

Seguimiento colaborativo del ruido ambiental utilizando dispositivos móviles y sistemas de información geográfica

Juan Humberto Juárez Hipólito*
Marco Antonio Moreno Ibarra**
Miguel Jesús Torres Ruiz***

Recibido el 4 de junio de 2017, aceptado el 20 de septiembre de 2017

Resumen

El ruido ambiental, es una de las principales formas de contaminación en las ciudades y que afecta la calidad de vida de las personas. En este artículo, se presenta una metodología que utiliza un enfoque basado en la Información Geográfica Voluntaria (VGI, por sus siglas en inglés) para el monitoreo, análisis y predicción del ruido ambiental, lo cual puede resultar muy útil para plantear alternativas que mejoren la vida en una ciudad. El presente trabajo considera las fases de adquisición de los datos, análisis y procesamiento de los datos, así como la visualización de la información, considerando la temporalidad de los mismos y tomando en cuenta niveles de análisis a nivel macro y micro para la superficie de estudio. Se presentan algunos detalles del diseño y desarrollo de un sistema de información geográfica, compuesto por una aplicación web de mapas, una aplicación para dispositivos móviles denominada "NoiseMonitor", el análisis geoespacial y los métodos de aprendizaje automático (máquinas de soporte vectorial y redes neuronales artificiales) para la predicción de ruido ambiental; utilizando información contextual; es decir, algunos datos relacionados con la ciudad. Se busca aprovechar la disposición de los ciudadanos de participar colaborativamente para monitorear su entorno y ser considerados como sensores humanos, lo cual a diferencia con los enfoques tradicionales, el costo asociado al desarrollo e implementación de este proyecto es mucho menor.

* Centro de Investigación en Computación (UPALM), Instituto Politécnico Nacional (IPN), Zacatenco, Av. Miguel Othón de Mendizabal s/n, Col. La Escalera, 07320, Ciudad de México, México, correo electrónico: a140097@sagitario.cic.ipn.mx

** UPALM, IPN, correo electrónico: marcomoreno@cic.ipn.mx.

*** UPALM, IPN, correo electrónico: mtorres@cic.ipn.mx.

Asimismo, se presenta y discute un caso de estudio basado en la Ciudad de México, particularmente el cuarto cuadrante del Centro Histórico, el cual es muy representativo por la variedad de ruido ambiental que se genera en esa zona de la ciudad. Este enfoque tiene aplicación en áreas tales como big data desde una perspectiva colaborativa, Internet de las Cosas y ciudades inteligentes.

Palabras clave: *Mapas de Ruido Ambiental, VGI, Aprendizaje Automático, Máquinas de Soporte Vectorial, Redes Neuronales Artificiales.*

Resumo

O ruído ambiental é uma das principais formas de contaminação ambiental nas cidades afetando a qualidade de vida das pessoas. Neste artigo, apresenta-se uma metodologia que utiliza um enfoque baseado na Informação Geográfica Voluntária (VGI, sigla em inglês) para o monitoramento, análise e previsão do ruído ambiental, que poderá ser muito útil para estimar alternativas que melhorem a vida em uma cidade. Este trabalho considera as fases de aquisição, análise e processamento dos dados, assim como a visualização da informação, considerando sua temporalidade e tomando em conta análises a nível macro e micro para a superfície de estudo. Também se apresentam alguns detalhes do desenho e desenvolvimento de um sistema de informação geográfica, composto por uma aplicação web de mapas e uma aplicação para dispositivos móveis denominada “NoiseMonitor”. Também é apresentada a análise geoespacial e os métodos de aprendizagem automática (máquinas de suporte vetorial e redes neurais artificiais) para a previsão de ruído ambiental; utilizando informação contextual com alguns dados relacionados com a cidade. Com este tipo de trabalho se busca aproveitar a disposição dos cidadãos de participar colaborativamente para monitorar seu entorno e serem considerados como sensores humanos, sem os enfoques tradicionais, sendo o custo associado ao desenvolvimento e implementação deste projeto muito menor. Desta forma, se apresenta e discute um estudo de caso baseado na Cidade do México, particularmente o quarto quadrante do Centro Histórico da Cidade, muito representativa pela variedade de ruído ambiental que se faz nessa região. Nesta abordagem tem-se aplicação em áreas como big data desde uma perspectiva colaborativa, Internet das coisas e cidades inteligentes.

