



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA



Evaluación de la Exactitud Posicional Horizontal e Inferencia de la Escala sobre Datos Libres OSM en Administraciones Zonales del DMQ

Autor: Roger Rafael Castro Zambrano

Tutor: Alfonso Tierra, PhD.

Junio de 2019

Antecedentes: OpenStreetMap

- Un “esfuerzo internacional para crear una fuente de datos libre, disponibles en un mapa, a través del esfuerzo voluntario” (Goodchild, 2007)
- Más de 300 aplicaciones dependen completa o parcialmente de estos datos (List of OSM-based services, 2019)
- leaflet.js, wikipedia.org, weather.com, flickr.com, geocaching.com, drivebestway.com, Snapchat, Uber, Strava, TripAdvisor...



Antecedentes: OpenStreetMap

Potencialidades:

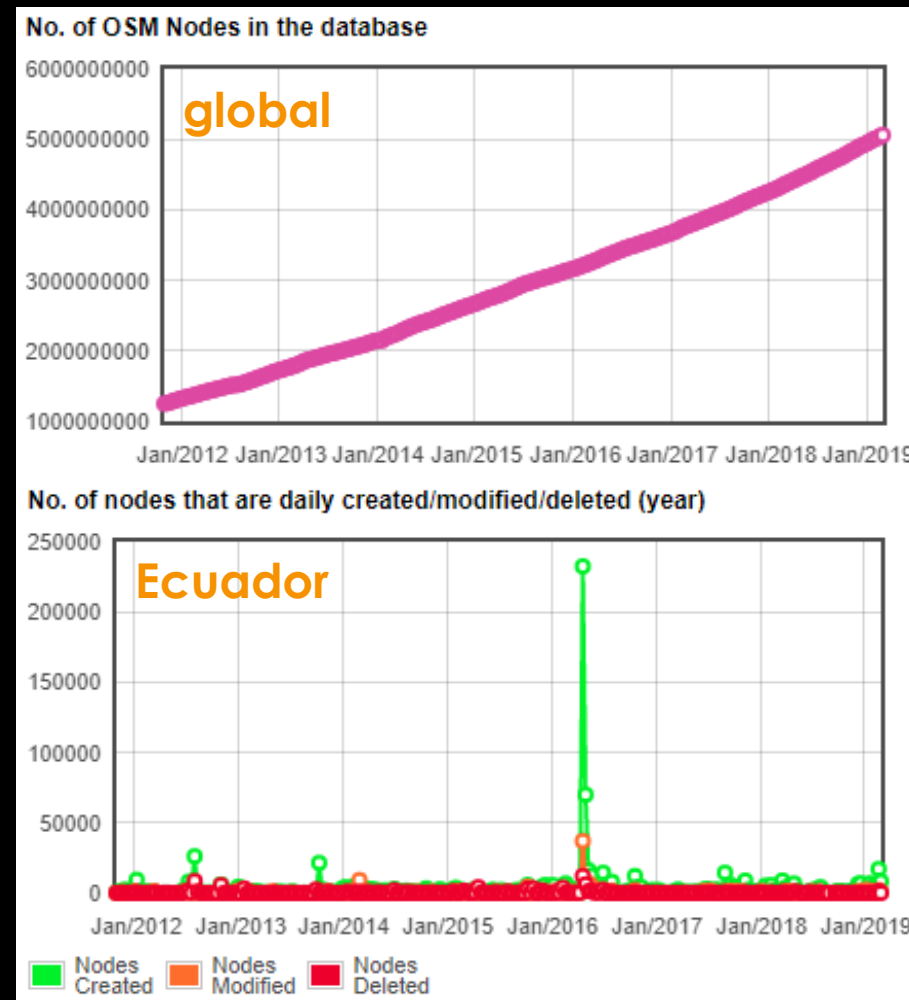
- sensores humanos
- poblaciones participativas
- infraestructuras de datos espaciales (IDEs)
- gestión de emergencias

Debilidades:

- carencia de una reputación
- desconocimiento del motivo del usuario
- inexperiencia del usuario
- división digital
- terrorismo digital

Planteamiento del Problema

- Las características de OSM varían considerablemente a través del tiempo y del espacio
- Las contribuciones en algunos países son emergentes
- Aún existen grandes vacíos (Jokar, Zipf, Mooney, & Helbich, 2015)



Planteamiento del Problema

- Investigaciones realizadas sobre OSM son eurocéntricas
- De todas maneras ya se ha mapeado en Latinoamérica
- Ecuador es un caso especial, muchas contribuciones corresponden a un evento de emergencia (sismo del 16 de abril de 2016)
- Mapping Ecuador tuvo lugar (IMPAQTO, 2016)
- Se desconoce la calidad de estos datos



Objetivos

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.

- Medir la exactitud posicional de vías urbanas mediante comparación de coordenadas horizontales (E, N) en intersecciones, respecto a las oficiales
- 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, con el propósito de respaldar el uso de estos datos

Zona de Estudio

Administraciones Zonales:

1. Calderón
2. Eloy Alfaro (excepto Lloa)
3. Eugenio Espejo
4. Manuela Sáenz
5. Quitumbe



Base Conceptual: Calidad de Datos

Exactitud Posicional Horizontal:

- ISO 19113
 - Componente Posicional
 - Exactitud
 - absoluta ←
 - relativa
 - de datos de rejilla
 - Componente Temático
 - Componente Temporal
 - Coherencia Lógica
 - Compleción

Escala:

- Estandarización de productos en instituciones oficiales
- Ambiente de uso
- Organización de la temática
- Calidad (metadato crucial)

Estado del Arte

Acercamiento intrínseco:

