p(h)} \sum_{i=1}^n (Z(x_i) - Z(x_i + h))^2$$

Dónde:

$N_p(h)$: Es el número de pares a la distancia h .

h : Es el incremento.

$Z(x_i)$: Son los valores experimentales.

x_i : Son las localizaciones donde son medidos los $Z(x_i)$.

Su cálculo no consiste en una simple evaluación de su expresión, esta operación está relacionada con los elementos siguientes:

- La dirección en la que será calculado el semivariograma, uno o dos ángulos que definen una dirección en el espacio α y/o β con tolerancias angulares $d\alpha$ y/o $d\beta$.
- El incremento o paso en el cálculo del semivariograma h y su tolerancia lineal dh , se recomienda que el valor de dh sea la mitad del incremento inicial.
- Una distancia, que representa la distancia máxima a la que pueden estar alejados los segundos puntos del par con respecto a la línea que define la dirección de cálculo, conocido como ancho de banda.
- Se recomienda que la distancia máxima hasta la cual será calculado el semivariograma sea la mitad de la distancia entre las muestras más alejadas.

Definido los criterios anteriores, se evalúa la expresión del semivariograma para todos los pares de localizaciones separadas a la distancia h que cumplan las siguientes condiciones:

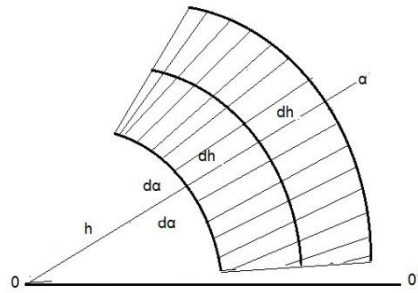


Figura 1.5: Espacio definido por tolerancia lineal

La distancia entre las localizaciones x_i y $x_i + h$ debe ser mayor que $h - dh$ y menor que $h + dh$, o lo que es lo mismo, el segundo punto del par esté incluido en el espacio definido por $h - dh$ y $h + dh$ encontrándose el primer punto del par en el origen 0 (figura 1.6), este origen se mueve entre las muestras a analizar.

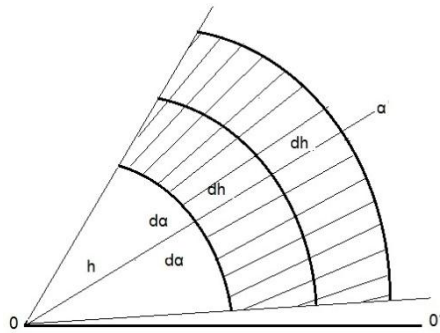


Figura 1.6: Espacio definido por tolerancia lineal y angular.

- e) El ángulo formado entre la línea que une los dos puntos del par y la dirección 0^0 debe estar incluido entre $\alpha - d\alpha$ y $\alpha + d\alpha$.
- f) La distancia entre el segundo punto del par y la línea que define la dirección de cálculo del semivariograma no debe superar el ancho de banda.
- g) Finalmente se representan gráficamente los valores del semivariograma $\gamma(h)$ en función de h .

h)

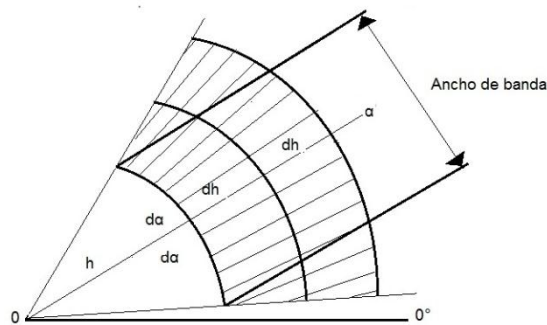


Figura 1.7: Espacio definido por tolerancia lineal

Isotropía y anisotropía: Hasta ahora se considera que la variación del valor de nuestra variable con el espacio es igual en todas las direcciones de éste (semivariograma omnidireccional). Si esto ocurre decimos que la variable tiene un comportamiento isotrópico. Pero no siempre es así, y puede ser que la variación espacial sea diferente en las distintas direcciones del espacio (anisotropía). Si tras una inspección visual sospechamos que puede ocurrir este fenómeno es interesante realizar la construcción de semivariogramas considerando por separado varias direcciones del espacio (semivariogramas direccionales). Tal como se muestra en la figura 1.8 podríamos sólo considerar los pares de puntos separados por distancias en el eje 90° - 270° , 45° - 225° , 0° - 180° y 135° - 315° para realizar cuatro semivariogramas diferentes y observar la posible existencia de anisotropía.

La construcción de semivariogramas anisotrópicos requiere un ángulo de tolerancia, de forma que todos los puntos de la parcela sean incluidos en el estudio.

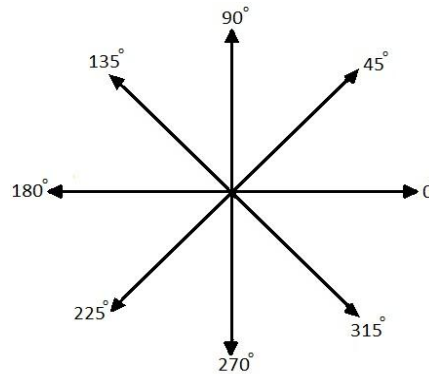


Figura 1.8: Direcciones del semivariograma

Tal como se observa en la figura 1.9, se muestra el modelo gráfico de la función semivariograma tiene una serie de componentes que permiten describir ciertos aspectos de la variabilidad espacial de la población objeto de estudio como la meseta/umbral, efecto pepita y rango.

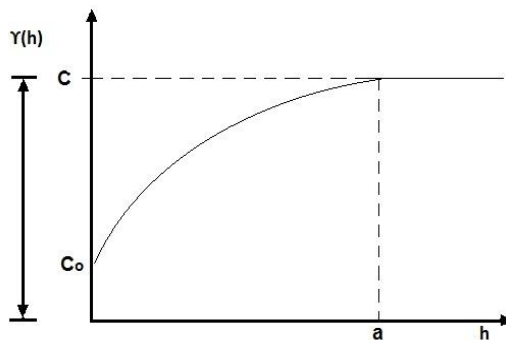


Figura 1.9: Modelo semivariograma experimental

Dónde:

C_o : Efecto pepita.

C : Meseta.

a : Rango.

Y se definen como:

Efecto Pepita: Se denota por C_o y representa una discontinuidad puntual del semivariograma en el origen (Fig. 1.9). Puede ser debido a errores de medición en la variable o a la escala de la misma. En algunas ocasiones puede ser indicativo de que parte de la estructura espacial se concentra a distancias inferiores a las observadas.

Meseta: La meseta se denota por C y es la cota superior del semivariograma. También puede definirse como el límite del semivariograma cuando la distancia h tiende a infinito. La meseta puede ser o no finita. Los semivariogramas que tienen meseta finita cumplen con la hipótesis de estacionariedad fuerte; mientras que cuando ocurre lo contrario, el semivariograma define un fenómeno natural que cumple sólo con la hipótesis intrínseca.

Si se interpreta la pepita como un error en las mediciones, esto explica porque se sugiere que en un modelo que explique bien la realidad, la pepita no debe representar más del 50% de la meseta. Si el ruido espacial en las mediciones explica en mayor proporción la variabilidad que la correlación del fenómeno, las predicciones que se obtengan pueden ser muy imprecisas.

Rango: En términos prácticos corresponde a la distancia a partir de la cual dos observaciones son independientes. El rango se interpreta como la zona de influencia. Existen algunos modelos de semivariograma en los que no existe una distancia finita para la cual dos observaciones sean independientes; por ello se llama rango efectivo a la distancia para la cual el semivariograma alcanza el 95% de la meseta. Entre más pequeño sea el rango, más cerca se está del modelo de independencia espacial.

1.4.3.2 Modelos teóricos del semivariograma

El semivariograma experimental proporciona bastante información del comportamiento espacial de la variable. Sin embargo, es necesario ajustar una función para cuantificar el grado y escala de variación espacial. Existen numerosos modelos que se utilizan en geoestadística, siendo los más comúnmente usados el modelo esférico, el modelo exponencial, el modelo gaussiano y el modelo lineal.

El ajuste a una función permite extraer una serie de parámetros que son los que van a ser usados para la interpolación geoestadística (kriging) y que definen el grado y escala de variación espacial.

Por otra parte, sólo está definido para ciertas distancias y direcciones, por lo que es incompleto. Además, está sujeto a ciertas aproximaciones, debido al número limitado de datos y a los parámetros de tolerancia utilizado en el cálculo.

Para subsanar esta situación, surge la idea de ajustar un modelo teórico del semivariograma en torno al semivariograma experimental. Esta etapa es la fase esencial de todo estudio geoestadístico, pues aquí es donde uno “interpreta” la continuidad espacial de la variable en estudio. Se estudian los distintos modelos teóricos para los ajustes del semivariograma.

1 - Efecto de Pepita: Este modelo de semivariograma alcanza inmediatamente su meseta, hay ausencia total de correlación espacial a cualquier distancia que las separe. Es un caso poco frecuente en las aplicaciones, a menos que los errores de medición sean muy grandes o que las distancias entre datos sean mayores que el alcance real. Su expresión matemática es la siguiente:

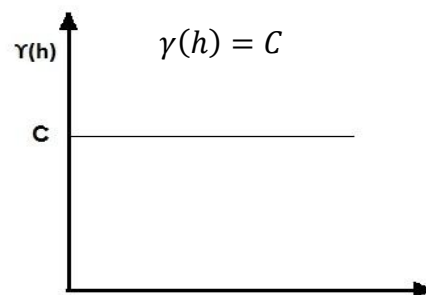


Figura 1.10: Modelo efecto pepita

2 - Modelo Esférico: Este modelo es probablemente el más utilizado. Es una expresión polinomial simple, su forma representada en la figura 1.11, se observa un crecimiento casi lineal y después a cierta distancia finita del origen se alcanza una estabilización, la meseta. Su expresión matemática es la siguiente:

$$\gamma(h) = \begin{cases} C[\frac{3}{2}(\frac{h}{a}) - \frac{1}{2}(\frac{h}{a})^3]; & \text{si } h \leq a. \\ C, & \text{en otro caso.} \end{cases}$$

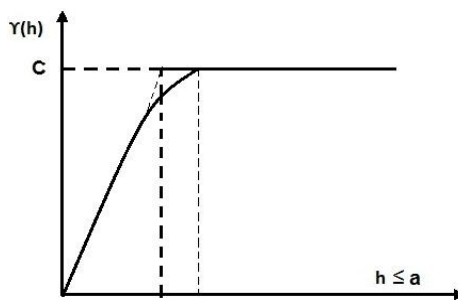


Figura 1.11 Modelo esférico

3 - Modelo Exponencial: Este modelo se aplica cuando la dependencia espacial tiene un crecimiento exponencial respecto a la distancia entre las observaciones. El valor del rango es igual a la distancia para la cual el semivariograma toma un valor igual al 95% de la meseta (Fig. 1.12).

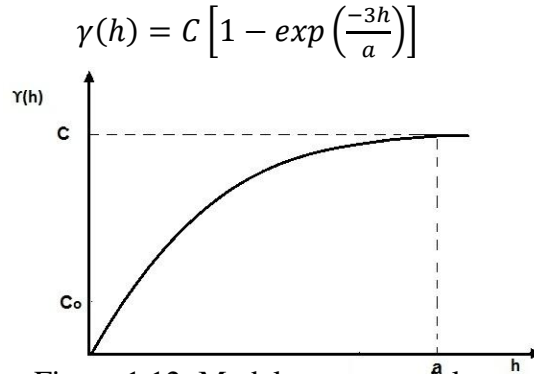


Figura 1.12: Modelo exponencial

4 - Modelo Gaussiano: Este es un modelo extremadamente continuo (figura 1.13), inicialmente presenta un comportamiento parabólico en el origen, después al igual que en el modelo Exponencial se alcanza la meseta de forma asintótica. Su expresión matemática es la siguiente:

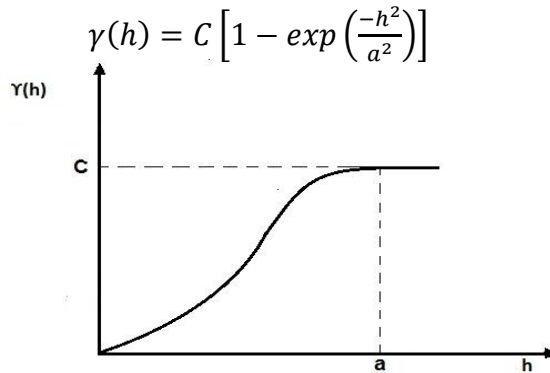


Figura 1.13: Modelo gaussiano

5 - Modelo Lineal: Este es un modelo sin meseta. Muchos programas informáticos denotan la pendiente del modelo lineal con la relación $\frac{C}{a}$. Su expresión matemática es la siguiente:

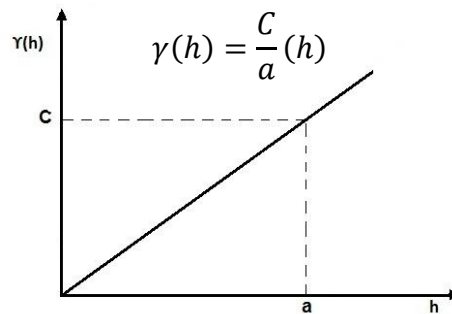


Figura 1.14: Modelo lineal.

La selección del modelo y los parámetros apropiados a las características del semivariograma empírico es de suma importancia para ser usados en la interpolación geoestadística que veremos posteriormente que es el punto más importante en el proceso plantearlo. Al respecto se refieren muchos autores sobre el efecto negativo que puede tener en la estimación el uso del krigeaje sin un estudio de estructura espacial y la selección adecuada del modelo de semivariograma y sus parámetros.

Los criterios para seleccionar un modelo u otro dependen de los objetivos del trabajo. Si el objetivo es encontrar el modelo que mejor se ajuste al semivariograma experimental para cada variable, y no tenemos información a priori, podemos dejar que un determinado software nos ajuste automáticamente el mejor modelo basado, bien en el mínimo de la suma de los cuadrados de los residuales, o bien en el R^2 de la ecuación. Si tenemos información a priori del comportamiento de nuestra variable, puede ser interesante realizar un ajuste manual de los modelos al semivariograma experimental. A veces pequeñas diferencias en los menores residuales o en el R^2 no justifican la elección de un modelo u otro. Si el objetivo del trabajo es comparar los parámetros de los semivariogramas entre distintas variables o cambios en el semivariograma con el tiempo o en el espacio, la utilización de modelos diferentes resulta poco útil.

1.4.4 Modelos de interpolación kriging

La estimación local busca estimar o predecir el valor de la variable regionalizada en un sector del campo estudiado, a partir de los datos disponibles, este es el objetivo de la interpolación espacial¹⁷. En este caso estudiaremos el método de interpolación kriging (bautizado así en honor de uno de los precursores de la geoestadística, Daniel Krige). El kriging es un término designado para el “mejor estimador lineal insesgado” de un punto y al mejor promedio lineal móvil ponderado de un bloque. El objetivo del método de interpolación kriging es simple obtener nuevos valores partiendo del conocimiento de un conjunto de puntos muestrales $\{Z(x_i), i = 1, \dots, n\}$. Con frecuencia estos valores están definidos en soportes puntuales o casi puntuales, en otros casos, son los valores medios $Z(x_i)$, definidos en los soportes centrados en los puntos x_i donde los n soportes pueden ser diferentes.

1.4.4.1 Construcción del kriging

La resolución de un problema de estimación por kriging se articula siempre en torno a una serie de etapas. Las diferentes variantes sólo radican en las hipótesis realizadas sobre la función aleatoria $Z = \{Z(x) \in D\}$ que representa la variable regionalizada. Dentro de la construcción del modelo kriging debemos considerar las siguientes restricciones:

- a) Restricción de Linealidad: El estimador tiene que ser una combinación lineal ponderada (promedio ponderado) de los datos, que denotamos de la siguiente forma:

¹⁷Aplicación de técnicas geoestadísticas en las ciencias ambientales, Asociación Española de Ecología Terrestre, 2004

$$Z(x_0) = a + \sum_{i=1}^n \lambda_i Z(x_i)$$

Donde $Z(x_0)$ es el sitio donde se busca tener una estimación, $\{x_i, i = 1, 2, \dots, n\}$ son los sitios con datos, mientras que $\{\lambda_i = 1, 2, \dots, n\}$ son los ponderadores, $Z(x_i)$ son los valores de la variable en los sitios x_i y el coeficiente a son las incógnitas del problema de kriging. En rigor, se debería escribir $\lambda_i(x_0)$ en lugar de λ_i puesto que los ponderadores de kriging dependerán del sitio x_0 a estimar.

Esta restricción se debe a la decisión de considerar sólo los primeros momentos de las distribuciones de probabilidad (esperanza y varianza). La construcción de estimadores más sofisticados que no sean combinaciones lineales de los datos, requeriría la especificación de la distribución espacial de la función aleatoria más allá de su semivariograma. Esto se puede realizar con métodos de geoestadística no lineal (kriging de indicadores, kriging disyuntivo, kriging lognormal, kriging multi-Gaussiano).

- b) Restricción de Insesgo: Esta etapa consiste en expresar que el error de estimación tiene esperanza nula, es decir:

$$E[Z^*(x_0) - Z(x_0)] = 0$$

Se puede interpretar esta restricción, reemplazando la esperanza matemática por una media en el espacio; si se calcula sobre numerosas configuraciones de kriging idénticas, la media de los errores de estimación cometidos se acerca a cero. La ausencia de sesgo no garantiza que los errores sean bajos, sino sólo que su media global es aproximadamente nula.

- c) Restricción de optimalidad: Al superar las etapas anteriores, el estimador está sometido a una o varias restricciones pero no está totalmente especificado. La última etapa consiste en buscar los ponderadores que minimizan la varianza del error de estimación:

$$Var[Z^*(x_0) - Z(x_0)] = 0, \text{ es mínima.}$$

En términos intuitivos, esta restricción significa que si se calcula sobre numerosas configuraciones de kriging idénticas, la varianza estadística de los errores de estimación cometido es la más baja posible. Este criterio de precisión equivale a la minimización del error cuadrático promedio.

1.4.4.2 Vecindad kriging

En general, los problemas de estimación local no involucran la totalidad del campo y tampoco utilizan todos los datos disponibles. Se define la vecindad kriging, como el dominio del espacio que contiene el sitio a estimar y los datos utilizados en la estimación. El usuario puede considerar varias posibilidades.

- a) Vecindad única: Se habla de vecindad única cuando se efectúa el kriging conservando todos los datos. En este caso, incluso los datos muy alejados intervendrán en la estimación. Sin embargo, salvo excepciones, su influencia será muy baja (intuitivamente, un sitio alejado no aporta demasiada información al sitio a estimar y se verá afectado por un ponderador de kriging bajo). Cuando los datos son muy numerosos, es inútil conservarlos todos para una estimación local, puesto que se corre el riesgo de aumentar considerablemente los tiempos de cálculo. Por lo tanto, es necesario reducir el tamaño de la vecindad de kriging.
- b) Vecindad móvil: El kriging se realiza en una vecindad móvil cuando sólo utiliza los datos “cercaños” al sitio a estimar. Ahora, en general, uno no se limita a una sola estimación local, sino que busca estimaciones en los nodos de una grilla regular que cubre la zona estudiada. Falta definir el tamaño y la forma de la vecindad, que se centra en el sitio a estimar y que se desplaza a través del campo, a medida que se realiza las estimaciones (de donde viene el adjetivo móvil).

Cabe notar que no hay justificación particular para limitar el tamaño de la vecindad al alcance del modelo variográfico, bajo el pretexto que los datos localizados más allá de este alcance no tienen correlación con el sitio a estimar. De hecho, en la mayoría de los casos, estos datos intervienen indirectamente en la estimación del valor promedio y mejoran la precisión, a veces de manera no despreciable. El factor a considerar en la elección del tamaño de la vecindad es la cantidad de datos disponibles en la vecindad más que el alcance del semivariograma.

- c) Forma de la vecindad: En la medida de lo posible, la forma de la vecindad debe tomar en cuenta la anisotropía de la variable regionalizada, revelada por el análisis exploratorio espacial y variográfico. Así, en el caso de una anisotropía geométrica, se considerará una vecindad en forma de elipse (o elipsoide) cuyas características (orientación y excentricidad) sean idénticas a las de la elipse (elipsoide) de anisotropía.

A menudo, también, se divide esta elipse en varios sectores (en general, en cuadrantes u octantes), en cada uno de los cuales se trata de buscar un número fijo de datos, con el fin de repartir de mejor manera en torno al sitio que se quiere estimar, la información que se va a conservar. La figura 1.15 presenta un ejemplo de vecindad móvil en el espacio de dos dimensiones, en forma de elipse centrada en el sitio a estimar.

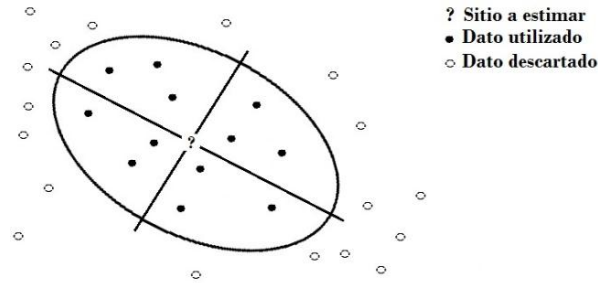


Figura 1.15: Vecindad Elíptica

En caso de anisotropía más compleja que la anisotropía geométrica, se suele conservar una vecindad en forma de elipse o elipsoide, aunque idealmente se debería escoger una forma más sofisticada (por ejemplo, una vecindad en forma de banda en caso de anisotropía zonal pura). Hay que buscar entonces un patrón en forma de elipse que se acerca lo mejor posible a las curvas de los valores del modelo variográfico, que indican el nivel de correlación en función de la distancia geográfica.

1.4.4.3 Ecuaciones kriging

Todos los estimadores lineales kriging pueden ser interpretados como proyecciones de un valor desconocido $Z^*(x_o)$ en el conjunto de los posibles estimadores. Mientras más amplio sea el conjunto en el cual es hecha la proyección más cercano estará el estimador kriging correspondiente del valor desconocido. A partir de estos criterios los kriging más usuales son:

a) Kriging Lineal con media constante en todo el dominio: Kriging ordinario

Hipótesis: Asume que el proceso es estacionario de media (desconocida). Se supone que la variable regionalizada z es la realización de una función aleatoria Z estacionaria tal que

$$x \in D, E[Z(x)] = m, \text{ desconocida.}$$

$$x, (x + h) \in D, Cov[Z(x + h), Z(x)] = C(h).$$

Linealidad: Se asegura esta restricción al tomar como estimador en x_o , su expresión es:

$$Z(x_o) = a + \sum_{i=1}^n \lambda_i Z(x_i)$$

Insesgo: El valor esperado del error de la estimación viene dado por:

$$E[Z^*(x_o) - Z(x_o)] = a + \sum_{i=1}^n \lambda_i E[Z(x_i)] - E[Z(x_o)] = a + \left(\sum_{i=1}^n \lambda_i - 1 \right) m$$

Como se desconoce el valor de la media, este valor esperado es nulo si:

$$a = 0 \quad y \quad \sum_{i=1}^n \lambda_i = 1.$$

La igualdad sobre la suma de los ponderadores asegura que, en el caso en que todos los datos son iguales a una misma constante, el valor estimado restituirá esta constante.

Optimalidad: Debemos calcular la varianza del error de estimación

$$Var[Z^*(x_o) - Z(x_o)] = \sum_{i=1}^m \sum_{j=1}^n \lambda_i \lambda_j C(x_i - x_j) + C(0) - 2 \sum_{i=1}^n \lambda_i C(x_i - x_j).$$

Se necesita minimizar esta expresión bajo la condición de insesgo, que impone que la suma de las incógnitas es igual a 1. Esto se logra introduciendo una incógnita adicional llamada multiplicador de Lagrange, que denotaremos como μ . Se escribe

$$Var[Z^*(x_o) - Z(x_o)] = C(0) + \sum_{i=1}^m \sum_{j=1}^n \lambda_i \lambda_j C(x_i - x_j) - 2 \sum_{i=1}^n \lambda_i C(x_i - x_o) + 2\mu \left(\sum_{i=1}^n \lambda_i - 1 \right)$$

Y se minimiza la función de las $n + 1$ variables $\lambda_1, \lambda_2, \dots, \lambda_n$. Calculando las $n + 1$ derivadas parciales de esta función y luego anulándolas, se obtiene el sistema:

$$\begin{cases} \frac{\partial}{\partial x_i} = 0: \sum_{j=1}^n \lambda_j C(x_i - x_j) + \mu = C(x_i - x_o); \quad \forall \lambda_i, i = 1, 2, \dots, n \\ \frac{\partial}{\partial \mu} = 0: \sum_{i=1}^n \lambda_i = 1 \text{ (condición de insesgo)} \end{cases}$$

Este sistema contiene una incógnita y una ecuación más que el sistema de kriging simple. Se puede escribir en notación matricial:

$$\begin{pmatrix} C(x_1 - x_1) & \cdots & C(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & 1 \\ C(x_n - x_1) & \cdots & C(x_n - x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}$$

Este kriging se denomina “kriging ordinario”. Siendo el semivariograma una herramienta equivalente a la covarianza, a partir de la relación $\gamma(h) = C(0) - C(h)$, se puede elegir utilizarlo en lugar de la función de covarianza. Las ecuaciones de kriging pasan a ser:

$$\left\{ \begin{array}{l} \sum_{j=1}^n \lambda_j \lambda_i (x_i - x_j) - \mu = \lambda_i (x_i - x_0); \forall i = 1, 2, \dots, n \\ \sum_{i=1}^n \lambda_i = 1 \end{array} \right.$$

Esto es

$$\begin{pmatrix} C(x_1 - x_1) & \dots & C(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & 1 \\ C(x_n - x_1) & \dots & C(x_n - x_n) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ -\mu \end{pmatrix} = \begin{pmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{pmatrix}$$

El estimador será de la forma:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) + \lambda_2 Z(x_2) + \dots + \lambda_n Z(x_n) = \sum_{i=1}^n \lambda_i Z(x_i)$$

Dónde:

$Z^*(x_0)$: La localización a estimar.

x_i : La localización i .

$Z(x_i)$: Valor de la variable en la localización i .

λ_i : El inverso de la distancia del punto a pronosticar al punto i .

Varianza de la estimación:

$$\sigma_{K0}^2(x_0) = \sigma^2 - \sum_{i=1}^n \lambda_i C(x_i - x_0) - \mu$$

Donde $\sigma^2 = C(0)$ es la varianza a priori de la función aleatoria Z , o sea, la meseta de su semivariograma.

b) Kriging lineal con media local conocida: Kriging simple.

Hipótesis: Se supone que la variable regionalizada z es la realización de una función aleatoria Z estacionaria tal que

$$x \in V, E[Z(x)] = m$$

$$x, (x + h) \in V, Cov[Z(x + h), Z(x)] = C(h)$$

Donde V representa la vecindad de la variable regionalizada.

Linealidad: Se asegura esta restricción al tomar como estimador en

$$Z(x_0) = a + \sum_{i=1}^n \lambda_i Z(x_i)$$

Insesgo: El valor esperado del error de la estimación es

$$E[Z^*(x_0) - Z(x_0)] = a + \sum_{i=1}^n \lambda_i E[Z(x_i)] - E[Z(x_0)] = a + \left(\sum_{i=1}^n \lambda_i - 1\right)m$$

Este valor esperado es nulo si:

$$a = \left(1 - \sum_{i=1}^n \lambda_i\right)m$$

Optimalidad: Debemos calcular la varianza del error de estimación

$$Var[Z^*(x_0) - Z(x_0)] = Var\left[\sum_{i=1}^n \lambda_i Z^*(x_i) - Z(x_0)\right]$$

El término constante a no influye en la varianza, por lo cual se omite de la expresión. Ahora se tiene el siguiente cálculo:

$$Var[\lambda_1 Z_1] = \lambda_1^2 Cov[Z_1, Z_1]$$

$$Var[\lambda_1 Z_1 + \lambda_2 Z_2] = \lambda_1^2 Cov[Z_1, Z_1] + \lambda_2^2 Cov[Z_2, Z_2] + 2\lambda_1 \lambda_2 Cov[Z_1, Z_2]$$

Generalizando:

$$Var\left[\sum_{i=1}^n \lambda_i Z_i\right] = \sum_{i=1}^n \sum_{l=1}^n \lambda_i \lambda_l Cov[Z_i, Z_l]$$

Aplicando esta fórmula, se obtiene:

$$Var[Z^*(x_0) - Z(x_0)] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j Cov[Z(x_i), Z(x_j)] + Cov[Z(x_0), Z(x_0)] - 2 \sum_{i=1}^n \lambda_i Cov[Z(x_i), Z(x_0)]$$

$$Var[Z^*(x_0) - Z(x_0)] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(x_i - x_j) + C(0) - 2 \sum_{i=1}^n \lambda_i C(x_i - x_0)$$

Es un sistema lineal, en el cual el número de ecuaciones y de incógnitas es igual a la cantidad de datos utilizados. En escritura matricial, este sistema es:

$$\begin{pmatrix} C(x_1 - x_1) & \cdots & C(x_1 - x_n) \\ \vdots & \ddots & \vdots \\ C(x_n - x_1) & \cdots & C(x_n - x_n) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} C(x_1 - x_0) \\ \vdots \\ C(x_n - x_0) \end{pmatrix}$$

Lo que permite determinar los ponderadores de kriging $\{\lambda_i, i = 1, \dots, n\}$. Es interesante notar que, debido a la condición de insesgo, el estimador se pone bajo la forma:

$$Z^*(x_0) = m + \left[\sum_{i=1}^n \lambda_i Z(x_i) - m \right] = m + \sum_{i=1}^n \lambda_i \varepsilon(x_i)$$

Dónde:

$Z^*(x_0)$: La localización a estimar.

x_i : La localización i .

$Z(x_i)$: Valor de la variable en la localización i .

λ_i : El inverso de la distancia del punto a pronosticar al punto i (Ponderaciones).

m : La media local.

$\varepsilon(x_i)$: Diferencia entre la media local y el valor de la variable en la localización x_i .

De modo que el valor de la media aparece como si fuera un dato adicional, al cual se asigna una ponderación igual al complemento de la ponderación acumulada de los otros datos.

Mientras menos ponderación le damos a los datos (en la práctica, esto ocurre cuando uno se aleja de estos datos), más ponderación recibe la media. El rol de la media es de compensar la falta de información cuando los datos son escasos o alejados. Este kriging que supone conocida la media fue bautizado kriging simple.

Varianza de la estimación:

$$\sigma_{KS}^2(x_0) = \sigma^2 - \sum_{i=1}^n \lambda_i C(x_i - x_0)$$

Donde $\sigma^2 = C(0)$ es la varianza a priori de la función aleatoria Z . Se puede mostrar que la varianza de kriging simple siempre es menor o igual a la varianza.

c) Kriging lineal en presencia de tendencia: Kriging Universal.

Hipótesis: Hasta ahora se ha asumido que la variable regionalizada es estacionaria (al menos se cumple con la hipótesis intrínseca). En muchos casos, la variable no satisface estas condiciones y se caracteriza por exhibir una tendencia. Este modelo considera la

media desconocida y permite que varíe dentro de cada vecindad situada en el área de estudio.

La tendencia puede expresarse mediante:

$$m(x) = \sum_{i=1}^p a_i f_i(x)$$

Donde las funciones $f(x)$ son conocidas y p es el número de términos empleados para ajustar $m(x)$.

Linealidad: Se asegura esta restricción al tomar como estimador en x_o .

$$Z(x_o) = a + \sum_{i=1}^n \lambda_i Z(x_i)$$

Insesgo: El valor esperado del error de la estimación es

$$E[Z^*(x_o)] = m(x_o)$$

$$E \left[\sum_{i=1}^n \lambda_i Z(x_i) \right] = m(x_o)$$

$$E \left[\sum_{i=1}^n m(x_i) \right] = m(x_o)$$

$$\sum_{i=1}^n \lambda_i \left[\sum_{l=1}^p a_l f_l(x_i) \right] = \sum_{l=1}^p a_l f_l(x_o)$$

$$\sum_{l=1}^p a_l \left[\sum_{i=1}^n \lambda_i f_l(x_i) \right] = \sum_{l=1}^p a_l f_l(x_o) \Rightarrow \sum_{i=1}^n \lambda_i f_l(x_i) = \sum_{l=1}^p f_l(x_o)$$

Este valor esperado es nulo si:

$$a = \left(1 - \sum_{i=1}^n \lambda_i \right) m$$

Optimalidad: Debemos calcular la varianza del error de estimación

$$Var[Z^*(x_o) - Z(x_o)] = Var \left[\sum_{i=1}^n \lambda_i Z^*(x_o) - Z(x_o) \right]$$

El estimador será de la forma:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i f^i Z(x_i)$$

Dónde:

$Z^*(x_0)$: La localización a estimar.

x_i : La localización i .

$Z(x_i)$: Valor de la variable en la localización i .

λ_i : El inverso de la distancia del punto a pronosticar al punto i (Ponderaciones).

f^i : Son las funciones $f(x)$ que describen la tendencia dentro de la vecindad.

Varianza:

$$\sigma_{KU}^2(x_0) = \sigma_{oo} - \sum_{i=1}^n \lambda_i \sigma_{io} + \sum_{l=1}^n \mu \phi_l(x_0)$$

1.4.4.4 Validación del modelo kriging

Es conveniente validar el modelo seleccionado y sus parámetros, al respecto se discute la validación cruzada¹⁸. El método de validación cruzada ha sido ampliamente utilizado para evaluar el grado de bondad de un modelo de predicción kriging y reconocido como un método óptimo de estimación de sus parámetros.