Palavras chave: *Mapas de Ruído Ambiental, VGI, Aprendizagem Automática, Máquinas de Suporte Vetorial, Redes Neurais Artificiais.*

Abstract

Environmental noise is a big problem related to the environmental pollution in cities, which affects the quality of people life. In this paper, a methodology that uses an approach based on Volunteered Geographic Information (VGI) for the monitoring, analysis and prediction of environmental noise is proposed. It can be very

useful to propose alternatives and initiatives that improve the life in a city. So, this work is composed of the following stages: data acquisition, analysis and, data processing, as well as the information visualization, considering the temporality of the same and taking into account macro and micro levels of analysis for the study surface. In addition, some details of the design and development of a geographic information system are presented, consisting of a web-mapping system, an application for mobile devices called “NoiseMonitor”, geospatial analysis and machine learning methods (support vector machines and artificial neural networks) for the prediction of environmental noise; by using contextual information; that is, some data related to the city. This kind of work seeks to take advantage of the willingness of citizens to participate collaboratively to sense their environment and be considered as human sensors, which unlike traditional approaches, the cost associated with the development and implementation of this project is much lower. Likewise, a case study based on the Mexico City is presented and discussed, particularly the fourth quadrant of the Historic Center of the City, which is very representative for the variety of environmental noise that is generated in that area.

The application domain of this approach is oriented towards big data from a collaborative perspective, Internet of Things and smart cities.

Key words: *Environmental Noise Maps, VGI, Machine Learning, Support Vector Machines, Artificial Neural Networks.*

Introducción

El ruido ambiental no es solamente una forma de contaminación ambiental, sino es un problema que afecta la calidad de vida de las personas, el cual genera estrés y que puede provocar la pérdida progresiva de la audición y otros problemas a la salud como hipertensión y problemas cardíacos (Antón y Anton, 2012). Se ha encontrado, que la exposición prolongada al tránsito vehicular está asociada con un pequeño aumento en el riesgo de mortandad por complicaciones cardiovasculares, especialmente en la población de avanzada edad (Halonen *et al.*, 2015). De acuerdo con la Organización Mundial de la Salud (OMS), el ruido es la segunda causa de problemas de salud, después de la mala calidad del aire. Los elevados niveles de ruido se dan principalmente en las zonas más densamente pobladas, tales como la Ciudad de México, una de las ciudades más pobladas del mundo, por lo que es importante responder a las preguntas sobre dónde, cuándo y por qué se generan los diferentes niveles de ruido.

Las tecnologías geoespaciales han dado a los ciudadanos la oportunidad de participar de forma activa e intencional en actividades que alguna vez estuvieron relegadas al mundo de geógrafos y cartógrafos expertos. Esto puede ser interpretado como el surgimiento de una nueva geografía pero sin geógrafos, lo cual se ha denominado como neogeografía; la cual representa un fenómeno social que describe

el uso de herramientas y técnicas geográficas por parte de la gente común, que voluntariamente producen su propia información, donde ésta tiene la cualidad de ser información con contenido geográfico (Connors *et al.*, 2012; Goodchild, 2009).

De este fenómeno derivan conceptos como la información geográfica voluntaria (VGI), que es aquella generada con un esfuerzo intencional y consiente por la comunidad de usuarios con poca o nula educación, experiencia o entrenamiento en geografía o cartografía; es decir, por los neogeógrafos (Goodchild, 2007). Esta información ayuda a ampliar la información de los expertos en el análisis de información geográfica, tales como la comunidad científica u organismos privados dedicados a la investigación, de tal forma que se pueda aprovechar tanto el conocimiento de expertos como de la comunidad no experta (Connors *et al.*, 2012; Elwood *et al.*, 2012; Goodchild, 2009). Algunas aplicaciones incluyen Wikimapia, OpenStreetMap, GeoNames, GeoCommons y TierraWiki (Elwood *et al.*, 2012; Goodchild, 2007; Sui, 2008), análisis de delincuencia y seguridad pública, actividades recreativas y de aventura (Parker *et al.*, 2013) y respuesta oportuna a desastres naturales (Connors *et al.*, 2012; Elwood *et al.*, 2012; Goodchild y Glennon, 2010; Middleton *et al.*, 2014).