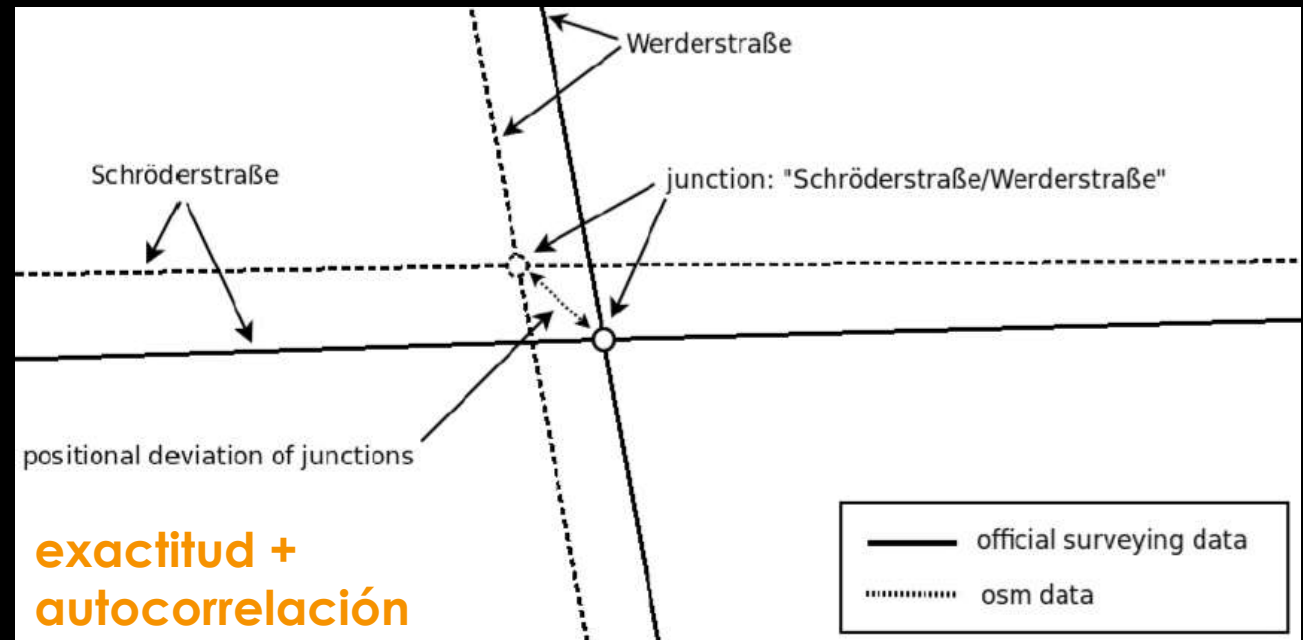
- Touya & Reimer (2015)



- Noskov (2018)
- Jokar & Vaz (2015)
- ...

Acercamiento extrínseco:

- Helbich, Amelunxen, Neis, & Zipf (2012)



- Neis, Zielstra & Zipf (2012)
- ...

Base Conceptual: Estadístico Getis-Ord

Pertenece a una familia de estadísticos dedicados a medir el grado de autocorrelación espacial. A diferencia de otros que se aplican globalmente, como la I de Moran, la G de Getis & Ord (1992) se aplica localmente, localizando puntos con alta correlación (clustering).

- Estadístico:
$$G_i(d) = \frac{\sum_j^n w_{ij}(d)x_j}{\sum_j^n x_j}; j \neq i$$
 - Esperanza:
$$E[G_i(d)] = \frac{W_i}{n-1}$$
 - Varianza:
$$V[G_i(d)] = \frac{W_i(n-1-W_i)}{(n-1)^2(n-2)} \frac{Y_{i2}}{Y_{i1}^2}$$
- $$W_i = \sum_j^n w_{ij}(d)$$
- $$Y_{i1} = \frac{\sum_j^n x_j}{n-1}$$
- $$Y_{i2} = \frac{\sum_j^n x_j^2}{n-1} - Y_{i1}^2$$

Base Conceptual: Estadístico Getis-Ord

Getis & Ord (1992) también definen:

- G_i^* es una versión de G_i en la cual se permite $j = i$
- G y G^* son las versiones globales de G_i y G_i^* respectivamente

Ord & Getis (1995) redefinen el estadístico, de manera que en cada punto se calcula el valor Z_i (valor normalizado) del estadístico original:

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

Los valores Z_i se comparan con un valor crítico, positivo y negativo, que se calcula considerando una corrección de Bonferroni.

Base Legal

Norma Internacional:



- ISO 19113
- Metodologías de Control Posicional por Puntos (Ariza, 2013)
 - NMAS
 - EMAS
 - NSSDA

Norma Nacional:

- Especificaciones Técnicas para la Producción de Cartografía Escala 1:5000 (Instituto Geográfico Militar, 2016)
 - Exactitud Posicional Horizontal no debe exceder al módulo de escala multiplicado por
 - 0.3 mm (90% muestra aleatoria)
 - 0.4 mm (10% restante)