La operación de validar un modelo ajustado no requiere de mucho tiempo y este se considera como el último de los pasos importantes del análisis de un estudio geoestadístico, debido a que una vez obtenido este resultado será utilizado en la estimación por krigeaje en cualquiera de sus variantes. El método consiste en extraer un elemento de la muestra y estimar el valor en ese punto utilizando el modelo kriging obtenido. De forma análoga se actúa para el resto de elementos de la muestra. Como resultado se obtiene un mapa de todas las diferencias $\{Z^*(x_i) - Z(x_i), i = 1, 2, \dots, n\}$, entre el valor real y el estimado.

De forma tal que si el modelo kriging es el óptimo reflejara adecuadamente la estructura espacial implícita en el conjunto de datos, entonces los valores estimados deben ser cercanos a los valores observados.

¹⁸ Journel y Huijbregts (1978)

Esta cercanía puede ser caracterizada según los siguientes estadígrafos:

- a) El error medio, dado por:

$$T_1 = \frac{1}{n} \sum_{i=1}^n [Z(x_i) - Z^*(x_i)]$$

Debe ser aproximadamente igual a cero.

- b) El error medio estandarizado, dado por

$$T_2 = \frac{1}{n} \sum_{i=1}^n \frac{[Z(x_i) - Z^*(x_i)]}{\sigma}$$

Debe ser lo más cercano a cero.

- c) La media del cuadrado del error, dado por

$$T_3 = \frac{1}{n} \sum_{i=1}^n [Z(x_i) - Z^*(x_i)]^2$$

Debe ser lo más pequeño posible.

- d) La correlación muestral, dada por:

$$T_4 = \frac{[Z(x_i) - Z^*(x_i)]}{\sigma_i}$$

Debe ser cercana a cero.

Siendo:

$Z(x_i)$: Los valores muestrales de la propiedad en el punto x_i .

$Z^*(x_i)$: Los valores estimados de la propiedad en el punto x_i .

σ_i : Es la desviación estándar de la estimación en el punto x_i .

n : Numero de observaciones.

Idealmente todos los estadígrafos anteriores deben satisfacerse simultáneamente, pero en la práctica una mejoría en uno de ellos puede degradar a otro. Por lo que es recomendable hacer un análisis integral de los estadígrafos de las diferencias.

Aplicación y análisis de los métodos geoestadístico a los accidentes de tránsito en el municipio de San Salvador 2006 – 2010.

Introducción:

La posibilidad de obtener valores cercanos a los datos muestreados con indicación del error cometido en la estimación de las variables en estudio es el éxito de la geoestadística, éxito que depende en gran parte al estudio estructural previo a la interpolación. El siguiente capítulo en estudio resume la aplicación de las técnicas geoestadísticas a los accidentes de tránsito en el municipio de San Salvador, está estructurado en tres partes esenciales, lo que es el análisis estructural, la fase de estimación y la validación del modelo a utilizar. El análisis estructural involucra realizar pruebas del fenómeno regionalizado mediante dos puntos de vistas, el primero, cuando se analiza qué tipo de patrón es el que está describiendo tal proceso y el segundo, conocer la variabilidad del fenómeno en relación a la distancia. Se analizan los estadísticos como el índice del vecino más próximo para contrastar que el fenómeno de los accidentes de tránsito presenta un patrón de tipo aleatorio, uniforme o agrupado; esto, para tener la evidencia de la existencia de regiones donde se presenta una mayor o menor frecuencia de accidentalidad vial. Por otro lado, se analiza dentro del análisis estructural, la caracterización del fenómeno, es decir, estudiar la correlación y variabilidad espacial mediante el contraste de I de Moran y la función semivariograma.

En la parte de la estimación se tiene que a partir del análisis estructural tener la capacidad de interpolar el fenómeno y generar un modelo teórico con el estimador kriging, para la obtención de una superficie o mapa de frecuencias de accidentes de tránsito ocurridos en el municipio de San Salvador. La última fase es el validar el modelo de kriging que mejor se ajuste a la naturaleza de los datos. El método a utilizar es el de la validación cruzada que consiste en calcular las diferencias entre los valores experimentales y los estimados para todos los sitios muestreados. Las diferencias se resumen mediante los estadísticos: media de los errores de estimación (MEE), error medio estandarizado (EME), error cuadrático medio (ECM) y el Coeficiente de correlación muestral (ECA).

En síntesis el objetivo principal de este capítulo es seleccionar y evaluar los diferentes modelos de interpolación kriging para obtener una superficie continua de datos estimados.

Se aplican las técnicas geoestadísticas a los accidentes de tránsito ocurridos en el municipio de San Salvador en el periodo 2006 – 2010.

2.1 Análisis estructural

El analizar la estructura espacial del fenómeno de los accidentes de tránsito en el municipio de San Salvador demanda en primer lugar analizar el patrón de las localizaciones en donde ha ocurrido un accidente de tránsito, lo que permitirá definir una primera idea de la estructura espacial que organiza la ocurrencia de los accidentes de tránsito.

Conociendo dicha descripción del proceso hay que inferir si el fenómeno presenta un patrón aleatorio, uniforme o agrupado. Una vez establecido el patrón espacial de los accidentes debe definirse cómo varía la intensidad o frecuencia de los accidentes en función de su localización, es decir, medir la correlación espacial, el cual mide el grado de similitud de una variable entre puntos vecinos sobre el área muestral.

La teoría geoestadística demanda que es necesario realizar las pruebas del patrón espacial y autocorrelación ya que de lo contrario a la hora de poder definir un modelo teórico al fenómeno de los accidentes de tránsito para su estimación se caería en un posible error de no conocer a ciencia cierta su patrón espacial, es decir, no habría forma de modelar el fenómeno cuando el patrón presente sea aleatorio en el área de estudio.

A continuación se analizan los accidentes de tránsito de forma espacial a partir de estas presunciones definidas anteriormente.

2.1.1 Análisis del patrón espacial

Comprender la distribución espacial de los accidentes de tránsito que ocurren en el municipio de San Salvador constituye un paso importante para el esclarecimiento del comportamiento de estos, es por ello, que la percepción visual de la distribución espacial del fenómeno es muy útil para traducir los patrones existentes en consideraciones objetivas y medibles, que nos dan respuesta a una parte de la dinámica del espacio, mas no a su totalidad. En este apartado se pretende confirmar el supuesto de no aleatoriedad de los accidentes de tránsito ocurridos en el municipio de San Salvador. Para ello, se considera necesario abordar dicho análisis mediante la técnica del análisis del vecino más próximo, el cual permite estimar la forma de distribución de cualquier variable regional sobre el área investigada a una o varias escalas seleccionadas, de tal manera que se cuantifique en una medida global esa variación espacial en distintas direcciones del espacio.

H₀: Los accidentes de tránsito presentan un patrón espacial aleatorio

H_i: Los accidentes de tránsito presentan un patrón espacial no aleatorio

En la tabla 2.1 se presenta en resumen los resultados del análisis del vecino más próximo estimado mediante el programa Crimestat 3.0¹⁹ para los años del 2006 hasta 2010. La herramienta del vecino más próximo devuelve cuatro valores: distancia media observada, índice de vecino más próximo, puntuación Z y valor P.

Año	Distancia media	Valor del NNI	P valor	Tipo de patrón
2006	75.20 m	0.3337	0.0001	Agrupado
2007	75.42 m	0.3584	0.0001	Agrupado
2008	76.79 m	0.3909	0.0001	Agrupado
2009	83.82 m	0.3630	0.0001	Agrupado
2010	90.06 m	0.3787	0.0001	Agrupado

Tabla 2.1: Resumen índice del vecino más próximo

Del que podemos destacar que el índice del vecino más próximo para cada año es menor que uno, el valor oscila entre 0.3337 y 0.3909 y el P valor es de 0.0001 menor que el nivel de significancia establecido, por lo que se rechaza la hipótesis nula (aleatoriedad) y se afirma que los accidentes de tránsito tienen un comportamiento no aleatorio, con tendencia al agrupamiento. Este resultado puede ser contrastado con los gráficos de K-orden del índice del vecino más próximo, el cual se muestra en la figura 2.1. La lectura y análisis de los gráficos es la siguiente: se tendrá una tendencia al comportamiento aleatorio si la función que representa la intensidad de accidentes de tránsito (línea azul) superara el límite del supuesto de aleatoriedad (color rojo). En el caso que la función de intensidad de los accidentes de tránsito se mantuviera por debajo del límite del supuesto de aleatoriedad, entonces afirmaríamos un comportamiento no aleatorio.

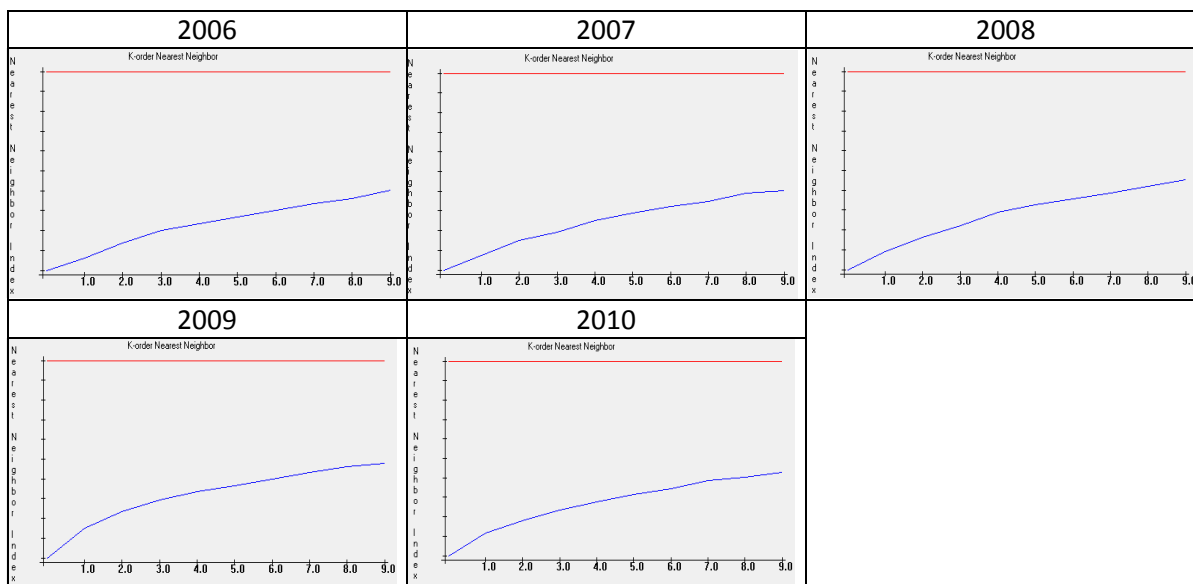


Figura 2.1: Salida de correlogramas vecino más próximo

¹⁹Crimestat 3.0: Programa de análisis espacial-estadístico diseñado para interactuar con sistemas de información geográfica y desarrollado por NedLevine & Associates, con financiamiento del departamento de justicia de Estados Unidos de Norteamérica.

Concluyendo así, que el patrón que describe el comportamiento de los accidentes de tránsito en el municipio de San Salvador entre los años 2006 y 2010 es de tipo no aleatorio con tendencia a la agrupación.

2.1.2 Análisis de correlación espacial

La forma más sencilla de realizar una medición de la similitud que presentan los distintos valores que toma la variable en distintas localidades del espacio, es mediante la estimación de la autocorrelación, que permite establecer una medida del grado de semejanza de esos valores entre localizaciones vecinas.

Esta perspectiva global de cómo están distribuidos estos valores es contrastada mediante el análisis de correlación espacial, el cual tiene como objetivo contrastar la hipótesis de que una variable se encuentre distribuida de forma totalmente aleatoria en un sistema espacial o, si por el contrario, existe algún tipo de asociación significativa de valores similares o distintos entre localidades vecinas.

Esta descripción de la correlación espacial es posible realizarla a través del test I de Moran cuya interpretación resulta ser en base al grado de significancia del P valor, al signo y valor del coeficiente de Moran. Por ello, es que a partir de comprobar la presencia de autocorrelación se invalida el supuesto de independencia espacial (la hipótesis de aleatoriedad), es decir, si existe autocorrelación significa entonces que las localidades vecinas muestran una intensidad de accidentalidad vial similar, lo que implica que la ocurrencia de accidentes a cierta distancia, dependerá de la intensidad de accidentes que ocurran en las localidades vecinas que resulten ser significativamente similares. En el caso que no existiese autocorrelación, se contrastaría que las localidades tendrán muy diferentes intensidades en la ocurrencia de accidentes, es decir, que en la mayoría de los casos existirán localidades con una mayor o menor intensidad que sus vecinos.

H_0 : Los accidentes de tránsito presentan un comportamiento espacial aleatorio

H_i : Los accidentes de tránsito presentan un comportamiento espacial no aleatorio

Los resultados obtenidos del test I de Moran mediante el programa Crimestat se resumen en la tabla 2.2.

Año	Índice de Moran	Z	P valor	Conclusión
2006	0.28702	96.139596	0.0001	Autocorrelación espacial
2007	0.27415	91.850369	0.0001	Autocorrelación espacial
2008	0.30814	103.181135	0.0001	Autocorrelación espacial
2009	0.27170	91.033784	0.0001	Autocorrelación espacial
2010	0.27512	92.172745	0.0001	Autocorrelación espacial

Tabla 2.2: Resumen índice de Moran

De la tabla anterior destacamos que el valor del índice de Moran es significativo para cada año y el P valor es de 0.0001, menor que el nivel de significancia establecido, por lo tanto rechazamos la hipótesis nula (la hipótesis de aleatoriedad) y afirmamos que existe correlación espacial en los accidentes de tránsito ocurridos en el municipio de San Salvador.

Tal afirmación puede ser verificada mediante la salida de los correlogramas de la función I de Moran. Su interpretación es la siguiente: la función de la distribución observada (color azul) se compara con los límites de las distribuciones teóricas (color verde y rojo) que representan la hipótesis de aleatoriedad.

La distribución observada será independiente si su recorrido se sitúa entre la menor (función en rojo) y la mayor (función en verde). Si el recorrido de la función observada se sitúa por encima de la condición de aleatoriedad se concluye que la distribución observada, tendrá correlación espacial positiva. Mientras que si se sitúa bajo la condición de aleatoriedad, la distribución observada tendrá correlación espacial negativa.

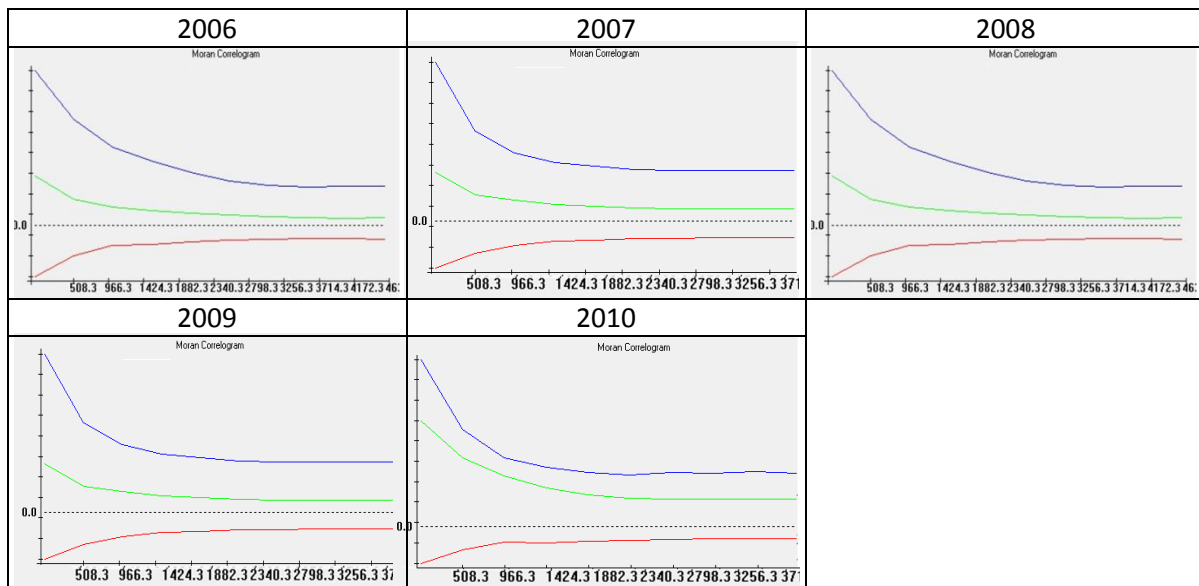


Figura 2.2: Salida de correlogramas I de Moran

Tal como se muestra en la figura 2.2, la distribución de accidentes de tránsito ocurridos en el municipio de San Salvador entre los años 2006 y 2010 presenta una naturaleza no aleatoria y con presencia de correlación espacial positiva. Una vez que se ha llegado a tal afirmación es importante ahora cuantificar la correlación existente entre los sitios muestreados. La herramienta mediante el cual se abordara dicho análisis será el cálculo del semivariograma.

2.1.3 Estimación de parámetros para el semivariograma

Uno de los objetivos primordiales del ajuste de un modelo teórico de semivariograma es que aparte de determinar la correlación espacial, también, es capaz de cuantificar la correlación de una variable regionalizada medida en serie de puntos (en este caso, los accidentes de tránsito en el municipio de San Salvador), es decir, puede conocer/predecir esa intensidad medida a distancias conocidas, pero que no ha sido muestreada.

En nuestro caso, ya que se comprobó que los accidentes de tránsito en el municipio de San Salvador presentan autocorrelación espacial positiva nos restringimos a estimar la variabilidad mediante los modelos esférico, exponencial y gaussiano, que son los modelos que garantizan el cumplimiento de al menos una de las hipótesis de estacionariedad.

Para elegir el modelo semivariograma teórico que mejor se adapte al experimental, es importante tener en cuenta que sobre todo, debe tomarse en cuenta el comportamiento en el origen y a grandes distancias de este último. Es decir no importa tanto la forma analítica concreta del semivariograma como el hecho de que respete las principales características del fenómeno.

Por orden de importancia son el efecto pepita, la pendiente en el origen, el rango, la meseta. En las tablas 2.3 a 2.7 se detallan los parámetros y gráficos de los modelos de los semivariogramas esférico, exponencial y gaussiano que mejor se ajustan a los accidentes de tránsito en el municipio de San Salvador entre los años 2006 y 2010.

2006				
Modelo	Ecuación	Efecto pepita	Rango	Meseta
Gaussiano	$3.1471(1 - e^{(-h^2/1.58)})$	0.131	1.26	3.1471
Exponencial	$3.0471(1 - e^{(-h^3/0.64)})$	0.012	0.86	3.0458
Esférico	$3.1065(1.11 * h - 0.40 * h^3)$	0.081	0.74	3.1065

Tabla 2.3: Parámetros de los modelos teóricos esférico, exponencial y gaussiano año 2006

2007				
Modelo	Ecuación	Efecto pepita	Rango	Meseta
Gaussiano	$3.1645(1 - e^{(-h^2/1.69)})$	0.145	1.30	3.1645
Exponencial	$3.0623(1 - e^{(-h^3/0.82)})$	0.013	0.91	3.0623
Esférico	$3.1456(1.23 * h - 0.55 * h^3)$	0.090	0.82	3.1456

Tabla 2.4: Parámetros de los modelos teóricos esférico, exponencial y gaussiano año 2007

2008				
Modelo	Ecuación	Efecto pepita	Rango	Meseta
Gaussiano	$3.1190(1 - e^{(-h^2/1.66)})$	0.147	1.29	3.1190
Exponencial	$3.0433(1 - e^{(-h^3/0.77)})$	0.012	0.88	3.0433
Esférico	$3.1123(1.18 * h - 0.24 * h^3)$	0.085	0.79	3.1123

Tabla 2.5: Parámetros de los modelos teóricos esférico, exponencial y gaussiano año 2008

2009				
Modelo	Ecuación	Efecto pepita	Rango	Meseta
Gaussiano	$3.1948(1 - e^{(-h^2/1.51)})$	0.144	1.23	3.1948
Exponencial	$3.0967(1 - e^{(-h^3/0.77)})$	0.012	0.88	3.0967
Esférico	$3.1385(1.21 * h - 0.32 * h^3)$	0.081	0.77	3.1385

Tabla 2.6: Parámetros de los modelos teóricos esférico, exponencial y gaussiano año 2009

2010				
Modelo	Ecuación	Efecto pepita	Rango	Meseta
Gaussiano	$3.1171(1 - e^{(-h^2/1.63)})$	0.101	1.28	3.1171
Exponencial	$3.0128(1 - e^{(-h^3/0.77)})$	0.019	0.88	3.0128
Esférico	$3.0099(1.06 * h - 0.17 * h^3)$	0.061	0.71	3.0099

Tabla 2.7: Parámetros de los modelos teóricos esférico, exponencial y gaussiano año 2010

En resumen en lo que respecta a la validación del mejor modelo del semivariograma es evidente que para cada año se alcanza una estabilidad a cierta distancia que no difiere mucho entre cada uno de los modelos para cada año.

Sin embargo hay que destacar que el modelo exponencial es el que reduce el valor del efecto pepita (error de medición) y el que a menor distancia se alcanza la estabilidad requerida.

Este comportamiento es más que significativo en lo que concierne a la variabilidad espacial, entre más corta sea la distancia a la que se alcance la estabilidad (rango) mayor es el grado de autocorrelación entre las observaciones, justificando así, que es factible lograr interpolar el fenómeno de los accidentes de tránsito a partir de los modelos de estimación que la teoría geoestadística ofrece.

Por lo tanto, se puede afirmar estadísticamente en base a las funciones de semivariograma para los accidentes de tránsito ocurridos en el municipio de San Salvador entre los años 2006 y 2010, que hay una evidencia de variabilidad, estacionariedad y existencia de correlación espacial, permitiendo así, definir un modelo teórico que pueda describir dicho proceso espacial.

Tal aseveración puede ser comprobada mediante los gráficos de los semivariogramas para los modelos esférico, exponencial y gaussiano correspondiente al periodo de años entre 2006 y 2010.

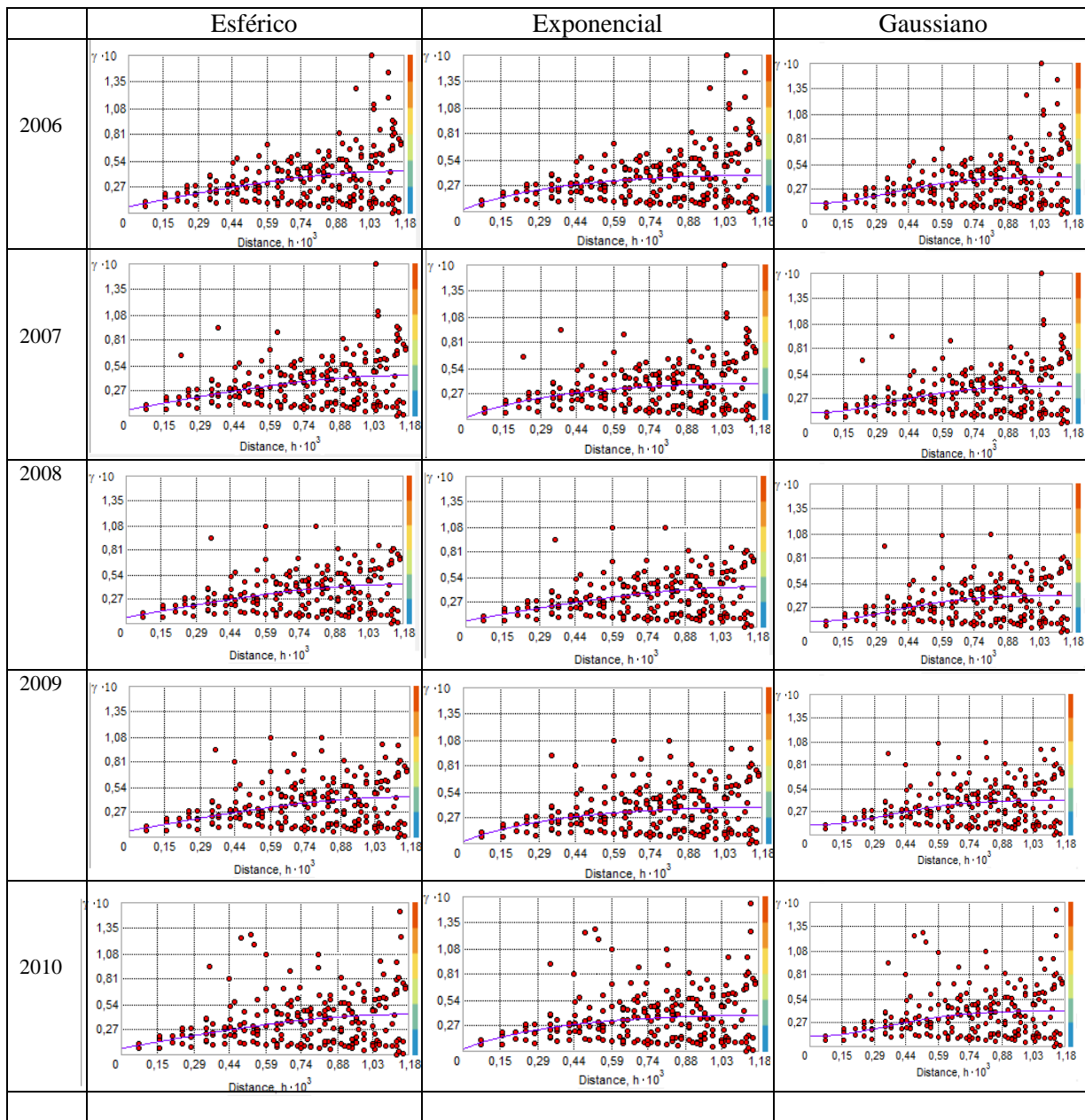


Figura 2.3: Modelos teóricos semivariogramas (esférico, exponencial y Gaussiano).

En los gráficos de los semivariogramas se observa que a pequeñas distancias el grado de similitud es muy alto, lo que significa que las observaciones vecinas mantienen un grado de correlación muy fuerte, esto es muy importante para nuestro objetivo, porque los métodos kriging asumen esta propiedad para realizar la estimación. Otro dato importante es que se observan muy pocos datos atípicos, esto significa que son pocas las localidades que tienen un comportamiento desigual al resto de mediciones realizadas y casi siempre estos datos atípicos se presentan a largas distancias del punto de interés. Por último se observa que la meseta o la cota superior del semivariograma, es rápidamente alcanzada para cada uno de

los años, esto indica que el fenómeno de los accidentes de tránsito en el municipio de San Salvador es muy homogéneo no en toda la área de estudio, pero si a nivel local o por zonas.

2.1.4 Validación del modelo kriging

Una vez comprobada la dependencia o autocorrelación espacial en los accidentes de tránsito, ahora se puede utilizar los datos para realizar predicciones. El kriging forma ponderaciones a partir de los valores medidos próximos para prever ubicaciones sin mediciones. Para crear una superficie continua del fenómeno, se realizan predicciones para cada ubicación en el área de estudio basadas en el semivariograma y la disposición espacial de los valores medidos que son cercanos. Existen tres métodos lineales del kriging: Simple, ordinario y universal. El kriging ordinario es el más general y menos utilizado de los métodos kriging. Presupone que el valor medio es constante en toda el área de estudio. El kriging simple es bastante similar al kriging ordinario, con la diferencia que el método simple asume que la media del fenómeno es constante y conocida a nivel local. Y el kriging universal permite cierta variabilidad en cada vecindad.

El método a utilizar para medir el grado de bondad de cada modelo de predicción kriging es el de la validación cruzada, que consiste en calcular las diferencias entre los valores experimentales y los estimados para todos los sitios muestreados. Las diferencias se resumen mediante los estadísticos: media de los errores de estimación (MEE), error medio estandarizado (EME), error cuadrático medio (ECM) y la correlación muestral (CM). En las tablas 2.8 al 2.12 se resumen los estadísticos utilizadas por la técnica de validación cruzada de los kriging simple, ordinario y universal para cada uno de los años.

Año 2006				
Kriging	MEE	EME	ECM	CM
Simple	0.0008125	0.0004834	1.665	0.105
Ordinario	0.0022870	0.0013301	1.681	0.210
Universal	0.0008125	0.0004699	1.665	0.105

Tabla 2.8: Validación cruzada año 2006

Año 2007				
Kriging	MEE	EME	ECM	CM
Simple	0.0005736	0.0003326	1.649	0.206
Ordinario	0.0018790	0.0011250	1.634	0.320
Universal	0.0005736	0.0003326	1.649	0.260

Tabla 2.8: Validación cruzada año 2007

Año 2008				
Kriging	MEE	EME	ECM	CM
Simple	0.0006129	0.003969	1.476	0.127
Ordinario	0.001735	0.001160	1.462	0,188
Universal	0.0006129	0.003969	1.476	0.127

Tabla 2.8: Validación cruzada año 2008

Año 2009				
Kriging	MEE	EME	ECM	CM
Simple	0.0006044	0.0004660	1.258	0.107
Ordinario	0.0017010	0.0013440	1.240	0.111
Universal	0.0006044	0.0004660	1.258	0.107

Tabla 2.8: Validación cruzada año 2009

Año 2010				
Kriging	MEE	EME	ECM	CM
Simple	0.0006044	0.0006121	1.258	0.108
Ordinario	0.0017010	0.0011940	1.240	0.111
Universal	0.0006044	0.0006121	1.258	0.108

Tabla 2.8: Validación cruzada año 2010

Finalmente, en la figura 2.4 se muestra el gráfico QQ de normalidad para los residuos.

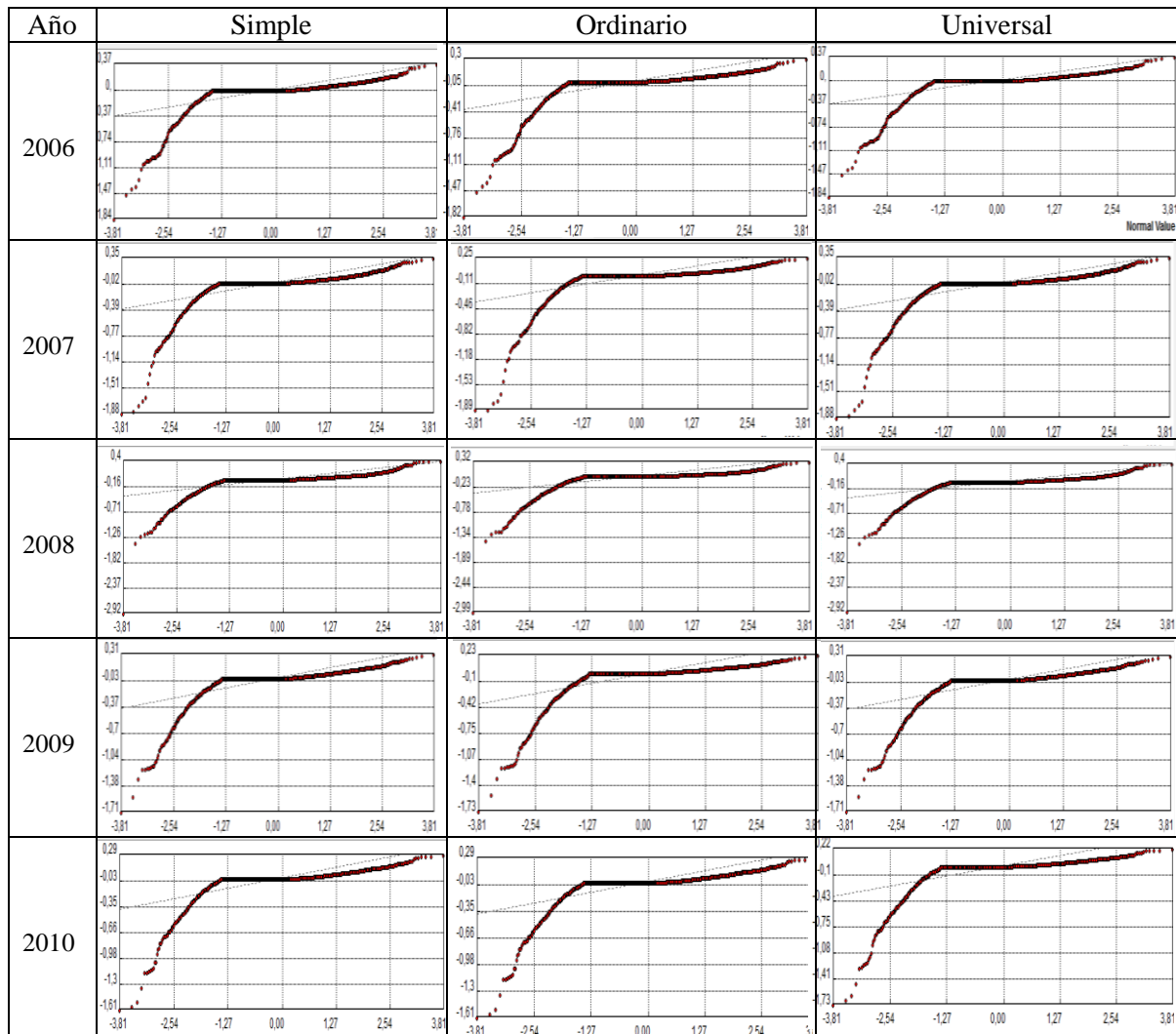


Figura 2.4: Grafico QQ residuales normalizados para los kriging ordinario, simple y universal.

