



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

**DEPARTAMENTO DE CIENCIAS DE LA TIERRA Y LA
CONSTRUCCIÓN**

CARRERA DE INGENIERÍA GEOGRÁFICA Y DEL MEDIO AMBIENTE

**TRABAJO DE TITULACIÓN, PREVIO A LA OBTENCIÓN DEL TÍTULO
DE INGENIERO GEÓGRAFO Y DEL MEDIO AMBIENTE**

**EVALUACIÓN DE LA EXACTITUD POSICIONAL HORIZONTAL E
INFERENCIA DE LA ESCALA SOBRE DATOS LIBRES OSM EN
ADMINISTRACIONES ZONALES DEL DMQ**

AUTOR: CASTRO ZAMBRANO, ROGER RAFAEL

DIRECTOR: DR. TIERRA CRIOLLO, ALFONSO RODRIGO

SANGOLQUÍ

2019



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE CIENCIAS DE LA TIERRA Y LA
CONSTRUCCIÓN

CARRERA DE INGENIERÍA GEOGRÁFICA Y DEL MEDIO AMBIENTE

CERTIFICACIÓN

Certifico que el trabajo de titulación, ***“EVALUACIÓN DE LA EXACTITUD POSICIONAL HORIZONTAL E INFERENCIA DE LA ESCALA SOBRE DATOS LIBRES OSM EN ADMINISTRACIONES ZONALES DEL DMQ”*** fue realizado por el señor ***Castro Zambrano, Roger Rafael*** el mismo que ha sido revisado en su totalidad, analizado por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, -razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 7 de junio de 2019

Una firma manuscrita en tinta azul que parece leer 'Alfonso Rodrigo Tierra Criollo'.

Dr. Tierra Criollo, Alfonso Rodrigo

C.C. 0601803273



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE CIENCIAS DE LA TIERRA Y LA
CONSTRUCCIÓN

CARRERA DE INGENIERÍA GEOGRÁFICA Y DEL MEDIO AMBIENTE

AUTORÍA DE RESPONSABILIDAD

Yo, *Castro Zambrano, Roger Rafael*, declaro que el contenido, ideas y criterios del trabajo de titulación: ***“EVALUACIÓN DE LA EXACTITUD POSICIONAL HORIZONTAL E INFERENCIA DE LA ESCALA SOBRE DATOS LIBRES OSM EN ADMINISTRACIONES ZONALES DEL DMQ”*** es de mi autoría y responsabilidad, cumpliendo con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, respetando derechos intelectuales de terceros y referenciando las citas bibliográficas. Consecuentemente el contenido de la investigación mencionada es veraz.

Sangolquí, 7 de junio de 2019

A blue ink handwritten signature of Roger Castro, written over a light blue rectangular background.

Castro Zambrano, Roger Rafael

C.C. 1313816637



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE CIENCIAS DE LA TIERRA Y LA
CONSTRUCCIÓN

CARRERA DE INGENIERÍA GEOGRÁFICA Y DEL MEDIO AMBIENTE

AUTORIZACIÓN

Yo, *Castro Zambrano, Roger Rafael*, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: ***“EVALUACIÓN DE LA EXACTITUD POSICIONAL HORIZONTAL E INFERENCIA DE LA ESCALA SOBRE DATOS LIBRES OSM EN ADMINISTRACIONES ZONALES DEL DMQ”*** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 7 de junio de 2019

A blue ink handwritten signature of Roger Rafael Castro Zambrano, written over a light blue rectangular background. Below the signature is a horizontal dotted line.

Castro Zambrano, Roger Rafael

C.C. 1313816637

AGRADECIMIENTOS

A mi mamá, quien me enseñó lo que significan las raíces, y a mi papá, quien me enseñó lo que significan las alas. A Sari y a Ema por alegrar mis días. Y Angelita y Robinson por todo el cariño que recibí.

A todos mis panas, a aquellos que me enseñaron precisamente las cosas que deseaba aprender, pero también a aquellos con los que aprendí lo que no esperaba. A todos los músicos y al profe Jumbo, con quienes creamos lo que de otra manera sería vacío.

A la Noe por ser valiente, por seguir aquí.

A los doctores Tierra y Luna quienes, aun en un sistema educativo con muchas falencias, me demostraron cómo debería ser la educación y cómo se comporta un verdadero docente. Finalmente a los y las que liberaron sus datos, sus algoritmos y sus conocimientos; no los conozco en persona, pero me hablaron a través de su trabajo.

ÍNDICE DE CONTENIDOS

v

CARÁTULA	
CERTIFICACIÓN.....	i
AUTORÍA DE RESPONSABILIDAD.....	ii
AUTORIZACIÓN.....	iii
AGRADECIMIENTOS.....	iv
ÍNDICE DE CONTENIDOS.....	v
ÍNDICE DE TABLAS.....	viii
ÍNDICE DE FIGURAS.....	ix
RESUMEN.....	x
ABSTRACT.....	xi
CAPÍTULO I ASPECTOS GENERALES.....	1
1.1. ANTECEDENTES.....	1
1.2. PLANTEAMIENTO DEL PROBLEMA.....	2
1.3. JUSTIFICACIÓN E IMPORTANCIA.....	3
1.4. ÁREA DE INFLUENCIA.....	5
1.5. OBJETIVOS Y METAS.....	6
1.5.1. Objetivo Principal.....	6
1.5.2. Objetivos Específicos.....	6
1.5.3. Metas.....	6
CAPÍTULO II MARCO TEÓRICO.....	7
2.1. ESTADO DEL ARTE.....	7

	vi
2.2. BASE TEÓRICA.....	8
2.2.1. Información Geográfica Voluntaria.....	8
2.2.2. OpenStreetMap.....	10
2.2.3. Big GeoData.....	11
2.3. BASE CONCEPTUAL.....	12
2.3.1. Calidad.....	12
2.3.2. Exactitud Posicional Horizontal.....	12
2.3.3. Escala.....	13
2.3.4. Estadístico Getis-Ord.....	13
2.4. BASE LEGAL.....	15
2.4.1. Norma Internacional.....	15
2.4.2. Norma Ecuatoriana.....	16
CAPÍTULO III MARCO METODOLÓGICO.....	17
3.1. OBTENCIÓN DE DATOS.....	17
3.2. SOFTWARE UTILIZADO.....	17
3.3. DIAGRAMA DE FLUJO.....	18
3.4. METODOLOGÍA POR SCRIPT.....	19
3.4.1. Primer Script de Evaluación.....	19
3.4.2. Primer Script de Reporte.....	24
3.4.3. Segundo Script de Evaluación.....	24
3.4.4. Segundo Script de Reporte.....	26
CAPÍTULO IV RESULTADOS Y DISCUSIÓN.....	27
4.1. RESULTADOS POR SCRIPT.....	27

	vii
4.1.1. Primer Script de Evaluación.....	27
4.1.2. Primer Script de Reporte.....	30
4.1.3. Segundo Script de Evaluación.....	31
4.1.2. Segundo Script de Reporte.....	32
4.2. DISCUSIÓN DE RESULTADOS.....	33
CAPÍTULO V CONCLUSIONES Y RECOMENDACIONES.....	38
5.1. CONCLUSIONES.....	38
5.2. RECOMENDACIONES.....	40
BIBLIOGRAFÍA.....	41

ÍNDICE DE TABLAS

Tabla 1 <i>Paquetes de código R utilizados.....</i>	18
Tabla 2 <i>Escalas para Errores Posicionales y Factores de Homogeneización variables.....</i>	22
Tabla 3 <i>Valores Z Críticos para Tamaños Muestrales y Niveles de Significancia variables....</i>	23
Tabla 4 <i>Objetos relevantes en el Espacio de Trabajo de R.....</i>	23
Tabla 5 <i>Detalles de la ejecución y de las coberturas viales.....</i>	27
Tabla 6 <i>Desempeño de los métodos para relacionar intersecciones.....</i>	28
Tabla 7 <i>Percentiles de los errores calculados con el Método de Unión por Nombres.....</i>	29
Tabla 8 <i>Detalles de la implementación del estadístico Getis-Ord.....</i>	32

ÍNDICE DE FIGURAS

Figura 1. Miembros activos (izquierda) y nodos creados/modificados/eliminados (derecha) diariamente en OSM para el caso de Ecuador.....	3
Figura 2. Ubicación de las Administraciones Zonales de interés. Proyección UTM 17S.....	5
Figura 3. Crecimiento en número de miembros (izquierda) y nodos (derecha) en OSM.....	10
Figura 4. Diagrama de Flujo de la Metodología aplicada.....	18
Figura 5. Captura de pantalla de las Secciones del primer Script.....	20
Figura 6. Corrección de una Vecindad con presencia de puntos calientes, antes (izquierda) y después (derecha).....	25
Figura 7. Gráficos de Caja de errores calculados con el Método de Unión por Nombres.....	29
Figura 8. Gráficos de Caja de errores calculados con el Método de Vecino más Cercano.....	29
Figura 9. Relación del valor p y la Distancia de Vecindad en los estadístico G y G*.....	30
Figura 10. Relación del valor p y el Factor de Homogeneización en el estadístico G*.....	31
Figura 11. Conteo de Puntos Calientes por Administración Zonal.....	33
Figura 12. Porcentaje de Polígonos de Voronoi por escala, por Administración Zonal.....	33
Figura 13. Ejemplo de error atípicamente grande en el Método de Unión por Nombres.....	34
Figura 14. Gráficos de densidad de errores calculados con el Método de Vecino más Cercano (izquierda) y de valores normalizados del estadístico G*i (derecha).....	36
Figura 15. Ubicación de los Puntos Calientes y Fríos.....	37

RESUMEN

OpenStreetMap es probablemente el proyecto de Información Geográfica Voluntaria más importante; sin embargo, su naturaleza colaborativa pone en duda aspectos de su calidad. Este problema se ha abordado en otras regiones, mas no en Ecuador; el presente trabajo pretende motivar tales investigaciones, iniciando con el aspecto posicional. La zona de estudio fueron varias Administraciones Zonales del Distrito Metropolitano de Quito. Dos métodos para calcular la exactitud posicional horizontal fueron implementados: el primero relaciona intersecciones viales en OpenStreetMap con sus equivalentes oficiales a través de los nombres; el segundo, a través de la búsqueda del vecino más cercano. Mientras que el primer método produjo escasos resultados y valores atípicos, el segundo se desempeñó mejor pero impuso un límite al error calculado. A continuación se calcularon las versiones local y global del estadístico Getis-Ord, con el propósito de evaluar la autocorrelación espacial. Fueron halladas escasas agrupaciones de errores pequeños, pero abundantes de errores grandes; las vecindades con tales errores fueron homogeneizadas con el propósito de incrementar su importancia. Después de esta corrección fue inferida la escala a la cual pertenecería cada intersección, generando polígonos de Voronoi. Finalmente, se discutieron las oportunidades y limitaciones de la vialidad OpenStreetMap en las zonas estudiadas.

PALABRAS CLAVES:

- **OPENSTREETMAP**
- **EXACTITUD POSICIONAL**
- **INFERENCIA DE ESCALA**
- **DATOS LIBRES**
- **LENGUAJE R**

ABSTRACT

OpenStreetMap is arguably the most important Volunteered Geographic Information project; nevertheless, its collaborative nature casts doubt upon its quality aspects. This problem has been addressed in other regions, but not yet in Ecuador; hence this work aims to motivate such researches, starting with the positional aspect. Study zone were several Administrative Zones of Quito's Metropolitan District. Two methods for computing the horizontal positional accuracy were implemented: the first one joins OpenStreetMap street intersections with their official relatives through names; the second one, through nearest neighbor search. Whereas the first method produced few results and atypical values, the second one behaved better but imposed a limit to the computed error. Then the local and global versions of Getis-Ord statistic were computed, to assess the spatial autocorrelation. Few clusters of little errors were found, but several of big errors; the neighborhoods with such errors were homogenized, so their importance increases. After this correction, the scale to which each intersection would belong was inferred, generating Voronoi polygons. Finally, opportunities and limitations of the OpenStreetMap highways in the studied zones were discussed.

KEYWORDS:

- **OPENSTREETMAP**
- **POSITIONAL ACCURACY**
- **SCALE INFERENCE**
- **OPEN DATA**
- **R LANGUAGE**

CAPÍTULO I

ASPECTOS GENERALES

1.1. ANTECEDENTES

En la primera década del siglo corriente, el desarrollo de ciertas tecnologías -la Web 2.0 principalmente - permitió al internet avanzar hacia un paradigma colaborativo; fruto de este avance fue la explosión de contenidos generados por usuarios (Goodchild, 2007; Sui, 2008; Coleman, Georgiadou, & Labonte, 2009). El ámbito geográfico no fue excepción: la generación y difusión de datos geoespaciales por la web han recibido considerable atención; de esta manera, Goodchild (2007) acuña el término “Información Geográfica Voluntaria” (VGI, por sus siglas en inglés).

Entre los varios proyectos VGI creados, OpenStreetMap (OSM) es el más reconocido; desde 2004 es “un esfuerzo internacional para crear una fuente de datos libre, disponible en un mapa, a través del esfuerzo voluntario” (Goodchild, 2007). La naturaleza OSM es comunitaria, de roles fluidos, de propiedad común, y similar a un palimpsesto: un producto incompleto pero que crece continuamente (Coleman, Georgiadou, & Labonte, 2009); en todo caso, es una alternativa gratuita e inmediata en los lugares del mundo donde un mapa digital no existe o no está liberado.