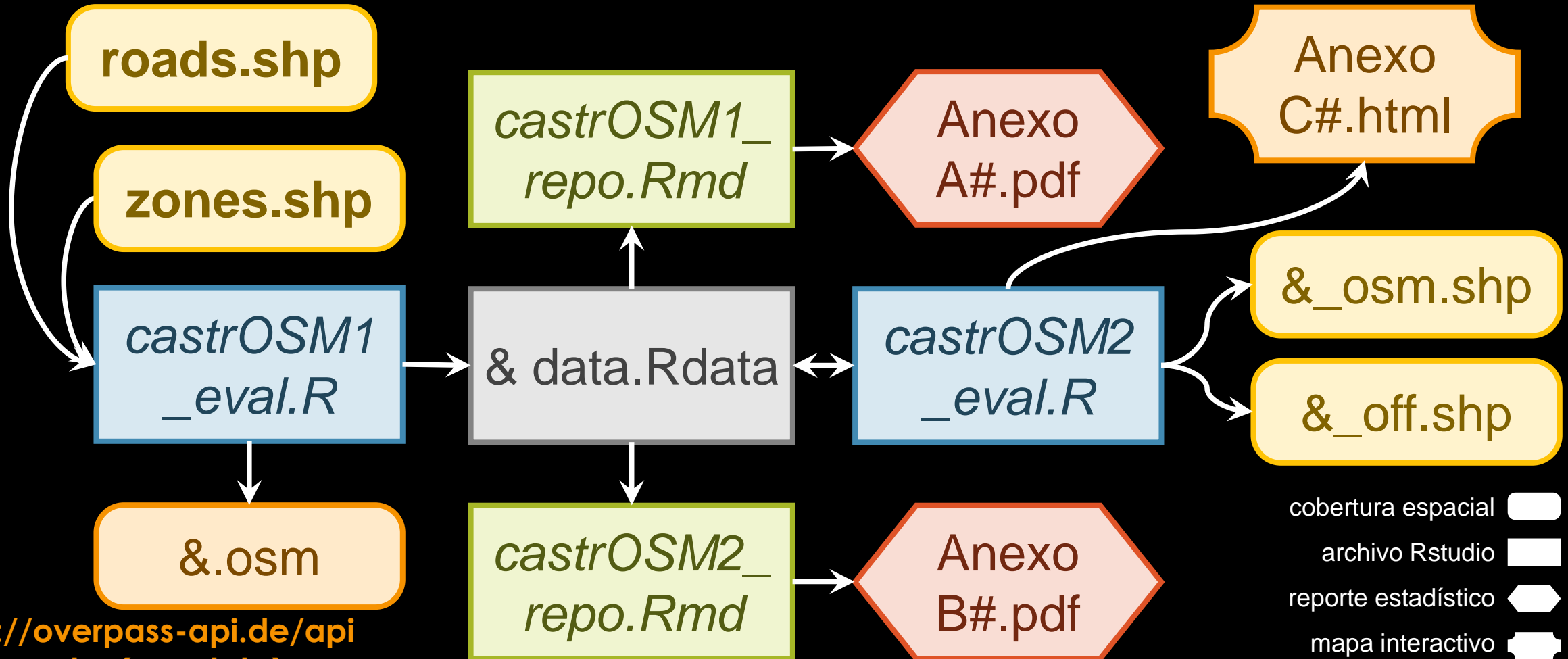
Metodología: Software Utilizado



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

Metodología: Diagrama de Flujo

http://gobiernoabierto.quito.gob.ec/?page_id=1122 (Municipio DMQ)



<http://overpass-api.de/api/interpreter> (osmdata)

- cobertura espacial
- archivo Rstudio
- reporte estadístico
- mapa interactivo

Metodología: Relacionar Intersecciones

Primer Método:
Unión por Nombres

```
# joining by names
inner_join(osmNam, shpNam) %>%
  mutate(Error = sqrt((osmE-shpE)^2
                      + (osmN-shpN)^2)) %>%
  group_by(osmE, osmN) %>%
  summarise_at(Error, min) %>%
  ungroup() %>%
  select(-osmE, -osmN) %>%
  arrange(desc(Error))
```

Segundo Método:
Vecino más Cercano

```
findDist = function(x, y, d = 10){
  z = as.data.frame(shpInt) %>%
    filter(shpE < x+d, shpE > x-d,
           shpN < y+d, shpN > y-d) %>%
    mutate(dist = sqrt((shpE-x)^2
                      + (shpN-y)^2)) %>%
    filter(dist <= d)
  if(nrow(z) == 0) return(NA)
  else return(min(z$dist))
} # finds nearest point returns distance
```

Metodología: Inferir la Escala

- Generar los polígonos de Voronoi
- Calcular el módulo de escala: $M = 1000 e / mm$
- Calcular G y G^* variando la distancia de vecindad

¿Cuál es la distancia adecuada?

- Calcular el módulo homogeneizado: $M = K \text{ ceiling}[1000 e / (K mm)]$
- Calcular G^* variando la distancia de vecindad y el factor de homogeneización

¿Cuál es el factor adecuado?

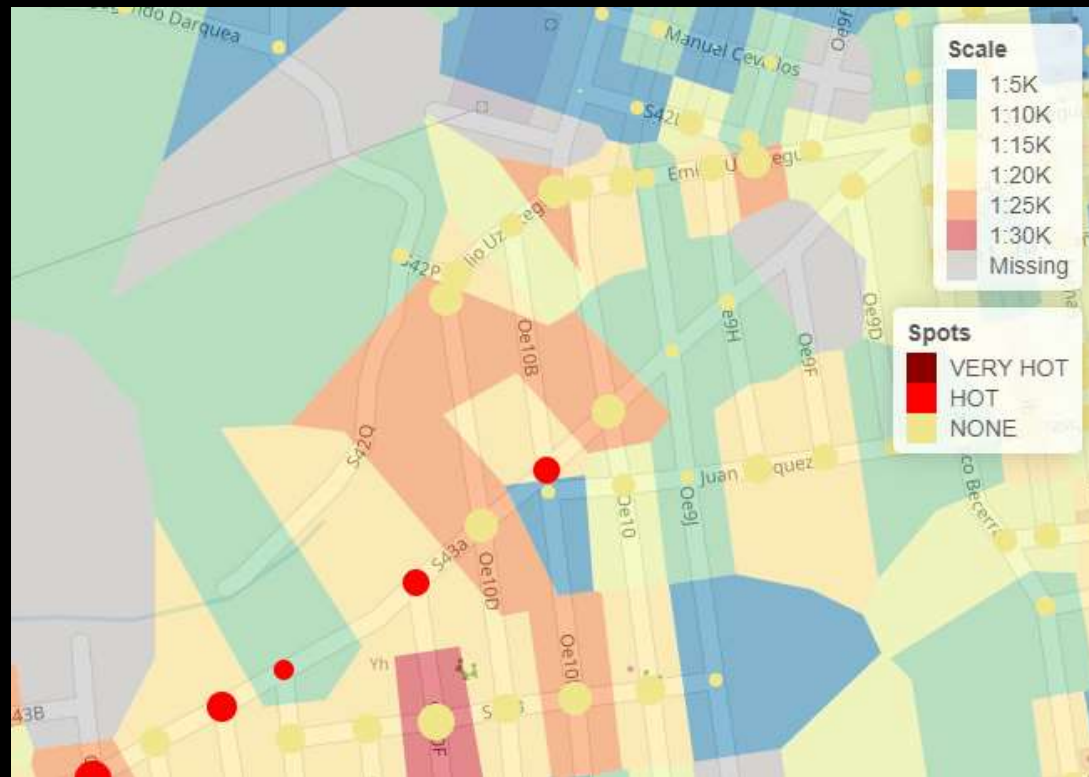
Metodología: Inferir la Escala

e [m]	1 : M	1 : M ₁	1 : M ₂	1 : M ₅	1 : M ₁₀
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