Como se observa, existen puntos o datos que se alejan considerablemente de la recta de los residuos normalizados, el cual puede ser producido por el mismo comportamiento natural de los accidentes de tránsito que ocurren con mayor intensidad en algunas intersecciones y en otras con menor intensidad. Para cada uno de los métodos kriging el comportamiento que se observa es similar para los cinco años en estudio, pero si retomamos el análisis de los residuos, concluimos que los mejores estimadores que se ajustan a la naturaleza de los accidentes de tránsito en el municipio de San Salvador para su estimación son el kriging simple y el kriging universal bajo la forma:

$$Z_S^*(x_0) = m + \left[\sum_{i=1}^n \lambda_i Z(x_i) - m \right] = m + \sum_{i=1}^n \lambda_i Z(x_i) \quad \text{y} \quad Z_U^*(x_0) = \sum_{i=1}^n \lambda_i f^i Z(x_i)$$

Presentación de resultados

Introducción:

En este capítulo se presentan los resultados finales del diagnóstico espacial de los accidentes de tránsito en el municipio de San Salvador 2006 – 2010, con lo que se pretende explicar el comportamiento y consecuencias que tienen estos en la sociedad. La información obtenida en este capítulo es el resultado de analizar el fenómeno mediante dos enfoques diferentes, primero se aborda el tema mediante la estadística clásica y segundo mediante un enfoque geoestadístico.

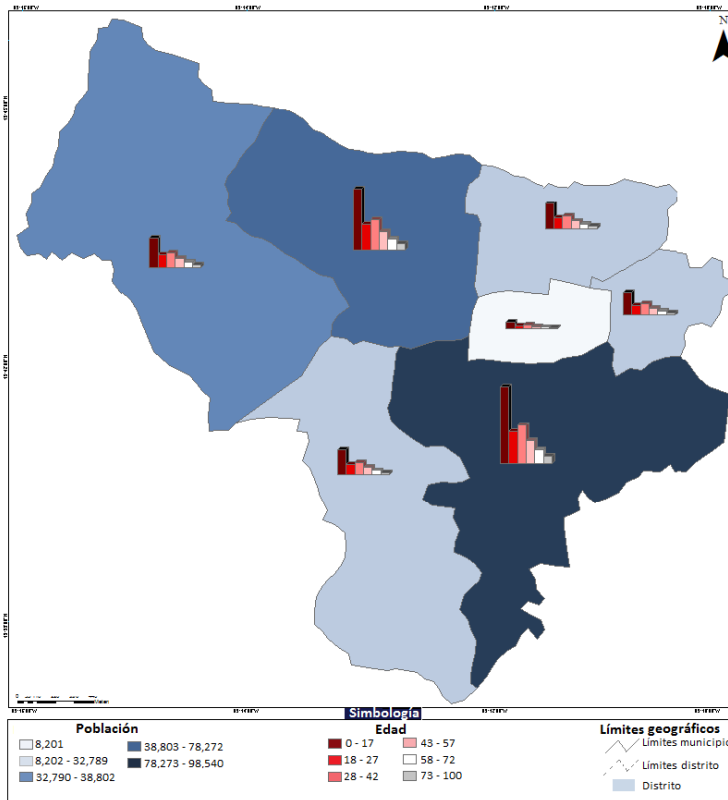
Es por ello que el capítulo está organizado básicamente en tres partes importantes. En el primero se detalla información básica para conocer algunas de las características geográficas, sociales y económicas relevantes a la accidentalidad vial en el municipio de San Salvador. Se detalla el área y población de cada uno de los distritos que conforman el municipio de San Salvador así como nivel de estudio, ocupación, sexo y edad de la población.

En el segundo apartado se analizan todos los factores que intervienen en los accidentes de tránsito como sus causas, tipos de accidentes, condiciones del conductor y consecuencias de los mismos. Se calculan tasas de mortalidad y accidentalidad vial en relación al número de pobladores de cada distrito.

Y un último apartado donde se presentan los mapas correspondientes a la ubicación geográfica de los accidentes de tránsito con el fin de conocer las intersecciones donde constantemente suceden y cuáles son las características más comunes de estos.

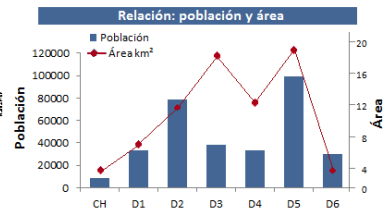
En síntesis esperamos que los resultados que a continuación se presentan sean de utilidad para tomar decisiones acertadas, que conlleven a reducir los accidentes de tránsito y disminuir sus altos y negativos impactos sociales y económicos en la población.

b) Distritos de San Salvador: Población y área, 2007.

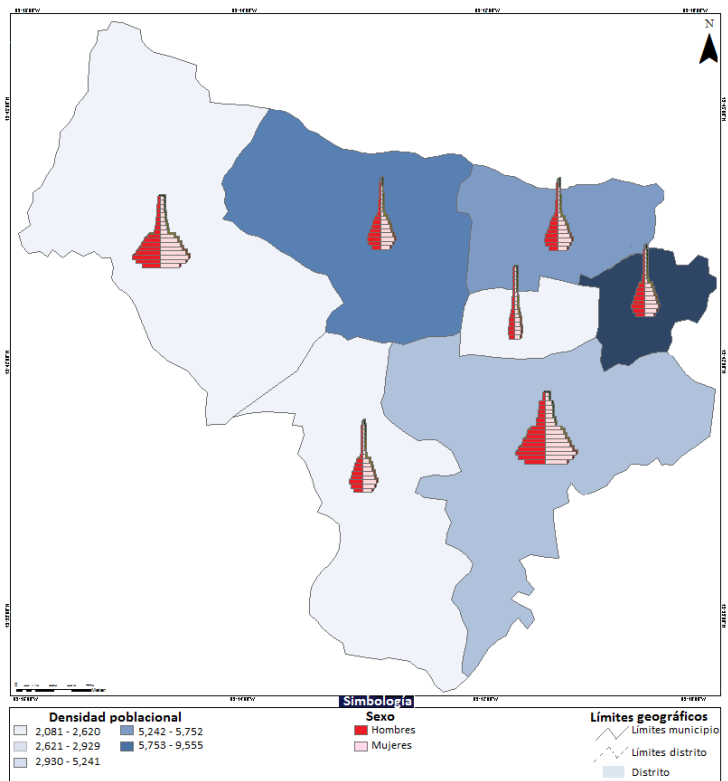


Población y área			
Código	Distrito	Población	Área km ²
CH	Centro Histórico	8,200	2.8
D1	Distrito 1	32,789	5.7
D2	Distrito 2	78,271	11.2
D3	Distrito 3	38,082	18.3
D4	Distrito 4	32,482	12.4
D5	Distrito 5	98,540	18.8
D6	Distrito 6	29,622	3.1
D7	Municipio de San Salvador	317,986	72.2

Fuente: Dirección General de Estadísticas y Censos

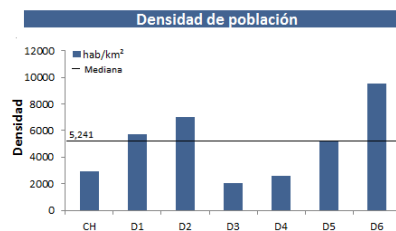


c) Distritos de San Salvador: Densidad de población, 2007.

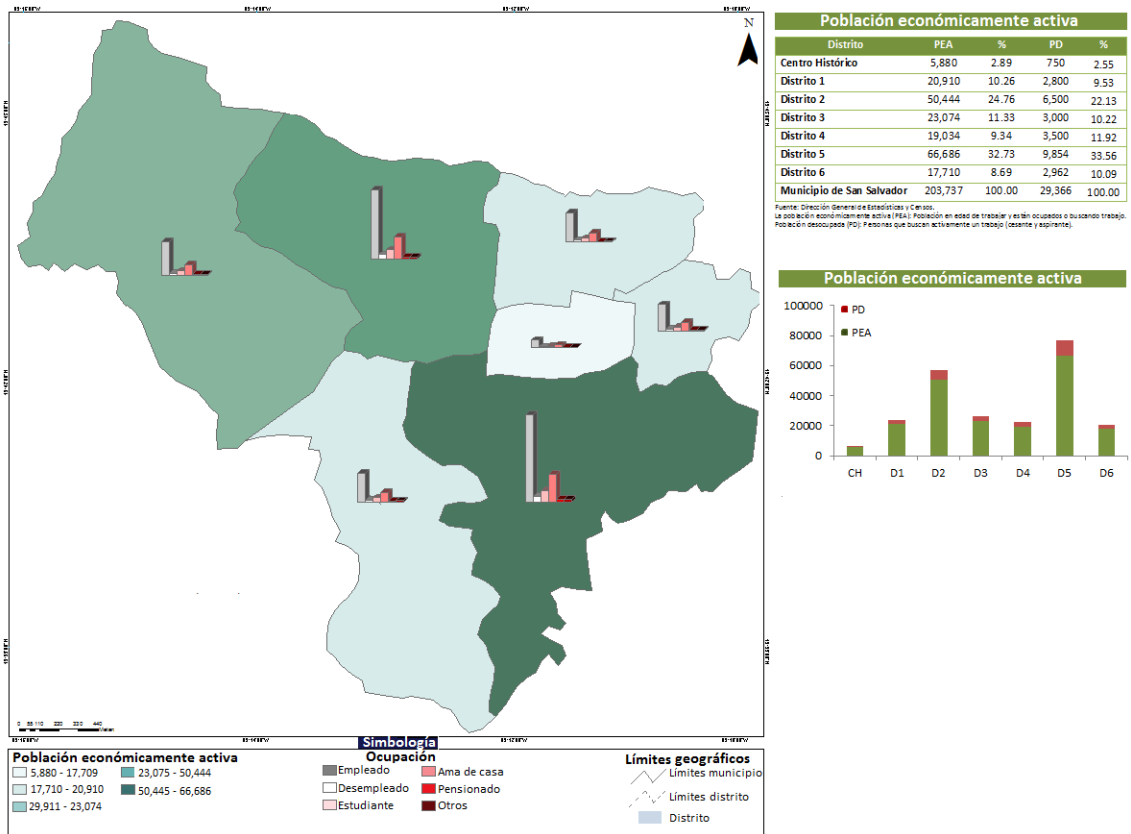


Densidad de población					
Distrito	Población	%	Área km ²	%	Hab./km ²
Centro Histórico	8,200	2.58	2.8	3.88	2,929
Distrito 1	32,789	10.31	5.7	7.89	5,752
Distrito 2	78,271	24.61	11.2	15.51	6,988
Distrito 3	38,082	11.98	18.3	25.35	2,081
Distrito 4	32,482	10.21	12.4	17.17	2,620
Distrito 5	98,540	30.99	18.8	26.04	5,241
Distrito 6	29,622	9.32	3.1	4.29	9,555
Municipio de San Salvador	317,986	100.00	72.2	100.00	4,404

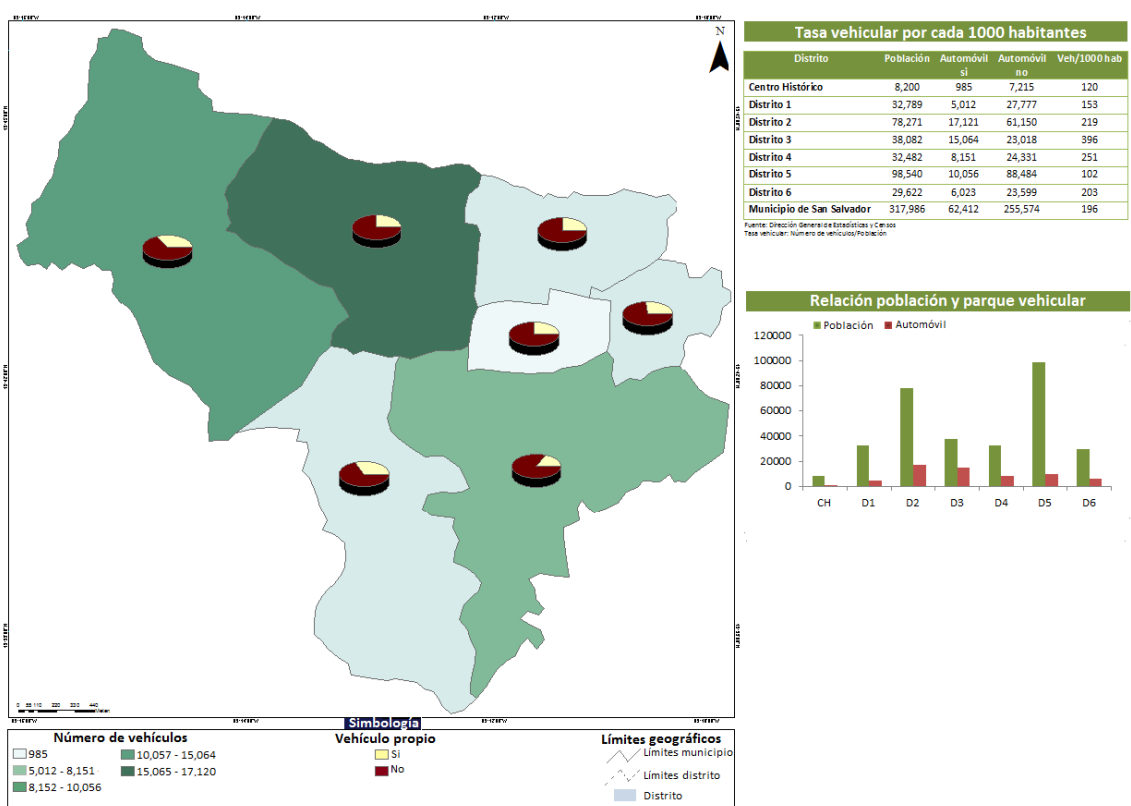
Fuente: Dirección General de Estadísticas y Censos
Densidad de población: Población/superficie km²



d) Distritos de San Salvador: Población económicamente activa, 2007.



e) Distritos de San Salvador: Tasa vehicular, 2007.



De la sección anterior destaca el caso del distrito tres como la jurisdicción más poblada, muy por encima del distrito dos que es el segundo más habitado. La distribución de la población con respecto a la edad en los distritos tres y dos es muy similar, en ambos se tiene un buen número de habitantes en edad escolar y productiva, el distrito menos poblado es el Centro Histórico con un poco más de 8,000 habitantes. Otro dato importante es la relación que existe entre la población y área del distrito, En este contexto el distrito seis es el que cuenta con mayor densidad de población casi 10,000 hab/km², su tamaño de población no está en equilibrio con su área territorial, el segundo con mayor densidad de población es el distrito dos con cerca de los 6,000 hab/km². El que cuenta con menor densidad de población es el Centro Histórico.

En lo que concierne a la población económicamente activa, el distrito cinco es el que cuenta con mayor número de pobladores que se dedican a alguna actividad productiva y el distrito seis es el que cuenta con el mayor porcentaje de población con empleo. Esto puede parecer un dato muy impreciso o general, pero es un primer acercamiento a la idea de cómo es el flujo de personas que se desplazan de sus hogares a sus trabajos y viceversa, que puede ser un factor fundamental en el aumento del riesgo vial.

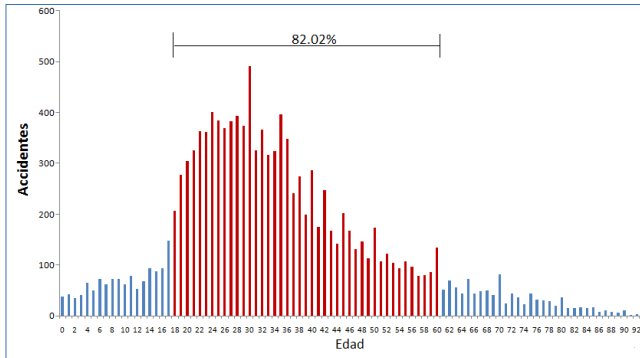
Finalmente, otro aspecto primordial en el análisis de los accidentes de tránsito es el número de vehículos que existen por distrito, en este aspecto el distrito dos es el que cuenta con el parque vehicular más grande y en el distrito tres se tiene que por cada cien habitantes existen cuarenta vehículos, lo que lo convierte en el distrito más denso con respecto a la flota vehicular.

3.2 Principales características de los accidentes de tránsito en el municipio de San Salvador 2006 - 2010.

En esta sección se analiza la evolución de los accidentes de tránsito entre los años 2006 y 2010 para cada uno de los distritos. Se obtienen frecuencias, tasas de accidentalidad vial, tasas de mortalidad y principales características relacionadas a los accidentes de tránsito como día, hora, sexo y edad de los accidentados, causas y tipo de accidentes. Se construyen mapas que explican la dinámica de los accidentes por distrito y así acentuar una primera idea de las principales zonas con mayor intensidad o riesgo de accidentes vehicular.

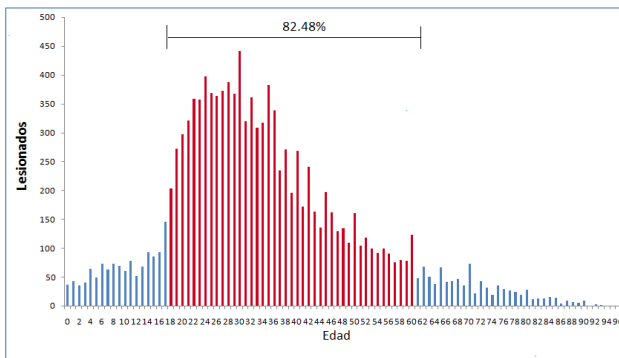
3.2.1 Relación: Accidentes de tránsito y edad.

a) Frecuencia de personas accidentadas por edad.



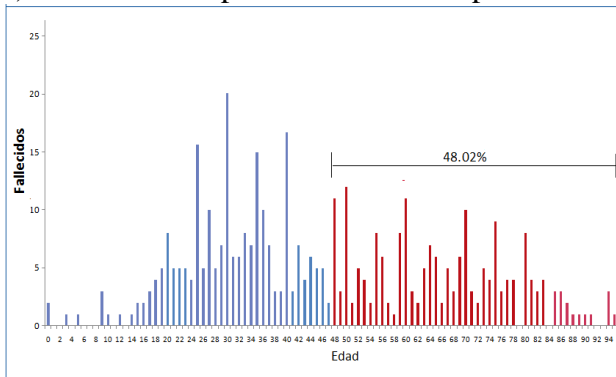
En lo que concierne al factor edad en los accidentes de tránsito el 82.02% de las personas involucradas se encuentran en el periodo de edad más productiva (de 18 a 55 años).

b) Frecuencia de personas lesionadas por edad.



De manera similar al caso que engloba a todos los involucrados el 82.48% de las personas lesionadas se encuentran entre las edades de 18 y 55 años.

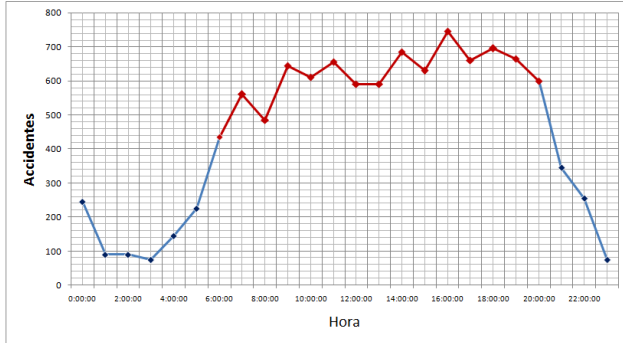
c) Frecuencia de personas fallecidas por edad.



Un hecho importante a destacar es el que de las personas fallecidas en los accidentes de tránsito la mitad pertenece a la tercera edad. Es decir las personas adultas son los que menos involucrados están en accidentes, pero cuando lo sufren casi siempre es letal.

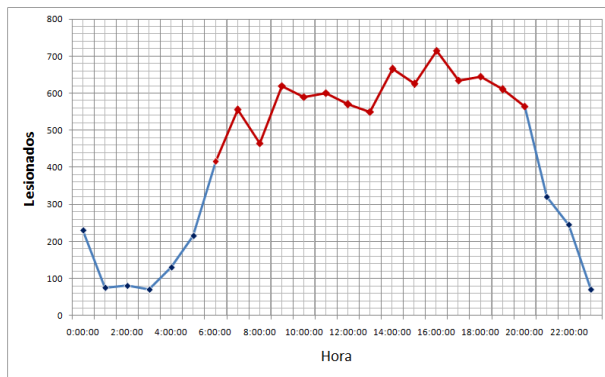
3.2.2 Relación: Accidentes de tránsito y hora

a) Frecuencia de personas accidentadas por hora.



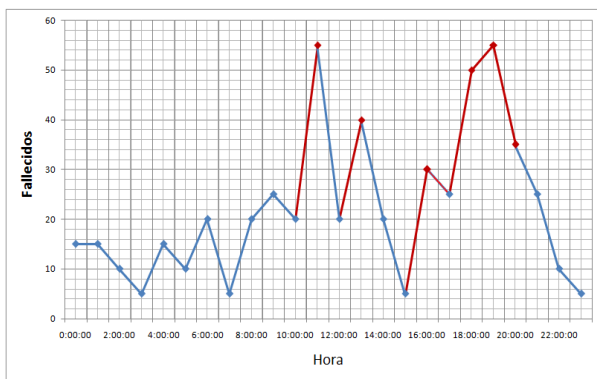
Referente a la dinámica de los accidentes de tránsito en relación a la hora, la concentración de estos van desde la 6:00 a.m. hasta las 8:00 p.m. y su punto más alto es alcanzado a las 4:00 p.m. que es donde se incrementa el movimiento de la ciudadanía al desplazarse de

b) Frecuencia de personas lesionadas por hora.



Al igual que en la gráfica anterior el mayor número de lesionados se registra entre las 6:00 a.m. y las 8:00 p.m. aparte de que se nota que desde las 4:00 a.m. hasta las 4:00 p.m. La dinámica de los accidentes casi siempre se mantiene en forma creciente. Es decir el número de accidentes aumenta según la hora

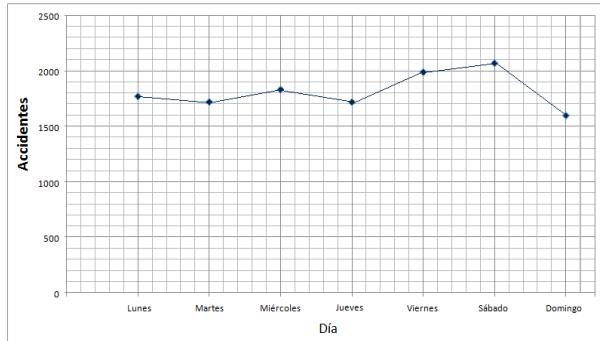
c) Frecuencia de personas fallecidas por hora.



En lo que concierne a la relación a los horarios de los accidentes y las defunciones, no se observa un patrón definido. Salvo lo que sí es claro es que los niveles más alarmantes se dan en los intervalos de horas 11:00 a.m. a 1:00 p.m. y 6:00 p.m. a las 8:00 p.m.

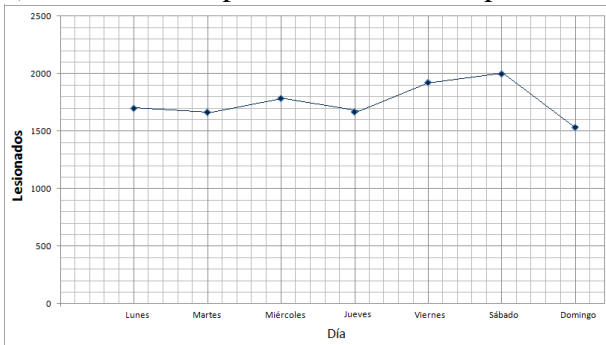
3.2.3 Relación: Accidentes de Tránsito y día.

a) Frecuencia de personas accidentadas por día.



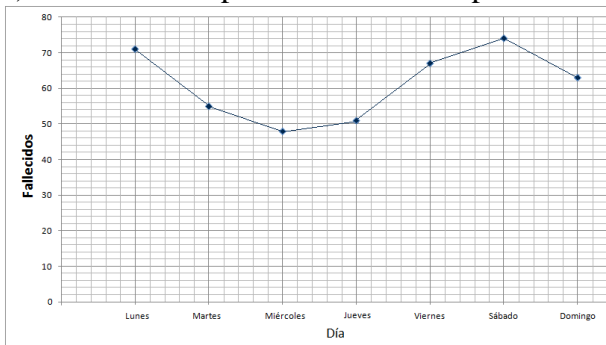
Se observa que la variabilidad de los accidentes entre los días es relativamente baja, a excepción de los días viernes y sábado, en el que se evidencia un mayor número de accidentes y los días domingos presentan una leve disminución.

b) Frecuencia de personas lesionadas por día.



El número de lesionados por día es conforme al número de accidentes que se den por día, la frecuencia de lesionados permanece casi de forma constante en todos los días de la semana.

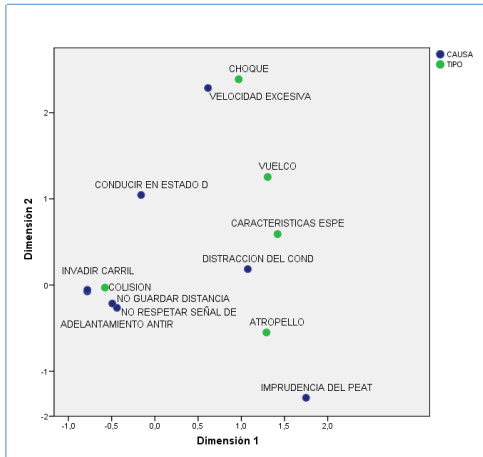
c) Frecuencia de personas fallecidas por día.



La frecuencia de fallecidos aumenta entre los días viernes, sábado, domingo y lunes (lo que puede considerarse como los fines de semana y el lunes). Contrario a lo anterior, de martes a jueves baja la intensidad de fallecimientos.

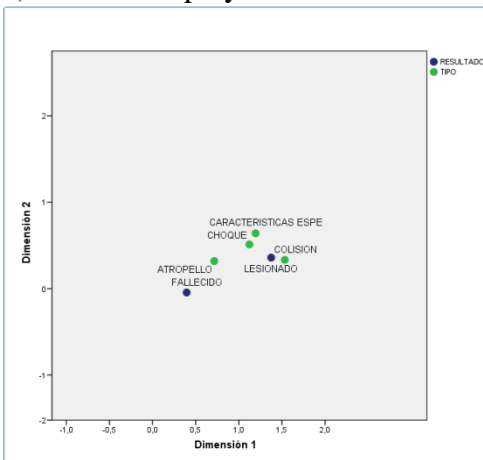
3.2.4 Relación: Causa del accidente y tipo de accidente.

a) Relación causa y tipo de accidente



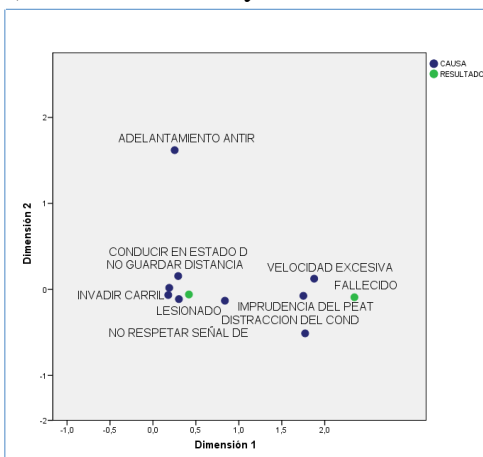
Al establecer la relación que existe entre las causas y el tipo de accidente, se puede apreciar que los atropellos están más relacionados a imprudencia del peatón y distracción del conductor, el choque está fuertemente ligado a la velocidad excesiva y la colisión a no guardar la distancia, no respetar la señal e invadir el carril.

b) Relación tipo y resultado del accidente.



Destaca el hecho que los atropellos son los accidentes que causan más fallecidos, lógicamente este resultado era de esperar, ya que un peatón luce totalmente desprotegido ante un evento como lo son los accidentes y las colisiones son los accidentes que más lesionados dejan.

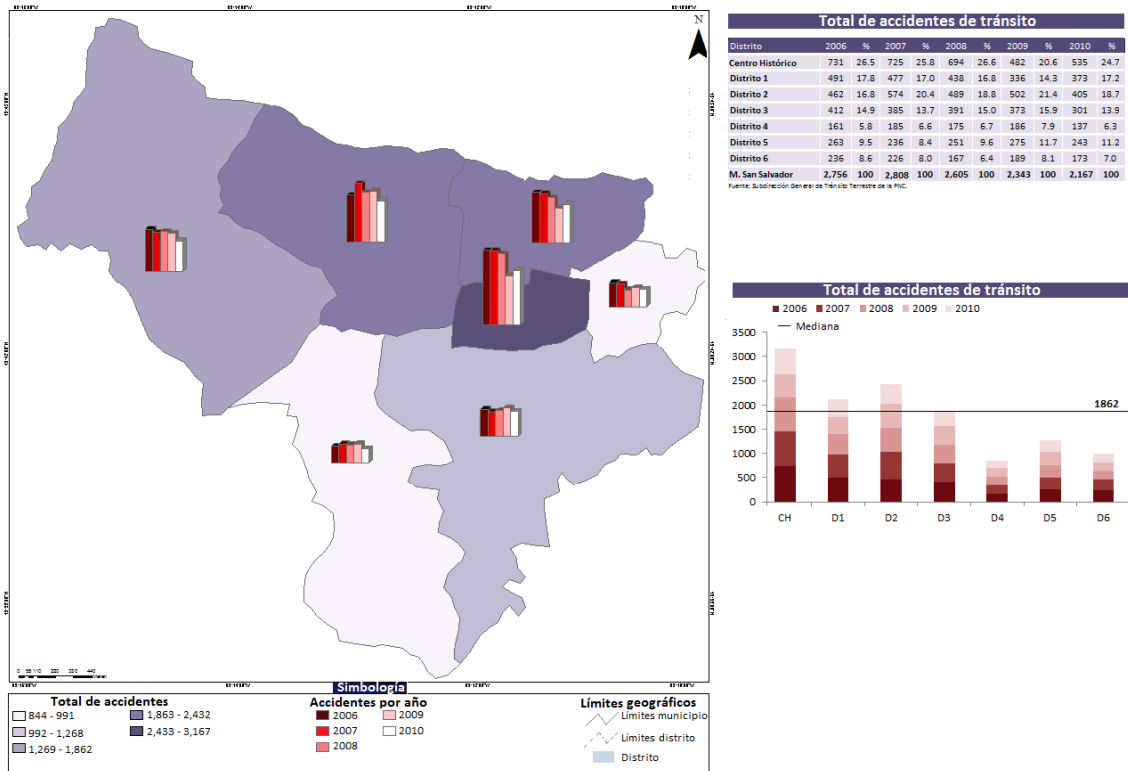
c) Relación causa y resultado del accidente.



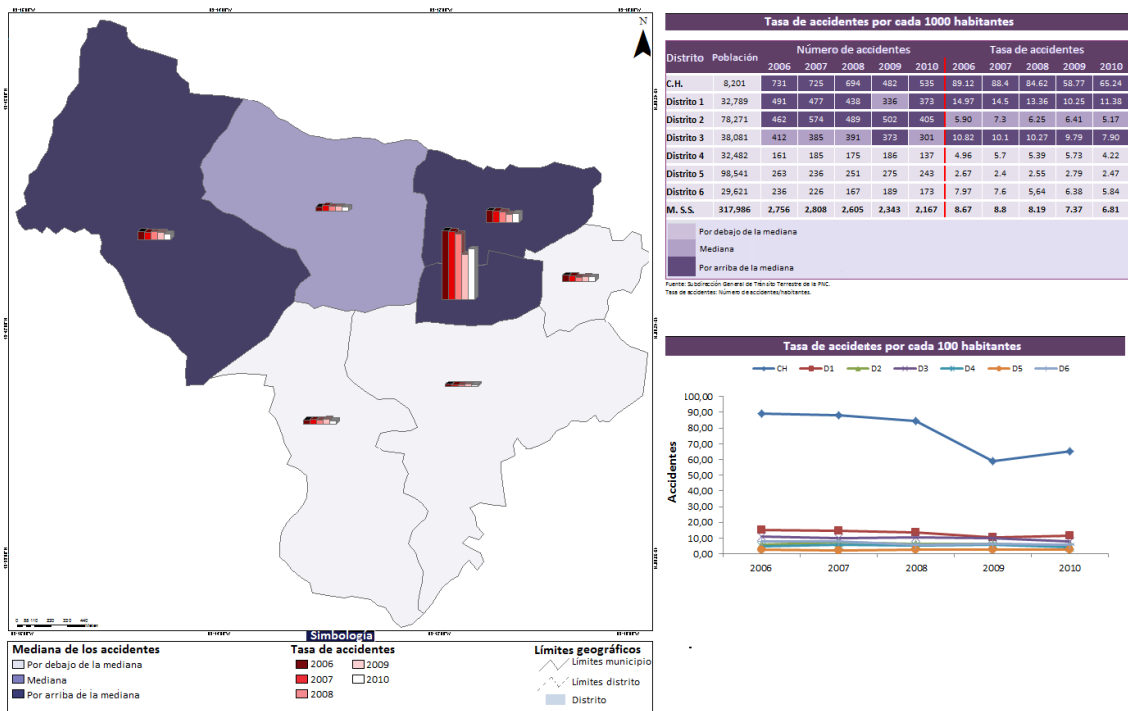
En relación a la causa y el resultado del accidente es notable que la velocidad excesiva, distracción del conductor y la imprudencia del peatón son las causas que mayor número de fallecidos dejan, esto en clara correspondencia a que estas son las causas que provocan mayor número de atropellos. El resto de causas están asociadas más que todo a dejar lesionados.

