

# Sistemas de Información para la minería de datos georeferenciada en ecología urbana: Perros callejeros como caso de estudio

Luis Ramos<sup>(1,2)</sup>, Diego E. Procopio<sup>(1,3)</sup>, Marta Lasso<sup>(1,2)</sup>, Maria Eugenia de San Pedro<sup>(1,2)</sup>

(1)Universidad Nacional de la Patagonia Austral-Santa Cruz, Argentina

(2)Instituto de Tecnologías Aplicadas, Unidad Académica Caleta Olivia

(3) Centro de Investigaciones Puerto Deseado, ICASUR- (UNPA-UACO)

## Abstract

*En el presente trabajo se explicita el proceso tecnológico, mediante una Arquitectura Web GIS que permita publicar información estadística obtenida y su posterior análisis aplicado en nuestro caso de estudio de ecología urbana en la distribución de los perros callejeros y su impacto en el ejido urbano y suburbano; de la ciudad de Puerto Deseado.*

*La necesidad de brindar accesibilidad a la información estadística resultante a través de la web, pondera la gestión del proceso a través de tecnologías que puedan dar soporte y ser sustentables a este requisito.*

*Además de soportar múltiples formatos de datos espaciales y la combinación de diferentes tipologías tecnológicas que deberían funcionar de manera integrada.*

## Palabras Claves

Arquitectura Web, GIS, Análisis Estadístico, Formatos datos espaciales.

## 1 Introducción

En la actualidad existe una alta demanda de implantación de Tecnologías Web-GIS, desde distintas organizaciones o entidades de la región patagónica.

El equipo de investigación se encuentra trabajando en identificar, diversas necesidades funcionales para profundizar en el marco tecnológico; y poder brindar soluciones en conjunto con otros campos de la ciencia como la sociología, biología entre otros. Con un perfil orientado a un escenario multidisciplinar.

Para ello es indispensable la integración de saberes, que permitan un intercambio bidireccional de los distintos campos de aplicación.

Ya que debido al auge de los Sistemas de Información Geográfica y su transversalidad a otros campos de la ciencia es relevante para la consecución de soluciones tecnológicas integradas.

En cuanto al dominio de datos, el mismo tiene sus orígenes en múltiples fuentes y formatos. Esto es una característica que prevalece en el manejo de datos espaciales.

## 2 Contexto del caso de estudio.

Los perros en varios países en desarrollo son vistos sobre las calles, especialmente en áreas urbanas [10]. Estos perros son llamados en general perros callejeros [13]. Ellos pasan gran parte de su vida en las calles como carroñeros, y a pesar que muchos no son propiedad de los humanos o no están bajo ninguna supervisión humana, dependen de los humanos para su sustento [16]. Estos perros típicamente tienen características mestizas y son un importante componente de la ecología urbana. Por lo tanto, son un muy buen sistema para estudios de modelos de ecología, etiología y para testear modelos de organización social [11], [16] y [10].

En una primera instancia se realizó un primer censo de perros callejeros para estudiar la distribución, abundancia y estado de la población; y en segunda instancia comenzar a definir un programa de seguimiento, con el fin de detectar cambios en la abundancia de la población así poder tomar medidas correctivas o de control dependiendo de la situación [15], en la ciudad de Puerto Deseado.

Para este fin hemos diseñado una planilla electrónica, diseñada para realizar un recuento completo de la

población y sus recursos, para ser utilizada en dos épocas al año, una en la época fría (otoño- invierno) y en la otra en la época cálida (primavera-verano).

En la localidad de Puerto Deseado, así como en tantas otras, existe una gran cantidad de perros que deambulan libremente en las calles, cuyas consecuencias afectan a la salud pública, bienestar animal y a la salud ambiental.

Obtener y analizar información geográfica sobre la distribución de perros callejeros, permitirá obtener métricas que faciliten poder clarificar la problemática, de manera que organismos municipales, de medio ambiente y demás organismos del estado, identifique parámetros para la toma de decisiones.

La construcción mediante una solución de tecnología GIS-Web, traerá consigo una visión más amplia y clara del fenómeno de la Distribución Geográfica de Perros Callejeros.