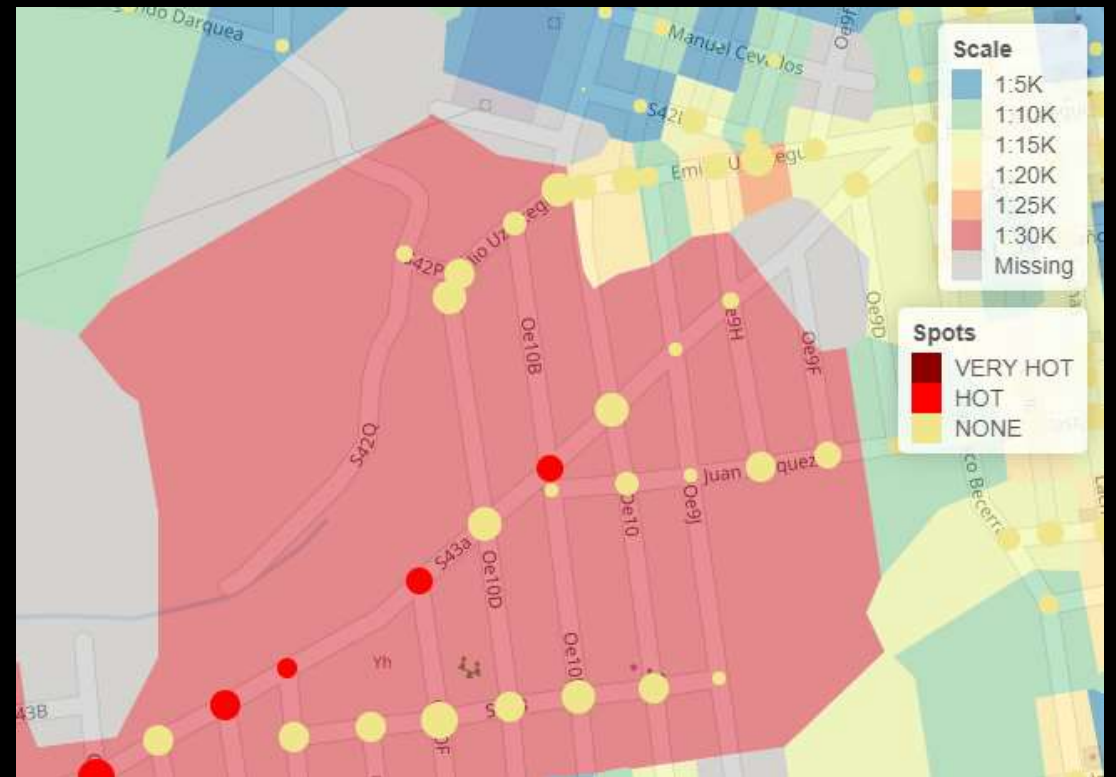
Escoger la distancia y el factor

- Calcular G^*_i con la distancia y el factor escogidos
- Comparar Z_i para identificar puntos calientes y fríos
 - muy caliente si $Z_i \geq Z_{0.999}$
 - caliente si $Z_i \geq Z_{0.900}$
 - frío si $Z_i \leq -Z_{0.900}$
 - muy frío si $Z_i \leq -Z_{0.999}$