3.2.5 Frecuencia y características de los accidentes de tránsito por distrito

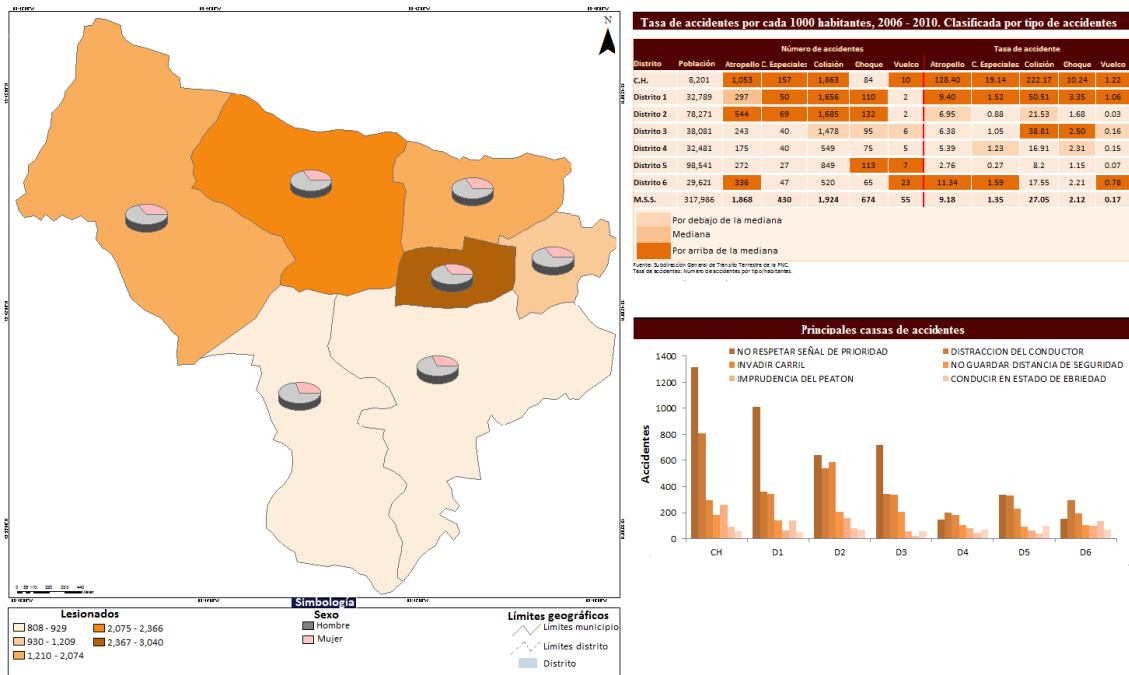
a) Distritos de San Salvador: Total de accidentes, 2006-2007.



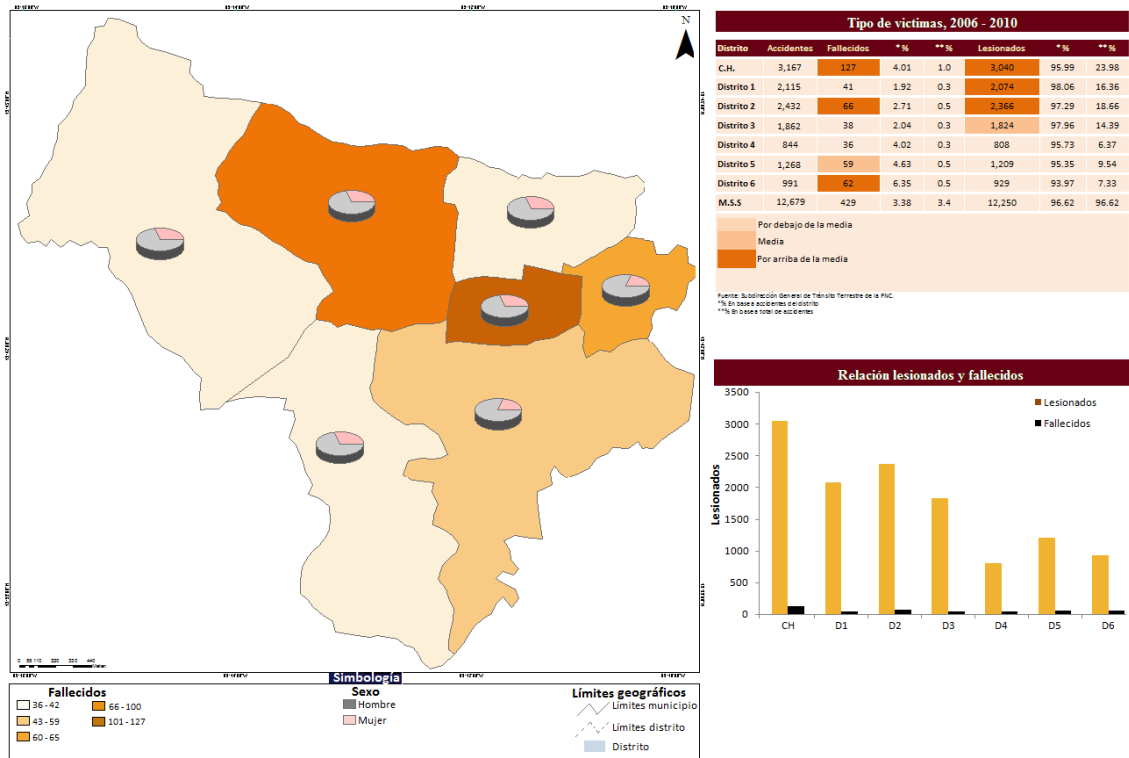
b) Distritos de San Salvador: Tasa de accidentes, 2006-2007.



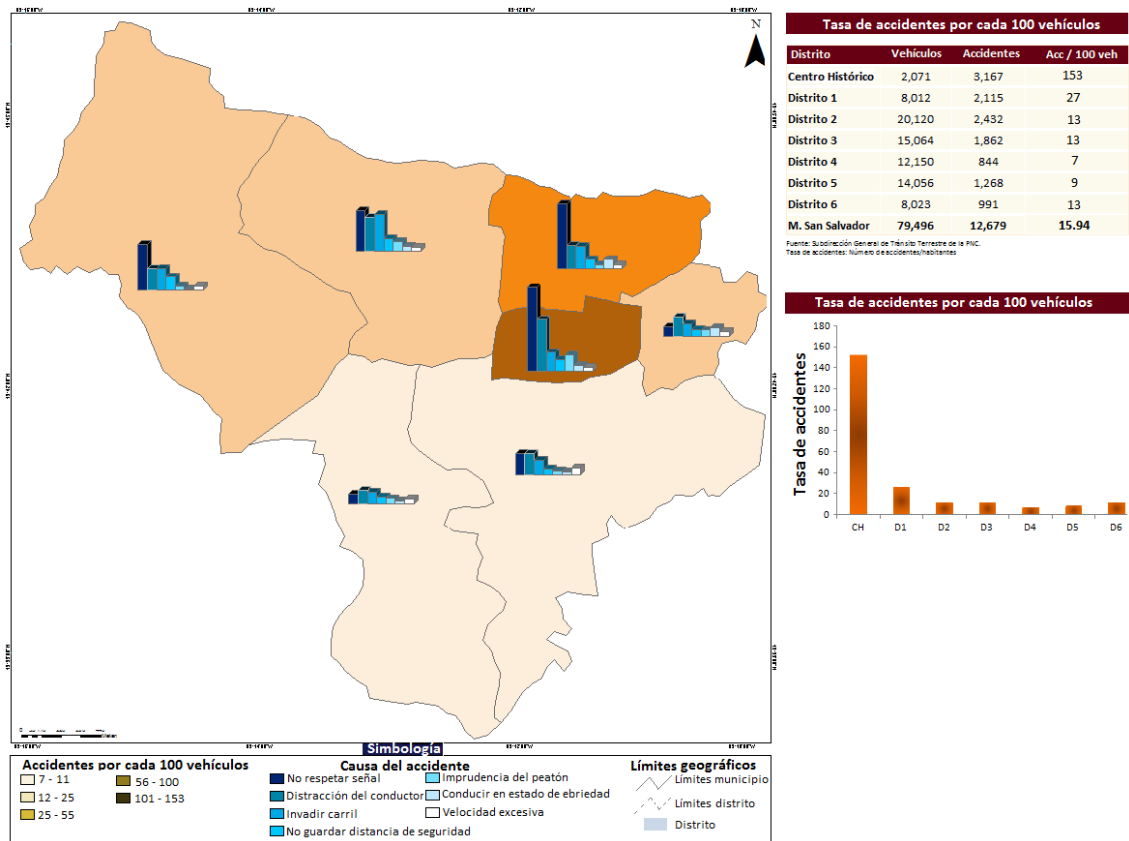
c) Distritos de San Salvador: Tasa de accidentes por tipo de evento, 2006-2007.



d) Distritos de San Salvador: Tasa de accidentes por tipo de resultado, 2006-2007.



e) Distritos de San Salvador: Tasa de accidentes por cada 100 vehículos, 2006-2007.



Con la información estadística procesada en esta sección, se considera que la frecuencia de accidentes de tránsito que presenta el Centro Histórico es muy alta comparada con el resto de distritos. La diferencia crece aún más si la comparación se hace en base a la tasa de accidentes por número de habitantes, número de lesionados y fallecidos, esto debido a que el Centro Histórico es el distrito menos poblado y el que mayor número de accidentes de tránsito presenta. Al Centro Histórico le siguen el distrito uno y tres como delegaciones de alto riesgo vial y como contraparte el distrito cuatro y cinco aparecen con las frecuencias y tasas de accidentalidad más bajas.

3.3 Distribución espacial de los accidentes de tránsito

3.3.1 Centro Histórico

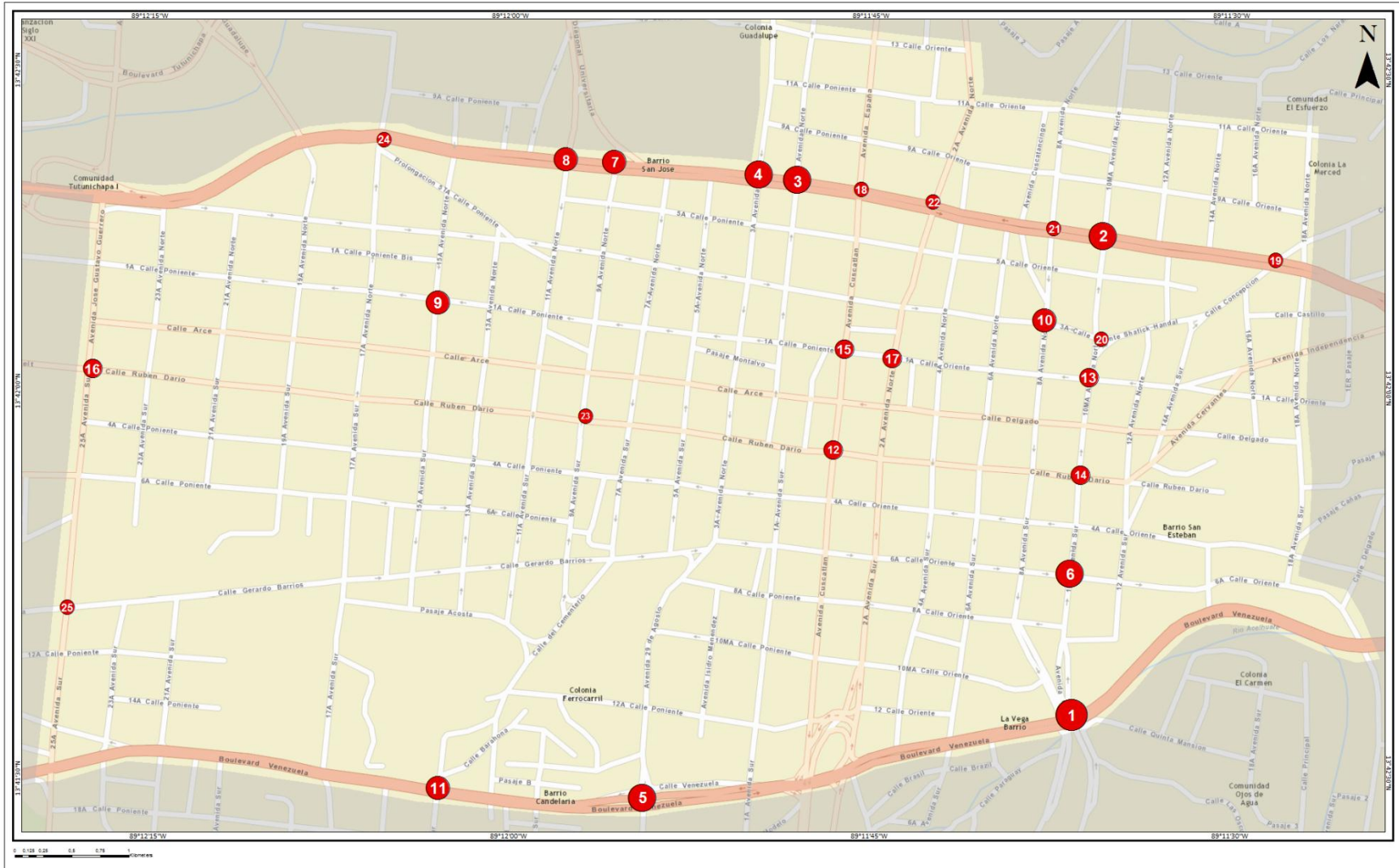
Datos generales

- Población Total: 8,201.
- Extensión Territorial: 2.8 km².
- Densidad de población: 2,929 hab/km².
- Límites:
 - Norte: Alameda Juan Pablo II y 11^a calle oriente.
 - Sur: Boulevard Arturo Castellanos (boulevard Venezuela).
 - Este: Calle Concepción y 18^a avenida norte.
 - Oeste: 25^a avenida norte.

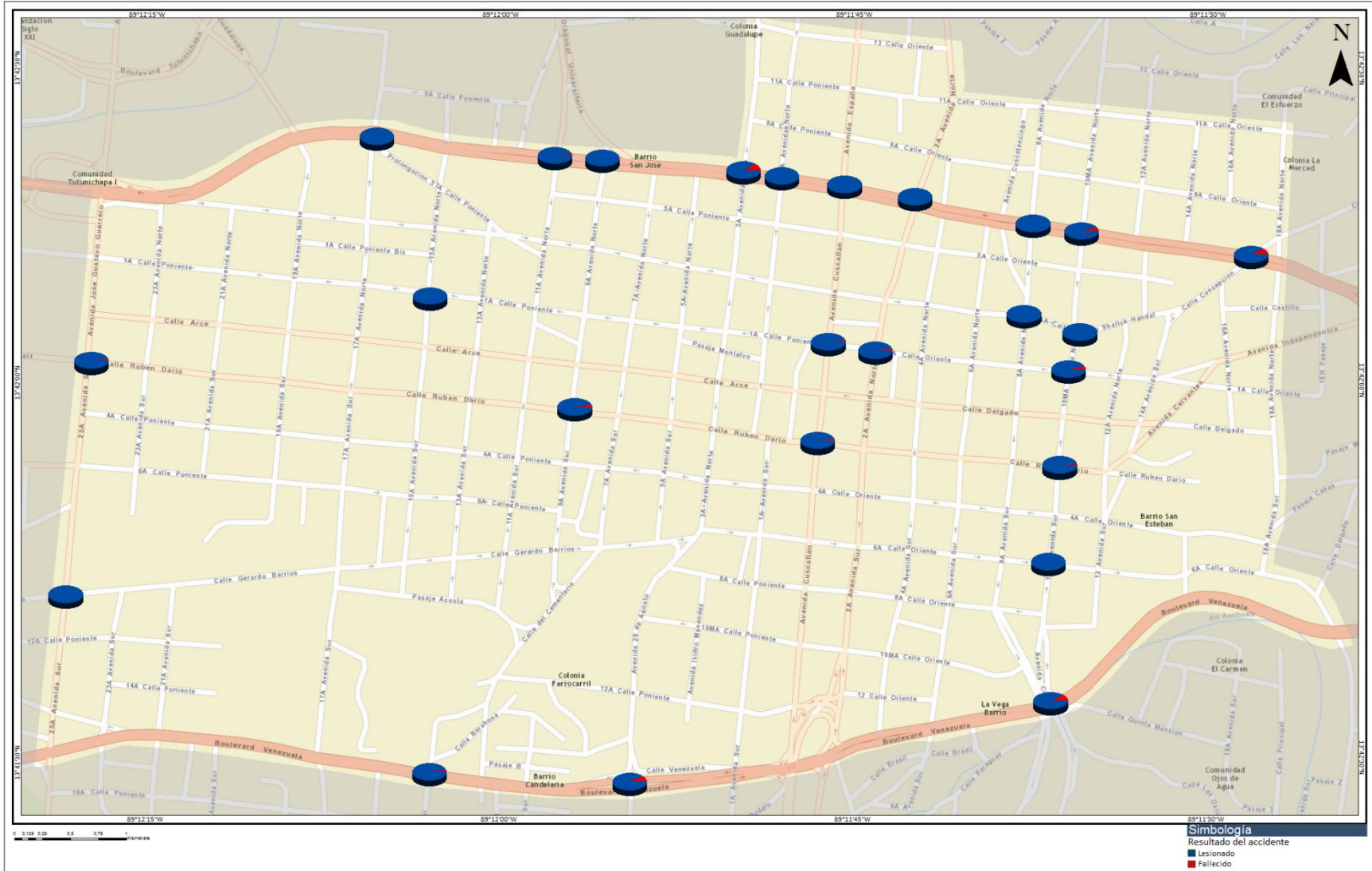
a) Intersecciones con alta frecuencia de accidentes de tránsito

Correlativo	Intersección	Frecuencia	Fallecidos	Lesionados	Atropellos	Choques	Colisiones
1	Boulevard Venezuela Y 10 ^a Avenida Sur	82	8	74	41	4	37
2	Alameda Juan Pablo II Y 10 ^a Avenida Norte	52	3	49	15	3	26
3	Alameda Juan Pablo II Y 1 ^a Avenida Norte	49	0	49	5	0	44
4	Alameda Juan Pablo II Y 3 ^a Avenida Norte	45	5	40	33	1	9
5	Bulevar Venezuela Y Avenida 29 de Agosto	45	4	41	19	0	26
6	6 ^a Calle Oriente Y 10 ^a Avenida Norte	44	0	44	2	0	42
7	Alameda Juan Pablo II Y 9 ^a Avenida Norte	42	0	42	5	1	34
8	Alameda Juan Pablo 2 ^a Y 11 ^a Avenida Norte	40	1	39	11	1	27
9	1 ^a Calle Poniente Y 17 ^a Avenida Norte	40	0	40	1	3	36
10	3 ^a Calle Poniente Y 8 ^a Avenida Norte	39	0	39	3	0	36
11	Boulevard Venezuela Y 13 ^a Avenida Sur	39	1	38	5	3	31
12	Calle Rubén Darío Y Avenida Cuscatlán	37	1	36	22	0	13
13	10 ^a Avenida Norte Y 1 ^a Calle Oriente	36	2	34	6	1	29
14	10 ^a Avenida Sur Y 2 ^a Calle Oriente	36	1	35	8	0	28
15	1 ^a Calle Oriente Y Avenida España	36	1	35	12	2	22
16	Calle Rubén Darío Y 25 ^a Avenida Sur	36	1	35	9	0	26
17	1 ^a Calle Oriente Y Avenida Monseñor Romero	33	1	32	17	0	11
18	Alameda Juan Pablo II Y Avenida España	32	0	32	12	3	17
19	Alameda Juan Pablo Y Calle Concepción	31	3	28	7	0	24
20	3 ^a Calle Poniente Y 11 ^a Avenida Sur	31	0	31	1	0	28
21	Alameda Juan Pablo II Y 8 ^a Avenida Norte	31	0	31	5	0	26
22	Alameda Juan Pablo II Y 2 ^a Avenida Norte	31	0	31	13	0	18
23	Calle Rubén Darío Y 9 ^a Avenida Sur	26	1	25	13	0	9
24	Alameda Juan Pablo II Y 17 ^a Avenida Norte	26	0	26	7	1	18
25	Calle Gerardo Barrios Y 25 ^a Avenida Sur	26	0	26	9	0	17
	Total	965	33	932	281	23	634

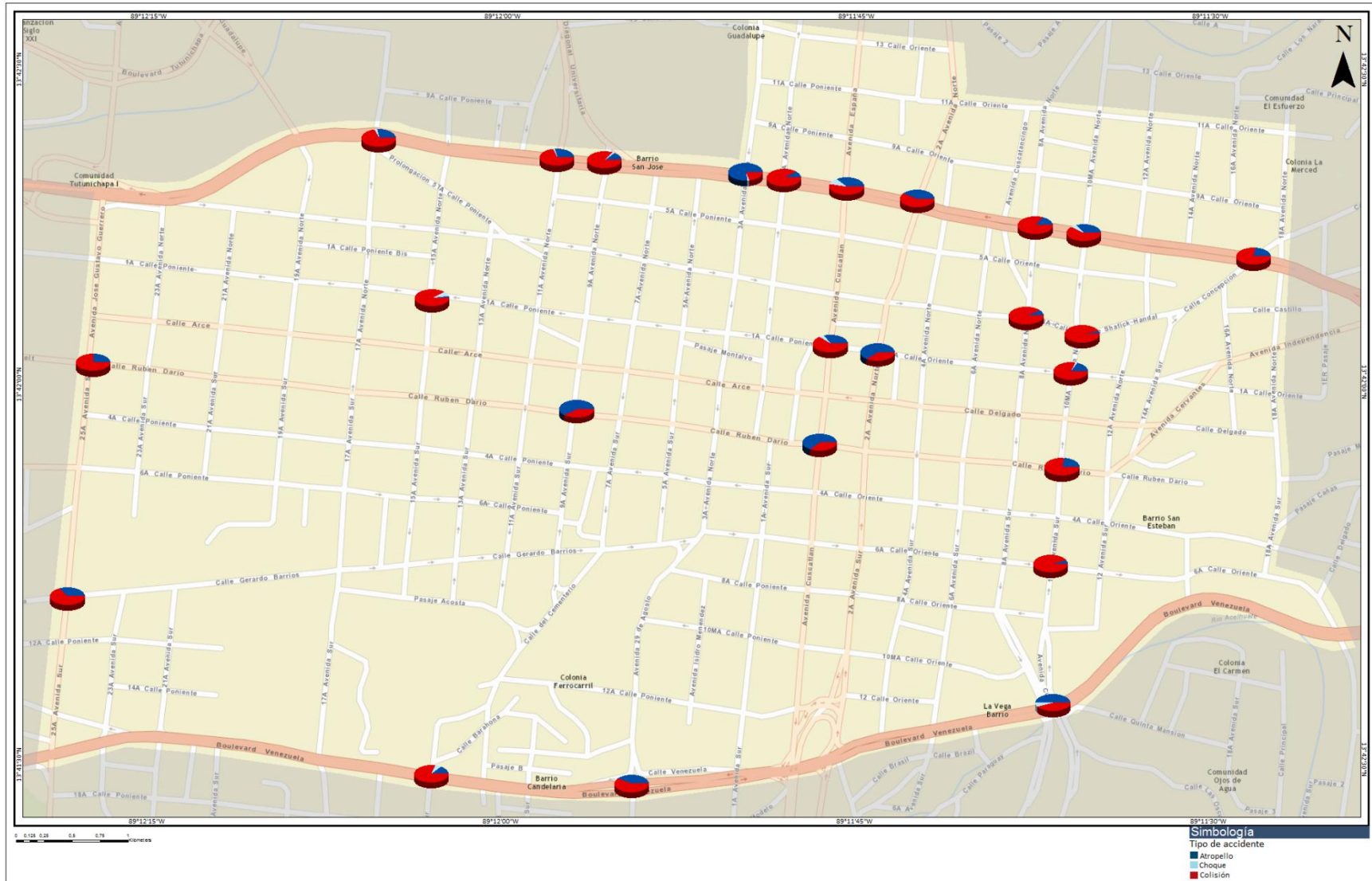
b) Ubicación geográfica de intersecciones con alta frecuencia de accidentes de tránsito



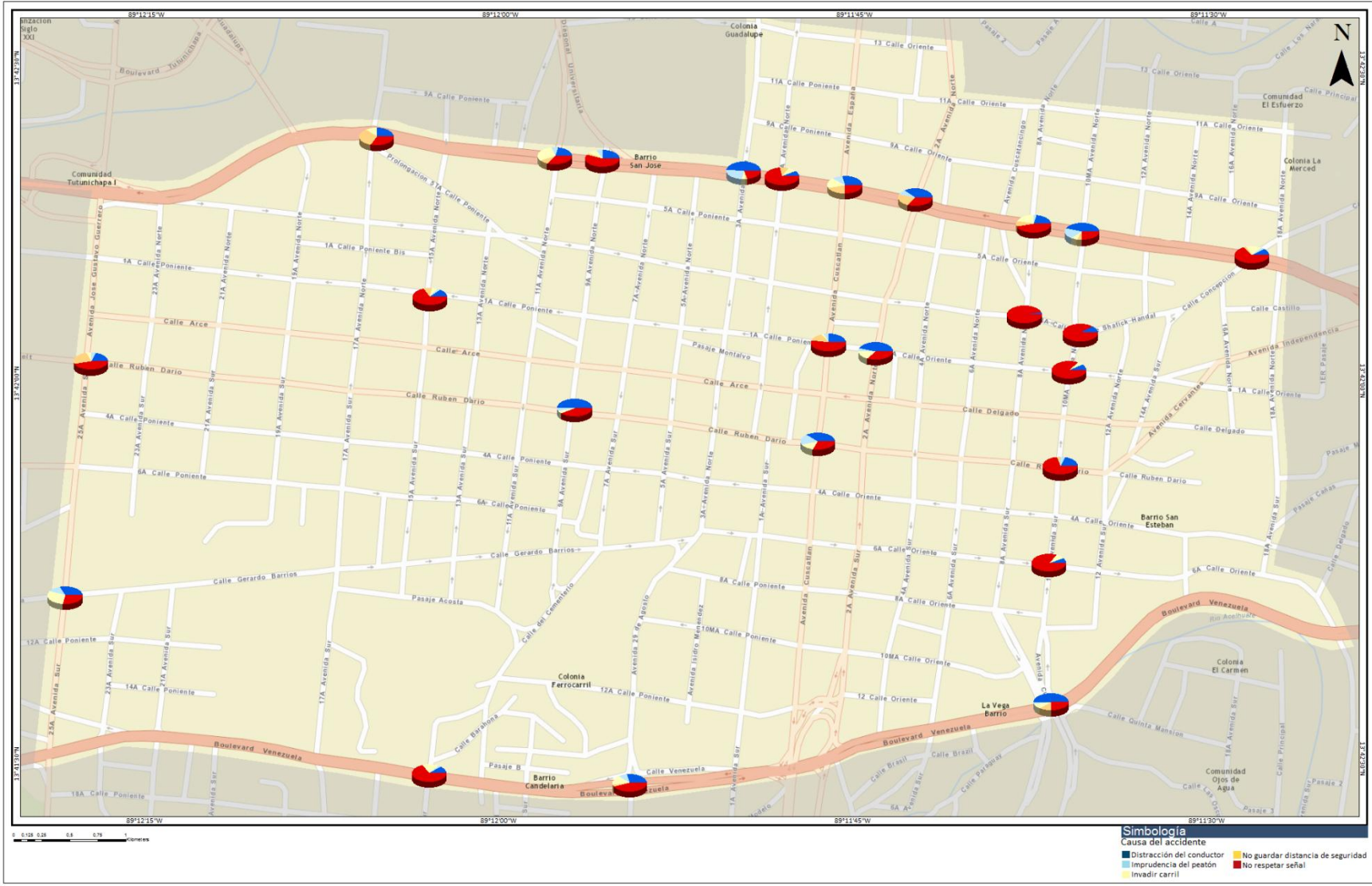
c) Ubicación geográfica: Resultado del accidente



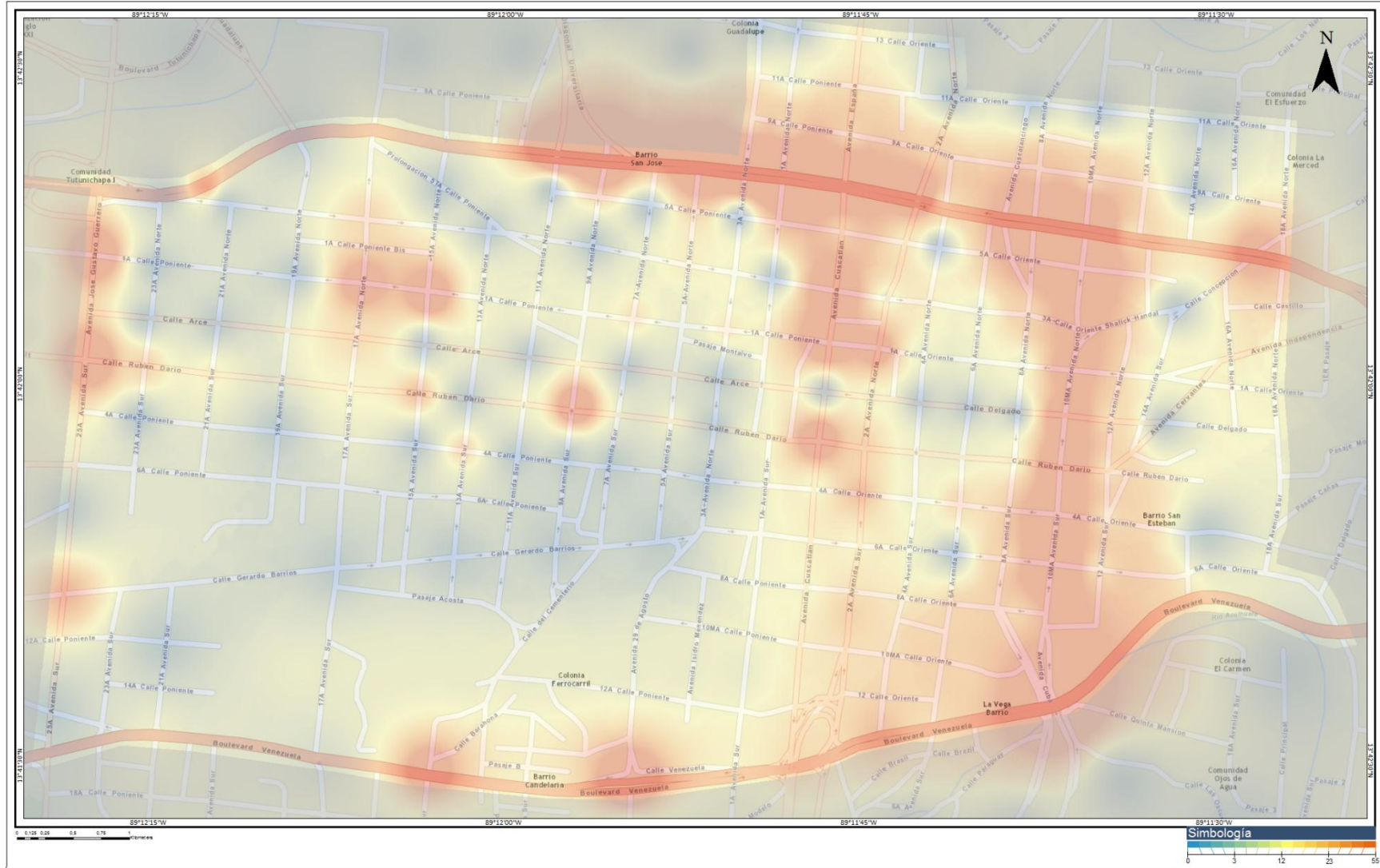
d) Ubicación geográfica: Tipo de accidente.



e) Ubicación geográfica: Causa del accidente



f) Frecuencia y proximidad de los accidentes de tránsito.



3.3.2 Distrito 1

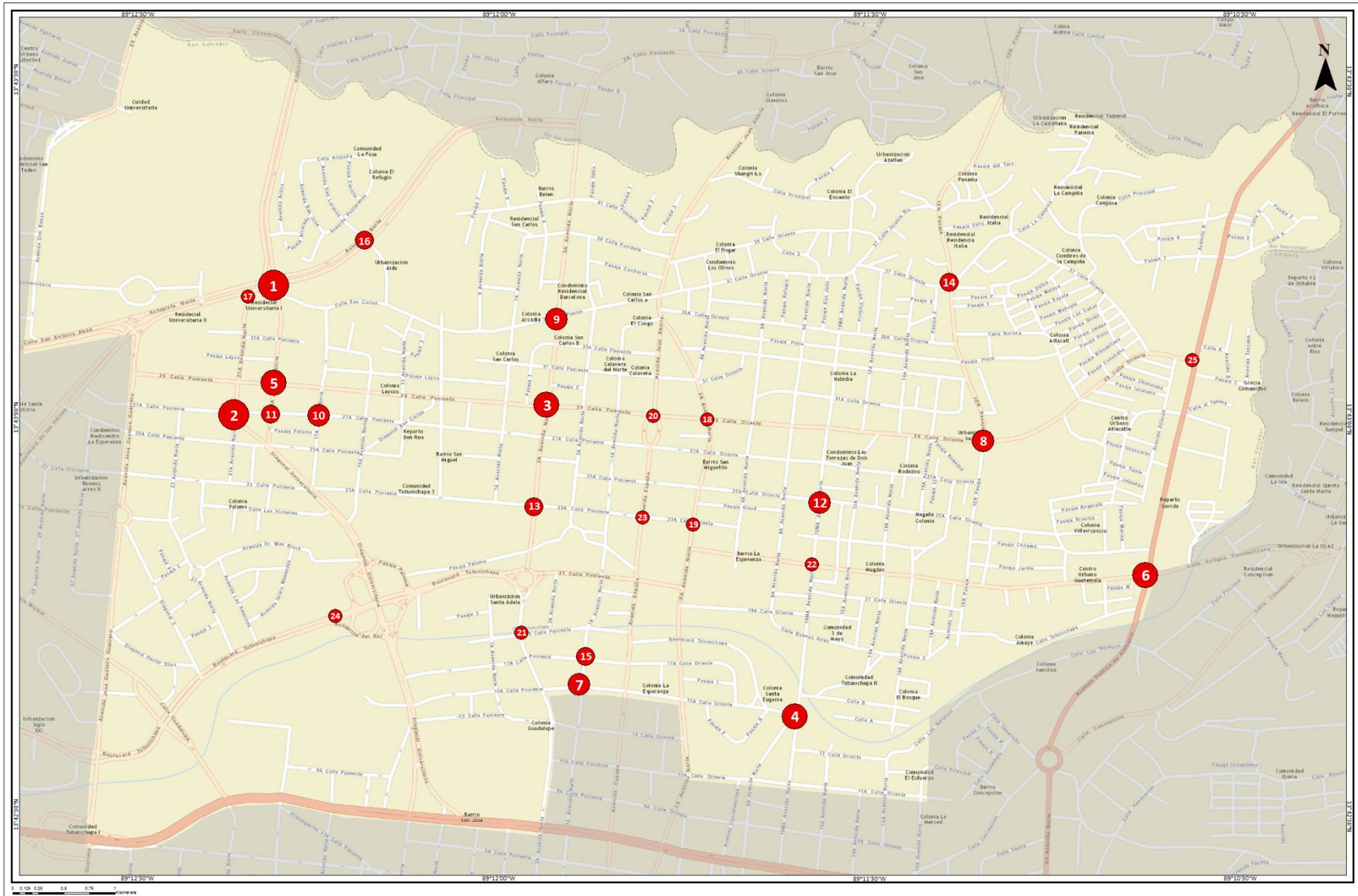
Datos generales

- Población Total: 32,789.
- Extensión Territorial: 5.2 km².
- Densidad de población: 5,752 hab/km².
- Límites:
 - Norte: Municipios de Mejicanos y Cuscatancingo.
 - Sur: Alameda Juan Pablo II y 13^a calle poniente.
 - Este: Municipios de Ciudad Delgado y Soyapango.
 - Oeste: Boulevard de los Héroes y avenida Don Bosco.