Para ello es preciso establecer un modelo de datos y modelo tecnológico que permita obtener una solución integral que cumpla satisfactoriamente, requerimientos tales como, visualizar la distribución espacial e identificar zonas geográficas con mayor y menor densidad de perros callejeros.



Para poder avanzar en estos lineamientos se analiza y selecciona las distintas herramientas que darán forma a la arquitectura tecnológica.

### 3 Conceptualización

#### Principales Definiciones GIS-Web

- Las aplicaciones GIS Web, mejoran las capacidades de los usuarios en tres formas distintas: la primera es el acceso a los datos espaciales y otra información. Con un ambiente basado en Web, el SIG se convierte en interactivo, dinámico y accesible a grandes grupos de usuarios como herramienta de comunicación visual; la segunda es la exploración de datos espaciales y geovisualización que ofrece a las personas de negocios apoyo para tomar decisiones; la tercera es el procesamiento, análisis y modelamiento de datos espaciales. [17].
- Un SIG Web se define como un SIG distribuido a través de una red de computadores para integrar, diseminar y comunicar información geográfica a través de la WWW [19].
- Un SIG Web es un tipo de sistema de información distribuido; de manera concisa, se considera que es cualquier SIG que utiliza tecnología Web para comunicar sus componentes. [20].

#### Secuencia de Flujo

- El usuario realiza ejecuta GISWeb a través del cliente, alternativamente a través de navegador web, un programa de escritorio o aplicación móvil.
- Cliente envía la petición de un requerimiento al Servidor a través de protocolo http.
- Servidor Web envía la petición al Servidor GIS, retorna los datos de la base de datos geográfica y procesa la petición, que puede ser renderizar un mapa, realizar una consulta o llevar adelante una tarea de análisis. Los datos, el mapa, o cualquier resultado se envían al Servidor Web.
- El Servidor Web se comunica con el cliente a través de protocolo http.
- Como último paso se visualizan los resultados por parte del cliente al usuario, completando la secuencia.

#### Arquitectura Gis Web (Figura 1)

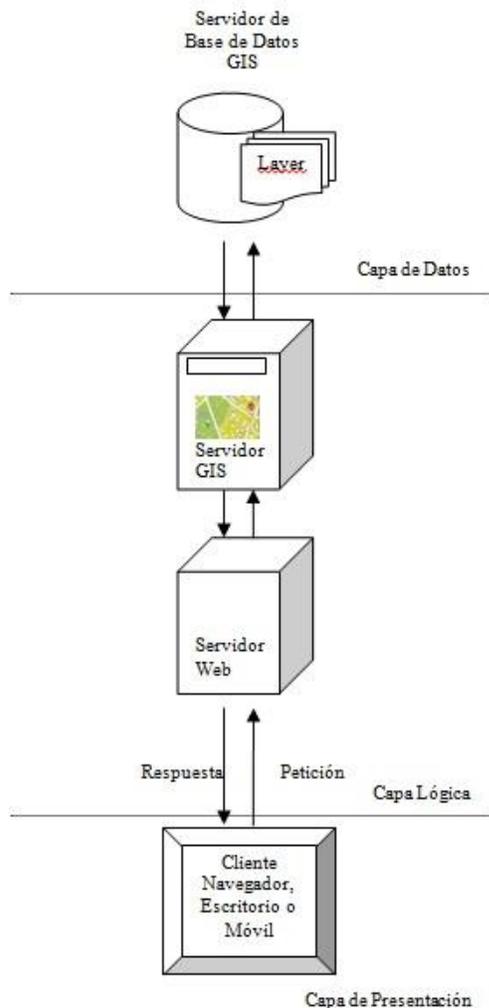


Figura 1: Arquitectura Gis Web

## 4 Desarrollo

El proceso de desarrollo del caso de estudio de Perros Callejeros, implica las siguientes etapas

- *Gestión de Datos*
  - a. Producción de Datos
    - i. Relevamiento de Datos en campo.
    - ii. Procesamiento Espacial y Obtención de capas geográficas. Aplicación de Modulo Estadístico.
  - b. Modelado de Datos
- *Infraestructura Tecnológica*

- *Gestión de Servicios*
  - a. Instalación , Configuración y Puesta a Punto de Servicios Web y Servicios GIS
- *Gestión de Casos de Uso*
  - a. Implementación de páginas dinámicas mediante Lenguaje de Script, Tablas de Estilo CSS, Framework, API.