Aunque la comunidad científica, las instituciones públicas y privadas, las industrias y la gente en general han reconocido su potencial, existe escepticismo respecto a la calidad de los datos OSM (Jokar, Zipf, Mooney, & Helbich, 2015), justificado por las propias características de este proyecto. En respuesta, en la última década se han investigado cuatro aspectos de OSM: administración y calidad de datos; contexto social; modelado de redes y rutas; y administración

territorial y urbana (Jokar, Zipf, Mooney, & Helbich, 2015). El presente trabajo pretende contribuir a la investigación de calidad de datos OSM, en el ámbito ecuatoriano. Específicamente se estudiará el componente horizontal de exactitud posicional de vías urbanas; esto, en varias Administraciones Zonales (es decir, divisiones administrativas) del Distrito Metropolitano de Quito (DMQ).

1.2. PLANTEAMIENTO DEL PROBLEMA

Entre las razones que se han aducido para poner en duda la utilidad de los datos VGI se encuentran: el desconocimiento de las motivaciones de los usuarios, la carencia de una reputación, la división digital, la inexperiencia de los usuarios, y la posibilidad de que exista terrorismo digital (Goodchild, 2007; Coleman, Georgiadou, & Labonte, 2009). En respuesta a estas dudas se ha investigado prácticamente todo lo relacionado a la VGI; sin embargo, en proyectos como OSM, varios aspectos aún no se comprenden en su totalidad. Lo que es más importante: dado que las características de OSM varían considerablemente a través del tiempo y del espacio, que las contribuciones en algunos países son emergentes, y que aún existen grandes vacíos (Jokar, Zipf, Mooney, & Helbich, 2015), es necesario continuar las investigaciones.

Mientras que en Europa existen importantes comunidades OSM, este no es el caso en Latinoamérica, África y la mayoría de Asia (Jokar, Zipf, Mooney, & Helbich, 2015); particularmente, las investigaciones realizadas sobre datos OSM son bastante eurocéntricas. En el ámbito latinoamericano, no se observan estudios relacionados con OSM en el estado del arte; no obstante, sí existen comunidades que en los últimos años han mapeado las urbes de estos países (Neis, 2019). En lo que respecta a contribuciones a OSM Ecuador constituye un caso especial pues, aunque existen registros desde 2012, la mayoría de las contribuciones corresponden a un único periodo en abril de 2016; la *Figura 1* demuestra este evento en la forma de un único pico.

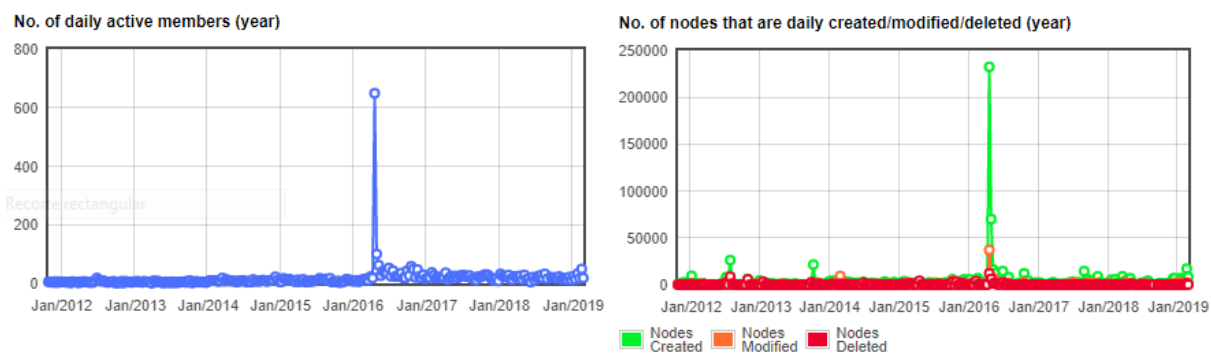


Figura 1. Miembros activos (izquierda) y nodos creados/modificados/eliminados (derecha) diariamente en OSM para el caso de Ecuador.

Fuente: Neis, 2019

La explicación de este fenómeno se remonta al sismo que el 16 de abril de 2016 tomó lugar en la costa ecuatoriana. En respuesta a este evento apareció la iniciativa Mapping Ecuador, dirigida por el Equipo Humanitario de OpenStreetMap (HOT, por sus siglas en inglés), con el propósito de mapear las regiones afectadas (IMPAQTO, 2016). Efectivamente, Mapping Ecuador consolidó las contribuciones a OSM en el país; sin embargo, siendo respuestas a una emergencia, muchos de estos datos presentan déficits en varios aspectos de calidad (Castro & Luna, 2018; Castro, Tierra, & Luna, 2019). En todo caso, los estudios realizados sobre OSM en Ecuador son incipientes y se desconoce la calidad de esos datos. En consecuencia, el estudiar un aspecto fundamental de calidad -la exactitud posicional horizontal- constituye el problema a abordar en el presente trabajo.

1.3. JUSTIFICACIÓN E IMPORTANCIA

A la fecha existen más de 300 aplicaciones que dependen completa o parcialmente de los datos OSM (List of OSM-based services, 2019); además, muchas plataformas aprovechan el mapa OSM, incluyendo: wikipedia.org, leaflet.js, flickr.com, pinterest.com, Uber, Strava, craigslist.org, TripAdvisor, entre otras (They are using OpenStreetMap, 2019). Lógicamente, los beneficios a obtenerse de dichas aplicaciones o plataformas van a venir en función de la calidad de los datos

OSM. En este sentido, Hashemi & Abbaspour (2015) demuestran cómo los problemas de calidad (errores posicionales, incoherencia lógica, etc.) pueden dirigir al usuario a realizar malas decisiones, e incluso perder tiempo y dinero.

La importancia del presente proyecto reside en que, mejorando su entendimiento de la VGI, el usuario será capaz de tomar decisiones acertadas respecto a dónde y cómo aprovecharla. Dichas decisiones dependerán de la calidad de los datos VGI, observando que, en el contexto geográfico, existen cinco componentes de calidad: posicional, temático, temporal, lógico y de compleción (Ariza, 2013); esto en concordancia con la norma ISO 19113. Si bien los cinco componentes poseen igual importancia y han sido objetos de diferentes estudios, este proyecto se limitará a evaluar la exactitud posicional horizontal; se trata del aspecto más estudiado, pues refleja la propia esencia de la información geográfica (Ariza, 2013).

Por otra parte, la escala de un conjunto de datos geográficos es un metadato crucial que las instituciones oficiales utilizan para estandarizar sus productos (Touya & Reimer, 2015). Además, la práctica de la geografía ha convertido a la escala en sinónimo de ambiente de uso, de organización de la temática descrita (Touya & Reimer, 2015), e incluso de calidad de un mapa. Mientras que las instituciones oficiales siguen procesos rigurosos para asegurar que sus productos se pueden utilizar hasta determinada escala, los proyectos VGI carecen de la misma, pues se nutren de contribuciones generadas de manera colaborativa. Por lo tanto, se observa la necesidad de inferir la escala para la cobertura OSM a través de algún acercamiento a sus datos.

Una de las premisas de esta investigación es que los usuarios de aplicaciones basadas en OpenStreetMap se sentirán más seguros y motivados si conocen la exactitud y la escala en el área que les interesa; entonces, a través de este proyecto, se respaldaría la utilización de la cobertura vial OSM en varios sectores del DMQ.

1.5. OBJETIVOS Y METAS

1.5.1. Objetivo Principal

Evaluar el componente posicional de la calidad y la escala de datos libres OpenStreetMap, a través del estudio de la exactitud posicional horizontal en vías urbanas contribuidas voluntariamente en Administraciones Zonales del Distrito Metropolitano de Quito, con el propósito de identificar las posibles aplicaciones de estos datos.

1.5.2. Objetivos Específicos

- Medir la exactitud posicional de las vías urbanas mediante la comparación de coordenadas horizontales (Este, Norte) en intersecciones equivalentes, con respecto a datos oficiales, obteniendo así los errores posicionales.
- Conocer la distribución de los errores posicionales mediante el cálculo de estadísticos locales, identificando zonas con distintos niveles de error.
- Inferir la escala de las vías urbanas, en función de los errores posicionales observados, generando valor agregado al conjunto de datos libres.
- Analizar las posibles aplicaciones y evidentes limitaciones, en función de la escala, de los datos OpenStreetMap para el caso del Distrito Metropolitano de Quito, con el propósito de respaldar el uso de los mismos.

1.5.3. Metas

- Una base de datos con las coberturas adquiridas y generadas.
- Un reporte de errores posicionales en cada Administración Zonal estudiada.
- Un reporte de estadísticos calculados en cada Administración Zonal estudiada.
- Un mapa de escalas inferidas en cada Administración Zonal estudiada.
- Un artículo científico con los resultados hallados.

CAPÍTULO II

MARCO TEÓRICO

2.1. ESTADO DEL ARTE

Desde su inicio en 2004, el proyecto OSM ha crecido progresivamente en número de usuarios y países alcanzados; de manera paralela, se ha evidenciado un incremento en investigaciones dedicadas a entender los aspectos que caracterizan a OSM. De acuerdo a Jokar et al. (2015) estas investigaciones se agrupan en cuatro aspectos; dos de ellos (modelado de rutas y administración urbana) se traducen en aplicaciones que aprovechan esta información. Ejemplos de tales aplicaciones incluyen el refinamiento a gran resolución de indicadores poblacionales (Bakillah, Liang, Mobasheri, Jokar, & Zipf, 2014), y la predicción del uso del suelo en diferentes metrópolis (Jokar & Vaz, 2015).

Otro aspecto de suma importancia en OSM, dada su naturaleza colaborativa, es el social. En este aspecto se hallan Coleman et al. (2009), con una discusión de los motivos para que un usuario contribuya o perjudique a un proyecto VGI, y Neis & Zipf (2012), con un análisis de niveles de contribución entre usuarios OSM registrados. Finalmente se encuentra el aspecto de calidad de datos, en donde el componente posicional ha recibido mayor atención, si bien sí existen estudios sobre los demás componentes; por ejemplo, el lógico en Hashemi & Abbaspour (2015).

Respecto al problema de conocer la exactitud posicional en OSM, en la literatura se encuentran dos acercamientos. El primero predice la exactitud a partir de otros datos que pueden ser explícitos, como la etiqueta de fuente (Touya & Reimer, 2015), o implícitos, como el número

de nodos detectados por un algoritmo (Noskov, 2018). El segundo acercamiento calcula la exactitud relacionando entidades en OSM con sus equivalentes en un conjunto de datos oficiales; Helbich, Amelunxen, Neis, & Zipf (2012) demuestran este acercamiento, el cual es de suma importancia para el desarrollo del presente trabajo.

2.2. BASE TEÓRICA

2.2.1. Información Geográfica Voluntaria

La Información Geográfica Voluntaria (VGI) es un fenómeno que ha revolucionado la manera como los usuarios de Sistemas de Información Geográfica (SIG) adquieren y utilizan datos. Su nacimiento ocurrió la primera década del siglo XXI, gracias al desarrollo de ciertas tecnologías: la Web 2.0, el Internet de Banda Ancha, el Sistema de Posicionamiento Global (GPS), los *geotags* y los avances en procesamiento de imágenes (Goodchild, 2007). Algunas de las iniciativas más reconocidas son Wikimapia, Flickr, *mash ups* de Google Earth, y por supuesto OpenStreetMap junto con sus aplicaciones derivadas. VGI fue en su momento una idea revolucionaria y, para comprenderla mejor, varios autores introdujeron nuevos conceptos:

- *neogeography* (Turner, 2006) se refiere a la evolución de la geografía, causada por la participación de personas que no poseen conocimientos científicos profundos.
- *humans as sensors* (Goodchild, 2007) se refiere al valor agregado que se gana cuando, a los sensores tradicionales, se suma la inteligencia y el conocimiento local de las personas.
- *wikification of GIS* (Sui, 2008) se refiere a las implicaciones de esta nueva tendencia para los SIG tradicionales, en cuanto a hardware, software, datos y personas.
- *producers* (Coleman, Georgiadou, & Labonte, 2009) se refiere al doble papel del usuario como productor y beneficiario de información colaborativa.

Entre las potencialidades de la VGI, los mismos autores señalan:

- sensores humanos; es posible pensar en el usuario como un sensor que, apoyado en ciertas tecnologías y añadiendo su conocimiento, genera mejor información que la convencional.
- poblaciones participativas; si bien las empresas generadoras de cartografía ya han trabajado antes con comunidades locales, este paradigma pasa de ser jerárquico a transversal.
- infraestructura de datos espaciales (IDE); las instituciones cartográficas oficiales pasan de generar la cartografía, a diseñar estándares por medio de los cuales otros la generan y alimentan a una IDE nacional o global.
- gestión de emergencias; esta clase de proyectos resultan oportunos durante las secuelas de emergencias o desastres, tal y como fue comprobado con Mapping Ecuador en 2016.

Por otra parte, también se han destacado debilidades de la VGI, las mismas que ponen en duda su calidad y limitan su utilización:

- carencia de una reputación; a diferencia de las instituciones oficiales, los proyectos colaborativos no poseen garantías con respecto a su calidad o veracidad.
- desconocimiento del motivo del usuario; aunque existen motivaciones evidentes -altruismo o intereses personales- en general se desconoce por qué una persona contribuye.
- inexperiencia del usuario; dado que las contribuciones son anónimas, lo más seguro es suponer que el usuario no posee conocimientos técnicos ni científicos avanzados.
- división digital; no todas las personas poseen acceso a Internet.
- terrorismo digital; aunque se han diseñado medidas con el propósito de controlar esta amenaza, siempre existe la posibilidad de que un usuario malintencionado altere o destruya la información.