Metodología: Tratar Puntos Calientes



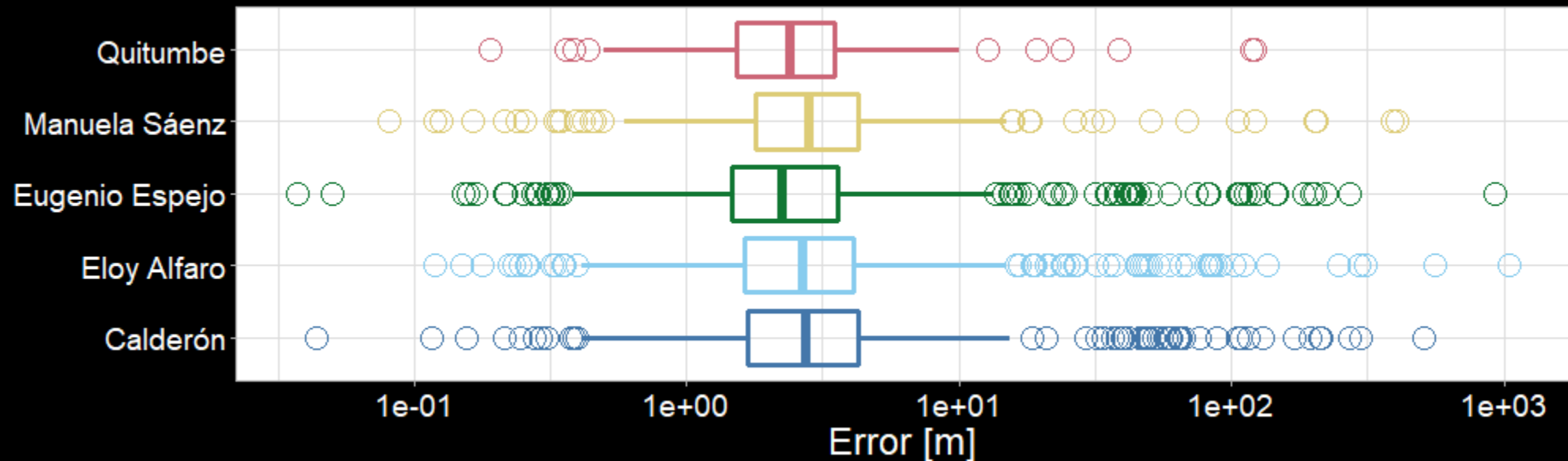
Antes del tratamiento



Después del tratamiento

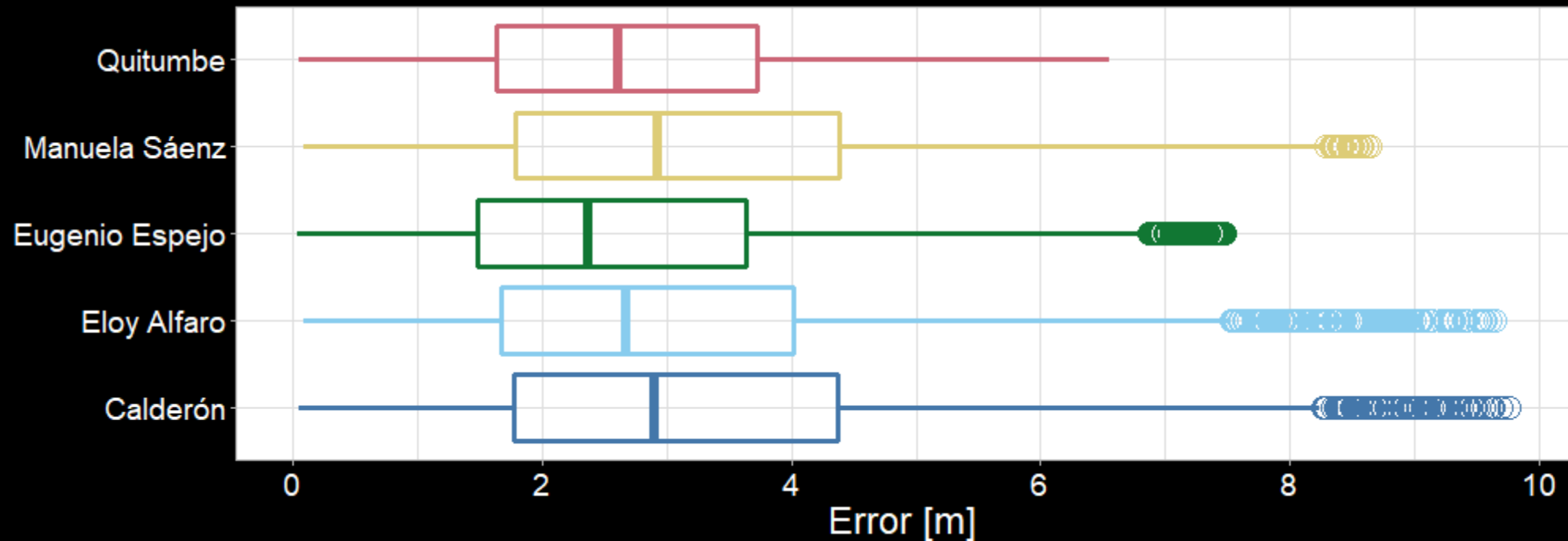
Resultados: Unión por Nombres

Variable	Calderón	Eloy Alfaro	Eugenio Espejo	Manuela Sáenz	Quitumbe
Intersecciones OSM	4710	5813	10268	3955	7774
Intersecciones OSM con error	711	877	1207	567	169
Porcentaje con error	15.10 %	15.09 %	11.75 %	14.34 %	2.17 %
Vías con nombre oficiales	72.37 %	64.84 %	82.21 %	70.68 %	25.51 %
Vías con nombre OSM	62.18 %	71.25 %	69.20 %	69.64 %	72.91 %

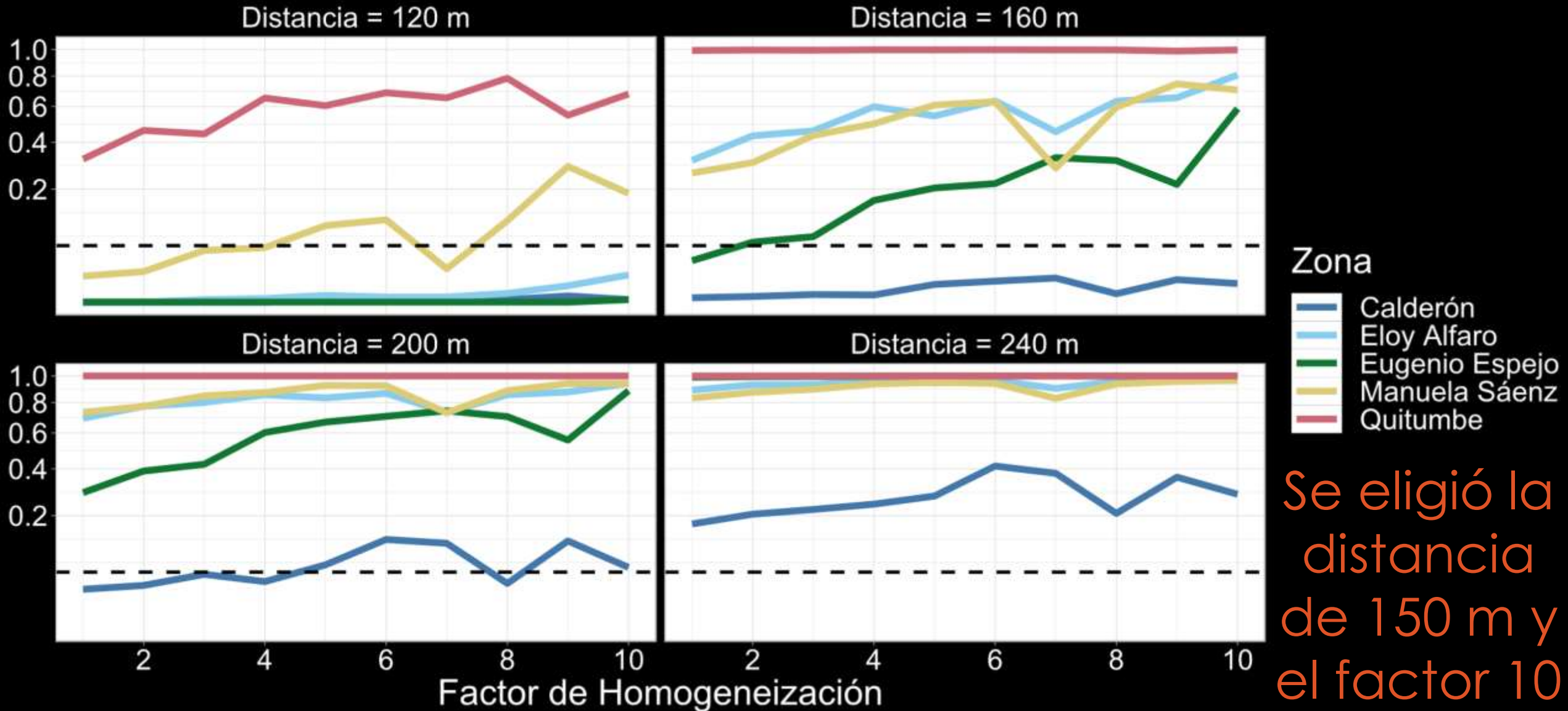


Resultados: Vecino más Cercano

Variable	Calderón	Eloy Alfaro	Eugenio Espejo	Manuela Sáenz	Quitumbe
Intersecciones OSM	4710	5813	10268	3955	7774
Intersecciones OSM con error	3635	4754	6260	2576	5925
Porcentaje con error	77.18 %	81.78 %	60.97 %	65.13 %	76.22 %
Porcentaje UN coincide VC	91.98 %	92.13 %	92.13 %	92.24 %	91.72 %