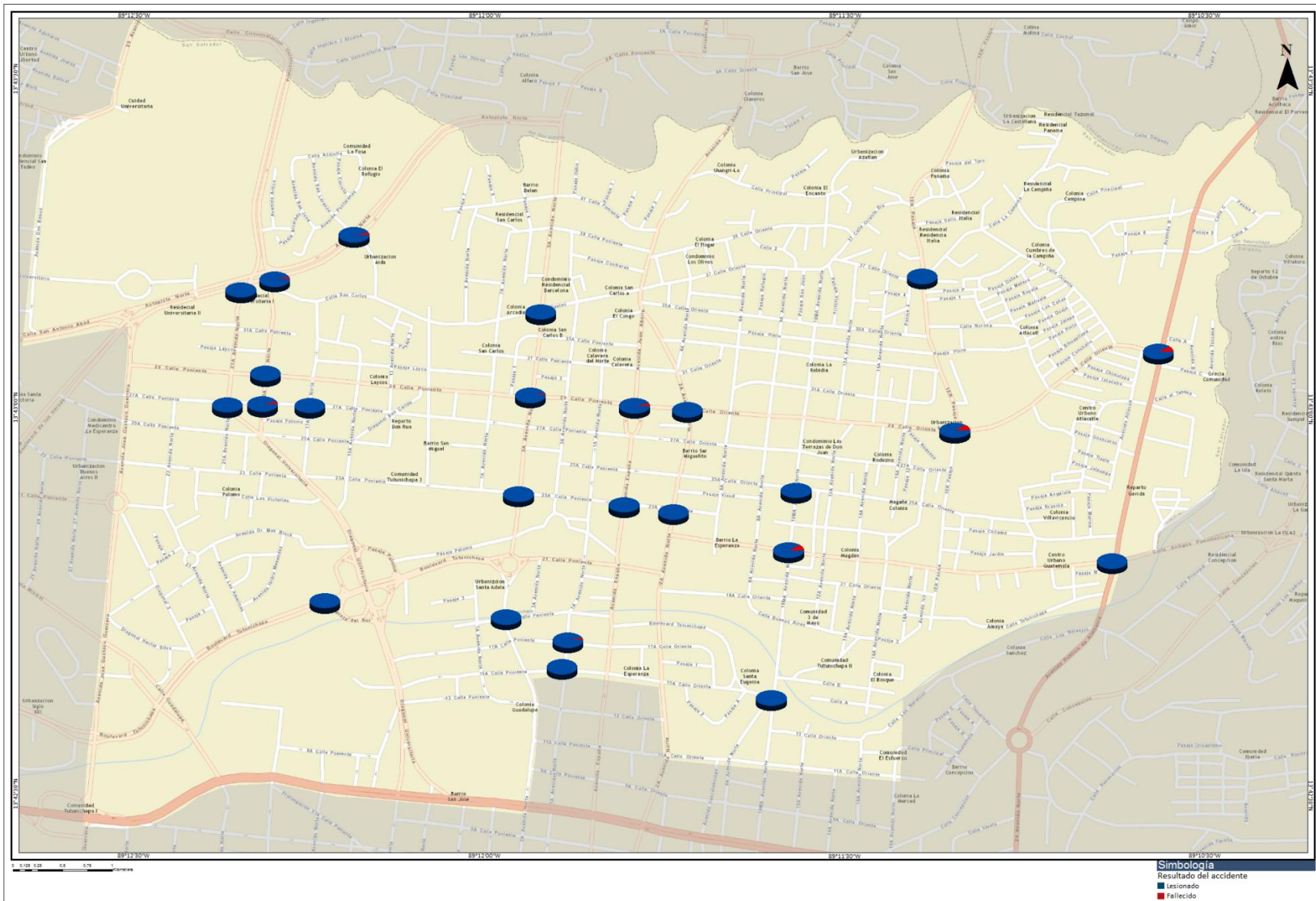
a) Intersecciones con alta frecuencia de accidentes de tránsito

Correlativo	Intersección	Frecuencia	Fallecidos	Lesionados	Atropellos	Choques	Colisiones
1	Autopista Norte Y 19ª Avenida Norte	49	2	47	3	1	44
2	27ª Calle Poniente Y 21ª Avenida Norte	48	0	48	1	0	43
3	29ª Calle Poniente Y 5ª Avenida Norte	38	1	37	10	1	25
4	10ª Avenida Norte y 15ª Calle Poniente	35	0	35	4	4	30
5	29ª Calle Poniente Y 19ª Avenida Norte	35	0	35	0	0	35
6	24ª Avenida Norte Y Calle 5 De Noviembre	33	0	33	10	0	23
7	15ª Calle Poniente Y 1ª Avenida Norte	30	0	30	0	0	30
8	20ª Avenida Norte Y 29ª Calle poniente	29	3	26	9	0	20
9	5ª Avenida Norte Y 37ª Calle poniente	28	0	28	8	0	20
10	27ª Calle Poniente Y 17ª Avenida Norte	28	0	28	0	0	28
11	27ª Calle Poniente Y 19ª Avenida Norte	26	1	25	0	1	24
12	10ª Avenida Norte y 25ª Calle Oriente	25	0	25	3	0	21
13	23ª Calle Poniente Y 5ª Avenida Norte	25	0	25	7	3	15
14	37ª Calle Oriente Y 20ª Avenida Norte	24	0	24	1	0	22
15	17ª Calle Poniente Y 1ª Avenida Norte.	24	1	23	1	0	22
16	Autopista Norte Y Avenida San Lorenzo	24	1	23	4	0	20
17	Autopista Norte Y Boulevard Universitario	23	0	23	2	0	20
18	29ª Calle Poniente Y Avenida Monseñor Romero	22	0	22	3	0	19
19	Avenida Monseñor Romero Y 23ª Calle Oriente	22	0	22	1	0	21
20	29ª Calle Oriente Y Avenida España	22	1	21	4	0	18
21	19ª Calle Poniente Y 5ª Avenida Norte	22	0	22	3	0	19
22	10ª Avenida Norte Y Calle 5 de Noviembre	21	2	19	6	0	15
23	23ª Calle Oriente Y Avenida España	21	0	21	1	0	20
24	Boulevard Tutunichapa Frente Al Centro Judicial Isidro Menéndez	21	0	21	9	1	12
25	Diagonal CIPACTLY Y 24ª Avenida Norte	20	2	18	8	2	10
Total		695	14	681	98	13	576

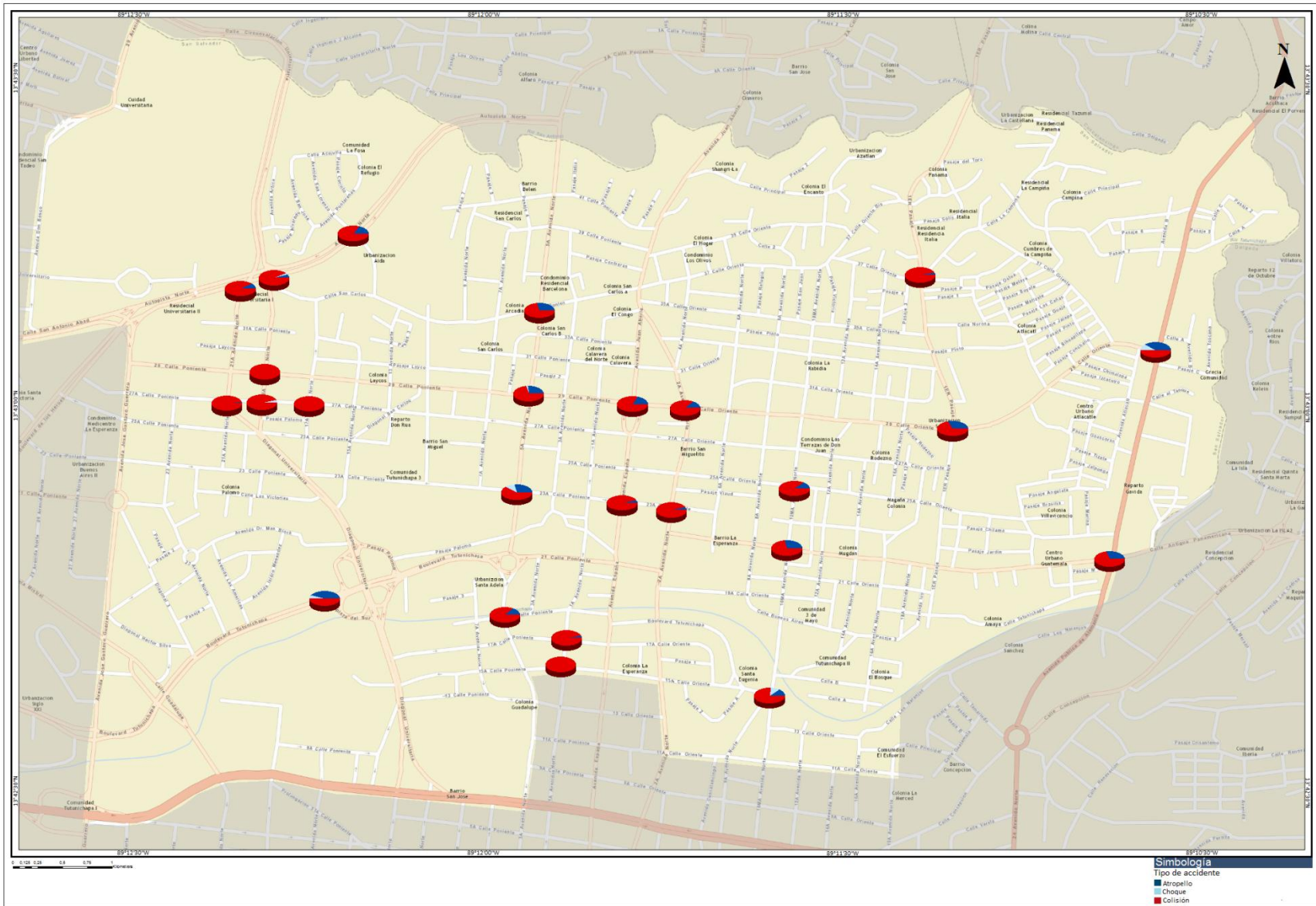
b) Ubicación geográfica de intersecciones con alta frecuencia de accidentes de tránsito



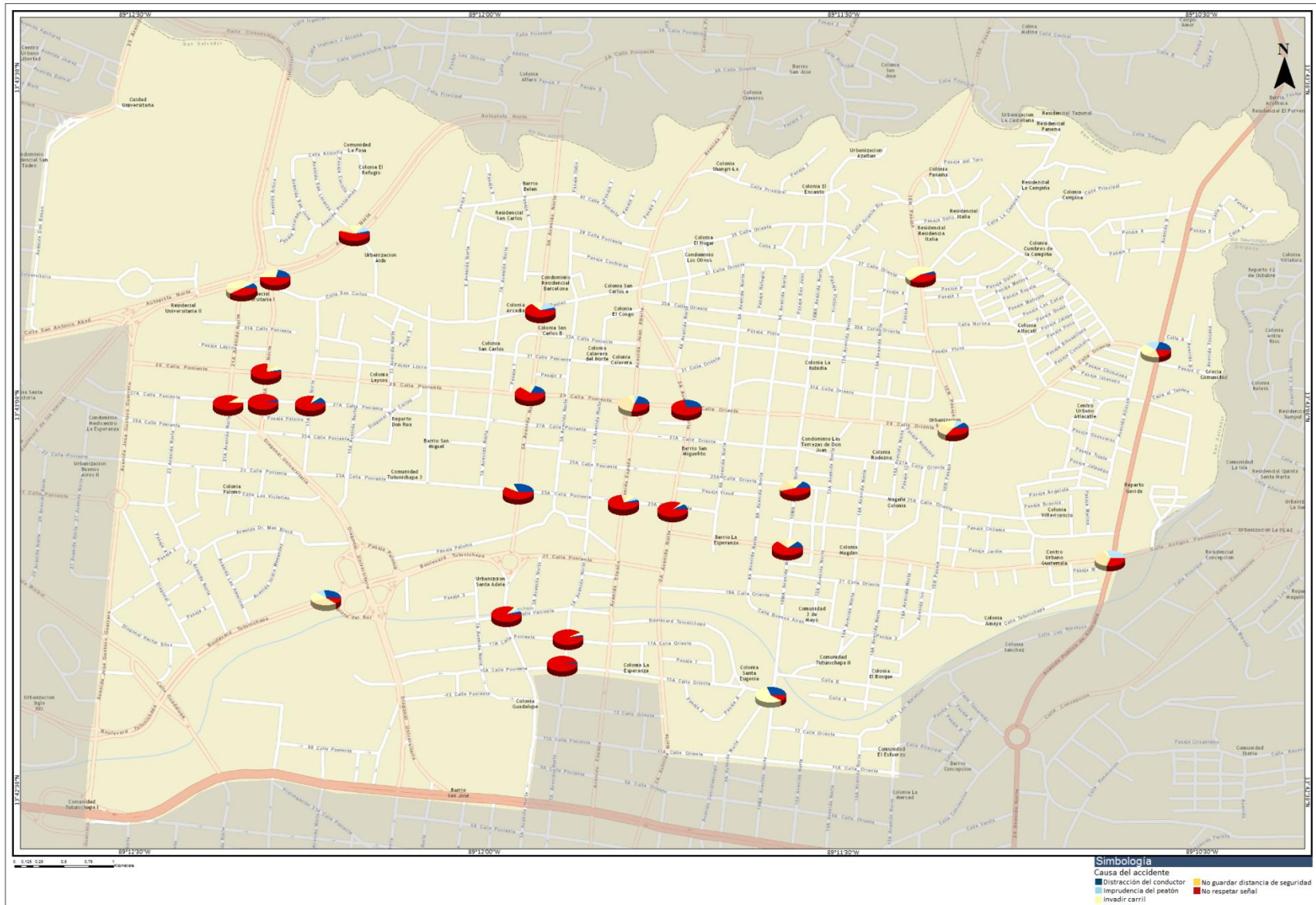
c) Ubicación geográfica: Resultado del accidente



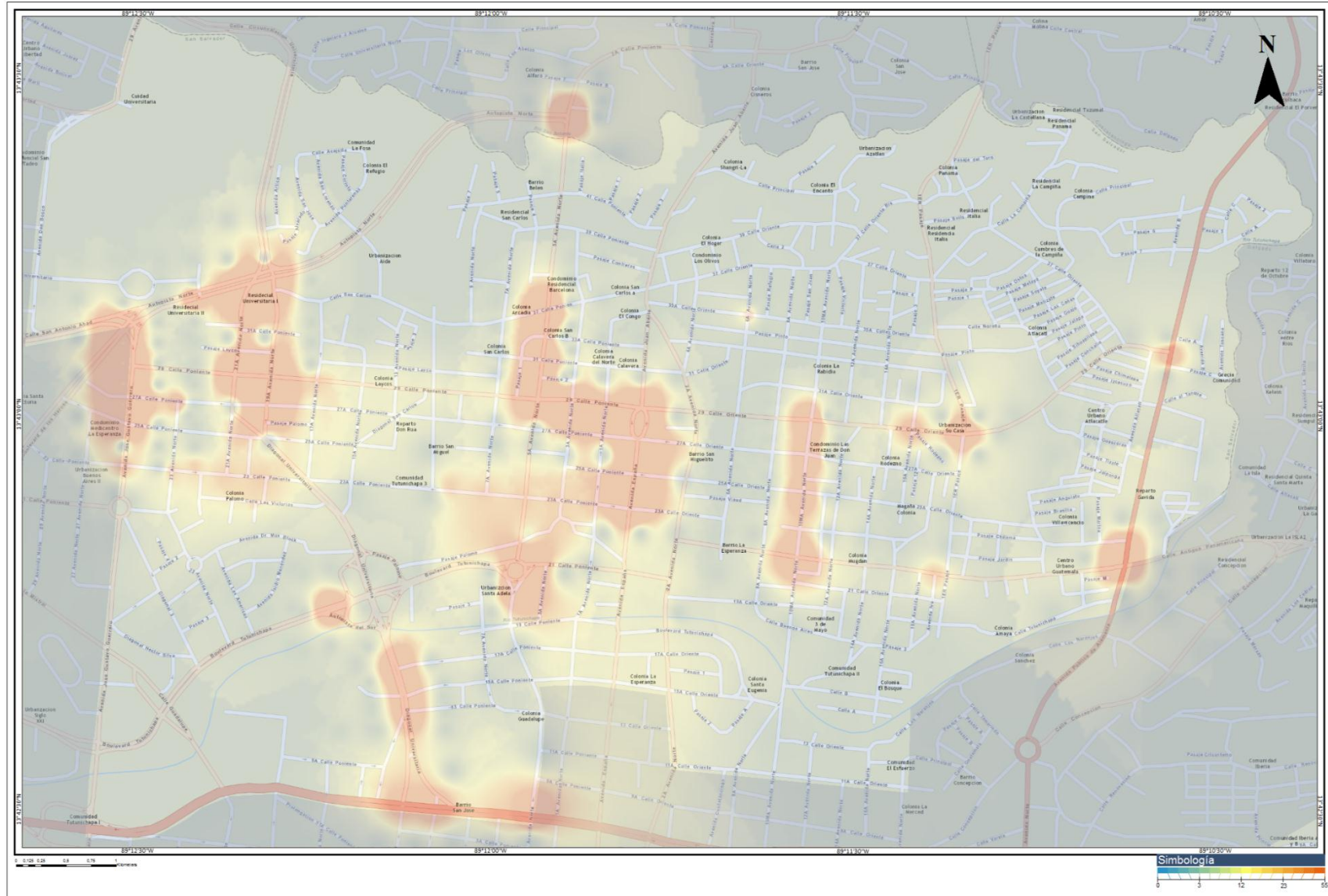
d) Ubicación geográfica: Tipo de accidente.



e) Ubicación geográfica: Causa del accidente



e) Frecuencia y proximidad de los accidentes de tránsito.



3.3.3 Distrito 2

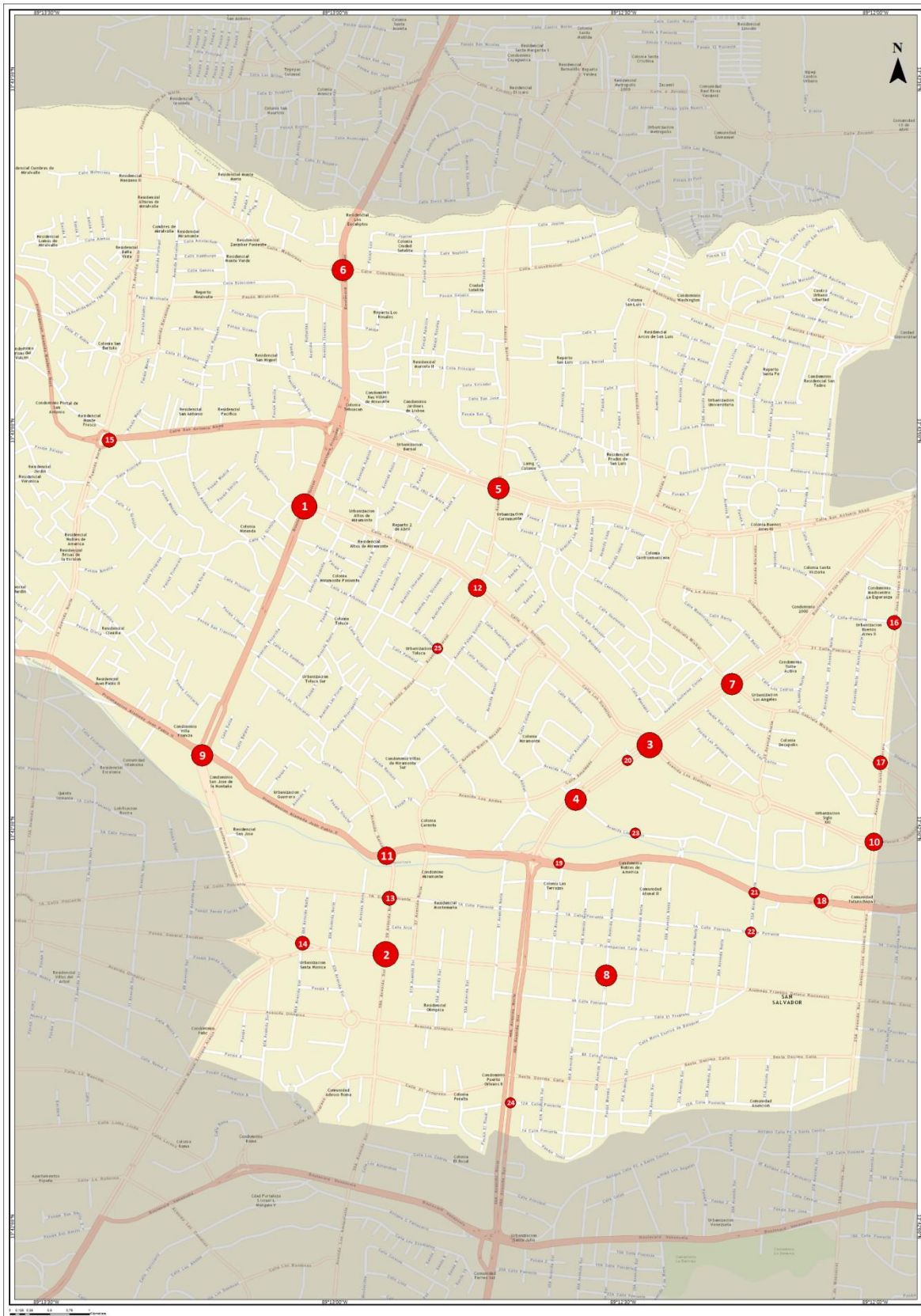
Datos generales

- Población Total: 78,271.
- Extensión Territorial: 11.2 km².
- Densidad de población: 6,988 hab/km².
- Límites:
 - Norte: Municipio de Mejicanos.
 - Sur: Alameda Enrique Araujo.
 - Este: 25^a avenida norte.
 - Oeste: Quebrada la mascota.

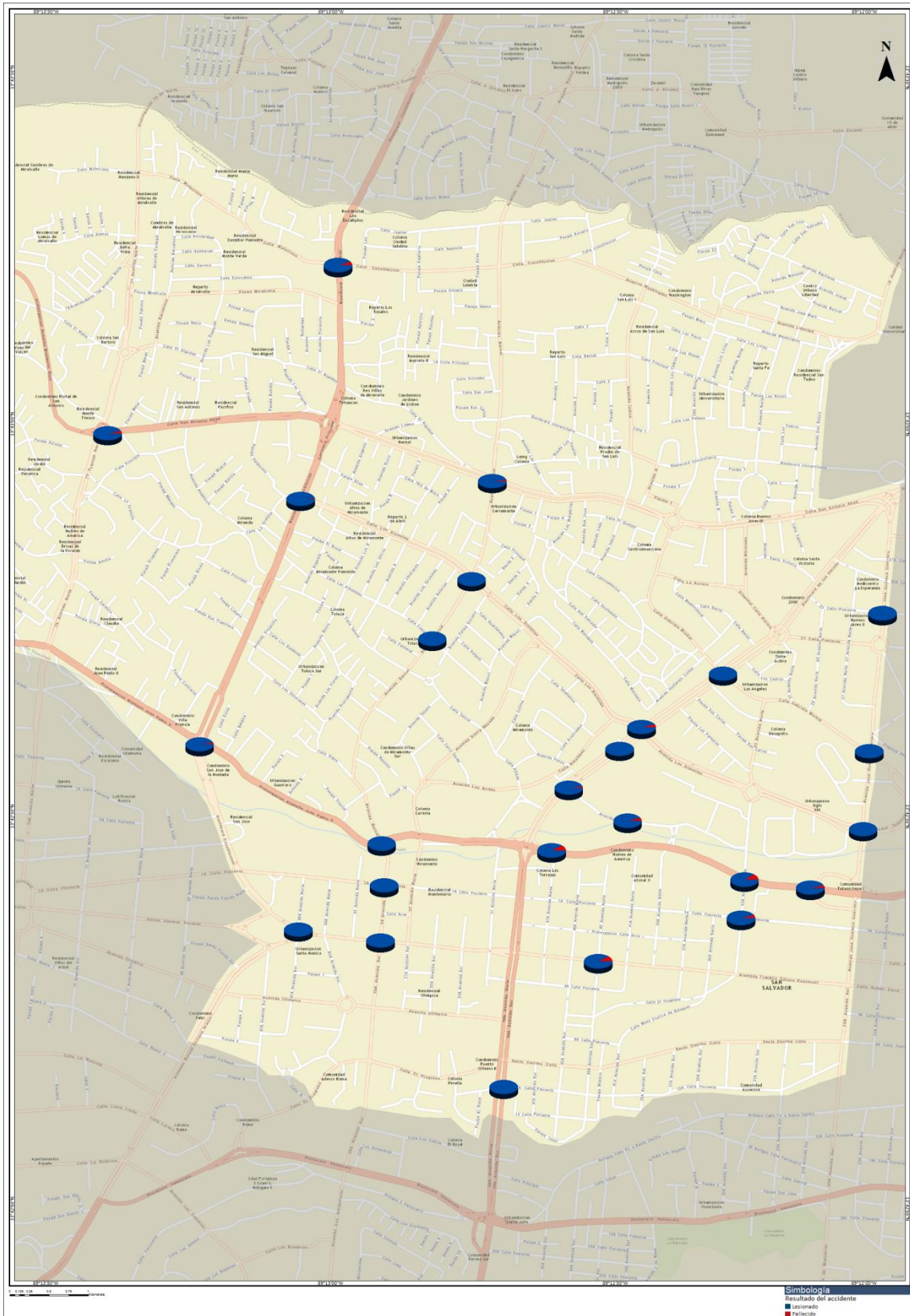
a) Intersecciones con alta frecuencia de accidentes de tránsito.

Correlativo	Intersección	Frecuencia	Fallecidos	Lesionados	Atropellos	Choques	Colisiones
1	Boulevard Constitución Y Calle Sisimiles	50	1	49	2	14	33
2	Alameda Roosevelt Y 59ª Avenida Norte	47	0	47	10	1	35
3	Boulevard Los Héroes Y Avenida Los Sisimiles	43	2	41	7	6	29
4	Boulevard Los Héroes Y Avenida Los Andes	34	1	33	15	0	18
5	Calle San Antonio Abad y Avenida Bernal	34	1	33	11	0	23
6	Boulevard Constitución y Calle Motocross	34	2	32	4	0	29
7	Boulevard Los Héroes y Calle Gabriela Mistral	33	0	33	8	2	23
8	Alameda Roosevelt Y 43ª Avenida Sur	32	3	29	3	0	29
9	Boulevard Constitución Y Alameda Juan Pablo II	32	1	31	7	2	22
10	Boulevard Tutunichapa y 25ª Avenida Norte	30	0	30	8	1	21
11	Alameda Juan Pablo II Y Avenida Bernal	27	0	27	4	1	20
12	Avenida Bernal Y Calle Sisimiles	26	0	26	2	0	23
13	1ª Calle Poniente y 59ª Norte	24	0	24	0	0	24
14	Alameda Roosevelt Y 65ª Avenida Sur	24	0	24	5	0	19
15	Calle San Antonio Abad Y 75ª Avenida Norte	24	1	23	3	2	19
16	23ª Calle Poniente Y 25ª Avenida Norte	24	0	24	2	0	21
17	Calle Gabriela Mistral y 25ª Avenida Norte	23	0	23	2	0	21
18	Alameda Juan Pablo II Frente Al Médico Quirúrgico	23	1	22	10	0	13
19	Alameda Juan Pablo II Y 47ª Avenida Norte	21	2	19	7	0	13
20	Boulevard Los Héroes Frente A Metrocentro	20	0	20	4	3	11
21	Alameda Juan Pablo II Y 33ª Avenida Norte	19	2	17	3	3	13
22	1ª Calle Poniente Y 33ª Avenida Norte	19	1	18	1	0	18
23	Avenida Los Andes Frente A Metrocentro	18	1	17	9	2	6
24	49ª Avenida Sur y 12 Calle Poniente	18	0	18	2	0	15
25	Calle Toluca y Avenida Bernal	18	0	18	1	0	17
Total		697	19	678	130	37	515

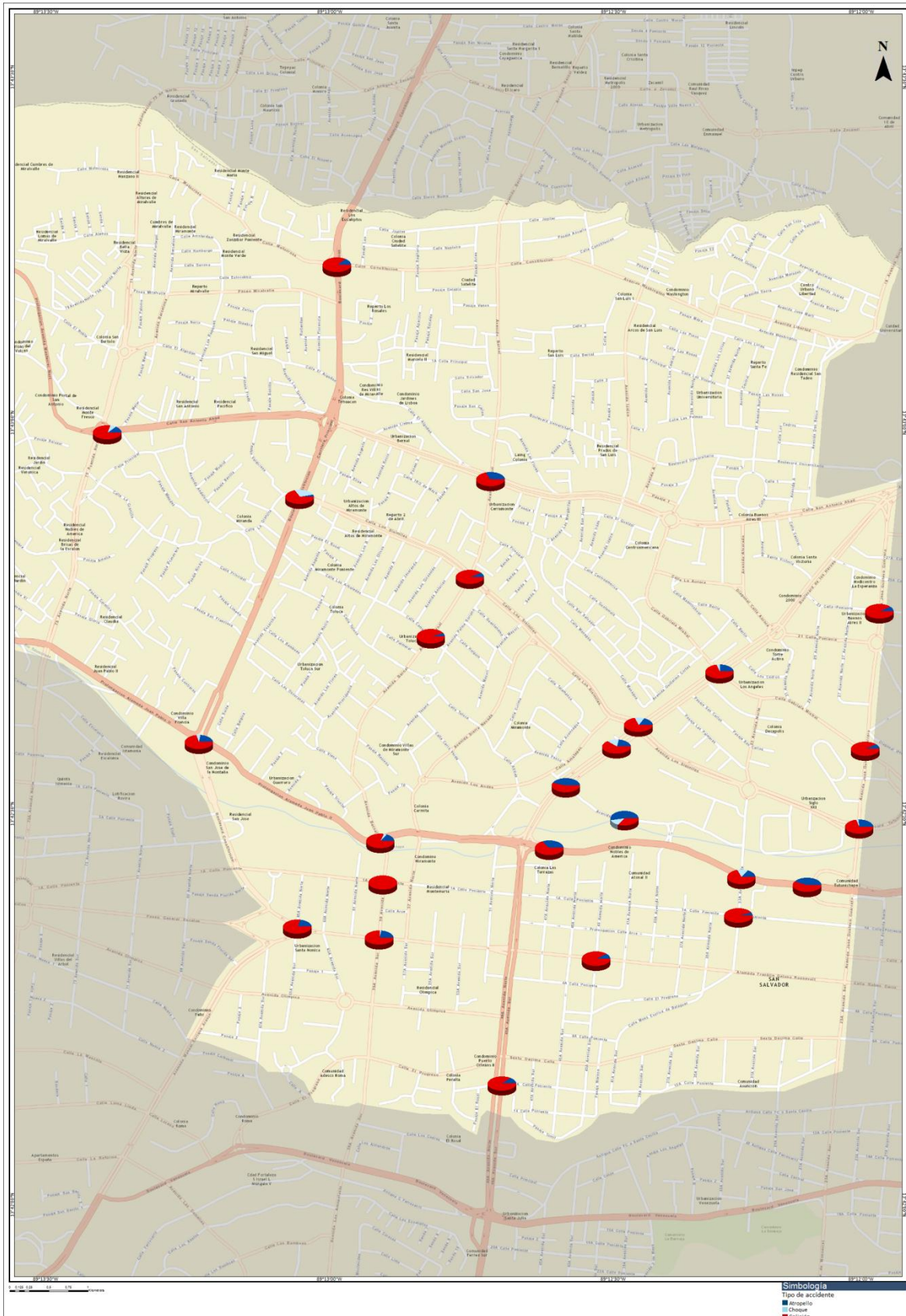
b) Ubicación geográfica de intersecciones con alta frecuencia de accidentes de tránsito



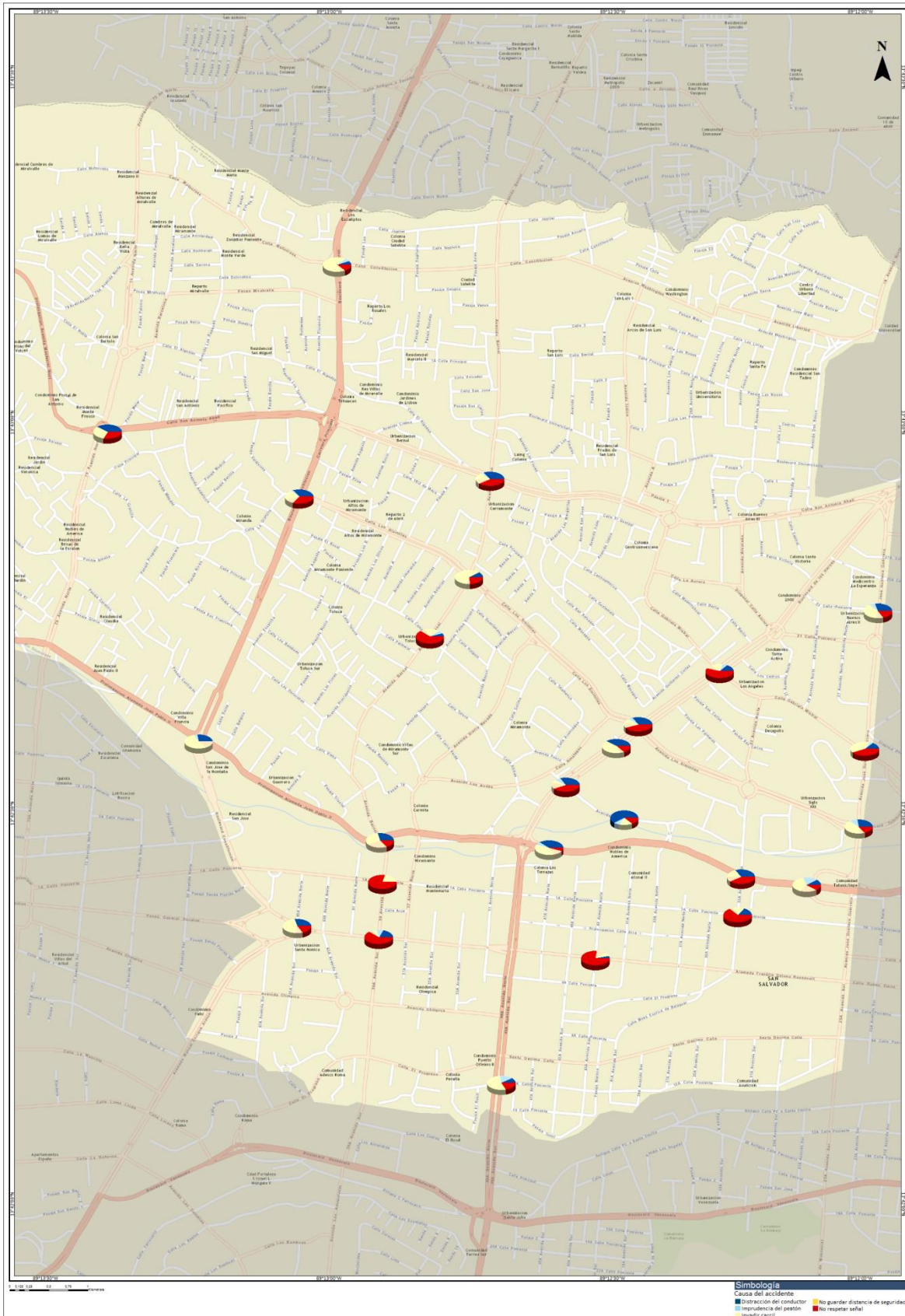
c) Ubicación geográfica: Resultado del accidente.