2.2.2. OpenStreetMap

OpenStreetMap (OSM) se puede considerar el proyecto VGI que mayor acogida ha tenido entre las comunidades SIG y el público en general, alrededor de todo el mundo (Jokar, Zipf, Mooney, & Helbich, 2015). Como su nombre lo indica, fue creado como un mapa para las vías de las urbes del mundo; no obstante, su continuo crecimiento (*Figura 3*) justificó la introducción de cambios que permitan a los usuarios añadir prácticamente cualquier clase de entidad geográfica.

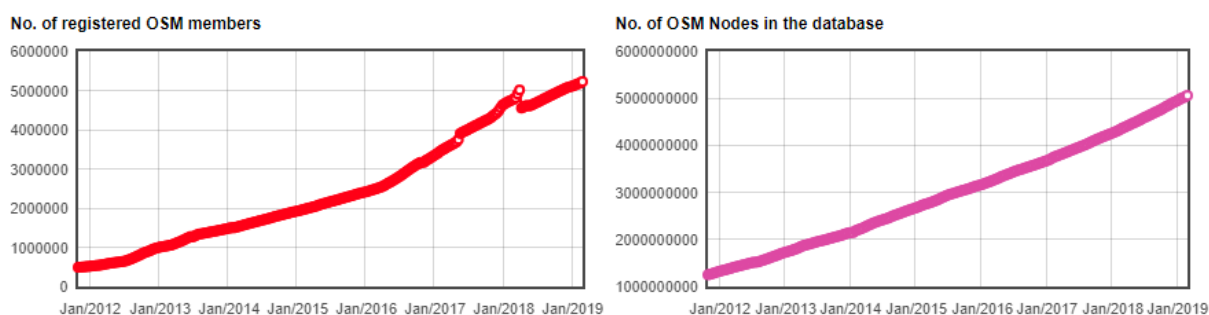


Figura 3. Crecimiento en número de miembros (izquierda) y nodos (derecha) en OSM.

Fuente: Neis, 2019

En OSM existen cuatro clases de elementos (Sax-Barnett, 2014):

- *nodes* son elementos puntuales (semáforos, monumentos...).
- *ways* son elementos lineales (calles, caminos, puentes...), formados por la unión de *nodes*.
- *areas* son elementos poligonales (edificaciones, parques...), formados por *ways* cerrados.
- *relations* son elementos que agrupan dentro de sí a otros elementos.

Cualquiera de estos elementos puede tener -o no- una serie de atributos almacenados en forma de etiquetas, compuestas de una clave y un valor: *tag: key = value* (Sax-Barnett, 2014). En áreas urbanas del DMQ, los keys más comunes son *name* (nombre), *highway* (autopista), *building* (edificio) y *source* (fuente) (Castro & Luna, 2018). Los elementos presentes en OSM pueden ser gratuitamente accedidos, editados, y descargados en un archivo .osm que posee estructura XML.

Es importante recordar que OSM no es un producto estático: está siendo aumentado, modificado y revisado en prácticamente todo momento. Actualmente existen varios métodos para agregar nuevos elementos: importación masiva de bases de datos, subida de *tracks* GPS, dibujo sobre imágenes aéreas Bing, entre otros. Estos métodos se han implementado a través de software libre conocido como editores; “iD” y “JOSM” son los editores más populares (Sax-Barnett, 2014). Si bien ahora es más sencillo para cualquier usuario el contribuir a OSM, ello supone más desafíos en el mantenimiento de los datos.

2.2.3. Big GeoData

Una característica adicional que puede adscribirse a la VGI es su condición de Big Data. Aunque este término no es nuevo, no posee una definición formal; con todo, es posible conceptualizarlo con la ayuda de unas pocas ideas. Tradicionalmente, estas son las tres Vs (Laney, 2001): Volumen: se trata de cantidades gigantescas de datos; Velocidad: estos datos son generados continuamente y están disponibles casi inmediatamente; y Variedad: estos datos provienen de diversas fuentes y presentan diferentes características.

OSM refleja claramente estas tres ideas: es una gran base de datos geográfica -volumen- que crece constantemente -velocidad- con las contribuciones de usuarios -variedad- en todo el mundo. A diferencia de la información generada por una institución oficial, donde existen procesos definidos y estrictos, la Big GeoData no atraviesa un procedimiento normal de control de calidad; esto se suma a las debilidades de la VGI descritas anteriormente.

En este sentido, existen tres maneras de fortalecer la calidad de la Big GeoData (Goodchild, 2013): la primera, aprovechar las observaciones independientes de las mismas entidades; la segunda, utilizar una jerarquía de usuarios para moderar las contribuciones (OSM ya utiliza este

control); y la tercera, implementar algoritmos automatizados de evaluación de calidad. En definitiva, solo el tercer método es sólido y rápido (Goodchild, 2013). El aseguramiento de la calidad de la Big GeoData es un asunto crucial en el mundo contemporáneo, donde las decisiones y los modelos dependen cada vez más de grandes volúmenes de datos.

2.3. BASE CONCEPTUAL

2.3.1. Calidad

Según Ariza (2013), la calidad “se refiere a la totalidad de las características de un producto o servicio tal que le confieren su aptitud para satisfacer unas necesidades explícitas e implícitas”. Las necesidades explícitas son aquellas que deben ser respondidas manifiestamente; por ejemplo - en el contexto cartográfico- la escala o la simbología de un mapa. Las necesidades implícitas, en cambio, son aquellas que no se expresan, pero son obvias; en este caso, un mapa responde a la necesidad implícita de poseer una representación de la realidad.

2.3.2. Exactitud Posicional Horizontal

Es uno de los aspectos fundamentales de calidad, exigible para cualquier tipo de dato (Olson, 2003; citado por Ariza, 2013), pues cuantifica cuán cerca del valor real se encuentran las mediciones generadas por un sistema de medición. Según Ariza (2013), la Exactitud Posicional Absoluta “es la proximidad entre los valores de coordenadas indicados y los valores verdaderos o aceptados como tales”, o la distancia euclidiana entre las coordenadas del mapa y las reales. Es importante destacar que, al utilizar Sistemas de Referencia muy diferentes, las coordenadas horizontales (Este, Norte) y verticales (altura elipsoidal u otro) se estudian de manera separada. La Exactitud Posicional Vertical no forma parte del presente estudio.

2.3.3. Escala

La escala es la razón entre distancias, tal y como son medidas en el mapa, y como son medidas en la realidad (International Cartographic Association, 1972; citado por Touya & Reimer, 2015). Es uno de los elementos constituyentes de un mapa e indica varias características de este: la densidad de la información, los contenidos significativos y el área de aplicación (Freitag, 1962).

2.3.4. Estadístico Getis-Ord

La G_i de Getis & Ord (1992) pertenece a una familia de estadísticos dedicados a medir el grado de autocorrelación o asociación espacial presente en datos puntuales. A diferencia de otros estadísticos que se aplican globalmente, como la I de Moran, la G_i se aplica localmente; esta ventaja permite localizar puntos con alta correlación, los cuales no aparecerían con un estadístico global. Sea un área dividida en n regiones, cada una identificada con un punto de coordenadas conocidas y con una medida x de una variable X positiva. El estadístico Getis-Ord es:

$$G_i(d) = \frac{\sum_j^n w_{ij}(d)x_j}{\sum_j^n x_j}; j \neq i \quad (1)$$

En (1) w_{ij} es una matriz de pesos simétrica y binaria, con unos representando todos los puntos dentro de una vecindad -determinada por la distancia d - del punto i , y ceros representando todos los demás puntos, incluido el propio i (Getis & Ord, 1992). G_i es local en el sentido de que otorga un valor a cada uno de los n puntos. Bajo consideraciones normales, el estadístico G_i de un determinado punto posee esperanza E y varianza V conocidas:

$$E[G_i(d)] = \frac{\sum_j^n w_{ij}(d)}{n-1} = \frac{W_i}{n-1} \quad (2)$$

$$V[G_i(d)] = \frac{W_i(n-1-W_i)}{(n-1)^2(n-2)} \frac{Y_{i2}}{Y_{i1}^2} \quad (3)$$

En (3) Y_{i1} y Y_{i2} son variables utilizadas para abreviar las siguientes expresiones:

$$Y_{i1} = \frac{\sum_j^n x_j}{n-1} \quad (4)$$

$$Y_{i2} = \frac{\sum_j^n x_j^2}{n-1} - Y_{i1}^2 \quad (5)$$

Getis & Ord (1992) también definen el estadístico derivado G^*_{i} en el cual se permite $j = i$, es decir que el i ésimo punto es incluido en su propia vecindad; si bien su ecuación es igual a G_i , G^*_{i} posee diferentes esperanza y varianza:

$$E[G^*_i(d)] = \frac{W_i^*}{n}; \quad W_i^* = W_i + w_{ii}(d) \quad (6)$$

$$V[G^*_i(d)] = \frac{W_i^*(n - W_i^*)}{n^2(n-1)} \frac{Y_{i2}}{Y_{i1}^2} \quad (7)$$

Además, Getis & Ord (1992) definen G y G^* como las versiones globales de G_i y G^*_{i} respectivamente; dichas versiones son similares a la I de Moran y arrojan un único estadístico para todos los puntos estudiados. Posteriormente (Ord & Getis, 1995) se redefine el estadístico Getis-Ord, de manera que en cada punto se calcula el valor Z_i (valor normalizado) del estadístico original:

$$Z_i(d) = \frac{G_i(d) - E[G_i(d)]}{\sqrt{V[G_i(d)]}} \quad (8)$$

La ecuación (6) también aplica para G^*_{i} (obteniendo Z^*_{i}), utilizando su respectivas esperanza y varianza. En adición a esta redefinición, el estadístico ahora permite evaluar cualquier tipo de variable X y asignar pesos no binarios en w_{ij} . Si la variable X es normal, la distribución del estadístico será exactamente normal; si es marcadamente sesgada, la distribución del estadístico no será normal, pero puede aproximarse a la normalidad conforme incrementa d (Ord & Getis, 1995). Los valores Z_i se pueden comparar con un valor crítico, positivo y negativo, con el fin de poner en evidencia la concentración significativa de medidas grandes o pequeñas de la variable X . Cuando

Z_i es mayor que el valor crítico positivo, el punto es caliente; es decir, existe concentración de medidas grandes en su vecindad. Por el contrario, cuando Z_i es menor que el valor crítico negativo, el punto es frío; existe concentración de medidas pequeñas en su vecindad. El conjunto de estadísticos locales sigue una distribución multinormal; por esta razón, al calcular valores Z críticos se debe aplicar una corrección de Bonferroni, descrita en Ord & Getis (1995).

Independientemente de qué versión se utilice, la utilidad del estadístico Getis-Ord es verificar la Hipótesis Nula de ausencia de autocorrelación espacial. En consecuencia, al rechazar la Hipótesis Nula se posee suficiente evidencia de la existencia de autocorrelación en determinado punto. Se ha demostrado que en un modelo espacial es importante estudiar la autocorrelación; el no hacerlo puede derivar en errores durante la interpretación del modelo (Getis & Ord, 1992). Así, el cálculo del estadístico Getis-Ord ya se ha incluido antes en la evaluación de la exactitud posicional (Helbich, Amelunxen, Neis, & Zipf, 2012), y es incluido en el presente estudio.

2.4. BASE LEGAL

2.4.1. Norma Internacional

Las normas ISO poseen aceptación internacional y establecen estándares en cada disciplina. Así, en el ámbito geográfico, la norma ISO 19113 define los elementos de calidad para datos espaciales: exactitud posicional, exactitud temática, exactitud temporal, consistencia lógica y compleción (Ariza, 2013). El posicional se divide en tres subelementos: absoluto, relativo y de datos de rejilla; solamente la Exactitud Posicional Absoluta interesa en este trabajo. En la práctica existen tres Metodologías de Control Posicional por Puntos (MCP) con amplio uso en ingeniería: NMA, EMA y NSSDA (Ariza, 2013); no obstante, las MCP no se han considerado en el estado del arte. Por lo tanto, desde el punto de vista legal, solo se considerará la norma nacional.

2.4.2. Norma Ecuatoriana

La calidad de la cartografía en el Ecuador se encuentra normada por el Instituto Geográfico Militar (IGM); particularmente, para escalas grandes, existen las “Especificaciones Técnicas para la Producción de Cartografía Escala 1:5000”. En ellas se estipula que, en un producto cartográfico, la exactitud posicional horizontal no debe exceder al módulo de escala multiplicado por un factor de 0.30 milímetros; esto para el 90% de entidades en una muestra aleatoria. Para el 10% restante, la exactitud posicional horizontal no debe exceder al módulo de escala multiplicado por 0.40 milímetros (Instituto Geográfico Militar, 2016).

Este reglamento es de suma importancia para obtener la escala pues, a diferencia de otros trabajos en los que la escala se infiere en función de varios factores (Touya & Reimer, 2015), en el presente trabajo solo dependerá de la exactitud posicional horizontal. Antes de describir la metodología es necesario advertir que va a operar de manera inversa al reglamento (Instituto Geográfico Militar, 2016). El reglamento determina, para una determinada escala, cuál es el máximo error posicional admisible; en la metodología se calculará, para un determinado error posicional, cuál es la escala adecuada.