### Relevamiento de Datos

La recolección de datos en el terreno suele ser costosa, consume gran cantidad de tiempo y no se encuentra exenta de errores. (Figura 2)

Tradicionalmente, el traspaso de datos de las planillas de campo a una base de datos involucra varios pasos. Si existe una acumulación de errores en cada uno de ellos tendremos finalmente muchos más que con la metodología que a continuación proponemos.

Existe una necesidad de contar con un método seguro y eficiente para incorporar los registros de forma ordenada, georreferenciada y reducir errores durante el procesamiento posterior. Para ello se estableció el diseño de una “planilla de terreno digital” que permite incorporar toda la información necesaria en cada registro del “censo de perros callejeros”, para una posterior incorporación a una base de datos [6].

La “planilla de terreno digital” se creó utilizando una versión gratuita del software Cybertracker cargada en un dispositivo de Android Smartphones. Además, el dispositivo estaba sincronizado con un GPS interno, de forma que cuando se completaba un registro el mismo era almacenado en forma secuencial y ordenada conjuntamente con su posición en el espacio. Cada registro consistió en el avistaje o contacto visual de perros callejero (sin capturarlos), así como también los conteos y geo-posicionamientos de los recursos de alimento, agua y refugio disponibles. En el caso de los perros callejeros los mismos era fotografiados para evitar contarlos más de una vez y se registró conjuntamente: el agrupamiento (solitario o en jauría), raza, edad (adulto, juvenil, cachorro), sexo (hembra o macho), tamaño (grande, mediano, pequeño), si poseía propietario (con collar y sin collar), salud (desmejorado, amputado, herido, Dermatitis, inanición, etc.), comportamiento (neutral, agresivo, relajado etc.) y actividad (Ambulando, comiendo, bebiendo, reproduciéndose, etc). En relación a los recursos, se registraron recursos tróficos disponibles (basura, comederos, contenedores, etc.), agua (charcos, grifos abiertos, pérdidas de agua, etc.) y refugio (baldíos, obras abandonadas, terreno con escombros, etc.).

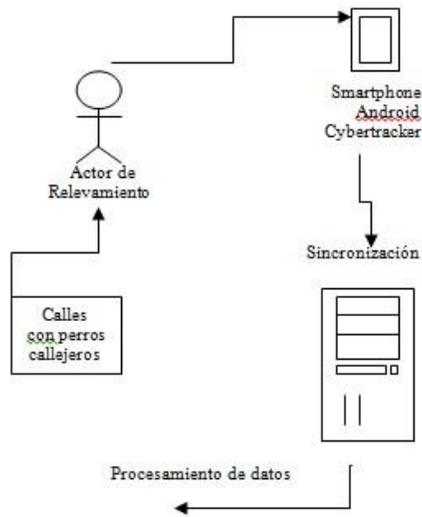


Figura 2: Relevamiento de Datos

Para la recolección de datos, la ciudad de Puerto Deseado fue dividida previamente en 12 zonas, teniendo en cuenta las arterias principales para tener una buena organización de las calles durante los recorridos. En una experiencia piloto, se realizó un ensayo para poder determinar un porcentaje de error en los perros callejeros detectados durante los recorridos para luego ser tenido en cuenta en la estimación de abundancia; el mismo consistió en 5 tramos al azar de 1 km recorriéndolos 2 veces, mostrando una diferencia entre el primer recorrido y el segundo un 10.4 %. El esfuerzo del censo de perros callejeros y el registro de los recursos tróficos y de refugio, llevó 6 días de recorrido en el mes de mayo, en vehículo a una velocidad de 5 km/h, recorriendo en total 110 km.