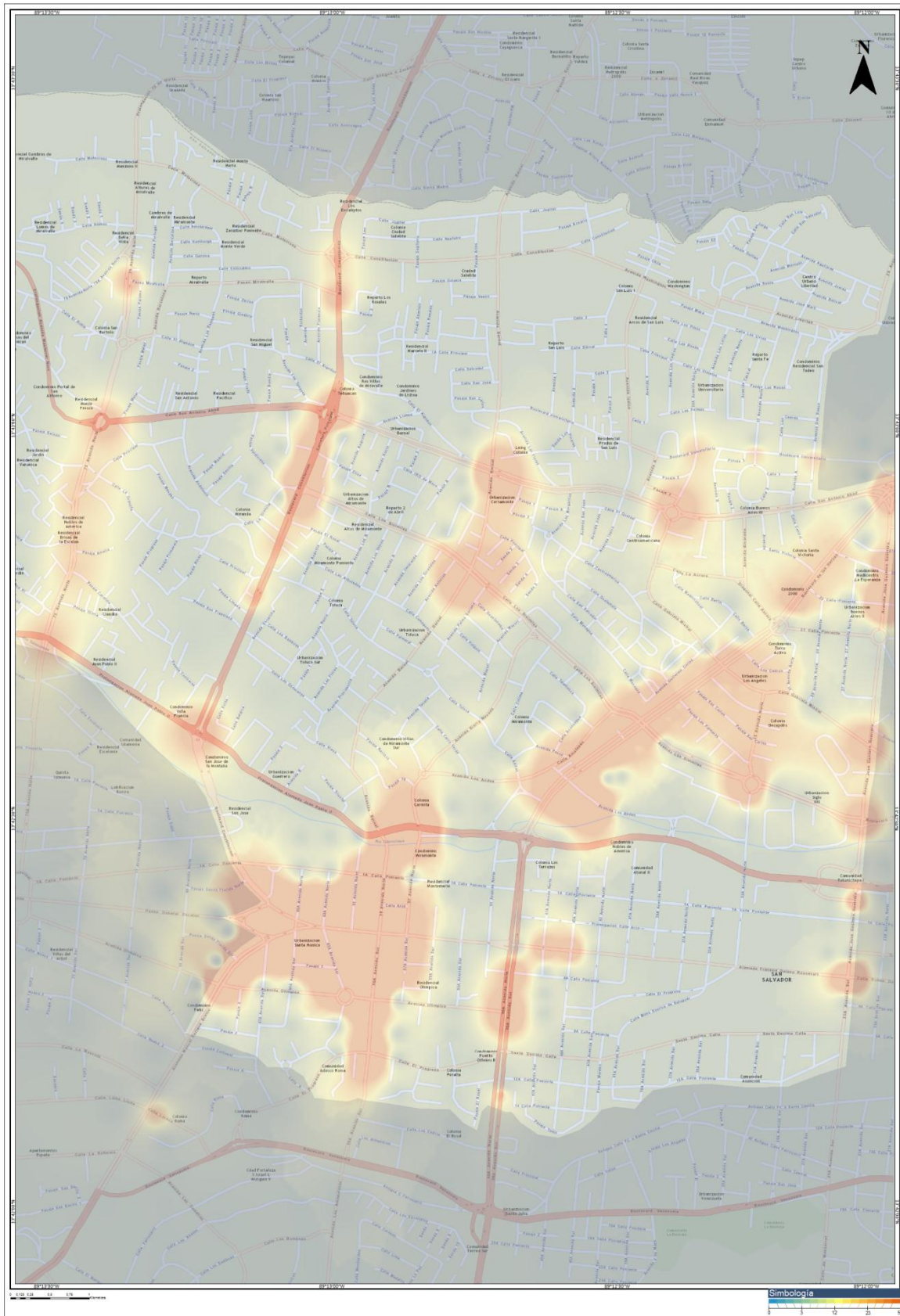
d) Ubicación geográfica: Tipo de accidente.



e) Ubicación geográfica: Causa del accidente.



f) Frecuencia y proximidad de los accidentes de tránsito.



3.3.4 Distrito 3

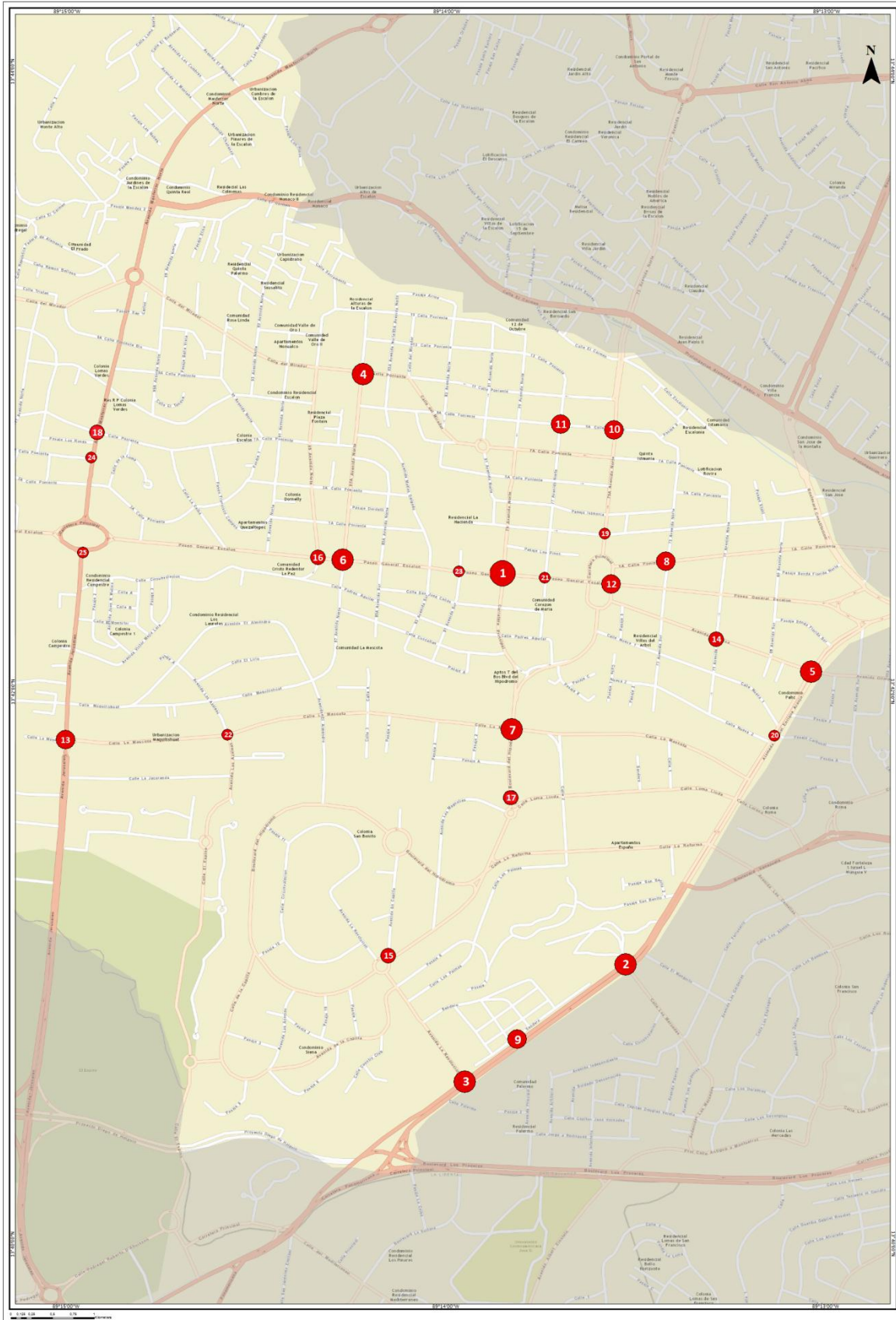
Datos generales

- Población Total: 32,081.
- Extensión Territorial: 18.3 km².
- Densidad de población: 2,081 hab/km².
- Límites:
 - Norte: Municipios de Mejicanos y Nejapa.
 - Sur: Municipios de Santa Tecla y Antiguo Cuscatlán.
 - Este: Alameda Manuel Enrique Araujo.
 - Oeste: Santa Tecla.

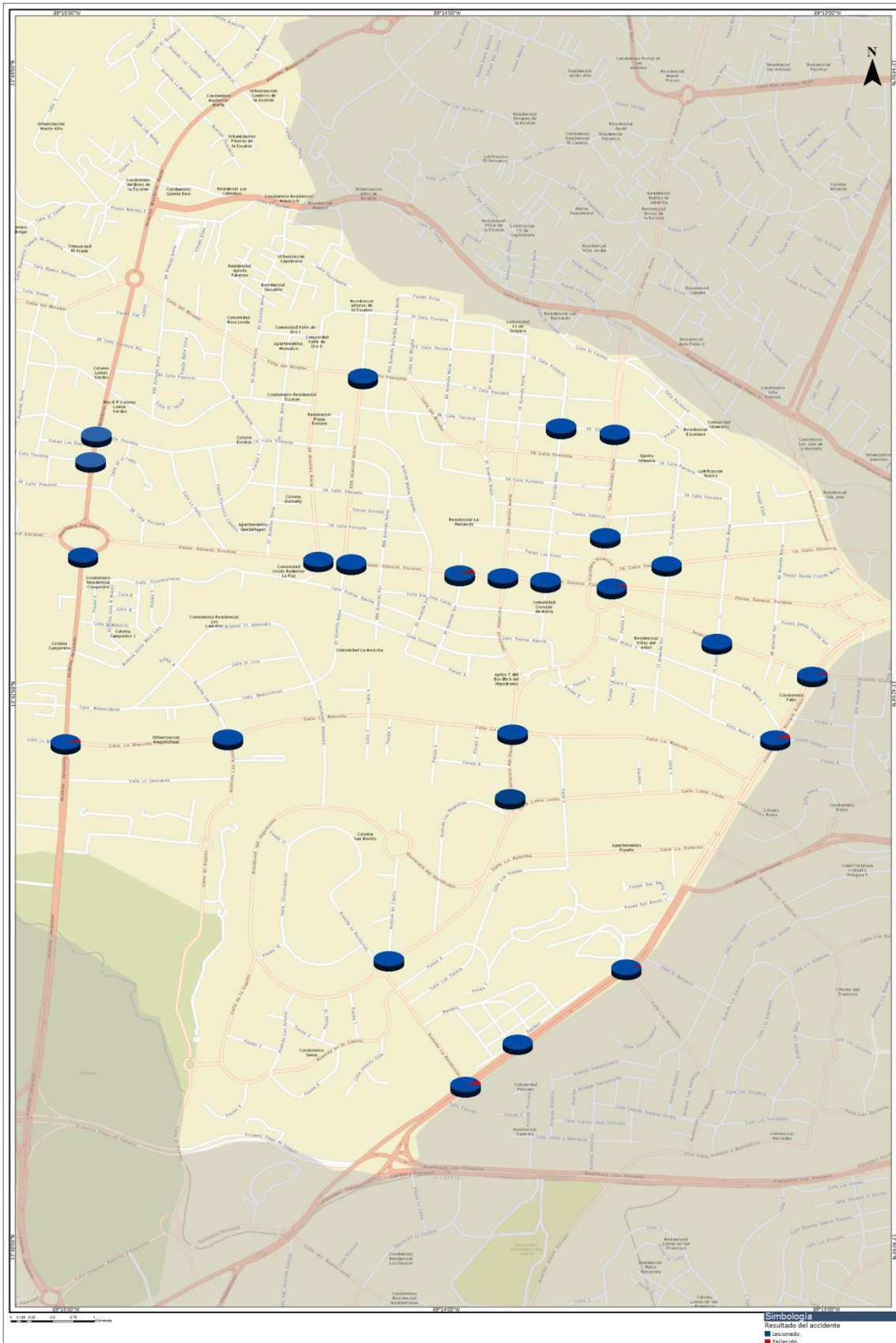
a) Intersecciones con alta frecuencia de accidentes de tránsito.

Correlativo	Intersección	Frecuencia	Fallecidos	Lesionados	Atropellos	Choques	Colisiones
1	Paseo General Escalón Y 79ª Avenida Norte	43	0	43	4	2	37
2	Alameda Manuel Araujo Y Calle Las Mercedes	32	1	31	2	3	26
3	Alameda Manuel Araujo Y Avenida Revolución	31	2	29	7	2	22
4	Calle Al Mirador Y 87ª Avenida Norte	31	0	31	0	0	31
5	Alameda Manuel Araujo Y Avenida Olímpica	29	1	28	6	0	22
6	Paseo General Escalón Y 87ª Avenida Norte	29	0	29	3	0	25
7	Calle La Mascota Y 79ª Avenida Sur	28	0	28	3	0	25
8	1ª Calle Poniente Y 73ª Avenida Norte	27	0	27	6	0	21
9	Alameda Manuel Araujo Frente A La Feria Internacional	27	0	27	2	10	15
10	9ª Calle Poniente Y 75ª Avenida Norte	26	0	26	1	0	24
11	9ª Calle Poniente Y 77ª Avenida Norte	25	0	25	0	0	25
12	Paseo General Escalón Y 75ª Avenida Norte	24	1	23	3	0	19
13	Avenida Jerusalén Y Calle la Mascota	24	1	23	2	3	17
14	Avenida Olímpica Y 71ª Avenida Sur	22	0	22	0	1	19
15	Boulevard El Hipódromo Y Avenida Capilla	21	0	21	1	2	18
16	Paseo General Escalón Y 89ª Avenida Norte	21	0	21	3	0	16
17	Carretera Panamericana Frente A La Ceiba De Guadalupe	20	1	19	5	1	13
18	7ª Calle Poniente Y 75ª Avenida Norte	20	0	20	0	2	17
19	3ª Calle Poniente Y 75ª Avenida Norte	19	0	19	1	0	18
20	Alameda Manuel Araujo Y Pasaje Carbonel	18	1	17	3	3	12
21	Paseo General Escalón Y 77ª Avenida Norte	18	0	18	3	0	14
22	Calle La Mascota Y Avenida Las Azaleas	18	0	18	0	1	17
23	7ª Calle Poniente Y Avenida Masferrer Norte	18	0	18	4	1	13
24	5ª Calle Poniente Y Avenida Masferrer Norte	18	0	18	3	0	15
25	Redondel Masferrer	17	0	17	1	8	7
Total		606	8	598	63	39	488

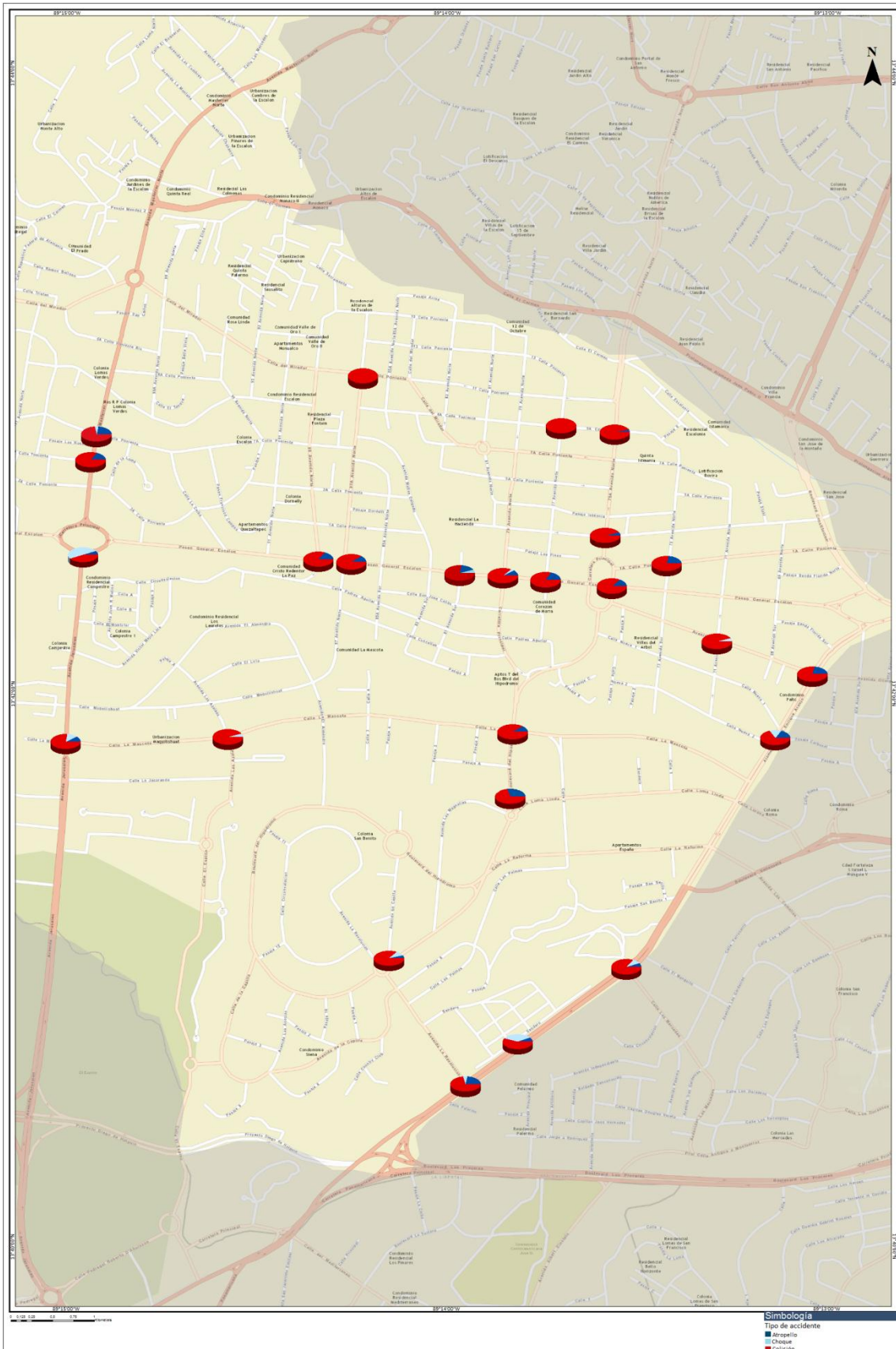
b) Ubicación geográfica de intersecciones con alta frecuencia de accidentes de tránsito



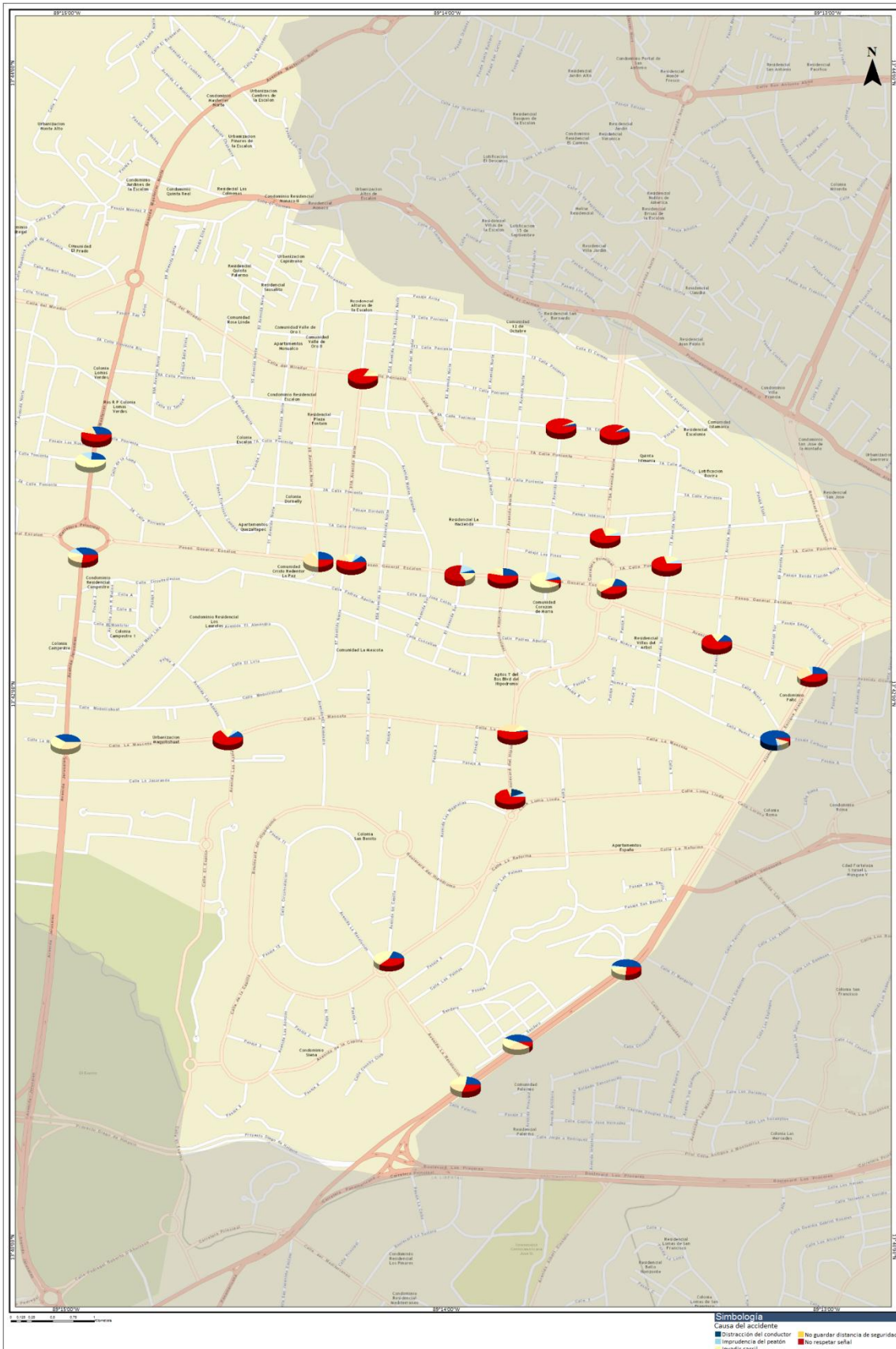
c) Ubicación geográfica: Resultado del accidente



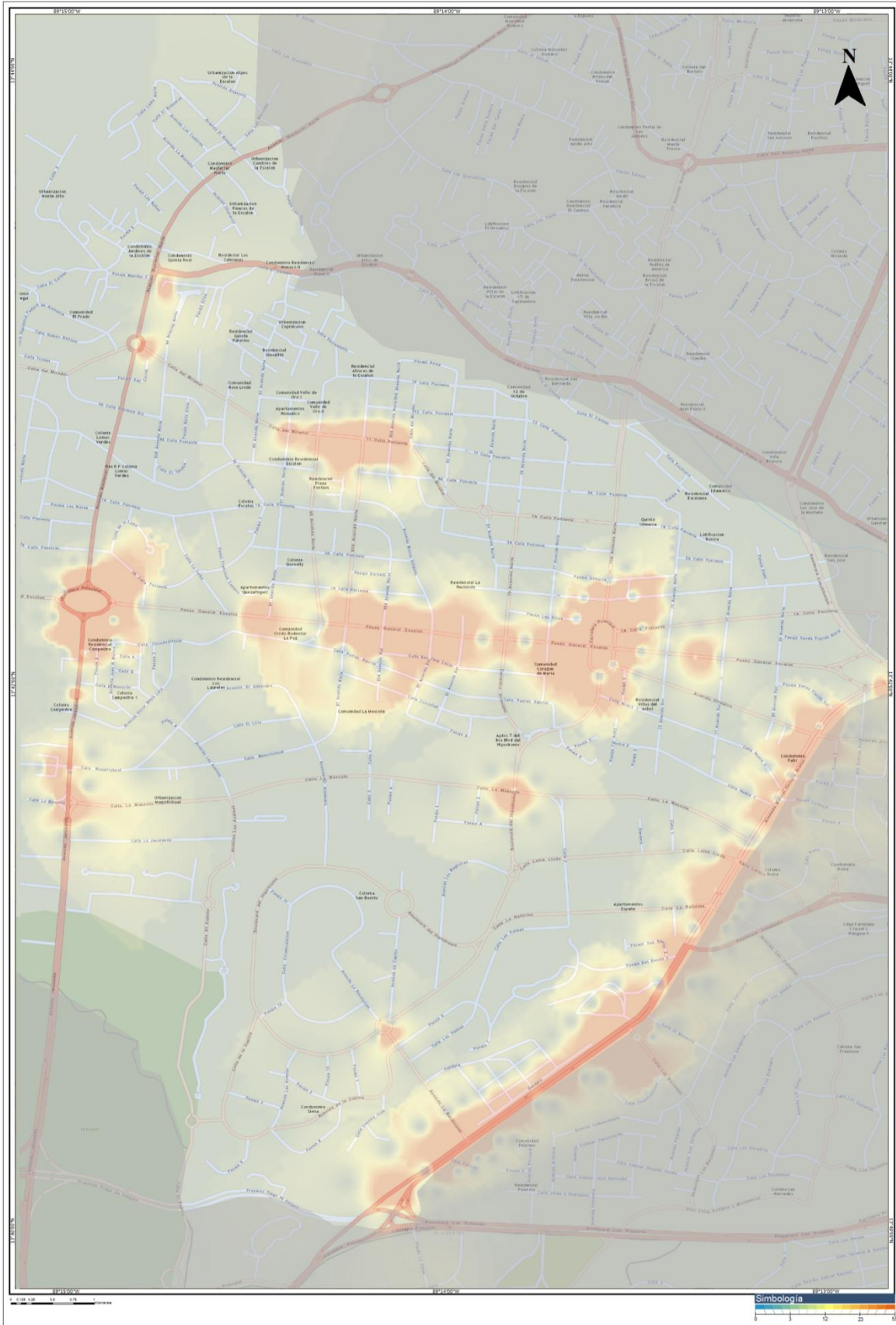
d) Ubicación geográfica: Tipo de accidente



e) Ubicación geográfica: Causa del accidente



f) Frecuencia y proximidad de los accidentes de tránsito.



3.3.5 Distrito 4

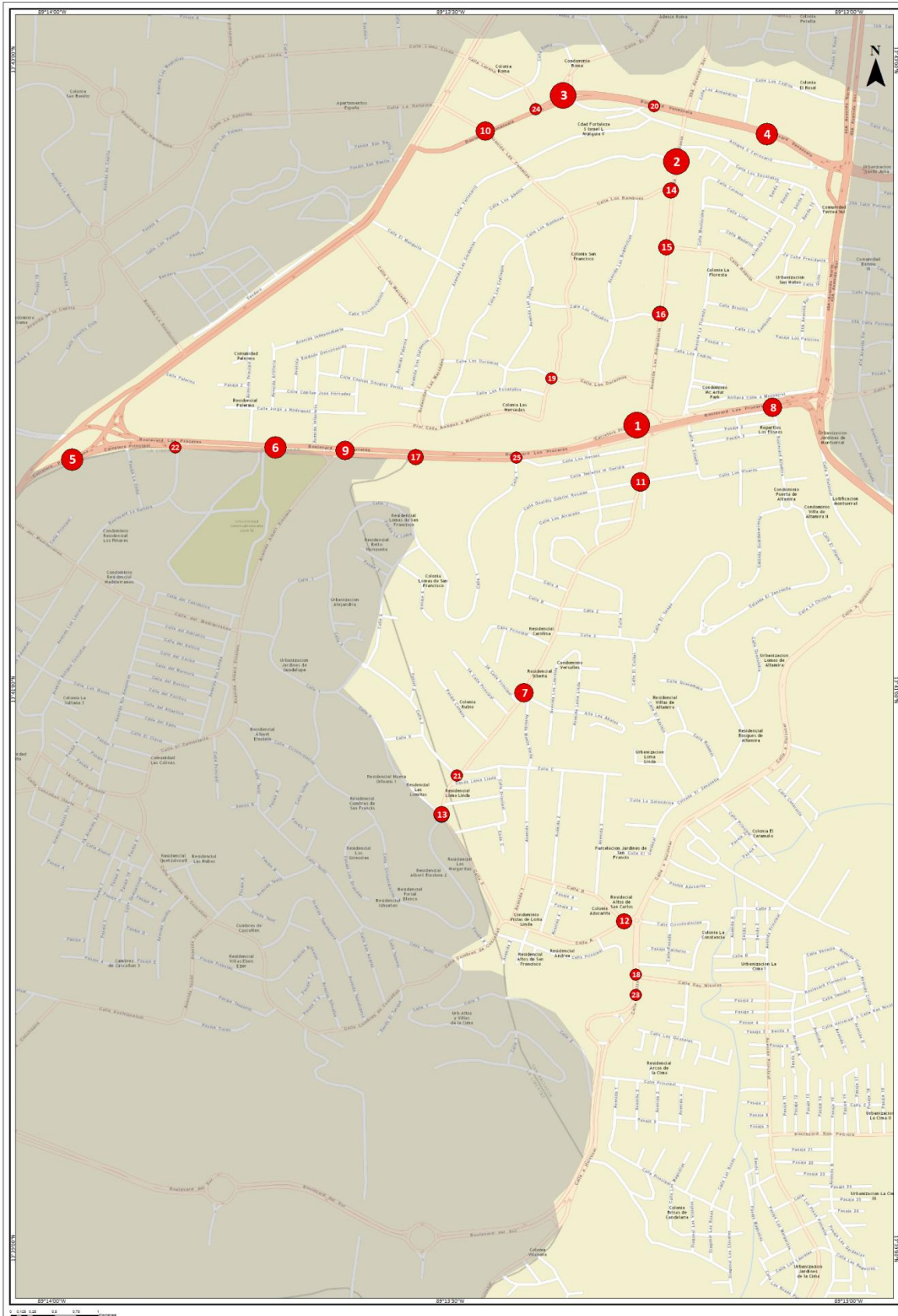
Datos generales

- Población Total: 32,482.
- Extensión Territorial: 12.4 km².
- Densidad de población: 2,620 hab/km².
- Límites:
 - Norte: Calle El Progreso.
 - Sur: Municipio de Antiguo Cuscatlán.
 - Este: 49ª Av. Sur y Autopista a Comalapa
 - Oeste: Alameda Manuel Enrique Araujo.