A la vez, la metodología también difiere en el sentido de que no utilizará una muestra de puntos -como lo establecen las Metodologías de Control Posicional por Puntos- sino la población completa: en este caso, todas las intersecciones viales en una Administración Zonal. En una MCPP se muestrea porque el producto a evaluar se debe contrastar con las coordenadas más exactas posibles; estas son las obtenidas a través de posicionamiento GNSS de alta precisión (Ariza, 2013). En cambio, en la metodología presente se considerará a los puntos presentes en la cartografía oficial como los más exactos posibles; esto, con la premisa de que la cartografía oficial ha atravesado el proceso correspondiente de fiscalización (Instituto Geográfico Militar, 2017).

CAPÍTULO III

MARCO METODOLÓGICO

3.1. OBTENCIÓN DE DATOS

La metodología posee dos insumos principales: la cobertura vial correspondiente a la base de datos oficial, y la correspondiente a la base de datos libre. En el caso oficial, tanto el *shapefile* (.shp) de vías como de Administraciones Zonales del DMQ son descargados del sitio web http://gobiernoabierto.quito.gob.ec/?page_id=1122 (Municipio Metropolitano de Quito, s.f.). Ambos *shapefiles* son modificados de modo que posean un campo NAME con los nombres de las entidades; además, se observó que se encuentren proyectados (archivos .prj incluidos). En el caso de la información libre existen dos opciones: utilizar la API de OSM (Overpass API, 2019) o el paquete en lenguaje R diseñado para este fin (*osmdata*, Tabla 1); en ambos casos se descarga un archivo con extensión .osm, para cada Administración Zonal. La segunda opción es escogida, integrando la obtención de datos libres al código R; la misma es ejecutada el 9 de mayo de 2019.

3.2. SOFTWARE UTILIZADO

La totalidad del presente trabajo fue desarrollado en código R 3.5.3 (R Core Team, 2019), incluyendo gráficos, mapas y reportes generados. Para escribir, ejecutar y corregir el código se utilizó el software RStudio 1.1.463 (RStudio Team, 2016), cuya interfaz de usuario facilita considerablemente estas tareas. Paquetes de código R -que extienden su funcionalidad y se distribuyen gratuitamente- utilizados en el presente trabajo se detallan en la Tabla 1; la

documentación de cada paquete está disponible en la Red Comprehensiva del Archivo R (CRAN, 2019); específicamente, en páginas web con formato “https://CRAN.R-project.org/package=”, añadiendo al final de la dirección el nombre de cada paquete.

Tabla 1

Paquetes de código R utilizados

Paquete	Versión	Autor	Descripción
<i>knitr</i>	1.22	Yihui Xie	Genera reportes en lenguaje <i>markdown</i>
<i>osmdata</i>	0.0.9	Mark Padgham	Descarga datos OpenStreetMap
<i>rmarkdown</i>	1.12	Yihui Xie	Incorpora el lenguaje <i>markdown</i> para reportes
<i>sf</i>	0.7-3	Edzer Pebesma	Incorpora clases y funciones para manipulación de entidades espaciales
<i>spdep</i>	1.0-2	Roger Bivand	Calcula vecindades y estadísticos espaciales, entre ellos el de Getis-Ord
<i>tidyverse</i>	1.2.1	Hadley Wickham	Importa varios paquetes para manipulación de datos, entre ellos <i>dplyr</i> y <i>purrr</i>
<i>tmap</i>	2.2	Martijn Tennekes	Genera mapas estáticos e interactivos

3.3. DIAGRAMA DE FLUJO

Con ayuda de los paquetes y el software descritos en la sección anterior, cuatro scripts (secuencias de código) fueron desarrollados: *castrOSM1_eval.R*, *castrOSM1_repo.Rmd*, *castrOSM2_eval.R* y *castrOSM2_repo.Rmd*; los cuatro scripts están diseñados para ser ejecutados en ese orden exacto. Las extensiones *.R* y *.Rmd* denotan scripts escritos en lenguaje R y *markdown*, respectivamente. La *Figura 4* presenta el diagrama de flujo con el uso de los cuatro scripts.

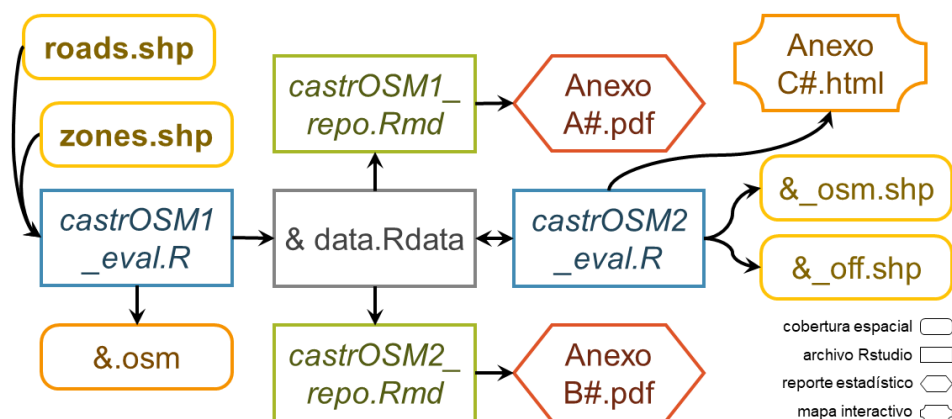


Figura 4. Diagrama de Flujo de la Metodología aplicada.

La *Figura 4* demuestra tres aspectos importantes de la metodología: primero, las coberturas vial y zonal son los únicos insumos; segundo, los Anexos A#, B#, C# se corresponden con las metas trazadas; tercero, en los archivos resultantes, los símbolos “&” y “#” identifican nombre y número de una Administración Zonal. Estas se numeran de la siguiente manera: Calderón 1, Eloy Alfaro 2, Eugenio Espejo 3, Manuela Sáenz 4, Quitumbe 5. Cabe aclarar que los mencionados Anexos A#, B#, C# y los cuatro scripts se hallan en la respectiva carpeta digital, mas no en el presente documento. En la siguiente sección se describe el funcionamiento de cada script.

3.4. METODOLOGÍA POR SCRIPT

3.4.1. Primer Script de Evaluación

El primer script se denomina `castrOSM1_eval.R`; es el más importante pues involucra la mayor parte del proceso y del tiempo de ejecución. Como entrada se emplean los dos *shapefiles* (vial y zonal), indicando además el nombre de la Administración Zonal de interés. Como salida se obtienen una cobertura OSM descargada y un *workspace* o Espacio de Trabajo de R; este último es un archivo con extensión `.Rdata` que contiene los objetos generados, entendiéndose que en lenguaje R cualquier elemento obtenido o generado es un “objeto”. El funcionamiento de este script depende intensivamente de funciones procedentes de los paquetes *tidyverse*, *sf* y *spdep*. De manera resumida, el funcionamiento contempla las siguientes acciones:

- indicar la Administración Zonal de interés e importar los paquetes necesarios.
- leer la cobertura zonal, seleccionando únicamente la zona de interés.
- leer la cobertura vial oficial y recortar la porción dentro de la zona de interés.
- intersecar la cobertura vial consigo misma para obtener las intersecciones viales, descartando intersecciones con coordenadas repetidas y de vías de igual nombre.

- descargar las polilíneas con etiqueta *highway* en OpenStreetMap, con los límites adecuados.
- leer la cobertura OpenStreetMap descargada y recortar dentro de la zona de interés.
- intersecar la cobertura OpenStreetMap consigo misma para obtener las intersecciones viales, descartando intersecciones con coordenadas repetidas y de vías de igual nombre.
- calcular el error posicional, relacionando las intersecciones oficiales y OpenStreetMap con las coincidencias de los nombres de las dos vías intersecadas.
- recalculer el error posicional, buscando el vecino más cercano entre las intersecciones.
- generar los polígonos de Voronoi para las intersecciones OpenStreetMap.
- calcular los estadísticos globales G y G^* sobre el error, variando el tamaño de la vecindad.
- calcular el estadístico local G^* sobre el error, variando el tamaño de la vecindad.
- guardar el *workspace*.

Cada uno de estos trece incisos identifica una sección en el script; la *Figura 5* es una captura de pantalla de estas secciones (los renglones que comienzan con tres numerales). Para mantener la uniformidad, todo el código fue redactado en inglés, incluyendo comentarios y mensajes.

```

1 tar = "Quitumbe" # target zone
2
3 message(tar, " is the Target Zone")
4
5 setwd("C:/datasets") # has NAMEd zones and roads shpfiles
6
7 > ### libraries functions
86 > ### offic zones reading
94 > ### offic roads reading
105 > ### offic roads intsect
122 > ### osm highway dwnload
134 > ### osm highway reading
146 > ### osm highway intsect
164 > ### joining in the name
180 > ### joining by distance
199 > ### voronoi tessellation
211 > ### getsord preparation
227 > ### getsord calculation
252 > ### the last messengers

```

Figura 5. Captura de pantalla de las Secciones del primer Script.

La obtención de intersecciones viales incluye las correcciones recomendadas para los inconvenientes presentados en Castro, Tierra, & Luna (2019). Una vez obtenidas las intersecciones viales en las dos coberturas, el problema de relacionar las intersecciones equivalentes se aborda con dos métodos. El primero -la Unión por Nombres, octava sección del script- equivale a una operación común en análisis de datos conocida como *join*; en ella, dos tablas se unen utilizando campos en común, los nombres de dos calles en este caso.

El inconveniente con este método es la generación de gran cantidad de errores atípicamente grandes, como se ha constatado ya en trabajos previos (Castro, Tierra, & Luna, 2019). La alternativa es el segundo método -el Vecino más Cercano, novena sección del script- donde simplemente se busca, para cada intersección en OpenStreetMap, el equivalente oficial más cercano. El único requisito es seleccionar un radio máximo de búsqueda, pues si se permitiese un radio infinito se relacionarían intersecciones que en realidad no son vecinas; este radio se calcula como un determinado cuantil de los errores hallados en el primer método.

Una vez calculados los errores posicionales a través del Método del Vecino más Cercano, se generan los polígonos de Voronoi (Bravo, s.f.) correspondientes a las intersecciones; el propósito es homogeneizar el error y obtener regiones con una determinada escala. El módulo de esta escala se calcula: $M = 1000 e/mm$; donde M es el módulo, e es el error posicional en metros, y mm es el factor de 0.30 milímetros determinado en la legislación (Instituto Geográfico Militar, 2016).

Las versiones global y local del estadístico de Getis-Ord son calculadas para conocer dónde existe agrupación significativa de valores altos o bajos del error posicional. En el caso global se calculan G y G* variando la distancia de vecindad, desde 80 hasta 280 metros, con intervalos de 20 metros. La evaluación de ambos estadísticos se basa en el *p.value* (valor p), es decir, en la probabilidad de observar un valor G o G* extremo si bien la Hipótesis Nula es verdadera; aunque

no es lo mismo, este valor está relacionado con la probabilidad de cometer el Error Tipo I. Si $p.value < 0.05$ entonces hay suficiente evidencia para afirmar la existencia de autocorrelación espacial, a nivel global y con un nivel de significancia del 5%.

En el caso local se calcula solamente el estadístico G^*_i , permitiendo así que el error en el punto i influya en su propia vecindad. El estadístico local se calcula variando la vecindad -utilizando distancias relevantes observadas en el caso global- y un factor de homogeneización de escala, introducido para evitar la aparición de escalas “inexactas”. Puesto que es práctica extendida el generar cartografía con módulos múltiplos de cierta cantidad, se introdujo una modificación al cálculo del módulo: $M_K = K \text{ ceiling}(1000 e / (K \text{ mm}))$; donde K es el número natural del cual los módulos serán múltiplos al nivel de miles, y *ceiling* es una función convencional que redondea números al inmediato superior (pues redondear un módulo al inmediato inferior supone un error). La *Tabla 2* ejemplifica el efecto de calcular la escala con diferentes errores y factores K .

Tabla 2

Escalas para Errores Posicionales y Factores de Homogeneización variables

e [m]	1 : M	1 : M₁	1 : M₂	1 : M₅	1 : M₁₀
0.10	1 : 333.333	1 : 1 000	1 : 2 000	1 : 5 000	1 : 10 000
0.30	1 : 1 000.000	1 : 1 000	1 : 2 000	1 : 5 000	1 : 10 000
0.50	1 : 1 666.666	1 : 2 000	1 : 2 000	1 : 5 000	1 : 10 000
1.00	1 : 3 333.333	1 : 4 000	1 : 4 000	1 : 5 000	1 : 10 000
2.00	1 : 6 666.666	1 : 7 000	1 : 8 000	1 : 10 000	1 : 10 000
5.00	1 : 16 666.666	1 : 17 000	1 : 18 000	1 : 20 000	1 : 20 000
10.00	1 : 33 333.333	1 : 34 000	1 : 34 000	1 : 35 000	1 : 40 000

Finalmente, del estadístico G^*_i se obtienen los valores Z y se identifican aquellos que son significativos. Comúnmente, a estos valores se los denomina puntos calientes o fríos, según si son mayores o menores que cierto valor Z crítico, calculado observando la corrección de Bonferroni (Ord & Getis, 1995). La consecuencia de dicha corrección es que los valores críticos crecen conforme crece el tamaño muestral n ; la *Tabla 3* demuestra los valores críticos corregidos para

varios tamaños muestrales y niveles de significancia, dado que la variable estudiada sea normal. En este caso fueron calculados los valores Z críticos correspondientes a los niveles de significancia del 10% y del 0.1% ($Z_{0.900}$ y $Z_{0.999}$); en consecuencia un punto iésimo, con estadístico normalizado Z_i , será designado: muy caliente (*very hot*) si $Z_i \geq Z_{0.999}$, caliente (*hot*) si $Z_i \geq Z_{0.900}$, regular si $-Z_{0.900} < Z_i < Z_{0.900}$, frío (*cold*) si $Z_i \leq -Z_{0.900}$, y muy frío (*very cold*) si $Z_i \leq -Z_{0.999}$.