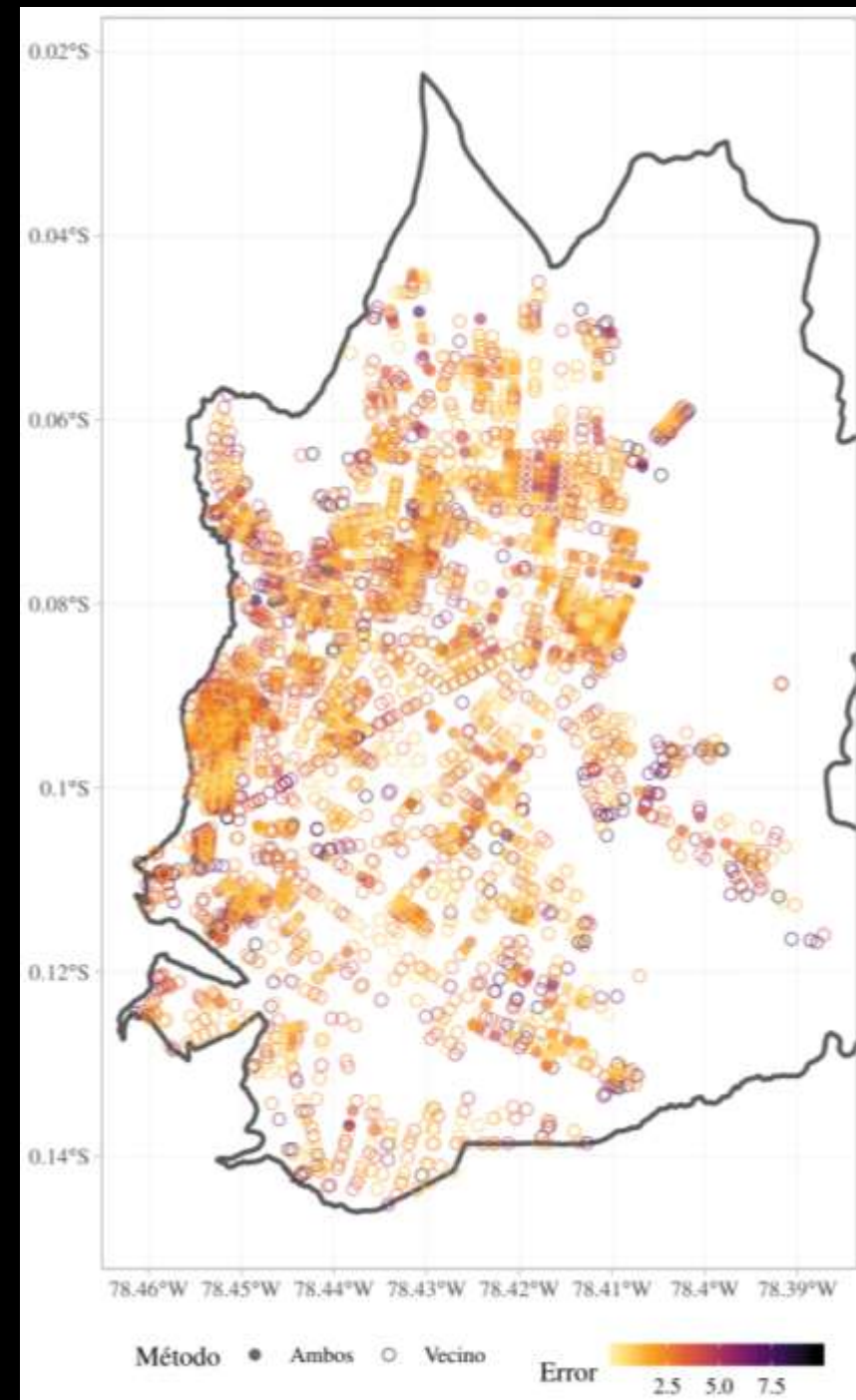
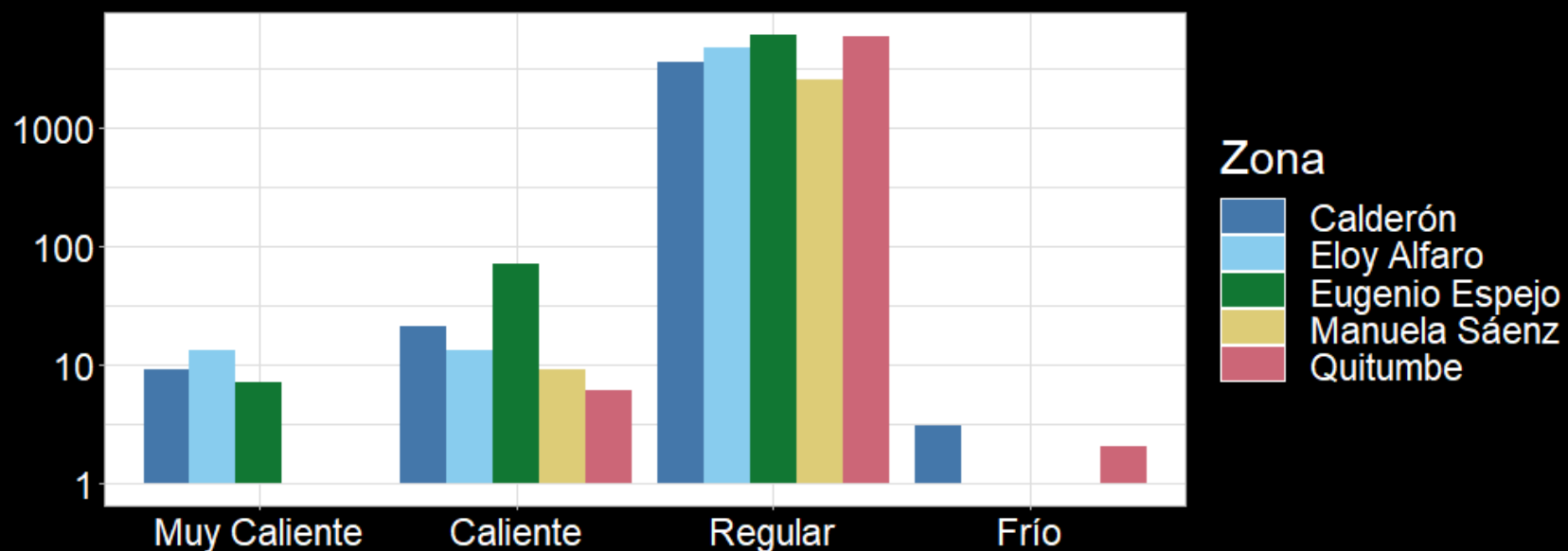
Resultados: Estadístico Global