La VGI intenta reducir la brecha que separa el trabajo conjunto del público, los investigadores y las autoridades responsables de tomar decisiones sobre políticas medioambientales de carácter público (Connors *et al.*, 2012), por lo que ha sido empleada en ciudades inteligentes para optimizar las operaciones y servicios relacionados con temas ambientales, sociales y económicos (Nakano *et al.*, 2015). Otras metodologías han sido desarrolladas para asegurar la calidad de la VGI (Fonte *et al.*, 2015). Diversos proyectos relativos al medio ambiente que se han realizado tomando en cuenta la participación de los ciudadanos, los cuales se relacionan con la restauración del hábitat, salud pública, planeación ambiental, monitoreo de la calidad del agua, salud y preservación de recursos forestales y animales, manejo de incendios forestales, mapeo y monitoreo de campos de cultivo, programas humanitarios, etc. De hecho, la VGI establece un nuevo paradigma para la investigación de fenómenos socio-espaciales, volviendo necesario el monitoreo continuo del comportamiento social cambiante y las interacciones en las sociedades urbanizadas que evolucionan rápidamente, haciendo de VGI un fenómeno de *big data* (grandes volúmenes de datos) que no puede concebirse aisladamente de otros conceptos de la ciencia de datos (Jiang y Thill, 2015). Este trabajo representa un enfoque para construir *big data* desde una perspectiva colaborativa.

Este artículo se organiza de la siguiente manera: en los “Antecedentes” se presentan algunos trabajos relacionados con el estado del arte, particularmente del monitoreo del ruido ambiental. Posteriormente se explica la metodología desarrollada para analizar y caracterizar el comportamiento del ruido ambiental. Luego se presenta un caso de estudio, así como los resultados y experimentos realizados,

utilizando dos enfoques de aprendizaje automático. Finalmente, se describen las conclusiones y trabajo a futuro.

Antecedentes

En esta sección se presentan algunos trabajos previos, relacionados con la generación de mapas de ruido ambiental y las aplicaciones para el monitoreo colaborativo del ruido ambiental.

Mapas de ruido ambiental en zonas urbanas

El primer mapa de ruido de la Zona Metropolitana del Valle de México fue desarrollado con la finalidad de ayudar a las autoridades a emitir políticas correctas contra los problemas de ruido que afectan a esta región, y con ellas poder mejorar la calidad de vida de sus habitantes; este proyecto realiza un análisis entre los niveles de ruido aceptables, tolerables, molestos e inaceptables (Rodríguez y Garay, 2012). En el 2014, se presentó un sistema para monitorear el ruido ambiental de Nueva York; el sistema analizaba datos históricos provenientes de las llamadas de quejas relativas al ruido, las redes de carreteras, datos de check-ins y puntos de interés provenientes de la red social Foursquare (Zheng *et al.*, 2014a; Zheng *et al.*, 2014b). Algunas de las principales distinciones de este proyecto, son su capacidad de inferir la composición del ruido, así como la tolerancia de las personas al ruido en diferentes horas. Una aplicación que establecía modelos de predicción de los niveles de ruido y que realizaba mapas de ruido con una exactitud del 80.9% fue presentada por Karolus (2013). Esta aplicación utilizaba técnicas de aprendizaje automático y mediciones voluntarias georeferenciadas tomadas de la aplicación NoiseMap. También se utilizó información voluntaria de OpenStreetMap y LinkedGeoData para obtener información acerca de calles, edificios y otras fuentes de ruido. Una de las principales aportaciones de este trabajo, fue la posibilidad de predecir niveles de ruido en zonas con pocas o nulas mediciones. Quintero (2013) presentó una propuesta para generar mapas acústicos basados en muestras tomadas en diferentes horarios por un sistema de micrófonos calibrados y un GPS. En 2011 fue creada una arquitectura escalable y paralela que empleaba el paradigma de programación MapReduce para el procesamiento paralelo de grandes colecciones de datos provenientes de diferentes flujos de datos, así como aprendizaje automático, particularmente métodos de kernel, para el reconocimiento de patrones. Con este proyecto, sus autores pudieron crear un mapa de ruido ambiental en tiempo real, con datos provenientes de una red fija de sensores y otra red de dispositivos móviles, así como crear modelos de predicción que permitían calcular los niveles de ruido provenientes de una simulación de flujo vehicular (Kaiser y Pozdnoukhov, 2013).