Tabla 3

Valores Z Críticos para Tamaños Muestrales y Niveles de Significancia variables

n	Z_{0.900}	Z_{0.950}	Z_{0.990}	Z_{0.999}
10	2.3091	2.5683	3.0888	3.7248
30	2.6964	2.9291	3.4050	3.9885
100	3.0778	3.2889	3.7238	4.2659
500	3.5375	3.7134	4.1075	4.6200
1000	3.7062	3.8855	4.2643	4.7667

El *workspace* es el resultado trascendental del primer script, pues constituye el insumo de los scripts restantes. La *Tabla 3* describe los objetos más importantes, de todos los presentes en el *workspace*; objetos de menor importancia son generados simplemente como intermediarios. Solo los objetos de clase *sf* (del paquete homónimo) son coberturas y, por ende, poseen geometría.

Tabla 4

Objetos relevantes en el Espacio de Trabajo de R

Objeto	Clase	Geometría	Descripción
osm	sf	polilíneas	Vías leídas desde los datos OpenStreetMap
osmInt	sf	puntos	Intersecciones viales obtenidas a partir de “osm”, con errores posicionales calculados con el Método de Vecino más Cercano y etiquetas de Puntos Calientes
osmNam	sf	puntos	Intersecciones viales obtenidas a partir de “osm”, con errores posicionales calculados con el Método de Unión por Nombres
osmNamQ	numeric		Cuantiles de los errores posicionales en “osmNam”
osmVor	sf	multipolígonos	Polígonos de Voronoi obtenidos a partir de “osmInt”
shp	sf	polilíneas	Vías leídas desde los datos oficiales
shpInt	sf	puntos	Intersecciones viales obtenidas a partir de “shp”
zon	sf	polígono	Límite de la Administración Zonal de interés
zonProj	crs		Sistema de Referencia Coordinado de “zon”
zonUbox	bbox		Límites de “zon” en latitud y longitud

3.4.2. Primer Script de Reporte

El *workspace* guardado por el script anterior es el único insumo del primer script de reporte, denominado `castrOSM1_repo.Rmd`. Su función es generar un reporte en español (Anexo A#) que permita visualizar los resultados obtenidos en el script anterior. El reporte contiene:

- una descripción de la ejecución, incluyendo el tiempo que demoró, el número de errores hallados en los dos métodos, el porcentaje de ambos con respecto al total de intersecciones viales en OpenStreetMap, y el porcentaje de coincidencia entre ambos métodos.
- dos gráficos (uno de caja y un histograma) demostrando la distribución de los errores posicionales tanto del Método de Unión por Nombres, como del Vecino más Cercano.
- un gráfico de líneas que permite comparar el valor p de los estadísticos globales G y G^* versus distancia de vecindad.
- un gráfico de líneas que permite comparar el valor p del estadístico local G^*_i versus distancia de vecindad y factor de homogeneización.
- un mapa con la ubicación y la magnitud de los errores posicionales hallados.

3.4.3. Segundo Script de Evaluación

El *workspace* guardado por el primer script es también el único insumo del segundo script de evaluación, `castrOSM2_eval.R`. La necesidad de analizar los estadísticos -antes de decidir cuál distancia de vecindad y factor de homogeneización se utilizará definitivamente- es la razón por la cual no se ejecuta toda la evaluación en un solo script. En consecuencia, antes de ejecutar este script, el usuario debería estudiar el reporte generado previamente. El funcionamiento de este script contempla las siguientes acciones:

- requerir del usuario una distancia de vecindad y un factor de homogeneización.

- corregir los errores calculados a través del Método del Vecino más Cercano en las intersecciones, considerando los resultados de $G*_i$ con la distancia de vecindad escogida.
- corregir de la misma manera los errores en los polígonos de Voronoi y calcular la escala, considerando el factor de homogeneización escogido.
- generar un mapa interactivo con la vialidad oficial, las intersecciones en OpenStreetMap con los errores corregidos y los polígonos de Voronoi con las escalas.
- exportar las coberturas finales como shapefiles y guardar nuevamente el *workspace*.

Los puntos calientes y fríos identificados en el primer script indican la agrupación significativa de errores posicionales grandes y pequeños, respectivamente. Los puntos fríos no fueron modificados; en cambio, en las vecindades de los puntos calientes (y muy calientes) se seleccionó solamente el mayor error para calcular la escala de toda la vecindad. De esta manera se consiguen tres efectos: primero, el polígono de Voronoi abarca toda la vecindad del punto caliente; segundo, aumenta la importancia de los errores significativamente grandes; y tercero, desaparece la autocorrelación, lo cual es verificable recalculando los estadísticos de Getis-Ord. La *Figura 6* ejemplifica una vecindad con varios puntos calientes, antes y después de esta corrección.

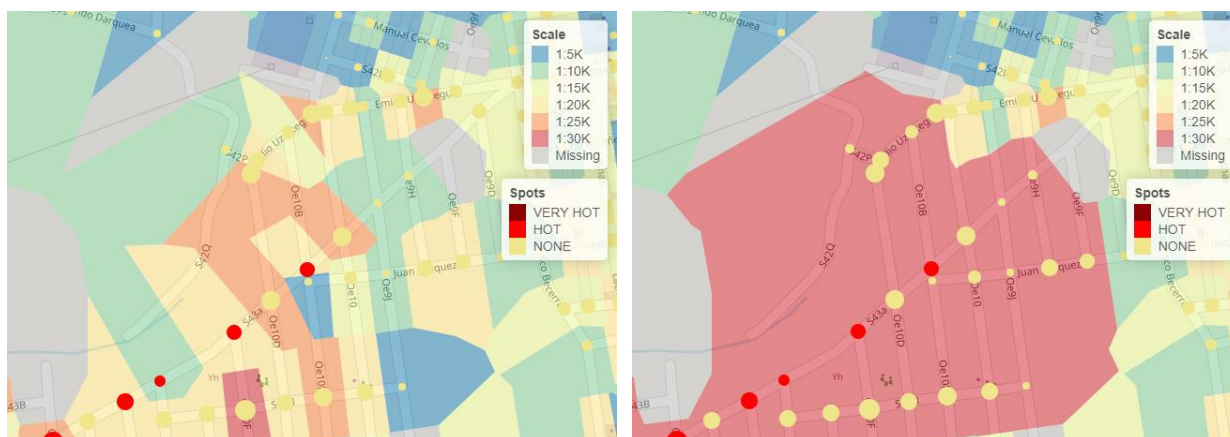


Figura 6. Corrección de una Vecindad con presencia de puntos calientes, antes (izquierda) y después (derecha).

Las capturas de pantalla presentadas en la *Figura 6* corresponden a un mapa interactivo generado con el presente script. Detrás de las cortinas, la tecnología que permite generar esta clase de mapas en formato HTML es *leaflet* (Cherniavskii, 2019), una librería javascript cuya implementación en R es posible con el paquete *tmap*. Debido a la gran densidad de puntos por Administración Zonal, y para cumplir mejor con la meta (Anexo C#) del presente trabajo, esta alternativa interactiva fue implementada.

Las coberturas finales son exportadas al formato shapefile de ESRI, el cual representa un estándar cartográfico. Esas coberturas son: la vialidad oficial (`&_off_roads.shp`); las intersecciones oficiales (`&_off_intersections.shp`); la vialidad en OSM (`&_osm_highways.shp`); las intersecciones en OSM, incluyendo errores y etiquetas de puntos calientes (`&_osm_intersections_errors.shp`); y los polígonos de Voronoi, incluyendo escalas (`&_osm_polygons_scales.shp`). Finalmente, el *workspace* es guardado nuevamente, actualizando los objetos corregidos.

3.4.4. Segundo Script de Reporte

El *workspace* actualizado es el único insumo del último script, `castrOSM2_repo.Rmd`. Este genera un segundo reporte (Anexo B#) que contiene:

- una descripción de la última ejecución, incluyendo la distancia de vecindad y el factor de homogeneización escogidos, el cálculo de G^* post correcciones, y los valores Z críticos.
- el histograma de G^*_i .
- dos gráficos de barras, uno del conteo de puntos calientes y fríos, y otro de las escalas.
- un mapa con la ubicación de los puntos calientes y fríos.
- un mapa con los polígonos de Voronoi y sus respectivas escalas.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. RESULTADOS POR SCRIPT

4.1.1. Primer Script de Evaluación

La *Tabla 5* contiene algunos detalles generales sobre la ejecución del primer script, y sobre las coberturas viales oficial y OpenStreetMap, para cada Administración Zonal estudiada.

Tabla 5

Detalles de la ejecución y de las coberturas viales

Variable	Calderón	Eloy Alfaro	Eugenio Espejo	Manuela Sáenz	Quitumbe
Tiempo de ejecución [seg]	183	269	450	169	396
Peso del workspace [Kb]	1618	2093	3218	1373	2487
Área de la zona [Ha]	8647	4449	11829	4784	8854
Vías oficiales (a)	9276	10792	13919	6367	14644
Vías oficiales con nombre (b)	6713	6998	11443	4500	3735
Proporción con nombre (b / a)	0.7237	0.6484	0.8221	0.7068	0.2551
Vías OSM (c)	3641	4644	9187	3663	5375
Vías OSM con nombre (d)	2264	3309	6357	2551	3919
Proporción con nombre (d / c)	0.6218	0.7125	0.6920	0.6964	0.7291

Los datos presentados confirman la relación directa existente entre el tiempo de ejecución y el peso del resultado; en cambio, la relación con los datos no es tan clara pues, mientras mayor sea la zona estudiada, ni el tiempo de ejecución ni el peso resultante son necesariamente mayores. Otros factores, como la densidad y la geometría viales parecen ejercer mayor influencia en la ejecución del script. Por otra parte, llama la atención la ausencia de nombres tanto en la vialidad OpenStreetMap como en la propia cartografía oficial; lo que es más importante, mientras que en

OSM el porcentaje de vías sin nombres es estable (entre 27% y 38%), en las vías oficiales varía considerablemente, siendo la Administración Zonal Quitumbe la más perjudicada con casi el 75% de segmentos viales sin nombrar. Como se observará a continuación, la ausencia de nombres supone el principal inconveniente en el primer método aplicado.

La *Tabla 6* presenta el desempeño de los dos métodos aplicados para relacionar las intersecciones: el Método de Unión por Nombres (MUN) y el de Vecino más Cercano (MVC).

Tabla 6

Desempeño de los métodos para relacionar intersecciones

Variable	Calderón	Eloy Alfaro	Eugenio Espejo	Manuela Sáenz	Quitumbe
Intersecciones viales OSM (a)	4710	5813	10268	3955	7774
MUN: errores hallados (b)	711	877	1207	567	169
MUN: proporción (b / a)	0.1510	0.1509	0.1175	0.1434	0.0217
MUN: coincidencias con MVC	0.9198	0.9213	0.9213	0.9224	0.9172
MVC: errores hallados (c)	3635	4754	6260	2576	5925
MVC: proporción (c / a)	0.7718	0.8178	0.6097	0.6513	0.7622
MVC: coincidencias con MUN	0.1799	0.1700	0.1776	0.2030	0.0262

Las intersecciones en las que se calculó un error con ayuda de los nombres de las vías son escasas, entre el 12% y el 15%, sin contar el caso excepcional de Quitumbe, donde la gran cantidad de nombres ausentes explica el valor del 2%. En contraste, el desempeño utilizando el vecino más cercano es aceptable, calculando el error entre el 61% y el 82% de las veces (72.25% en promedio). Resulta interesante constatar que un porcentaje elevado y estable del primer método coincide con el segundo; quiere decir que, en promedio, el 92.04% de las relaciones respaldadas por nombres son con el vecino más cercano. Con todo, esto también quiere decir que un 8% de intersecciones relacionadas - uniendo nombres de vías- no lo están con la más cercana. De ese 8% la mayoría son errores atípicamente grandes, mayores que el radio de búsqueda del vecino más cercano; en todos los casos, el radio se fijó en el percentil 92.5 de los errores del primer método (*Tabla 7*).

Tabla 7

Percentiles de los errores calculados con el Método de Unión por Nombres

Administración Zonal	mín.	Q₁	Q₂	Q₃	P₉₀	P_{92.5}	P₉₅	P_{97.5}	máx.
Calderón	0.044	1.678	2.731	4.251	6.98	9.76	38.82	63.48	507.9
Eloy Alfaro	0.120	1.634	2.665	4.091	6.66	9.68	14.52	46.95	1043.
Eugenio Espejo	0.037	1.461	2.248	3.567	5.79	7.48	10.46	35.25	929.3
Manuela Sáenz	0.081	1.805	2.808	4.266	7.67	8.66	11.04	15.83	406.3
Quitumbe	0.190	1.520	2.409	3.481	5.91	6.56	7.51	17.99	121.9

Las *Figuras 7 y 8* presentan los gráficos de caja de los errores calculados en el primer y segundo método respectivamente, por Administración Zonal. Es importante notar que el primer gráfico fue construido con escala logarítmica, para facilitar la visualización de errores atípicos.