Se eligió la distancia de 150 m y el factor 10

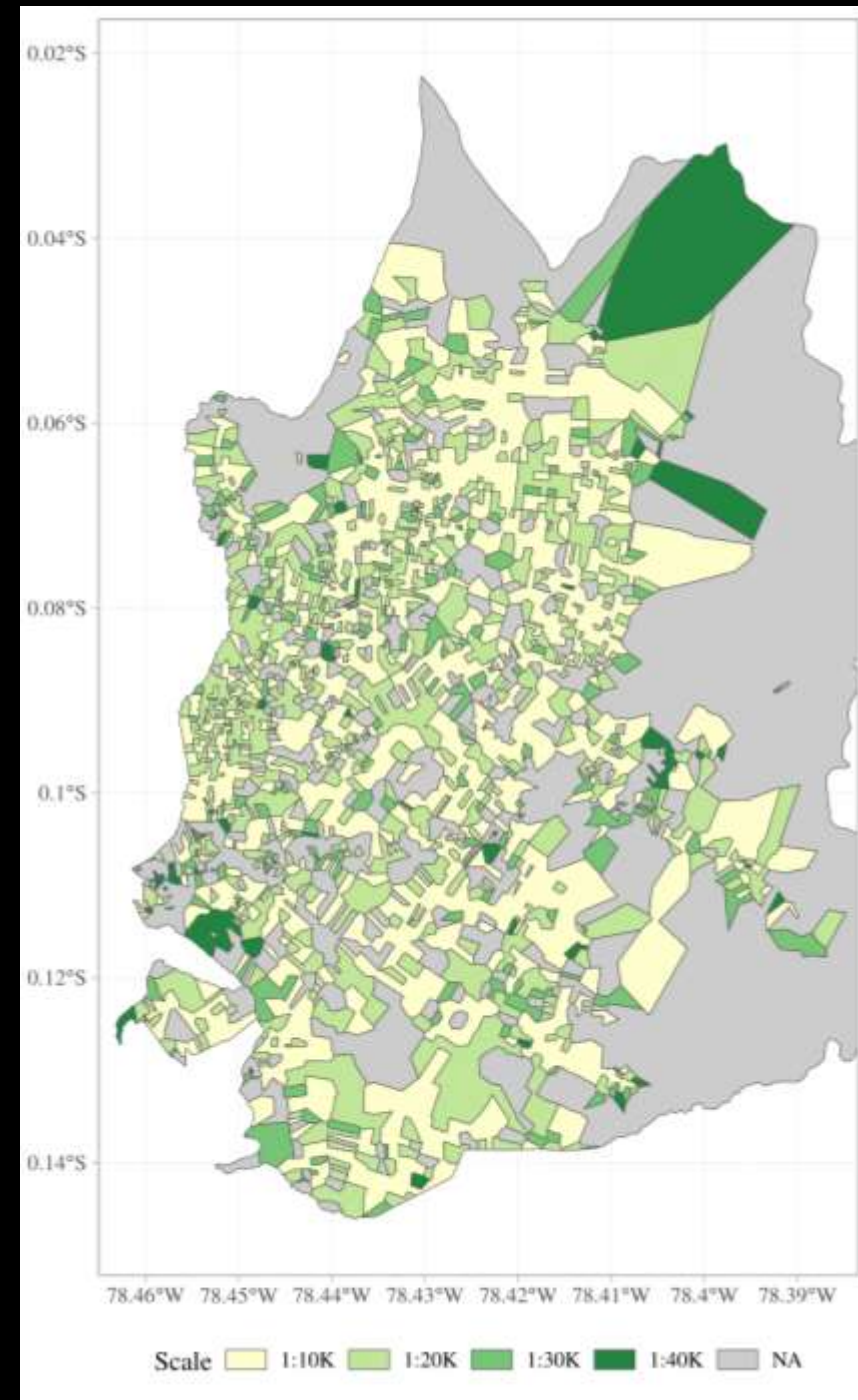
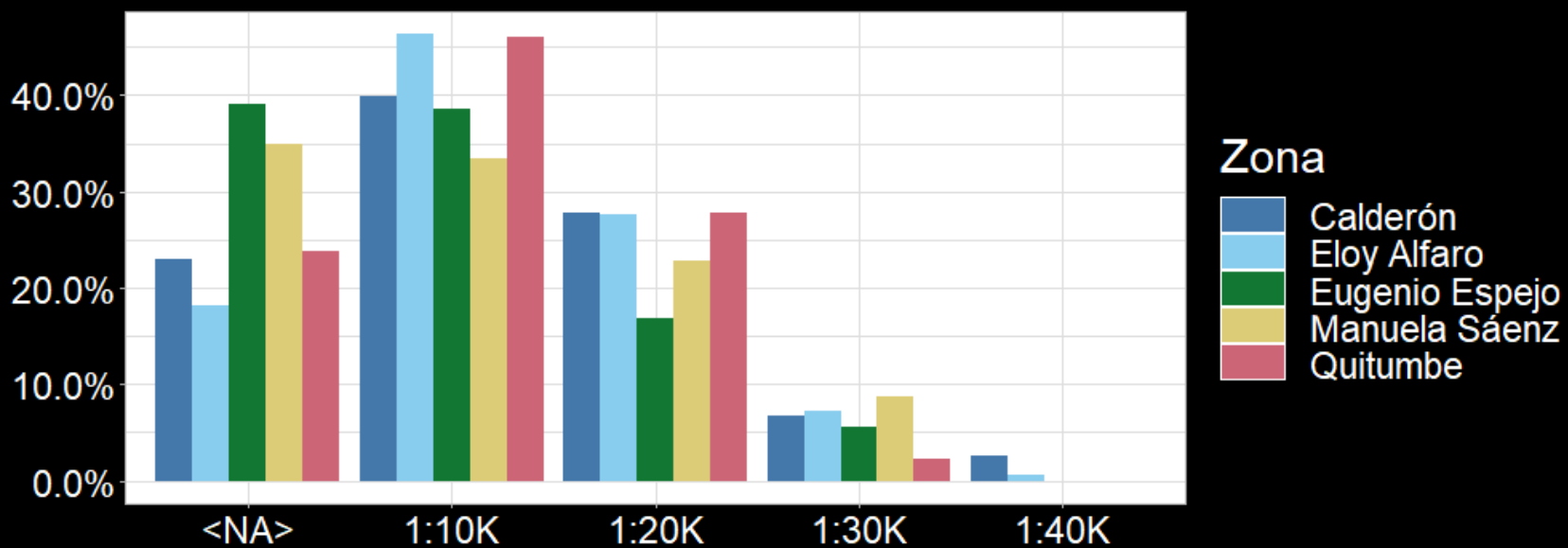
Resultados: Error Posicional