Aplicaciones colaborativas para el monitoreo de ruido ambiental

Laermometer es una aplicación que recolecta señales de audio de los dispositivos móviles y las envía a un servidor, permitiendo visualizar los mapas de ruido generados con las mediciones en una interfaz web. El objetivo de esta aplicación es ayudar al usuario a encontrar lugares tranquilos donde estar, o simplemente compartir los niveles de sonido en cierta área (Bilandzic *et al.*, 2008). NoiseSPY contiene un portal web en el que se despliegan mapas de ruido en tiempo real (Kanjo, 2010). NoiseTube permite a los usuarios capturar mediciones de ruido utilizando su dispositivo móvil, para crear con esas capturas mapas de ruido que son mostrados en un portal web (Maisonneuve *et al.*, 2009). Ear-Phone es un sistema que muestra en un portal web mapas de ruido generados por las mediciones capturadas con los dispositivos móviles de los usuarios y que fueron enviadas a un servidor central. Cuenta con un sistema de calibración de punto simple, el cual compara valores desconocidos del dispositivo con estándares de referencia de acuerdo con algoritmos determinados. Este sistema incluye métodos de reconstrucción para mediciones faltantes, con la finalidad de mejorar la predicción en los mapas de ruido generados (Rana *et al.*, 2010). NoiseMap recolecta mediciones de ruido, consta de una aplicación móvil y se apoya en otra plataforma web de monitoreo. Con la finalidad de motivar la participación voluntaria, este sistema premia a sus usuarios con rangos de acuerdo con su nivel de actividad (Schweizer *et al.*, 2011). NoiseWatch incluye mediciones de ruido tanto de fuentes científicas oficiales como de VGI, los usuarios voluntarios pueden ingresar desde una plataforma web o desde sus dispositivos móviles (Jirka *et al.*, 2012). También hay aplicaciones que usan técnicas de gamificación para motivar a la gente a participar voluntariamente, tales como Noise Battle, donde el objetivo del usuario es conquistar el mayor número de celdas posibles, lo cual se consigue tomando más y mejores mediciones de ruido en dichas áreas que otros jugadores y evitar perderlas, lo cual ocurre cuando otro usuario se vuelve quien conquiste el área de la misma manera (García *et al.*, 2013; Karolus, 2013).

Comparación con trabajos previos

En la Figura 1, muestra una comparación de los trabajos presentados para el monitoreo de ruido ambiental, en donde se pueden apreciar las diferencias de manera puntual con el trabajo propuesto y presentado en este artículo. En ésta, se destaca que el presente trabajo promueve la VGI, emplea técnicas de aprendizaje automático para predecir niveles de ruido considerando factores alrededor de las mediciones, se cuenta con una aplicación móvil que ha sido desarrollada utilizando los sensores del dispositivo móvil y se tiene además un Web-GIS para la visualización, análisis e interpretación de los datos de ruido ambiental.

Características	Propuestas															
	Laermómetro	NoisePY	NoiseTube	Noise Battle	NoiseQuest	Ear-Phone	NoiseWatch	NoiseMap	Oráculos con métodos de kernel	Ruido en aeropuerto (Sánchez, 2011)	Ruido en unidad habitacional (Quintero, 2013)	Ruido considerando zonas vecinas (Karolus, 2013)	Ruido de Nueva York (Sheng, 2014)	Ruido en la Ciudad de México	Sonómetros para dispositivos móviles	NoiseMonitor (nuestra propuesta)
Promueve la VGI	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Dispositivos móviles	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Muestra mapas de ruido en sitio web	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Mapas de ruido en dispositivos móviles	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Permite el intercambio de mediciones	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Implementa sus propios algoritmos de medición	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Clasifica la fuente y horario del ruido	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Considera factores alrededor de la medición	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Predice utilizando aprendizaje automático (ML)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

Figura 1. Comparación de varias aplicaciones existentes para monitoreo de ruido y sus funcionalidades.

Metodología propuesta

La metodología busca aprovechar la VGI generada por la participación de las personas quienes, a través de una aplicación móvil, capturan las mediciones de ruido en los lugares en que se encuentren. Después, usando técnicas de aprendizaje automático, tales como máquinas de soporte vectorial y redes neuronales, se predicen los niveles de ruido en puntos predefinidos y se generan mapas de ruido utilizando tanto las mediciones, como los valores predichos. Esta metodología está compuesta de las siguientes etapas (véase la Figura 2):

- Adquisición de datos
- Análisis y procesamiento de datos
- Visualización de la información