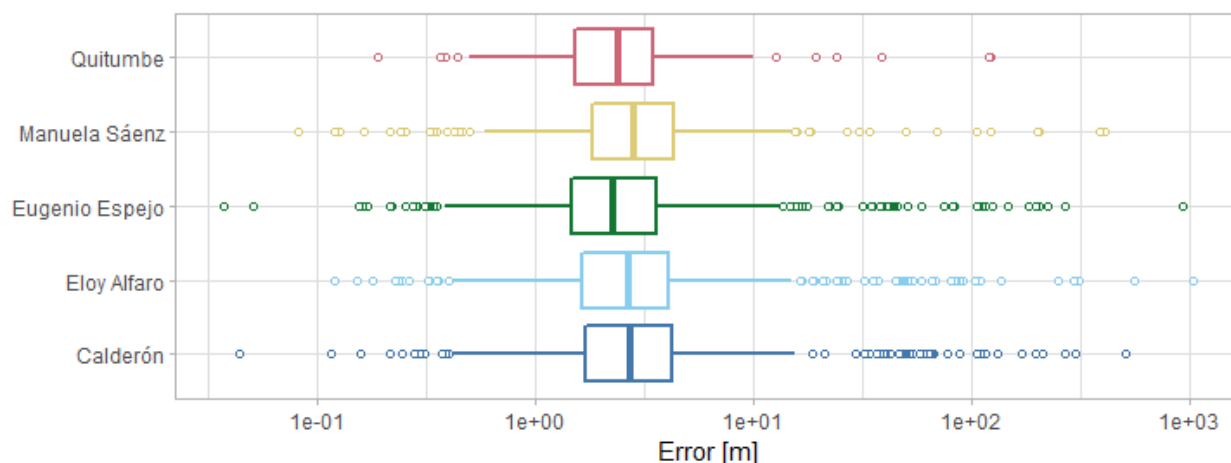


Figura 7. Gráficos de Caja de errores calculados con el Método de Unión por Nombres.

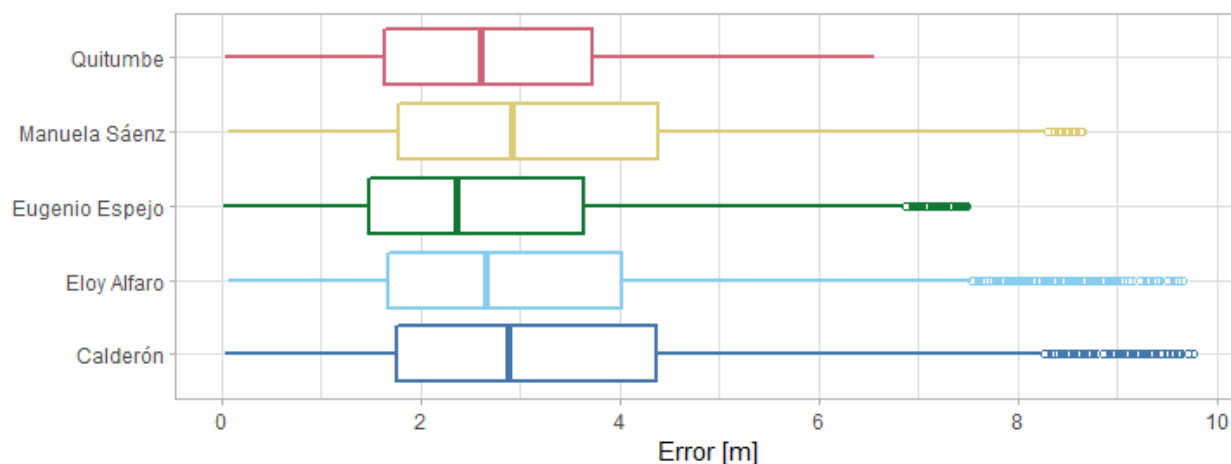


Figura 8. Gráficos de Caja de errores calculados con el Método de Vecino más Cercano.

4.1.2. Primer Script de Reporte

Al calcular los estadísticos globales, con varias distancias de vecindad, solo se conservaron los valores p ($p.value$). La *Figura 9* demuestra la evolución del valor p , conforme crece la distancia, para G y G^* en cada Administración Zonal; la línea entrecortada representa $p.value = 0.05$.

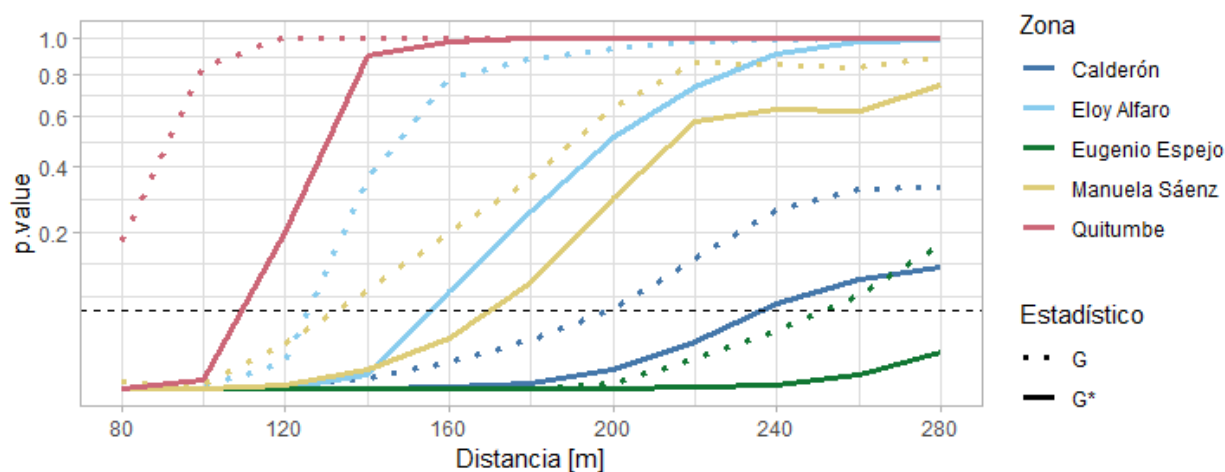


Figura 9. Relación del valor p y la Distancia de Vecindad en los estadístico G y G^* .

Como indica esta figura, el valor p exhibe comportamientos variados en las cinco Administraciones Zonales. Particularmente en Quitumbe no existe evidencia de autocorrelación espacial salvo al utilizar vecindades pequeñas, con distancias menores que 110 metros; no obstante, esto es un efecto de unos errores comparativamente pequeños (*Figura 8*) debido al bajo desempeño del primer método -que reduce el radio de búsqueda del segundo- en esta zona. Por el contrario, en Calderón existe evidencia de autocorrelación incluso con vecindades grandes, de hasta 250 metros; esto se relaciona con el hecho de poseer esta zona los cuantiles más grandes (*Tabla 7*).

De manera general, conforme crece la distancia la autocorrelación desaparece, siendo el estadístico G^* más estricto que G en todos los casos. Los mayores cambios del valor p se observan entre 120 y 240 metros; de esta manera, la *Figura 10* demuestra la relación del valor p con los factores de homogeneización, para el estadístico G^* a 120, 160, 200 y 240 metros.

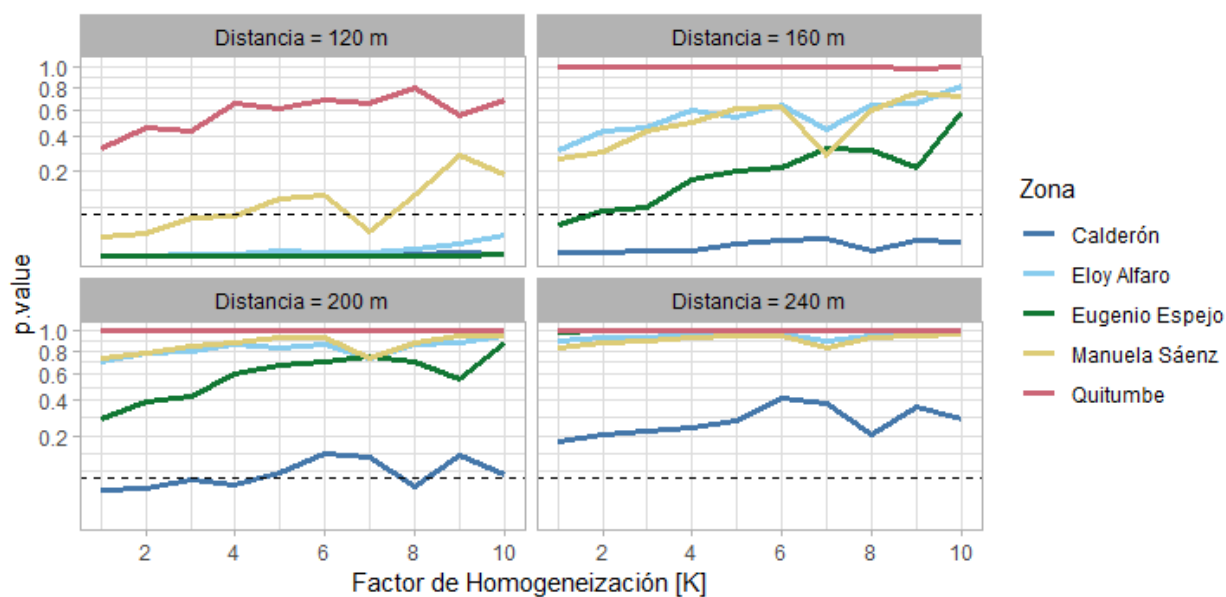


Figura 10. Relación del valor p y el Factor de Homogeneización en el estadístico G^* .

En esta figura se demuestra que, si bien el valor p también aumenta con el factor de homogeneización, la mayor cantidad de variación se debe a la distancia; particularmente al pasar de 120 a 160 metros se produce un gran cambio, después del cual persiste la autocorrelación solamente en Calderón. En consecuencia, se escogió 150 metros como la distancia de vecindad definitiva para identificar los puntos calientes y fríos en todas las Administraciones Zonales. Cabe destacar que, una vez desaparecida la autocorrelación, da igual escoger cualquier factor de homogeneización; no obstante, con el fin de simplificar los mapas resultantes, se escogió el factor de diez, obteniendo módulos múltiples de 10 000. Finalmente, los mapas de errores posicionales no se hallan en este documento sino en los Anexos A# recordando la numeración de las zonas.

4.1.3. Segundo Script de Evaluación

La Tabla 8 presenta un resumen de valores importantes utilizados en la corrección de errores posicionales, implementando lo aprendido con el estadístico de Getis-Ord.

Tabla 8*Detalles de la implementación del estadístico Getis-Ord*

Variable	Calderón	Eloy Alfaro	Eugenio Espejo	Manuela Sáenz	Quitumbe
Errores hallados (MVC)	3635	4754	6260	2576	5925
Valor crítico al 10% ($Z_{0.900}$)	4.021	4.084	4.147	3.939	4.135
Valor crítico al 0.1% ($Z_{0.999}$)	5.008	5.059	5.111	4.941	5.101
Valor p del estadístico G*	0.059	0.627	0.262	0.581	0.990
Número de puntos muy fríos	0	0	0	0	0
Número de puntos fríos	3	0	0	1	2
Número de puntos regulares	3602	4728	6183	2566	5917
Número de puntos calientes	21	13	70	9	6
Número de puntos muy calientes	9	13	7	0	0

Los valores p en esta tabla corresponden al estadístico global G* calculado con distancia de vecindad de 150 metros, y factor de homogeneización de 10. Después de la corrección, solamente Calderón presenta un valor p cercano a 0.05, indicando la persistencia de alguna autocorrelación espacial. En las demás zonas ya no se evidencia autocorrelación, siendo Eugenio Espejo donde se observó el mayor cambio, debido al gran número de puntos calientes hallados. Los puntos fríos, por otra parte, son pocos o inexistentes; esto indica que, a pesar de hallarse en gran cantidad, los errores posicionales pequeños no poseen una estructura de autocorrelación.

4.1.2. Segundo Script de Reporte

Las *Figuras 11 y 12* son gráficos de barra representando el conteo de puntos calientes y el porcentaje de polígonos de Voronoi con determinada escala, respectivamente, por Administración Zonal. En todas las zonas los polígonos escala 1 : 10 000 son los más abundantes; no obstante, en Eugenio Espejo y Manuela Sáenz, los polígonos sin escala (<NA>) son igual de abundantes debido al elevado porcentaje de intersecciones sin error (39% y 35%). Nuevamente, los mapas de puntos calientes y escalas se encuentran en los Anexos B# y no en este documento.