#### Procesamiento Espacial - Estadístico

Con los contactos de los perros callejeros georreferenciados y a través del software Crime Stat IV [25], la primera intención es determinar los descriptores más básicos de la distribución espacial, como son las estadísticas centrográficas ([18]; [21]; [22]; [23]; [24]), determinando el centro medio de distribución de localizaciones y el centro de mínima distancia, el cual este último nos permitiría por ejemplo poder decidir donde instalar una casilla de castración y/o control sanitario para que todos los propietarios que tengan sus perros en la vía pública tengan la misma distancia en llegar y ser atendidos. También fue necesario determinar si las localizaciones de los perros observados exhibieron algún tipo de patrón sistemático, en oposición a una distribución aleatoria y a que distancia ocurrieron estos patrones, para analizar dichos patrones se utilizó la estadística del índice del vecino más cercano [7] y la K

de Ripley respectivamente [12]. A posteriori se realizó el análisis de zonas calientes o de clúster, para corroborar si estas zonas calientes coinciden con las aglomeraciones de localizaciones o son producto del azar [9].

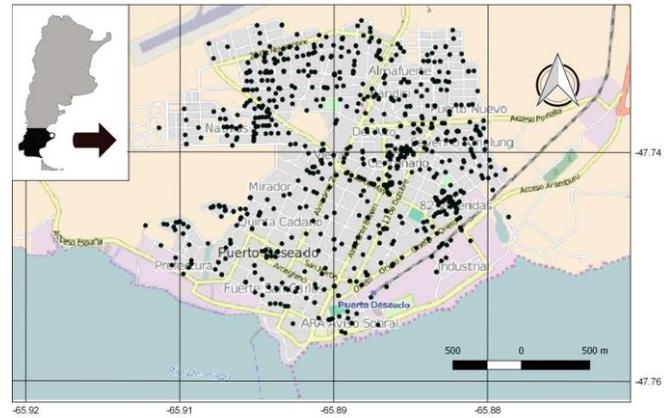


Figura 3: ciudad de Puerto Deseado y distribución de los perros censados

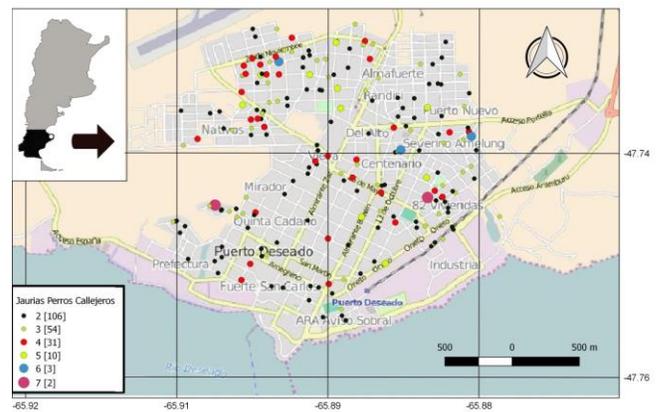


Figura 4: Agrupaciones de perros, identificadas por tamaño de jaurías y entre paréntesis la cantidad de las mismas por cada categoría.

Luego de ésta primera etapa, se exploró como la población de perros callejeros utilizan el área de estudio y se resolvió estimando la intensidad de uso o densidad relativa al área de estudio (perros/km<sup>2</sup>) mediante una interpolación utilizando la función Kernel [14].

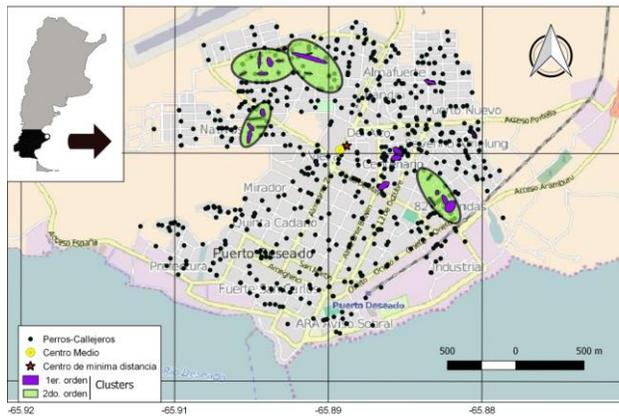


Figura 5: Muestra las distintas zonas calientes o Clusters, junto a medidas centrográficas que han sido establecidas como referencia.