Variable	Calderón	Eloy Alfaro	Eugenio Espejo	Manuela Sáenz	Quitumbe
n	3635	4754	6260	2576	5925
$Z_{0.900}$	4.021	4.084	4.147	3.939	4.135
$Z_{0.999}$	5.008	5.059	5.111	4.941	5.101
p value	0.059	0.627	0.262	0.581	0.990



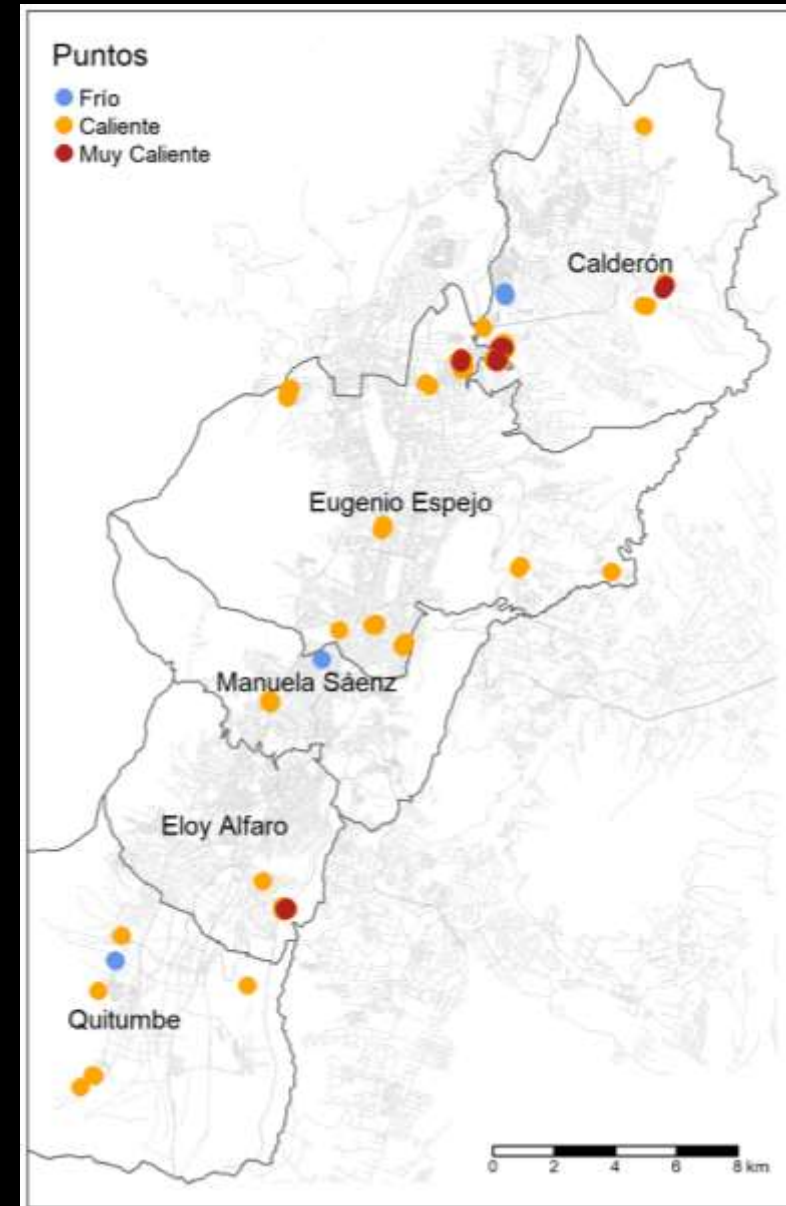
Resultados: Escala Inferida

Todas las zonas presentan dos inconvenientes: la granularidad (polígonos muy pequeños) y los polígonos sin escala. La manera más segura de utilizar esta cobertura es con la escala más pequeña en cada zona.



Conclusiones

- Se demostró la utilidad de un ambiente de programación para evaluar la VGI. El análisis se ejecuta progresivamente, involucrando al usuario.
- La gran cantidad de información ausente limita al método sencillo (join), que además genera errores atípicamente grandes. El método más complicado (vecino más cercano) arroja mejores resultados, pero limita la magnitud de los errores hallados.
- El cálculo de estadísticos locales permitió hallar autocorrelación espacial en varios puntos, de manera particular en los límites de las zonas o de la vialidad. La atención debe dirigirse a vecindades con gran error, donde son necesarias correcciones.



Conclusiones

- La escala de la cartografía a la cual pertenece cada intersección fue inferida, pues en los procedimientos convencionales la escala no es un resultado, sino un estándar.
- Otros datos disponibles no fueron estudiados. Las etiquetas *source* y *timestamp* permitirían, por ejemplo, ampliar la evaluación para que considere los aspectos temático y temporal, respectivamente.
- Las escalas inferidas presentan alta granularidad y vacíos. Con todo, la vialidad OSM se podría utilizar en cartografía hasta 1 : 40 000 en Calderón y Eloy Alfaro, y hasta 1 : 30 000 en las demás zonas.
- Estas escalas restringen ciertas aplicaciones -e.g. catastro urbano- pero permiten otras, a nivel zonal -planificación territorial, uso del suelo, etc. No obstante, es extremadamente importante recordar que estos resultados no son permanentes.

Recomendaciones

- Trabajar en la implementación de los resultados y de la metodología, para repetirla en el futuro y en otras zonas.
- Considerar otros métodos para relacionar intersecciones equivalentes en dos conjuntos de datos.
- Estudiar los efectos de la no normalidad en el estadístico Getis-Ord, particularmente para una distribución Poisson.
- Investigar la aparición de puntos calientes de manera especial en los límites de las zonas y de la vialidad.
- Diseñar estrategias para reducir la granularidad y rellenar los espacios vacíos.

Bibliografía citada

- Goodchild, M. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(1), 211-221.
- 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
- 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.
- 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
- 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.
- 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.
- Getis, A., & Ord, J. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24(3), 189-206.
- Ord, J., & Getis, A. (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*, 27(4), 286-306.
- Ariza, F. (2013). *Fundamentos de Evaluación de la Calidad de la Información Geográfica*. Jaén: Universidad de Jaén.
- IGM (2016). *Especificaciones Técnicas para la Producción de Cartografía Escala 1:5000*. Instituto Geográfico Militar, Quito.
- 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/>