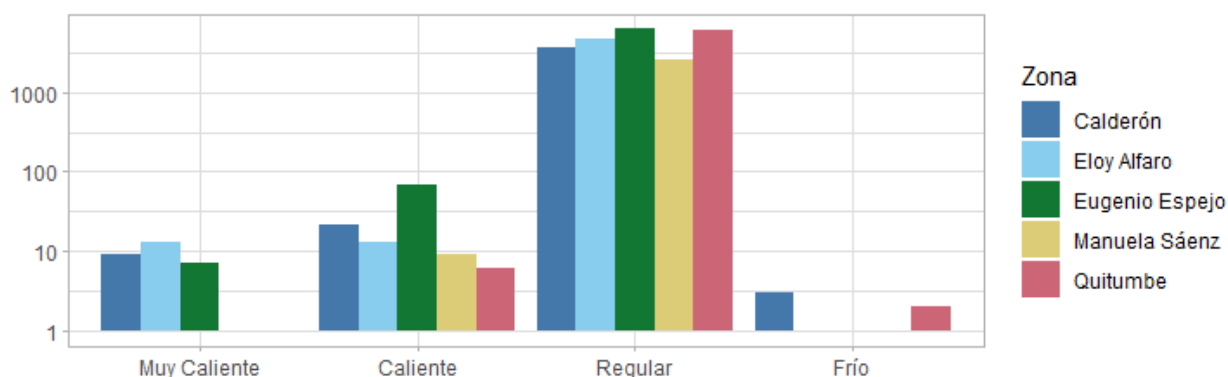


Figura 11. Conteo de Puntos Calientes por Administración Zonal.

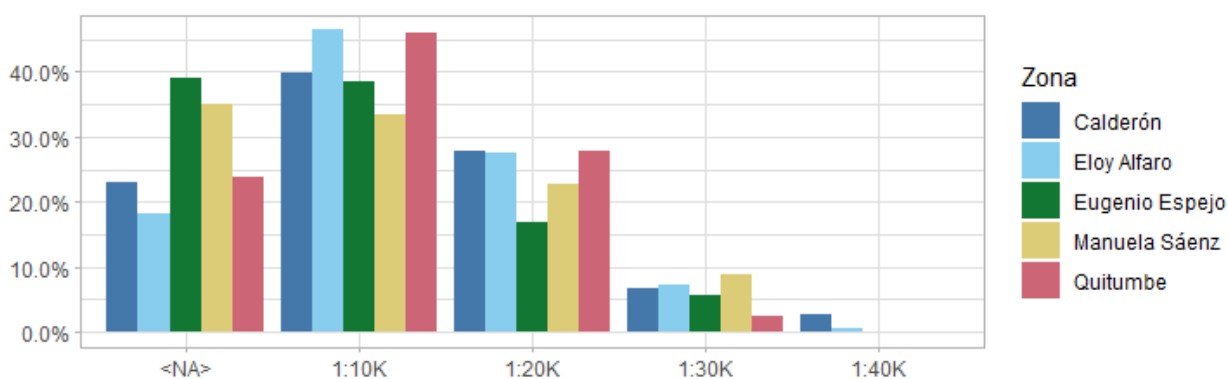


Figura 12. Porcentaje de Polígonos de Voronoi por escala, por Administración Zonal.

4.2. DISCUSIÓN DE RESULTADOS

La metodología presentada, al implementar programación, demuestra ventajas respecto a una metodología SIG tradicional. No solo el tiempo de ejecución varía entre 2.8 y 7.5 minutos en el primer script, y menos de un minuto en los demás; también la intervención del usuario es mínima: sus acciones se reducen a obtener los insumos, indicar la zona de interés y ejecutar el código. Con todo, existen tres requisitos para que la ejecución sea correcta: primero, los dos insumos (roads.shp y zones.shp) deben colocarse en el directorio “C:/datasets/” dentro de carpetas homónimas (roads y zones); segundo, ambos shapefiles deben poseer un campo NAME; tercero, debe existir una conexión a internet. Una consideración adicional es que los tiempos de ejecución reportados son

orientativos y cambiarán de manera natural si se repite el proceso, incluso con los mismos datos. Un verdadero análisis del desempeño incluye un procedimiento denominado *benchmarking*, donde el código se ejecuta repetidamente. Además de realizar dicho análisis, el desempeño podría mejorar ejecutándose el código en paralelo o en un mejor procesador.

Con respecto al Método de Unión por Nombres, existen dos grandes inconvenientes. El primero, ya mencionado, es la gran cantidad de vías sin nombrar tanto en la cartografía oficial como en OSM, imposibles de relacionar a través de este método. El segundo son las vías mal trazadas, mal nombradas o ausentes en OSM; en un trabajo anterior (Castro, Tierra, & Luna, 2019) se halló que tales vías pueden ser responsables de la aparición de errores atípicamente grandes. Considérese la *Figura 13* que ilustra el error más grande hallado en la Administración Zonal Eloy Alfaro: si bien en OSM la vía naranja va desde la intersección A hasta la C, oficialmente cambia de nombre en B. Al unir por nombres, el método falló al relacionar la vía naranja con la roja en A, pero sí lo logró en C, generando de esta manera un error de 1044.558 metros.

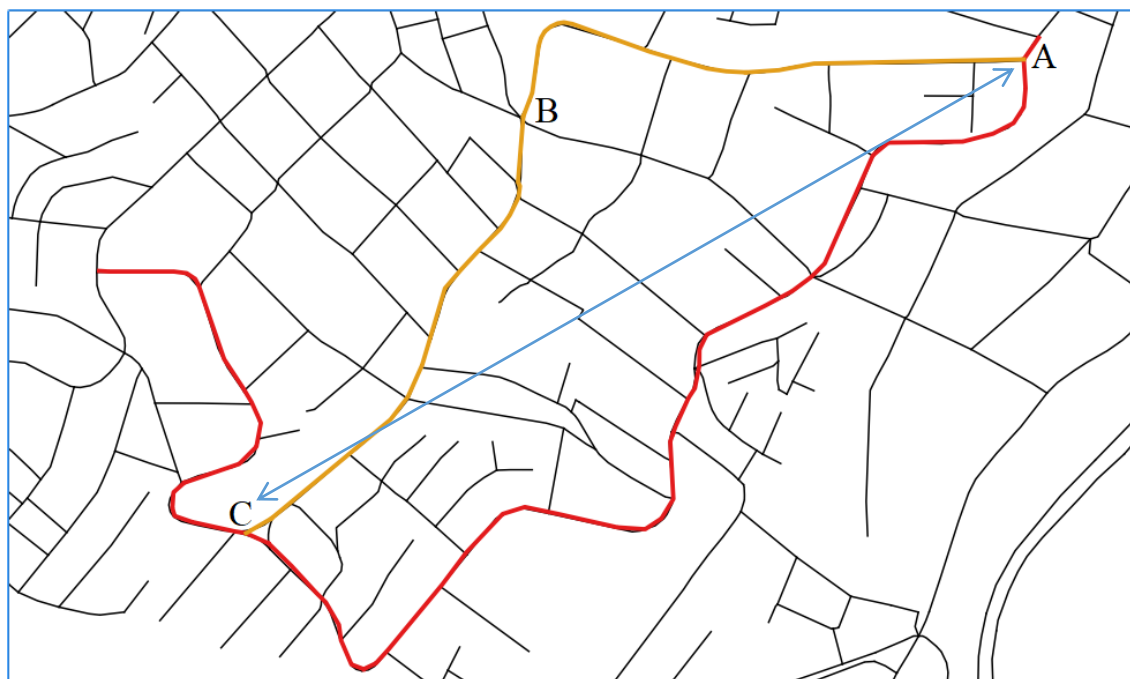


Figura 13. Ejemplo de error atípicamente grande en el Método de Unión por Nombres.

Los errores generados de esta manera imposibilitan la utilización de este método para la inferencia de escalas. Por esta razón el Método de Vecino más Cercano es introducido, con el único requisito de especificar un radio máximo de búsqueda. Recordando que la legislación establece un error máximo de 0.30 milímetros por el módulo de escala, para un 90% muestral (Instituto Geográfico Militar, 2016), es plausible calcular este percentil de los errores hallados en el primer método. De acuerdo a la *Tabla 7* el percentil P_{90} es comparativamente pequeño en todas las zonas, mientras que en el P_{95} la influencia de los valores atípicos ya es considerable. El $P_{92.5}$ fue escogido por esta razón; además, este valor no subestima el porcentaje declarado en la legislación.

No obstante, es crucial entender que escoger el radio de búsqueda de esta manera controla completamente la magnitud de los errores calculados y, en consecuencia, las escalas obtenidas; este detalle es la principal debilidad del segundo método. Además, ninguna consideración es hecha respecto al 10% muestral restante cuyo error, según la misma legislación, no debe superar los 0.40 milímetros por el módulo de escala. Finalmente, el 92% reportado -errores en el primer método que coinciden con el segundo- parece ser consecuencia del $P_{92.5}$ escogido; la manera de confirmar esto es comparar este porcentaje al utilizar otros radios de búsqueda.

Con respecto al estadístico Getis-Ord, un aspecto que se ha pasado por alto es el hecho de que los valores Z serán normales solamente si la variable X es normal (Ord & Getis, 1995). Los diagramas de caja ya presentados (*Figura 8*), o los de densidad a continuación (*Figura 14*), dan indicios de que el error posicional es una variable no normal con marcado sesgo derecho. Con todo, pruebas de Shapiro-Wilk fueron aplicadas a esta variable, obteniendo indicios altamente significativos de no normalidad en todas las Administraciones Zonales. El mismo resultado se obtiene aplicando la prueba a los valores Z , cuya distribución también se presenta en la *Figura 8*.

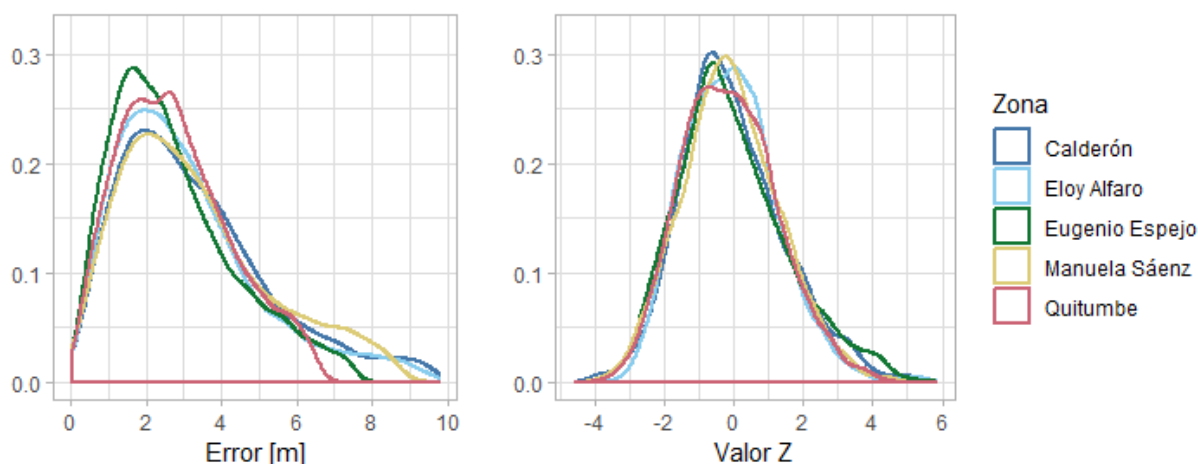


Figura 14. Gráficos de densidad de errores calculados con el Método de Vecino más Cercano (izquierda) y de valores normalizados del estadístico G^*i (derecha).

Las implicaciones de utilizar una variable no normal son mínimas, excepto al identificar los puntos calientes y fríos, pues los valores críticos provienen de una distribución normal. Parecería que el error posicional tuviese una distribución de Poisson, comportamiento típico de una variable que proviene de un conteo. En este sentido, es necesario estudiar la distribución del error -y sus implicaciones en el estadístico de Getis-Ord- con mayor profundidad. Además, como lo sugieren los propios Getis & Ord (1992), un estudio completo de la autocorrelación espacial involucra otros estadísticos -como la tradicional I de Moran- que presenten una visión diferente de los datos.

Con respecto a las escalas inferidas, las correcciones introducidas con el estadístico Getis-Ord, así como la homogeneización de los módulos a múltiplos de 10 000 (donde los errores son múltiplos de 3 metros), aseguran que cada polígono no está relacionado con los adyacentes. Sin embargo, en todas las Administraciones Zonales son evidentes dos inconvenientes: primero, la granularidad es alta, es decir que la mayoría de polígonos son demasiado pequeños para ser útiles; y segundo, los polígonos sin escala asignada se encuentran interfiriendo, como agujeros en medio de los datos. Debido a estos inconvenientes, por lo pronto, la vialidad OSM se puede utilizar de manera segura considerando la escala más pequeña hallada en cada zona. Así, se demostró que las

Administraciones Zonales Calderón y Eloy Alfaro son aprovechables en proyectos hasta la escala 1 : 40 000; mientras que Eugenio Espejo, Manuela Sáenz y Quitumbe, hasta 1 : 30 000.

A pesar de ello, obsérvese que, en Calderón y Eloy Alfaro, los polígonos escala 1 : 40 000 no alcanzan el 5% del total y que, en todas las Administraciones Zonales, los de escala 1 : 30 000 no alcanzan el 10%. Con el propósito de aprovechar las zonas de mayor calidad se deben investigar estrategias para corregir la información perjudicial y rellenar la faltante. Interesantemente, se observó que los puntos calientes casi siempre aparecen en los bordes de las zonas o de la vialidad (Figura 15), indicando que cierto fenómeno ignorado ocurre en esos lugares y debería ser investigado en futuros trabajos. Por supuesto, cualquier estrategia que se desee implementar debe recordar que estos datos no son estáticos, debido a la naturaleza de la VGI.