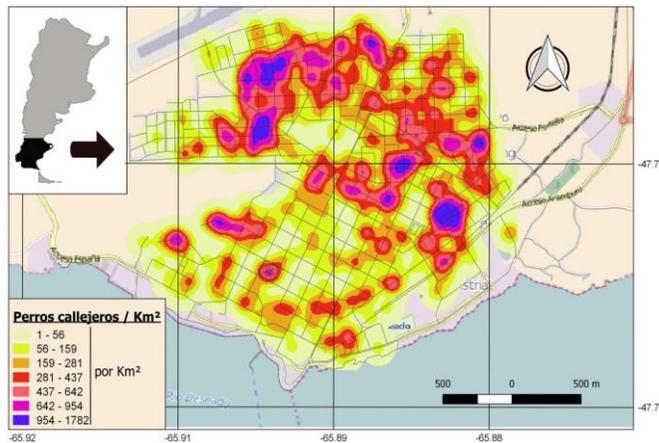


Figura 6: Densidad de perros callejeros por km<sup>2</sup>

Con los datos recolectados del estado de la población de perros callejeros, junto con los registros de recursos tróficos y de refugio, se pretende determinar mediante modelos estadísticos predictivos, la importancia que tienen estos en la distribución de la población.

Durante el censo de perros callejeros se almacenaron 2.030 registros georreferenciados en 4.38 km<sup>2</sup>, de los cuales  $1.034 \pm 103$  de los registros, fueron perros callejeros. Estos fueron representados en un 68.4 % por perros mestizos y un 31.6 % de perros de diferentes razas, siendo los más representativos los perros de las razas Caniche, Pastor alemán, Border collie, Labrador retriever, Golden retriever y Pequines.

En relación a las medidas centrográficas de distribución, patrones de distribución, zonas calientes y de densidades relativas de perros, se pudo determinar que

el centro de mínima distancia y centro medio ubicándose muy próximos unos de otros. Se identificaron 21 zonas calientes de primer orden y 4 de segundo orden coincidiendo estas con las mayores aglomeraciones de localizaciones dándonos a entender que éstas no son producto del azar. (Figura 5). El índice de vecino más cercano presento una distribución significativamente agrupada (NN: 0.468, Z:-33,  $p < 0.001$ ). En la estadística de k de Ripley se observó que la agregación espacial de los perros callejeros es agregada en todas las distancias, comenzando a ser está más significativa a partir de los 42 metros, utilizada ésta última como distancia de búsqueda para realizar la interpolación mediante la función Kernel. De esta manera, se pudieron representar 7 clases de intensidades de uso que van de 1-56 perros por Km<sup>2</sup> en las zonas menos densas, a zonas más densas de 954 a 1.782 perro/km<sup>2</sup> (Figura 6).

Por ultimo luego de esta primera experiencia en los meses fríos, se esperara a una segunda experiencia en los meses cálidos, para poder tener una evaluación de las condiciones iniciales, de partida, contra las cuales luego se compararán los cambios, comenzando en una segunda etapa con una experiencia piloto de seguimiento .

#### Infraestructura Tecnológica

El modelo arquitectónico está conformado por el siguiente escenario, orientado principalmente a soluciones libres y en menor medida de mercado. Debido a estándares de interoperabilidad que ha desarrollado el Open Geospatial Consortium, [1] es posible la integración de ambas tecnologías.

La Metodología utilizada es RUP - Proceso Unificado Racional (por sus siglas en inglés de Rational Unified Process), que es un proceso de desarrollo de software desarrollado por la empresa Rational Software, actualmente propiedad de IBM.