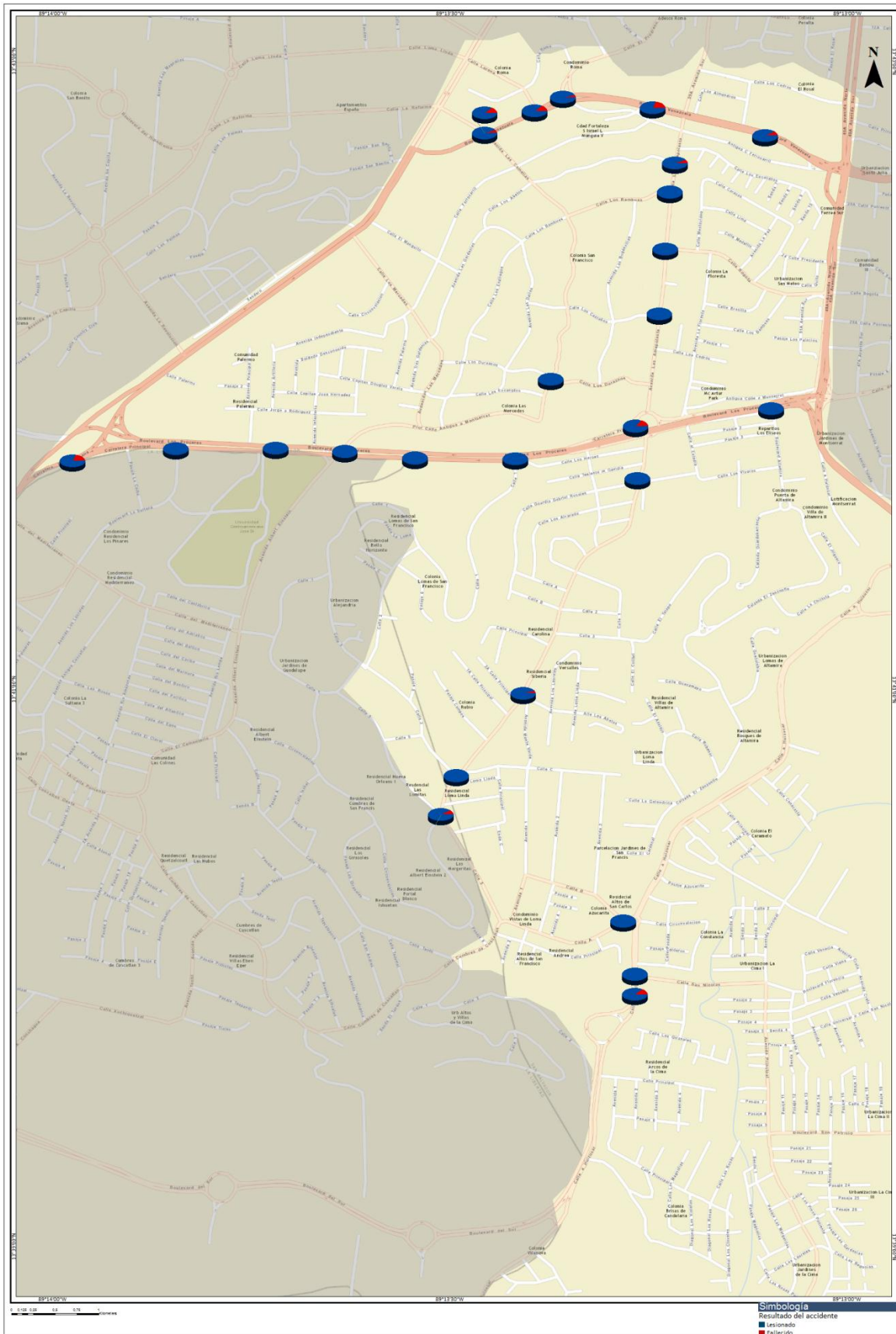
a) Intersecciones con alta frecuencia de accidentes de tránsito.

Correlativo	Intersección	Frecuencia	Fallecidos	Lesionados	Atropellos	Choques	Colisiones
1	Boulevard Los Próceres Y Árbol De La Paz	49	8	41	10	0	38
2	Calle Los Abetos Y Avenida Las Amapolas	43	3	40	7	1	35
3	Boulevard Venezuela Y Calle El Progreso	42	2	40	8	4	30
4	Boulevard Venezuela Frente A La Terminal De Occidente	32	3	29	21	2	7
5	Boulevard Los Próceres Y Alameda Araujo	26	5	21	8	5	13
6	Boulevard Los Próceres Frente A La UCA	24	0	24	5	1	15
7	Calle Antigua A Huizucar Y Calle Loma Linda	19	1	18	2	0	17
8	Boulevard Los Próceres Y Avenida Altamira.	18	0	18	3	1	14
9	Boulevard Los Próceres Frente A La Torre City	18	0	18	0	2	13
10	Boulevard Venezuela Y Avenida Las Camelias.	15	2	13	5	0	10
11	Boulevard Venezuela Y Calle Gabriel Rosales	15	1	14	4	3	8
12	Calle Antigua A Huizucar Y Calle A	14	0	14	4	0	10
13	Calle Antigua A Huizucar Y Calle Nº 5	13	0	13	0	3	10
14	Calle Los Bambúes Y Avenida Las Amapolas	12	0	12	0	0	12
15	Calle Bogotá Y Avenida Las Amapolas	11	0	11	2	0	9
16	Calle Antigua A Huizucar Y Calle Circunvalación	11	1	10	1	0	10
17	Boulevard Los Próceres Frente A Juguetón	11	0	11	2	0	9
18	Calle Antigua A Huizucar Y Calle San Nicolás	10	0	10	0	1	9
19	Calle Las Camelias Y Avenida Los Duraznos	10	0	10	0	1	9
20	Boulevard Venezuela Frente Al INTI	10	2	8	5	2	3
21	Calle Antigua A Huizucar Y Calle Nº 3	9	0	9	0	0	9
22	Boulevard Los Próceres Y Avenida La Sultana	8	0	8	1	4	3
23	Calle Antigua A Huizucar Frente A La Despensa De Don Juan	8	1	7	2	1	4
24	Boulevard Venezuela y Calle Lorena	8	1	7	4	1	3
25	Boulevard Los Próceres y Calle Nº 1	8	0	8	4	0	4
Total		444	30	414	98	32	304

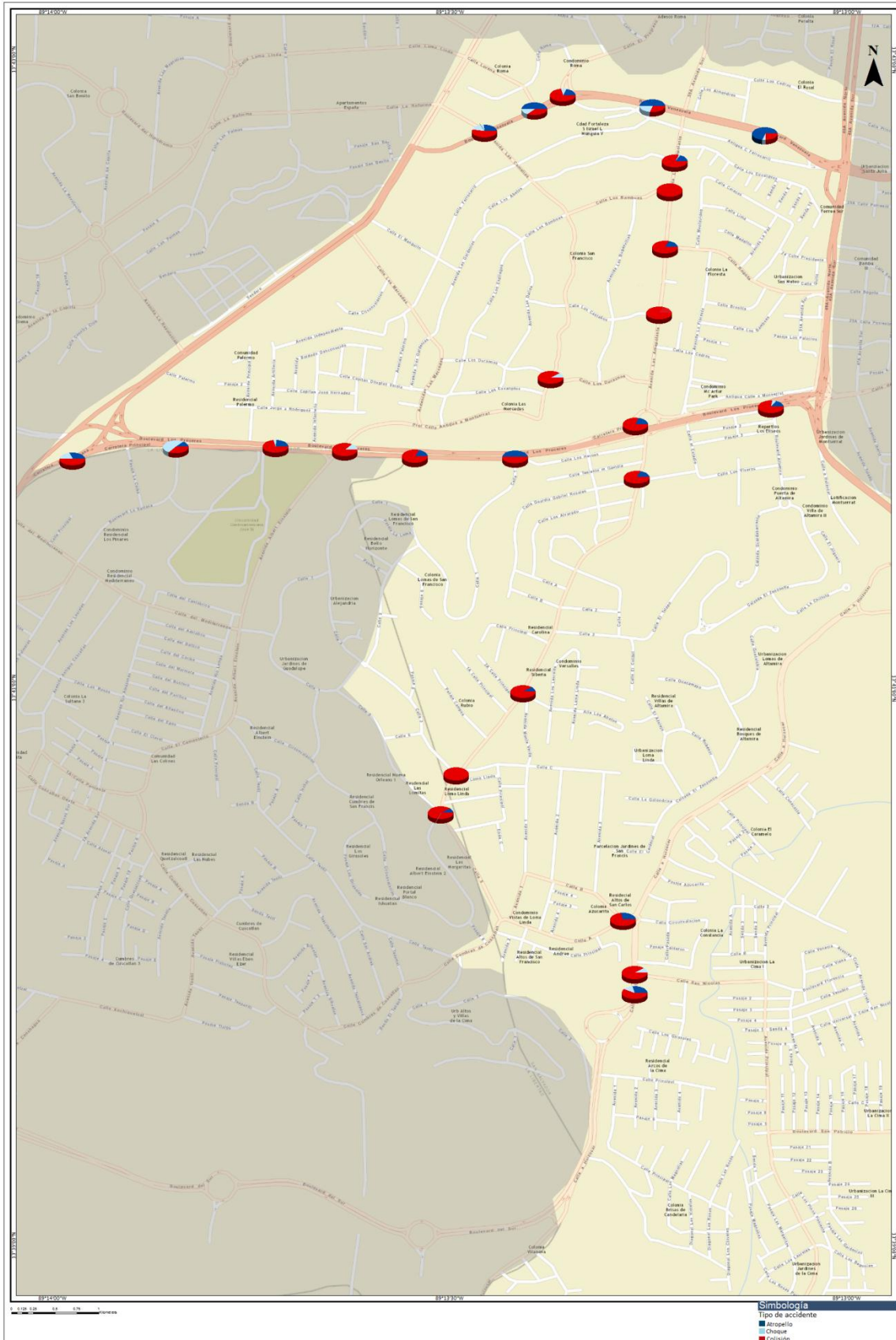
b) Ubicación geográfica de intersecciones con alta frecuencia de accidentes de tránsito



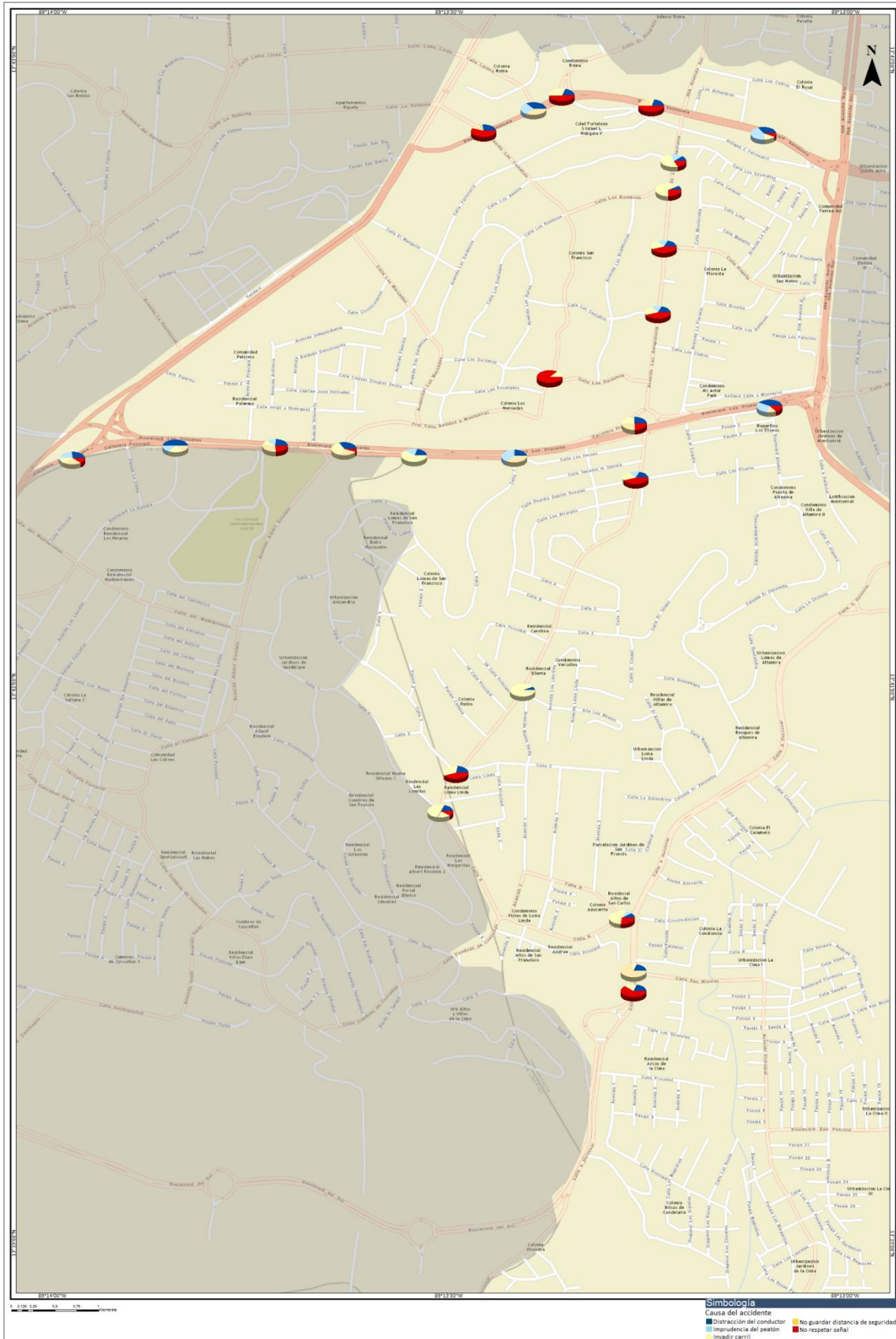
c) Ubicación geográfica: Resultado del accidente



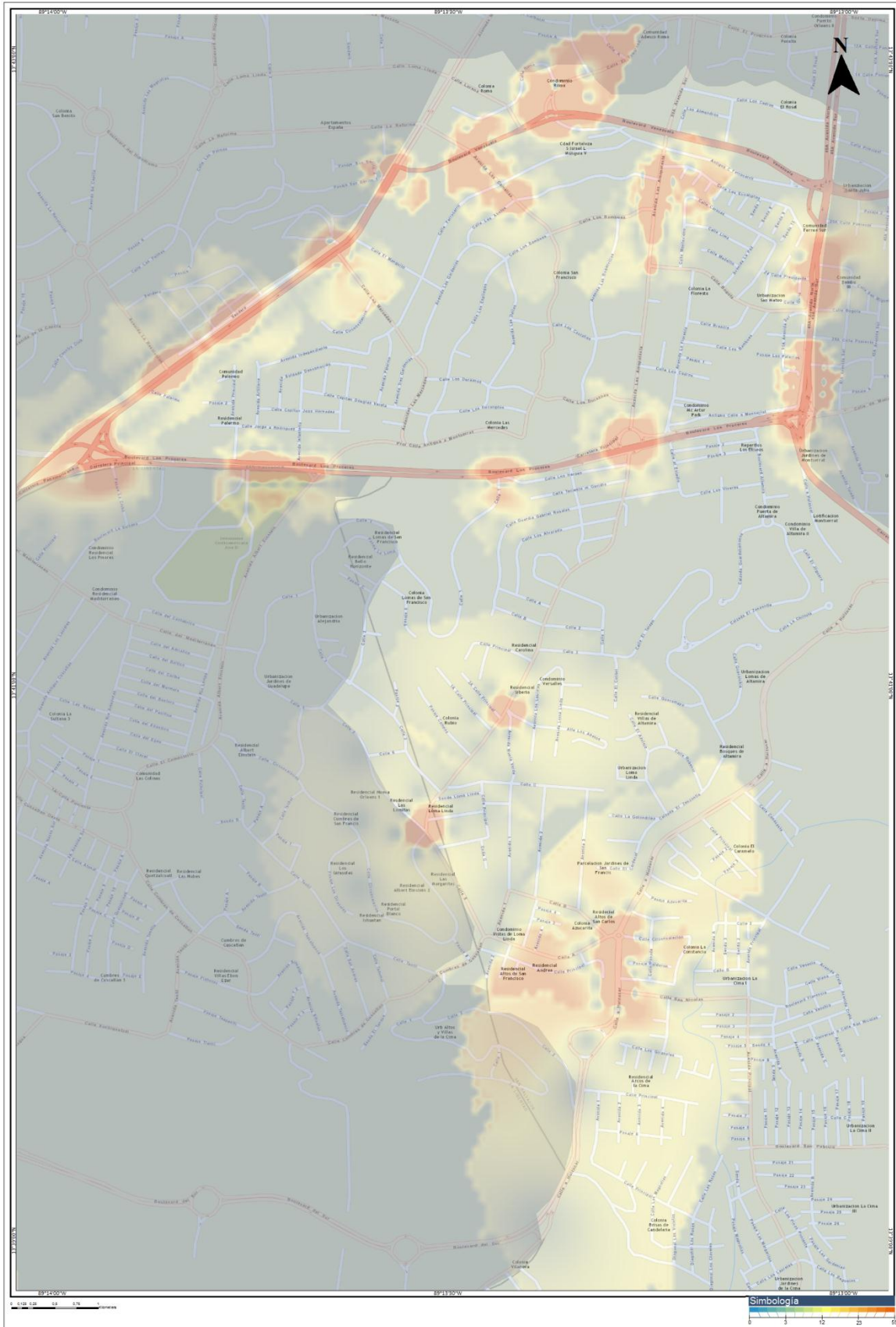
d) Ubicación geográfica: Tipo de accidente



e) Ubicación geográfica: Causa del accidente



f) Frecuencia y proximidad de los accidentes de tránsito.



3.3.6 Distrito 5

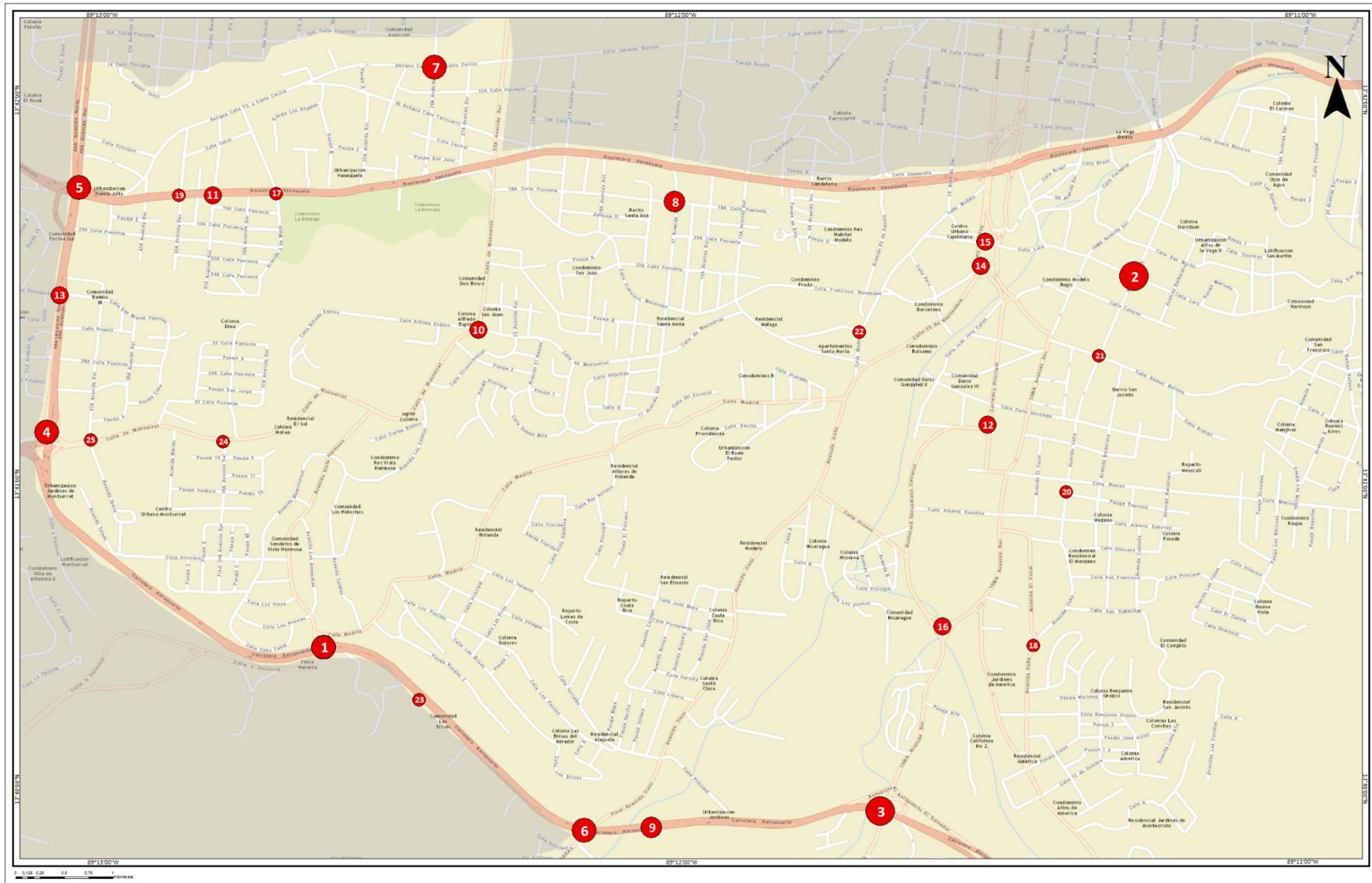
Datos generales

- Población Total: 98,541.
- Extensión Territorial: 18.8 km².
- Densidad de población: 5,242hab/km².
- Límites:
 - Norte: Calle Boulevard Venezuela.
 - Sur: Municipio de San Marcos.
 - Este: 24^a Av. Sur
 - Oeste: 49^a Av. Sur y Autopista a Comalapa.

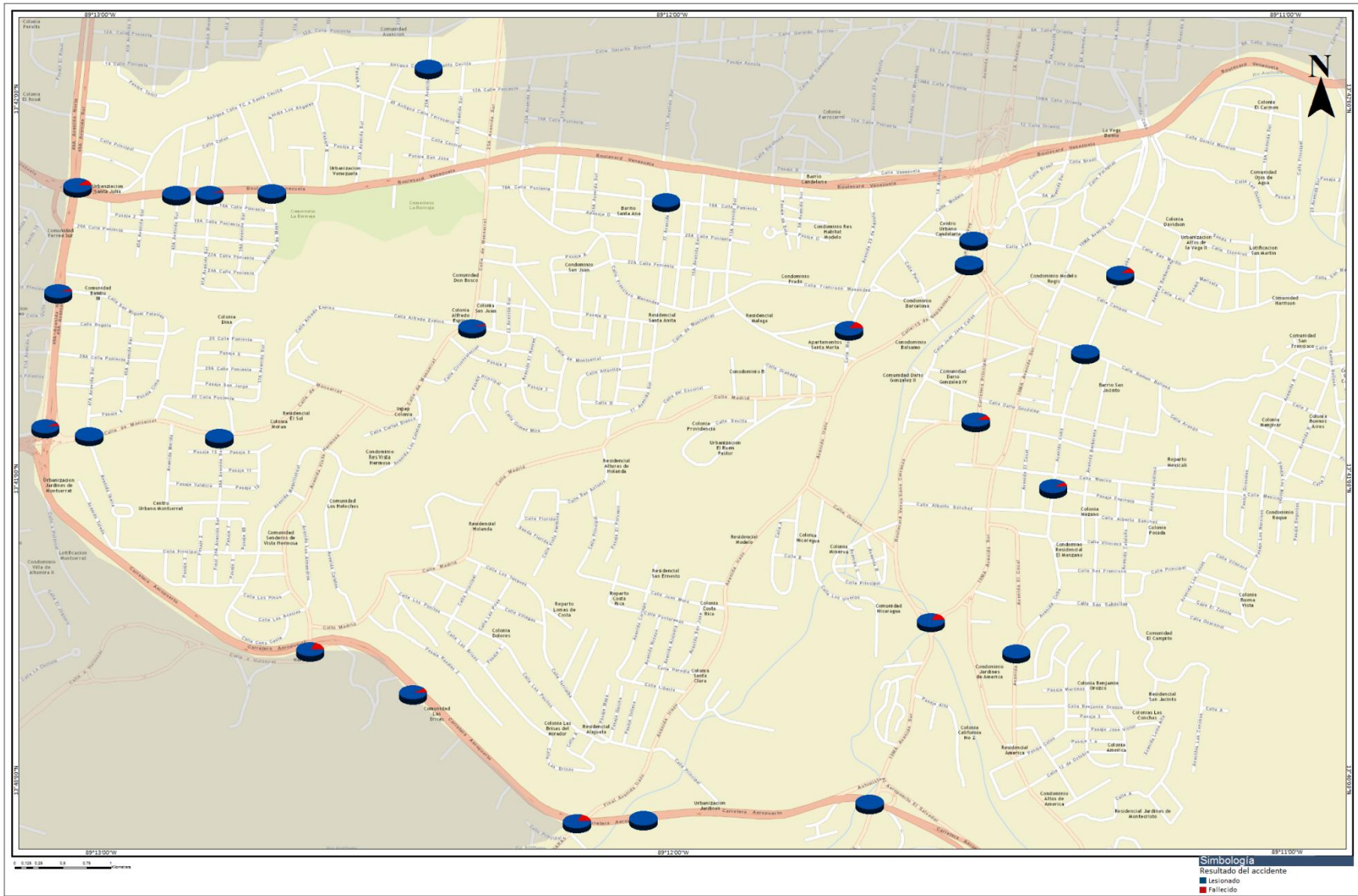
a) Intersecciones con alta frecuencia de accidentes de tránsito.

Correlativo	Intersección	Frecuencia	Fallecidos	Lesionados	Atropellos	Choques	Colisiones
1	Autopista A Comalapa Y Avenida Vista Hermosa	58	1	57	5	6	46
2	Calle Lara Y Avenida Cuba.	56	6	50	7	7	42
3	Autopista A Comalapa Y Calle A Los Planes De Renderos	52	0	52	1	13	32
4	Boulevard Los Próceres Y 49 ^a Avenida Sur A La Altura Del Hermano Lejano	49	3	46	0	20	25
5	49 ^a avenida Sur Y Boulevard Venezuela	46	5	41	17	6	21
6	Autopista Comalapa Y Calle Antigua A Huizucar	45	8	37	15	5	25
7	Calle Gerardo Barrios y 29 ^a Avenida Sur	42	0	42	0	0	42
8	18 ^a Calle Poniente Y 17 ^a Avenida Sur	28	0	28	0	0	28
9	Autopista A Comalapa A La Altura De Cementerio Jardines Del Recuerdo	26	0	26	2	8	14
10	Calle Monserrat Y 25 ^a Avenida Sur	23	1	22	9	2	12
11	Boulevard Venezuela Y 41 ^a Avenida Sur	22	1	21	5	1	16
12	Boulevard Venustiano Carranza Y Avenida Los Diplomáticos	22	2	20	8	3	10
13	24 ^a Calle Poniente Y 49 ^a Avenida Sur	21	1	20	5	5	8
14	Final Avenida Cuscatlán y Calle 15 de Septiembre	20	0	20	6	0	13
15	Calle Lara Y Avenida Cuscatlán	18	0	18	3	6	6
16	Boulevard Venustiano Carranza Y Calle Modelo	18	3	15	5	3	10
17	Boulevard Venezuela Y 39 ^a Avenida Sur	15	0	15	1	0	14
18	Avenida Cuba Frente Al Colegio Bautista	14	0	14	2	2	10
19	Boulevard Venezuela Y 43 ^a Avenida Sur	14	0	14	2	0	12
20	Calle México Y Avenida Cuba	14	1	13	2	2	10
21	Calle Ramón Beloso Y Avenida Cuba	13	0	13	2	0	9
22	Calle Madrid y Calle Modelo	13	2	11	3	0	10
23	Autopista A Comalapa Frente A Gasolinera Texaco Navarra	13	1	12	1	2	10
24	Calle Monserrat Y 39 ^a Avenida Sur	13	0	13	4	0	8
25	Calle A Monserrat Y Avenida Iberia	13	0	13	4	0	9
Total		668	35	633	109	91	442

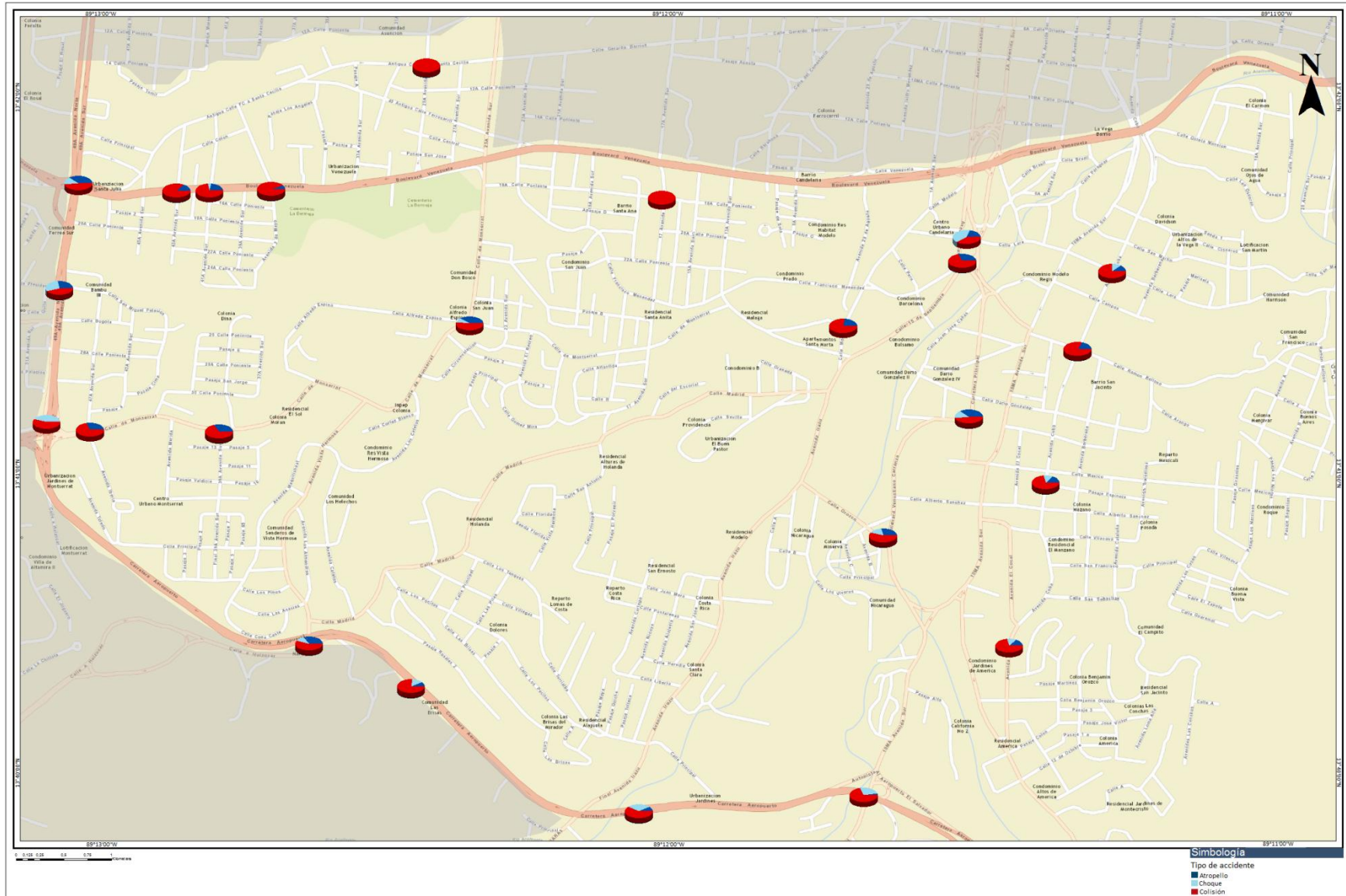
b) Ubicación geográfica de intersecciones con alta frecuencia de accidentes de tránsito