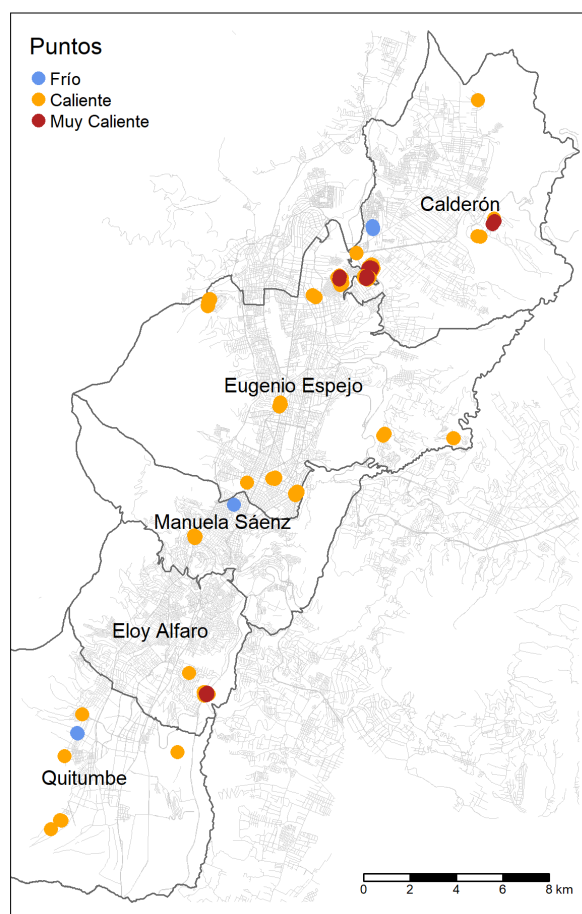


Figura 15. Ubicación de los Puntos Calientes y Fríos.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

En el presente proyecto se demostró la utilidad de un ambiente de programación estadística, el lenguaje R, para obtener información respecto de un conjunto valioso de VGI, como lo es OpenStreetMap, el cual no se ha estudiado antes en el país. Se utilizó de manera intensiva los paquetes *tidyverse*, *sf* y *spdep* para desarrollar cuatro scripts que permitiesen resolver el problema progresivamente, involucrando al usuario en el análisis de los resultados.

Se demostró que la gran cantidad de información ausente tanto en la cartografía oficial como en la libre, limita el uso de un procedimiento sencillo tal como un *join* de tablas. De manera particular, errores perjudicialmente grandes aparecen en esta clase de procedimiento; esto ya fue demostrado en un artículo científico preliminar (Castro, Tierra, & Luna, 2019), como parte de las metas trazadas. Por otra parte un método más complicado, tal como buscar el vecino más cercano, produce mejores resultados pero introduce una camisa de fuerza a la magnitud de los errores hallados, al momento de escoger un radio máximo de búsqueda.

Mediante el análisis de estadísticos locales se aprendió que existe una estructura de autocorrelación espacial en varios puntos, particularmente en los límites de las zonas estudiadas. Se determinó que 150 metros es la distancia de vecindad adecuada. La autocorrelación determina frecuentemente la aparición de vecindades con agrupación de errores grandes, y rara vez de errores pequeños. Dado que se está estudiando un aspecto de calidad, los errores posicionales pequeños no

representan una preocupación; más bien, la atención debe dirigirse a aquellas vecindades con gran error, pues allí son necesarias correcciones en los datos libres, con el propósito de permitir su uso en mayor cantidad de aplicaciones.

La escala de la cartografía a la cual pertenecería cada intersección vial con error fue inferida, en el sentido de que los procedimientos cartográficos tradicionales no consideran a la escala como un resultado, sino como un estándar sobre el cual se generará la cartografía. De esta manera, se podría decir que esta metodología es ingeniería inversa, pues va de los datos finales a los metadatos. Tales métodos son típicos en el estado del arte VGI pues, debido a la manera como es generada, esta información carece de elementos cruciales y requiere metodologías alternativas. No obstante, otros elementos que sí están presentes no se observaron en este estudio; por ejemplo, las etiquetas *source* y *timestamp* permitirían ampliar la evaluación para que también considere los aspectos de calidad temático y temporal, respectivamente, y sus relaciones con el posicional.

Finalmente, las escalas inferidas presentan altas granularidad y cantidad de espacios vacíos, lo cual dificulta la asignación de una utilidad a toda la cobertura vial OpenStreetMap de una zona. Con todo -y observando la legislación- esta cobertura se puede usar en cartografía hasta 1 : 30 000 en Eugenio Espejo, Manuela Sáenz, Quitumbe, y hasta 1 : 40 000 en Calderón y Eloy Alfaro. Si bien estas escalas restringen su aprovechamiento en ciertas aplicaciones -como el catastro urbano- la cobertura OpenStreetMap puede utilizarse con seguridad en muchas otras aplicaciones a nivel zonal o distrital, como planificación territorial y uso del suelo, por citar ejemplos. No obstante, es extremadamente importante recordar que estos resultados no son permanentes, puesto que los datos son susceptibles a cambios en cualquier momento. En consecuencia, se debería estudiar cómo implementar esta metodología en una plataforma que permita su reiterada evaluación.

5.2. RECOMENDACIONES

- Trabajar en la implementación, a través de alguna plataforma digital, de estos resultados; de la misma manera, en la implementación del proceso, de manera que se pueda repetir la evaluación en el futuro y en otras zonas.
- Considerar otros métodos, diferentes a la unión por nombres y el vecino más cercano, para relacionar las intersecciones equivalentes en dos conjuntos de datos, buscando maximizar el número de relaciones halladas.
- Estudiar las implicaciones de utilizar una variable no normal en el estadístico Getis-Ord; particularmente, estudiar cuál es el tratamiento que debe recibir una variable que sigue una distribución de Poisson.
- Investigar con mayor profundidad qué fenómeno explica la aparición de puntos calientes fundamentalmente en los límites de las zonas estudiadas y de la vialidad.
- Analizar las características de los polígonos de Voronoi trazados para cada escala, con el propósito de diseñar estrategias para reducir la granularidad y rellenar los espacios vacíos.

BIBLIOGRAFÍA

- Ariza, F. (2013). *Fundamentos de Evaluación de la Calidad de la Información Geográfica*. Jaén: Universidad de Jaén.
- Bakillah, M., Liang, S., Mobasher, A., Jokar, J., & Zipf, A. (2014). Fine-resolution population mapping using OpenStreetMap points-of-interest. *International Journal of Geographical Information Science*, 28(9), 1940-1963. doi:10.1080/13658816.2014.909045
- Bravo, J. (s.f.). *Diagrama de Voronoi*. (U. de Extremadura, Ed.) Recuperado el 16 de abril de 2019, de Departamento de Matemáticas: <http://matematicas.unex.es/~trinidad/mui/voronoi.pdf>
- Castro, R., & Luna, M. (2018). Exploratory Data Analysis of Volunteered Geographic Information: case study in the Manuela Sáenz administrative zone in Quito. *Latin American Journal of Computing*, 5(2), 21-26.
- Castro, R., Tierra, A., & Luna, M. (2019). Assessing the Horizontal Positional Accuracy in OpenStreetMap: A Big Data Approach. En Á. Rocha, H. Adeli, L. Reis, & S. Costanzo (Ed.), *New Knowledge in Information Systems and Technologies*. 931, págs. 513-523. Cham: Springer. doi:10.1007/978-3-030-16184-2_49
- Cherniavskii, A. (8 de mayo de 2019). *Announcing Leaflet 1.5.1*. Obtenido de Leaflet: an open-source JavaScript library for mobile-friendly interactive maps: <https://leafletjs.com/2019/05/08/leaflet-1.5.1.html>
- Coleman, D., Georgiadou, Y., & Labonte, J. (2009). Volunteered Geographic Information: The Nature and Motivation of Producers. *International Journal of Spatial Data Infrastructures Research*, 4(1), 332-358. doi:10.2902/1725-0463.2009.04.art16
- CRAN. (2019). *The Comprehensive R Archive Network*. Viena: R Foundation for Statistical

Computing. Obtenido de <https://cran.r-project.org/>

Freitag, U. (1962). Der Kartenmaßstab - Betrachtungen über den Maßstabsbegriff in der Kartographie. *Kartographische Nachrichten*, 5(12), 134-146.

Getis, A., & Ord, J. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24(3), 189-206.

Goodchild, M. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(1), 211-221. doi:10.1007/s10708-007-9111-y

Goodchild, M. (2013). The quality of big (geo)data. *Dialogues in Human Geography*, 3(3), 280-284. doi:10.1177/2043820613513392

Hashemi, P., & Abbaspour, R. (2015). Assessment of Logical Consistency in OpenStreetMap Based on the Spatial Similarity Concept. En J. Jokar, A. Zipf, P. Mooney, & M. Helbich (Edits.), *OpenStreetMap in GIScience* (Vol. 75, págs. 19-36). Basilea: Springer. doi:10.1007/978-3-319-14280-7_2

Helbich, M., Amelunxen, C., Neis, P., & Zipf, A. (2012). Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata. En T. Jekel, A. Car, J. Strobl, & G. Griesebner (Ed.), *GI_Forum 2012: Geovizualisation, Society and Learning* (págs. 24-33). Berlín: Herbert Wichmann Verlag.

IMPAQTO. (12 de mayo de 2016). *Mapping Ecuador consolida la logística en zonas afectadas por el terremoto*. Obtenido de IMPAQTO: https://www.impaqto.net/mapping_ecuador

Instituto Geográfico Militar. (2016). *Especificaciones Técnicas para la Producción de Cartografía Escala 1:5000*. Instituto Geográfico Militar, Quito. Obtenido de www.geoportaligm.gob.ec/portal/?wpfb_dl=851

Instituto Geográfico Militar. (2017). *Protocolo de Fiscalización para Proyectos de Generación de*

Cartografía Base con Fines Catastrales Escala 1:1000. Instituto Geográfico Militar, Quito.

Obtenido de www.geoportaligm.gob.ec/portal/

Jokar, J., & Vaz, E. (2015). An assessment of a collaborative mapping approach for exploring land use patterns for several European metropolises. *International Journal of Applied Earth Observation and Geoinformation*, 35(1), 329-337. doi:10.1016/j.jag.2014.09.009

Jokar, J., Zipf, A., Mooney, P., & Helbich, M. (2015). An Introduction to OpenStreetMap in Geographic Information Science: Experiences, Research, and Applications. En J. Jokar, A. Zipf, P. Mooney, & M. Helbich (Edits.), *OpenStreetMap in GIScience* (Vol. 75, págs. 1-15). Basilea: Springer. doi:10.1007/978-3-319-14280-7_1

Laney, D. (6 de febrero de 2001). *3D Data Management: Controlling Data volume, Velocity, and Variety*. (META Group Inc., Ed.) Obtenido de Application Delivery Strategies: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

Municipio Metropolitano de Quito. (s.f.). *Sistema de Información Metropolitano*. Recuperado el 7 de enero de 2019, de Gobierno Abierto: gobiernoabierto.quito.gob.ec/?page_id=1105

Neis, P. (7 de marzo de 2019). *Statistics of the free wiki world map*. Obtenido de OSMstats: <http://osmstats.neis-one.org>

Neis, P., & Zipf, A. (2012). Analyzing the Contributor Activity of a Volunteered Geographic Information Project - The Case of OpenStreetMap. *International Journal of Geo-Information*, 1(1), 146-165. doi:10.3390/ijgi1020146

Noskov, A. (2018). Computer Vision Approaches for Big Geo-Spatial Data: Quality Assessment of Raster Tiled Web Maps for Smart City Solutions. En T. Bandrova, & K. M (Ed.), *Proceedings, 7th International Conference on Cartography and GIS*, (págs. 296-305).

Sozopol. doi:10.5281/zenodo.1346671

OpenStreetMap Wiki. (3 de marzo de 2019). *List of OSM-based services*. Obtenido de

OpenStreetMap Wiki: wiki.openstreetmap.org/wiki/List_of_OSM-based_services

OpenStreetMap Wiki. (3 de febrero de 2019). *Overpass API*. Obtenido de OpenStreetMap Wiki:

wiki.openstreetmap.org/wiki/Overpass_API

OpenStreetMap Wiki. (7 de marzo de 2019). *They are using OpenStreetMap*. Obtenido de

OpenStreetMap Wiki: wiki.openstreetmap.org/wiki/They_are_using_OpenStreetMap

Ord, J., & Getis, A. (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an

Application. *Geographical Analysis*, 27(4), 286-306.

R Core Team. (2019). *R: A language and environment for statistical computing*. Viena: R

Foundation for Statistical Computing. Obtenido de <https://www.r-project.org/>

RStudio Team. (2016). *RStudio: Integrated Development for R*. Boston: RStudio, Inc. Obtenido de

<http://www.rstudio.com/>

Sax-Barnett, M. (2014). An Introduction to OpenStreetMap. *State of the Map US*. Washington DC.

Obtenido de pdxmele.com/FOSS4G_OSM_workshop.pdf

Sui, D. (2008). The Wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and

the future of GIS. *Computers, Environment and Urban Systems*, 32(1), 1-5.

doi:10.1016/j.compenvurbsys.2007.12.001

Touya, G., & Reimer, A. (2015). Inferring the Scale of OpenStreetMap Features. En J. Jokar, A.

Zipf, P. Mooney, & M. Helbich (Edits.), *OpenStreetMap in GIScience* (Vol. 75, págs. 81-

99). Basilea: Springer. doi:10.1007/978-3-319-14280-7_5

Turner, A. (2006). *Introduction to Neogeography*. Sebastopol, California: O'Reilly Media.