Se constituye un modelo Integral compuesto de diversas tipologías tecnológicas tales como:

- Bases de Datos: PostgreSQL con extensión PostGis Programación: Framework OpenLayer, lenguaje JavaScript, API Google Map integrado en OpenLayer.
- Procesamiento de datos espaciales: Quantum Gis.
- Servicios: Servidor Web Apache, Servidor Gis Geoserver, GeoWebCache integrado en Geoserver para optimización de visualización.
- Software de Relevamiento de datos: Cybertracker para móvil,
- Modulo Estadístico: Software Comercial Crime Stat IV.

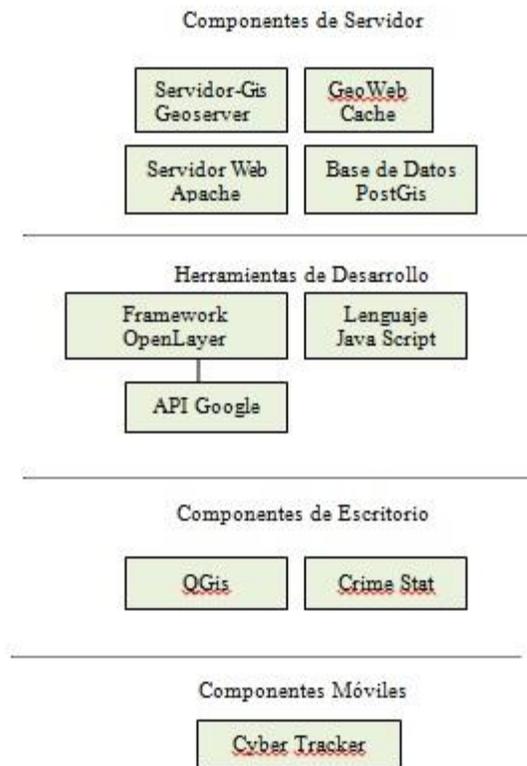


Figura 7: Infraestructura tecnológica

La selección de herramientas tiene en cuenta los siguientes factores como:

- La pertinencia de apoyo vinculado a proyectos de desarrollo de tecnologías por parte del OSGeo. GeoServer [2], OpenLayer [3], Qgis, son ejemplos de proyectos que apoyan el OSGeo.
- La facilitación de alto grado de Interoperabilidad
- La no dependencia de herramientas con licencias comerciales, sin que esto implique un alto decremento en la performance.
- Compatibilidad con distintos sistemas operativos.
- En cuanto a Servidores, que posean un amplio abanico de posibilidades funcionales

### Gestión de Servicios

Para esta primera producción de desarrollo, se llevó adelante la configuración y puesta en funcionamiento de un Servidor Web Apache.

En cuanto al manejo de datos se realizó la instalación y configuración de PostgreSQL y el módulo de soporte de datos espaciales PostGIS [26].

Luego del procesamiento de capas vectoriales en referencia a los perros, se incorporaron las mismas a la base de datos.

Para su posterior visualización a través de la Web, se dispusieron las capas accesibles a través de Geoserver y Servidor Web Apache.

### Implementación de Casos de Uso

Los casos de uso identificados son los siguientes.

- Visualización de capa de localización de perros callejeros
- Visualización de capas espaciales estadísticas
- Interfaz Gráfica para la selección de capas.
- Funcionalidad de zoom y desplazamiento
- Visualización de Capas provistas por Google Map [27].
- Visualización de Capa de Google Map Street
- Herramienta de mapa contextual.

## 4 Referencias a Futuro y Conclusiones

Se prevé un segundo trabajo de relevamiento para la incorporación de datos más actualizados y poder realizar un análisis comparativo con la referencia anterior. Se proyecta la identificación de patrones de comportamiento mediante la acumulación de datos históricos.

En cuanto al marco tecnológico, como estrategia se apunta a soluciones escalables que permitan el agregado de requerimientos que impliquen la adaptabilidad.

El proyecto de desarrollo se encuentra alojado en el laboratorio de informática, se prevé la transferencia para el acceso a través de la web mediante el sitio de la universidad.

También permitirá la transferencia con otras organizaciones de especial interés en organismo del estado.