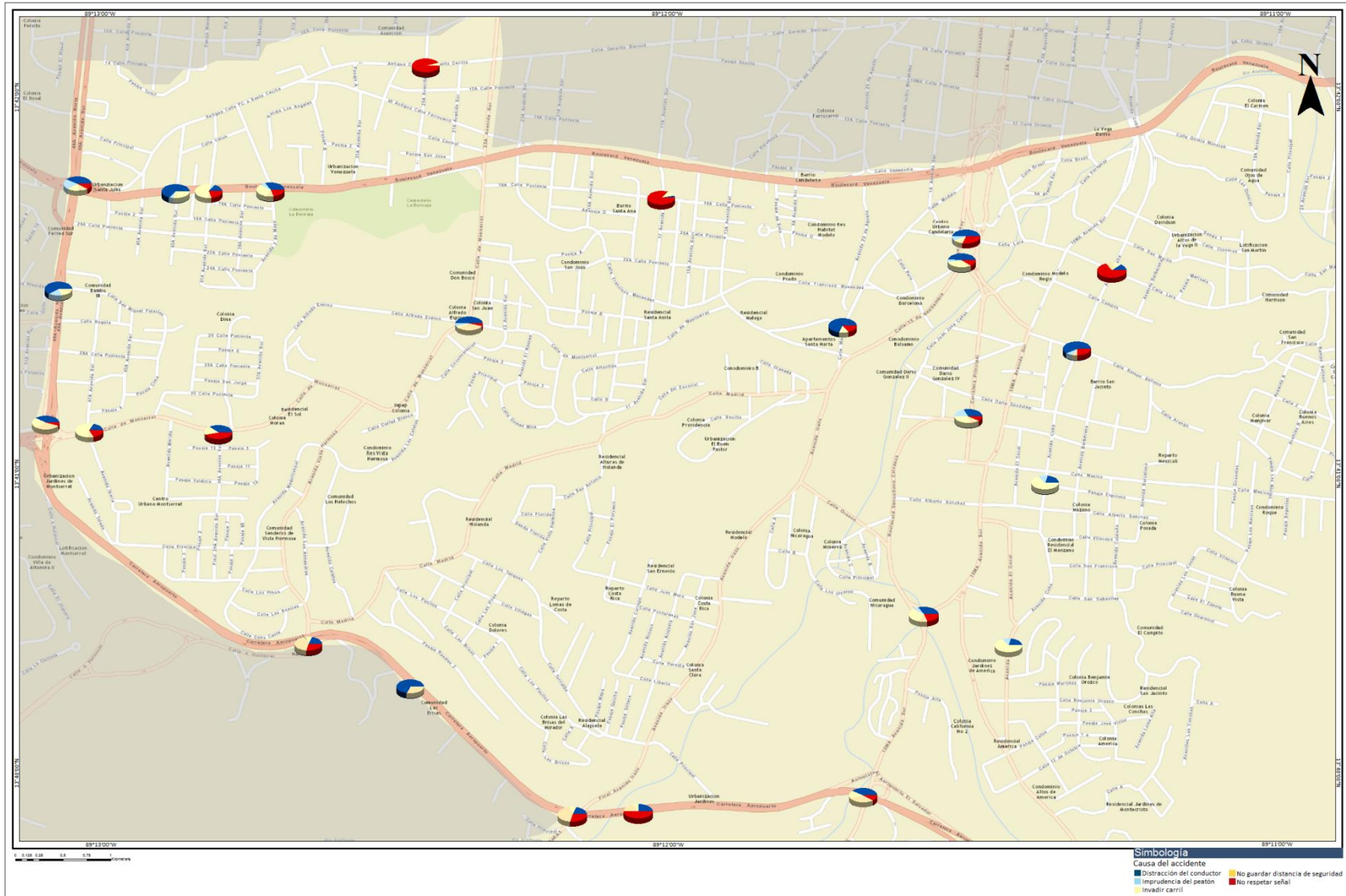
c) Ubicación geográfica: Resultado del accidente



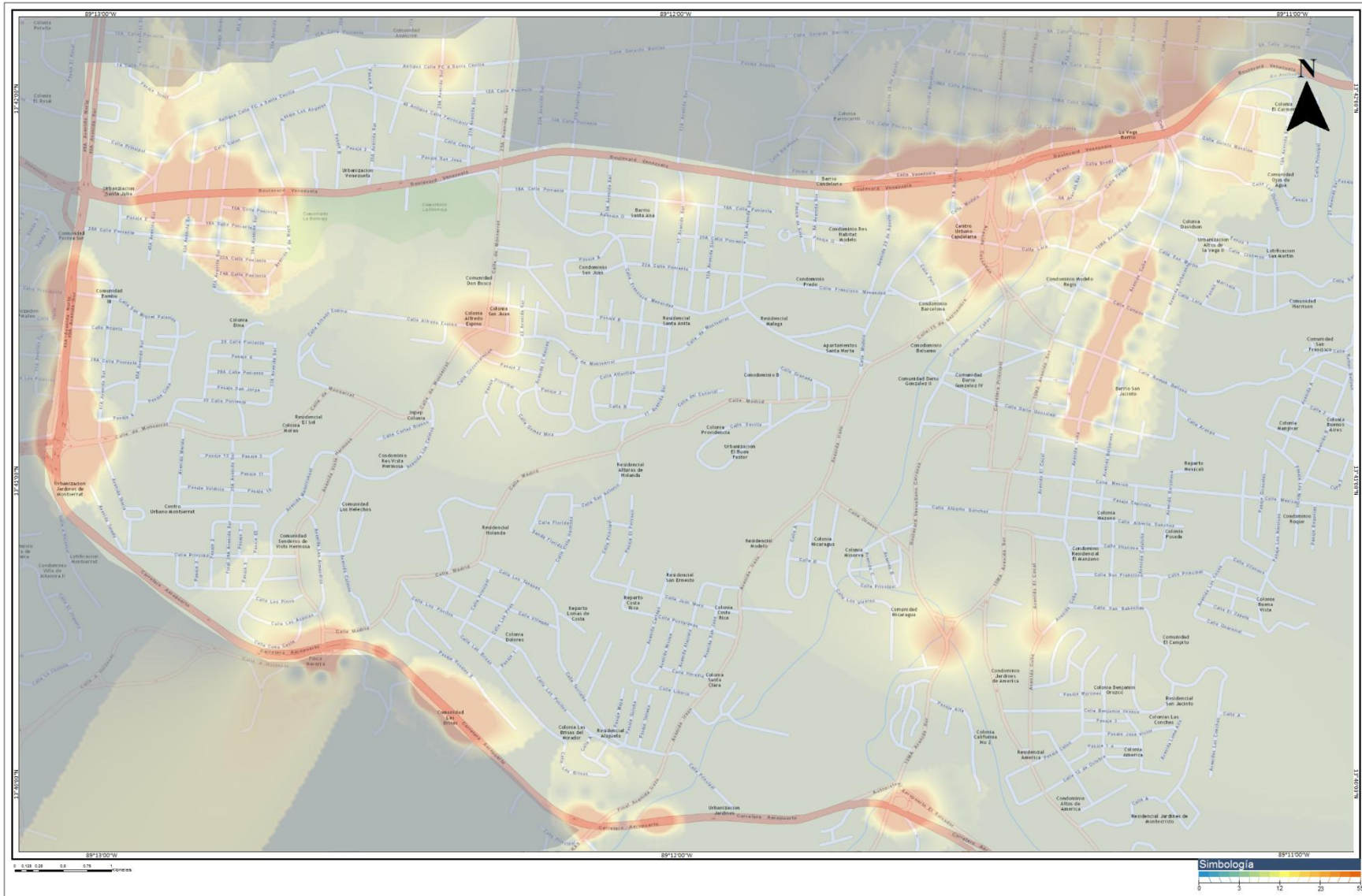
d) Ubicación geográfica: Tipo de accidente



e) Ubicación geográfica: Causa del accidente



f) Frecuencia y proximidad de los accidentes de tránsito



3.3.7 Distrito 6

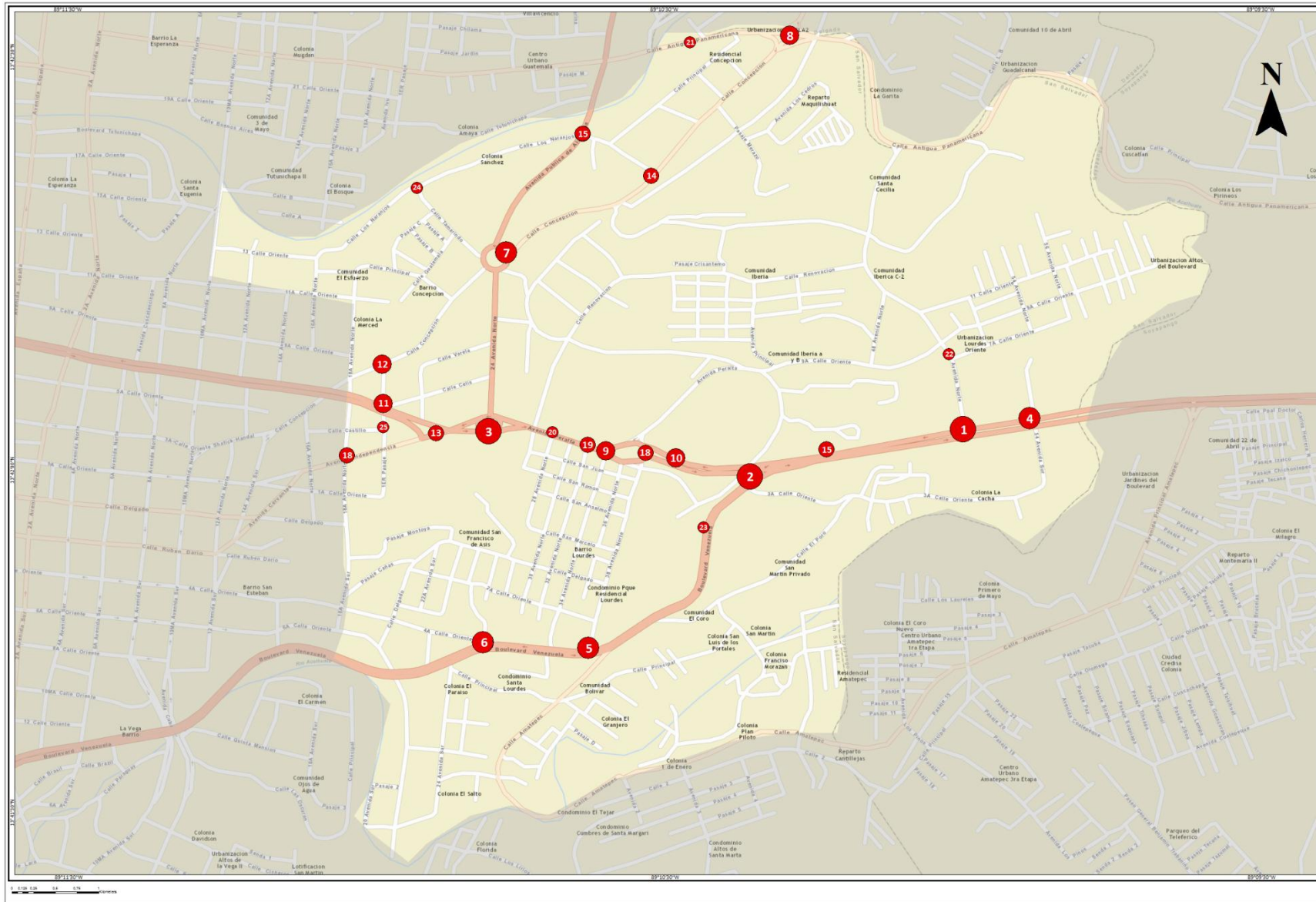
Datos generales

- Población Total: 98,541.
- Extensión Territorial: 18.8 km².
- Densidad de población: 5,242hab/km².
- Límites:
 - Norte: 10^a Av. Norte y Río Tutunichapa
 - Sur: Línea férrea San Salvador y Soyapango.
 - Este: Municipio de Soyapango.
 - Oeste: Calle Concepción y 18^a avenida norte.

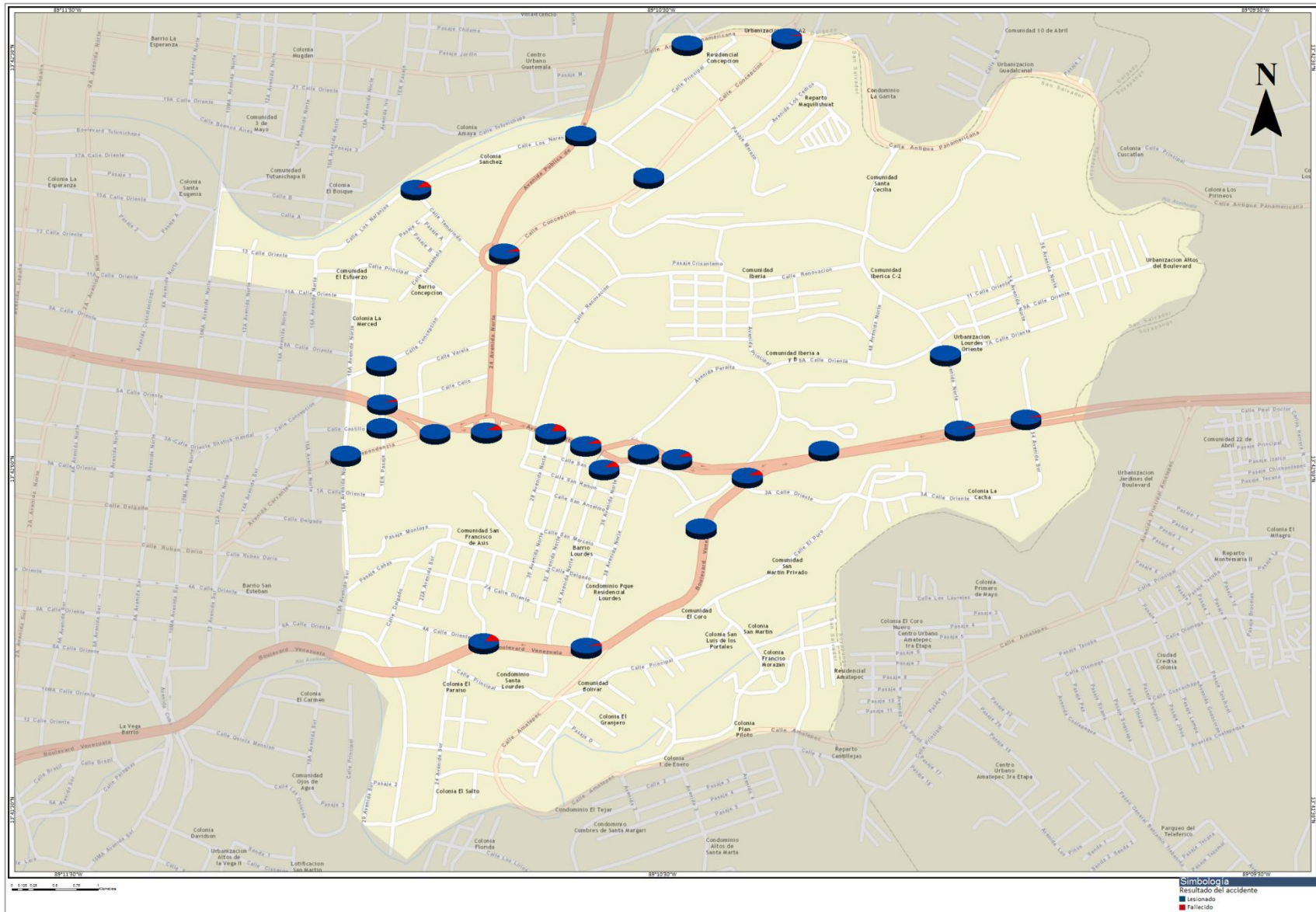
a) Intersecciones con alta frecuencia de accidentes de tránsito

Correlativo	Intersección	Frecuencia	Fallecidos	Lesionados	Atropellos	Choques	Colisiones
1	Boulevard Del Ejercito Y 50 ^a Avenida Norte	92	4	88	5	11	70
2	Boulevard Venezuela Y Boulevard Del Ejercito	84	8	76	12	20	46
3	Avenida Independencia y 24 ^a Avenida Norte	67	7	60	33	0	30
4	Boulevard Del Ejercito Y 54 ^a Avenida Norte	53	2	51	5	0	48
5	Boulevard Venezuela Y 38 ^a Avenida Sur	45	2	43	14	2	18
6	Boulevard Venezuela Frente A Gasolinera Esso Lourdes	41	2	39	11	1	28
7	Calle Concepción Y Redondel La Isla	39	2	37	12	3	20
8	Calle 5 De Noviembre Y Calle Concepción	29	1	28	8	0	19
9	Avenida Peralta Lagos Y 36 ^a Avenida Norte	25	3	22	15	1	7
10	Boulevard Del Ejercito Frente A La Terminal De Oriente.	23	2	21	10	1	9
11	Alameda Juan Pablo II Y 20 ^a Avenida Norte	21	1	20	19	0	2
12	Calle Concepción Y 20 ^a Avenida Norte	20	0	20	9	0	11
13	Alameda Juan Pablo II Y Avenida Independencia	19	0	19	7	0	12
14	Calle Concepción Y Calle los Naranjos	17	0	17	4	0	13
15	Boulevard Del Ejercito Frente A Molsa	16	0	16	4	4	8
16	Avenida Peralta Lagos Y 38 ^a Avenida Norte	14	0	14	9	0	5
17	Boulevard Venezuela Y 4 ^a Calle Oriente	14	2	12	2	0	6
18	Avenida Independencia Y 18 ^a Avenida Norte	14	0	14	8	1	5
19	Avenida Peralta Lagos Y 32 ^a Avenida Norte	13	1	12	3	0	10
20	Avenida Peralta Lagos Y 28 ^a Avenida Norte	11	2	9	8	0	3
21	Calle 5 De Noviembre Frente Al Cuartel CAM	10	0	10	1	0	8
22	50 ^a Avenida Norte Y 7 ^a Calle Oriente	10	0	10	2	1	7
23	Boulevard Venezuela Frente a Taller de la PNC	10	0	10	2	2	6
24	Calle Los Naranjos Y Calle El Tamarindo	8	1	7	4	1	3
25	Calle Celis Y 20 ^a Avenida Norte	8	0	8	4	0	4
	Total	703	40	663	211	48	398

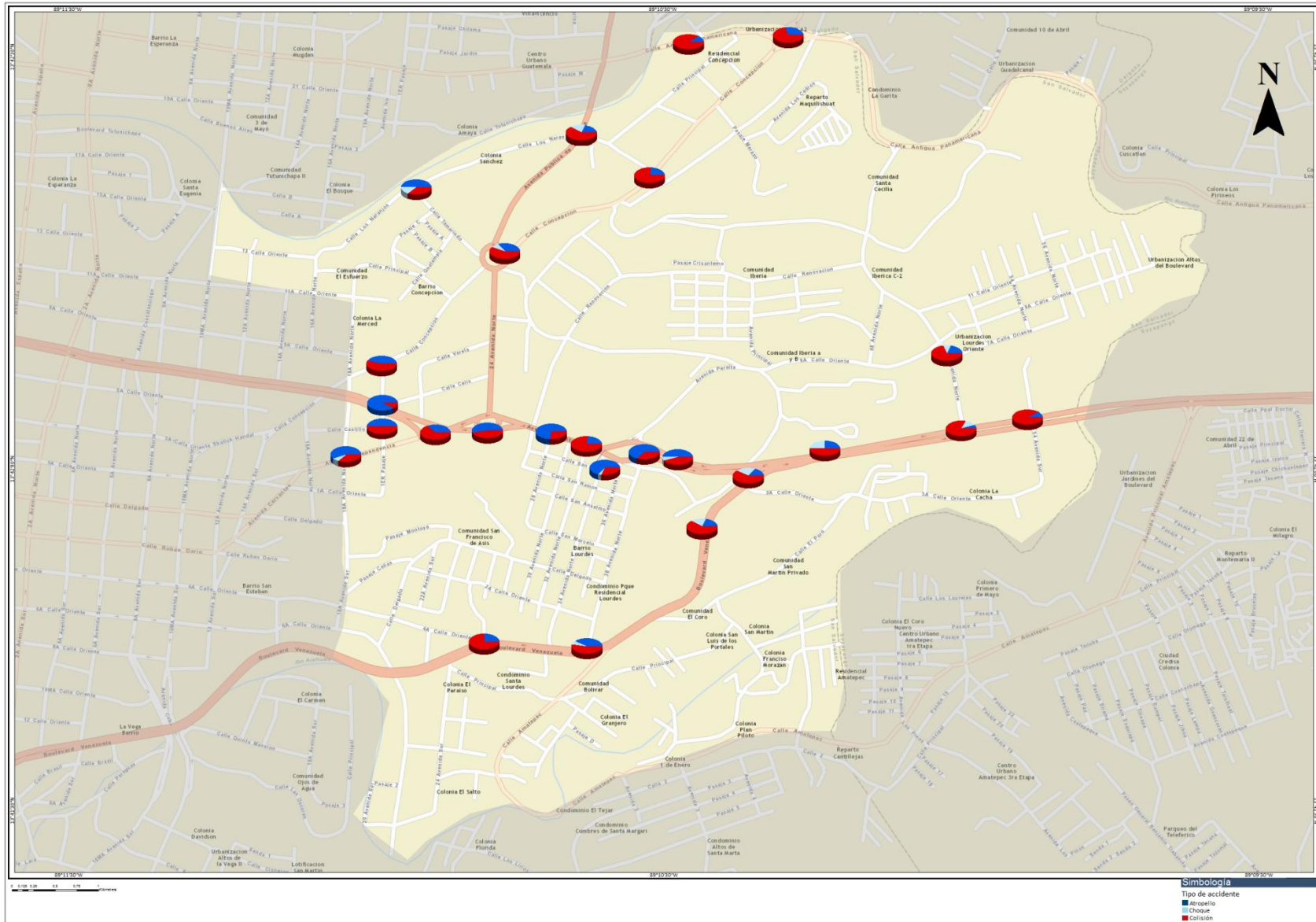
b) Ubicación geográfica de intersecciones con alta frecuencia de accidentes de tránsito



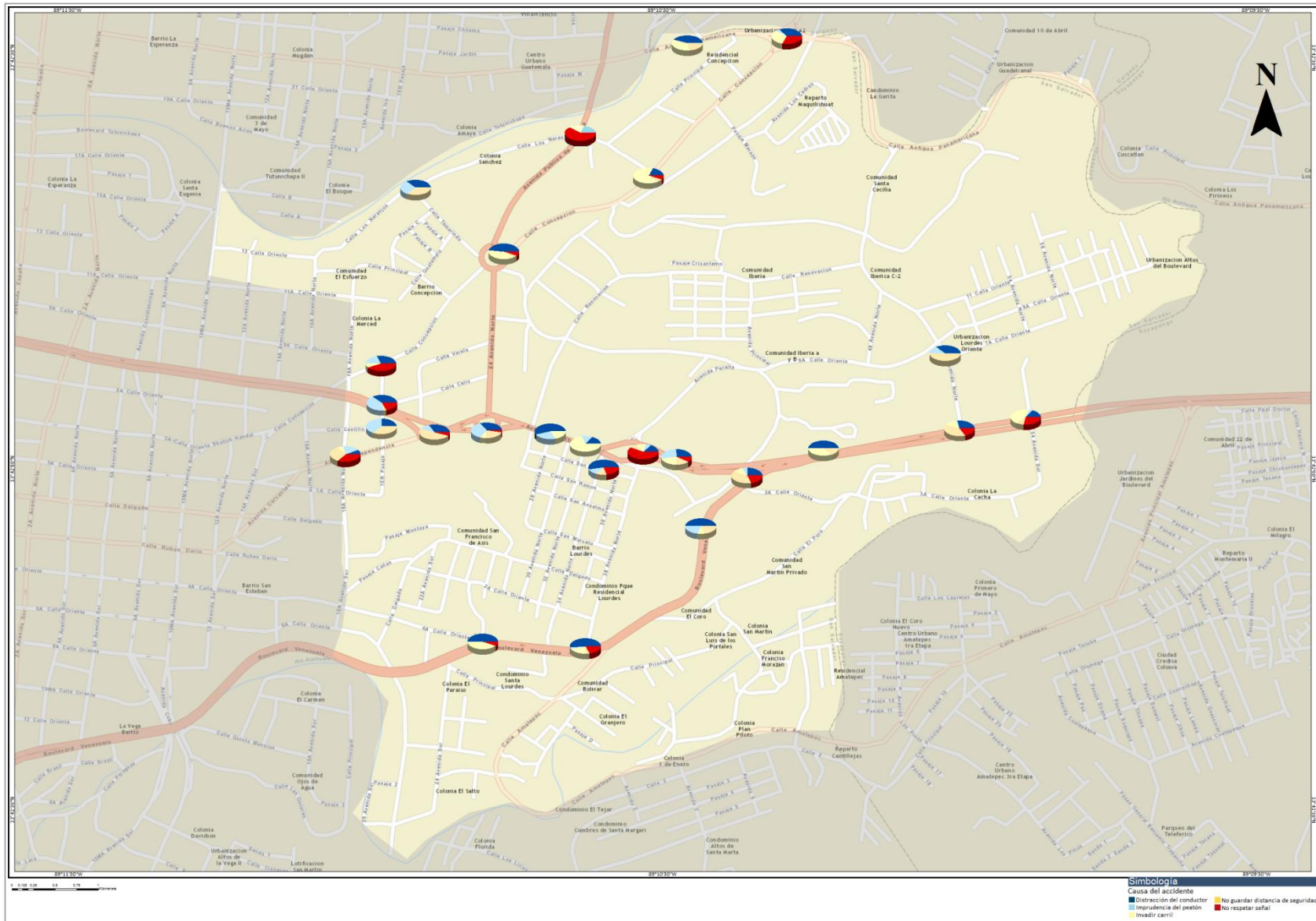
c) Ubicación geográfica: Resultado del accidente



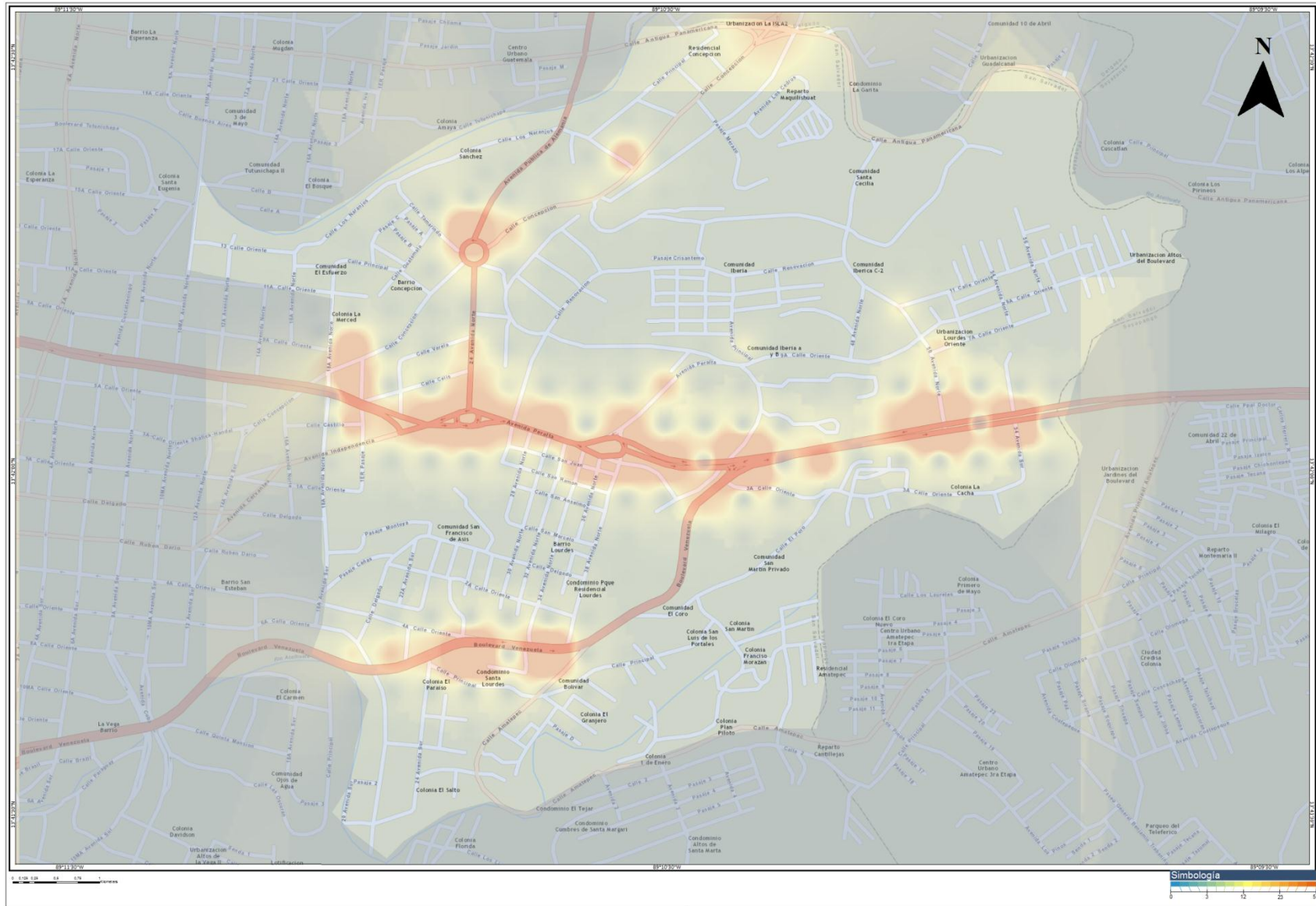
d) Ubicación geográfica: Tipo de accidente



e) Ubicación geográfica: Causa del accidente



f) Frecuencia y proximidad de los accidentes de tránsito



Conclusiones

1 - En lo que respecta al análisis de la estructura espacial de los accidentes de tránsito en el municipio de San Salvador se concluye que el fenómeno presenta un patrón no aleatorio con tendencia al agrupamiento, lo que significa que hay zonas donde suceden más accidentes que en otras, este análisis se hizo mediante la prueba del vecino más próximo y el test I de Moran. Además el fenómeno presenta la existencia de correlación espacial y mínima variabilidad en distancia cortas, es decir las intersecciones vecinas tienden a ser similares, condición requerida para poder aplicar el estimador lineal kriging.

2 - En cuanto a la interpolación espacial, esta fue posible realizarla por el hecho de la presencia de autocorrelación espacial en el campo de estudio, haciendo de esta manera que la estimación o predicción espacial identificara por medio del mapa de frecuencias las zonas más vulnerables a la ocurrencia de accidentes de tránsito. Los estimadores lineales kriging que mejor se ajustan al comportamiento propio de los accidentes de tránsito son el kriging universal y el kriging simple, estimadores que permiten que el fenómeno no sea constante en todo el área de estudio, pero si a nivel local.

3 - El Centro Histórico a pesar de ser el distrito menos poblado y el de menos extensión territorial es el que mayor número de accidente reporta, pareciera que esta situación se debe a dos razones de peso, la primera a la fuerte actividad de comercio informal que impera en San Salvador que se mueve de forma desordenada y se expande diariamente hacia otros lugares de la capital. El comercio informal ha provocado el cierre parcial o total de las principales vías de San Salvador, provocando que el tráfico vehicular se concentre solo en algunas vías del Centro Histórico. Y la segunda, las calles del Centro Histórico sirven de conexión entre municipios vecinos a la ciudad capital lo que deriva en que el Centro Histórico debe cargar con una fuerte circulación de unidades de transporte colectivo.

4 - En cuanto al patrón del horario de los accidentes de tránsito, la concentración de estos van desde la 6:00 a.m. hasta 8:00 p.m. horarios que corresponden tradicionalmente a las actividades cotidianas como el estudio o trabajo.

5 - Se ha demostrado que los accidentes de tránsito están fuertemente asociados a los patrones culturales de los salvadoreños como es la poca educación vial de los peatones, el comportamiento irresponsable de los conductores y las condiciones sociales, geográficas y económicas del país.

Recomendaciones

- 1- Promover el tema de seguridad vial en estudiantes del sistema educativo escolar a nivel nacional.
- 2- Supervisar de cerca a escuelas de manejo o tramitadoras de licencia de vehículos.
- 3- Realizar acciones de concientización dirigidas a la ciudadanía en general (campaña educativa a través de medios de comunicación, talleres, seminarios, conferencias, charlas).
- 4- Implementación de sanciones más drásticas contra los infractores de la ley
- 5- Contar con centros de revisión de vehículos debidamente certificados.
- 6- Promover el uso de bicicletas como transporte alternativo para distancias cortas.
- 7- Mejorar la marcación, señalización y estado de las calles.
- 8- Mejorar la calidad de los servicios del transporte público.
- 9- Ubicar agentes de tránsito como visores y ordenadores en las calles de mayor flujo vehicular.
- 10- Fortalecer la vigilancia con el uso de cámaras de seguridad.
- 11- Reordenar el recorrido del transporte colectivo.
- 12- Impulsar un plan de atención de los puntos críticos de la red vial.
- 13- Crear un centro de monitoreo nacional y proveerlo de los recursos necesarios para su funcionamiento.

Referencia bibliográfica

1. Castillejo, Maximiano (2007). Teoría Geoestadística aplicada al análisis de la Variabilidad Espacial Arqueológica Intra-Site.
2. G. Gneiting (1999). Geostatistical Space-Time Models, Stationarity, Separability and Full Symmetry. Department of Statistics, University of Washington Seattle, Washington, USA.
3. Ramón Giraldo Henao (2002). Introducción a la Geoestadística Teoría y Aplicación. Universidad Nacional de Colombia.
4. Frank Badu Olsen (2010). Spatial Statistics of Epidemic Data: The Case of Cholera Epidemiology in Ghana. University of Twente.
5. Informe de seguridad vial para América Latina. Organización Panamericana de la Salud OPS y Organización Mundial de la Salud (2010).
6. Jaime Hernández. Aplicaciones Técnicas de interpolación Espacial Kriging. Universidad de Chile Facultad de Arquitectura y Urbanismo Escuela de Geografía (2002).
7. Díaz, Martín. (2002). Geoestadística Aplicada.
8. M. Victoria Ibáñez. Modelos Estadísticos espacio-temporales en perimétrica. Escuela Superior de Tecnología y Ciencias Experimentales, Universitat Jaume I. Julio (2003).
9. Programa Nacional de Salud: Alianza para una mejor salud 2007 – 2012. Secretaría de salud de México (2012).
10. Dunn, R. Statistical mapping. The American statistician (1987).
11. Organización Panamericana de la Salud. Cooperación técnica de la OPS en sistemas de información geográfica aplicados a la epidemiología en las américas (1996).
12. Aplicación de la geoestadística a las ciencias ambientales, Asociación Española de Ecología Terrestre (2004).
13. Técnicas de análisis espacial de patrones de puntos aplicadas en ecología forestal, Unidad de Recursos Forestales CITA, Gobierno de Aragón, Zaragoza España (2005).
14. Apuntes de geoestadística, ingeniería en minas. Xavier Emery, Facultad de Matemática y Física Universidad de Chile (2007).