## 5 Referencias

- [1] OGC Open Gis Consortium, Sitio Oficial (1999) [Online]. Disponible <http://www.osgeo.org>
- [2] Sitio Oficial Geoserver [Online]. Disponible <http://www.geoserver.org>
- [3] Sitio Oficial OpenLayer [Online]. Disponible <http://www.openlayer.org>
- [4] Louis Liebenberg (1999) Sitio Oficial Cybertracker [Online]. Disponible <http://www.cybertracker.org>

- [5] Comité Director del Proyecto Qgis (2007) .Sitio Oficial Qgis [Online].Disponible <http://www.qgis.org/es/site>.
- [6] Calo A y E. Tyson (2012) Un Sistema de Monitoreo Comunitario Basado en Móviles: Un Estudio de Caso en la Sierra Madre de Chiapas. Tesis El Colegio de la Frontera Sur y Colorado State University.243 pp.
- [7] Clark, P.J., Evans, F.C. 1954. Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology* 35, 445-453.
- [8] CyberTracker 2016. (CyberTracker Conservation. NPC. Reg. No. 2000/016138/08), <http://www.cybertracker.org>.
- [9] Day, W.H.E. & Edelsbrunner, H. *Journal of Classification* (1984) 1: 7. doi:10.1007/BF01890115
- [10] Majumder S.S., Bhadra A., Ghosh A., Mitra S., Bhattachaejee D., Nandi A. N. and Bhadra A..(2014). To be or not to be social: foraging associations of free-ranging dogs in an urban ecosystem. *Acta ethol* 17:1-8.
- [11] Pal SK, Ghosh B, Roy S (1998) Agonistic behaviour of free-ranging dogs (*Canis familiaris*) in relation to season, sex and age. *Appl Anim Behav Sc* 59:331–348
- [12] Ripley, B.D. (1976). The second order analysis of stationary point processes. *Journal of Applied Probability*, 13, 255-266
- [13] Serpell J. (1995). The domestic dog: its evolution, behavior, and interaction with people. Cambridge UP,Chicago.
- [14] Silverman, B. W. Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1986. x+175 pp. ISBN: 0-412-24620-1.
- [15] Thompson, W.L.; G.C. White y C. Gowan. 1998. Monitoring Vertebrate Populations. Academic Press, San Diego. 365 pp.
- [16] Vanak AT, Gompper ME (2009). Dietary niche separation between sympatric free-ranging domestic dogs and Indian foxes in central India. *J Mamm* 90:1058–1065.
- [17] Draguicevic (2004) “The potential of Web-based GIS”*Journal of Geographical Systems* ,6,79-81.
- [18] Lefever, D. 1926. “Measuring geographic concentration by means of the standard deviational ellipse”. *American Journal of Sociology*, 32(1): 88-94.
- [19] Peng.Z.R.,Tsou.M.H.,(2003).”Internet GIS: distributed geographic information services for the Internet nd wireless networks”.Jhon Wiley & Sons Inc.
- [20] Pinde Fu, Jiulin Sun (2011) - “Web GIS: Principles and Applications”, ESRI Press, ISBN-13: 978-1589482456
- [21] Bachi, R. 1957. *Statistical Analysis of Geographical Series*. Central Bureau of Statistics, Kaplan School, Hebrew University: Jerusalem.
- [22] Neft D.S. 1962. *Statistical Analysis for Areal Distributions*. P h.D. dissertation, Columbia University: New York.
- [23] Hultquist J., L. Brown and J. Holmes. 1971. “Centro: a program for centrographic measures”. Discussion paper no. 21, Department of Geography, Ohio State University:Columbus, OH.
- [24] Ebdon, D..1988. *Statistics in Geography* (second edition with corrections). Blackwell: Oxford.
- [25] Ned Levine (2015). *CrimeStat IV: A Spatial Statistics Program for the Analysis of Crime Incident Locations* (version 3.3). Ned Levine & Associates, Houston, TX; National Institute of Justice, Washington, DC. July
- [26] Sitio Oficial PostGis [Online] - [www.postgis.net/](http://www.postgis.net/)
- [27] Sitio Oficial API Google Map [Online] <https://developers.google.com/maps/?hl=es-419>