Se puede observar que en la etapa de adquisición de datos, los usuarios, actúan voluntariamente como sensores, capturan los niveles de ruido ambiental así como sus ubicaciones geográficas, utilizando sus dispositivos móviles y envían esas mediciones a un servidor central, donde son almacenadas en una base de datos en PostGIS. Posteriormente, en la etapa de análisis y procesamiento de datos, las mediciones y alguna información contextual (datos acerca de los intervalos de tiempo, negocios cercanos, colonias y otras mediciones) son utilizadas para entrenar y probar el modelo de aprendizaje automático; en este caso, una máquina de soporte vectorial (SVM), para describir el comportamiento del ruido y predecir sus niveles,

almacenando los resultados en la base de datos de PostGIS. Finalmente, en la etapa de visualización de la información, tanto las mediciones como las predicciones son interpoladas, utilizando un servicio web de procesamiento (WPS). Para esta tarea se implementó y aplicó el algoritmo Barnes con un código estándar de colores; esta capa es sobrepuesta a otras capas en formato servicio web de mapas (WMS), tales como OpenStreetMap (OSM), dando por resultado un mapa de ruido ambiental que se presenta al usuario junto con algunas estadísticas relevantes a través de su navegador web. El sistema cubre las tres etapas de la metodología, posee a una arquitectura cliente-servidor, donde la aplicación de dispositivos móviles y los mapas de ruido pertenecen al lado del cliente y el resto al lado del servidor.

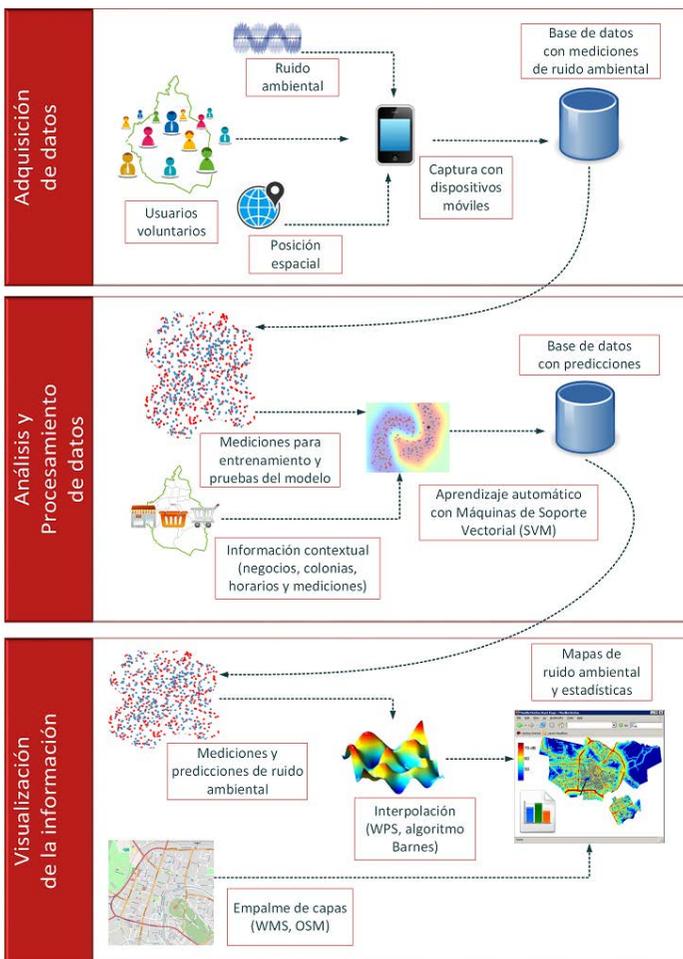


Figura 2. Etapas que componen la metodología propuesta.

Etapa de adquisición de datos

La adquisición de datos consiste en que los usuarios voluntarios, a través de sus dispositivos móviles, sean capaces de medir los niveles de ruido ambiental del lugar en el que se encuentren, para enviar esas mediciones a un servidor de base de datos, donde son almacenadas. Las actividades que se cubren en esta etapa son las siguientes: 1) Registro, 2) Recolección de niveles de ruido, 3) Almacenamiento temporal y 4) Almacenamiento en servidor.

En el Registro, cada usuario se inscribe en el servidor principal utilizando la aplicación instalada en su dispositivo móvil, proporcionando un nombre de usuario y una breve descripción sobre su dispositivo. Un mismo usuario puede registrarse para diferentes dispositivos móviles, y el registro es un proceso que sólo es necesario realizar en una única ocasión en cada dispositivo. El sensor del micrófono del dispositivo móvil del usuario se utiliza para monitorear el ruido ambiental y determinar la medición del sonido en esa ubicación geográfica. Cada medición es georreferenciada, utilizando el radio GPS o el radio Wi-Fi del dispositivo. La medición georreferenciada se almacena temporalmente dentro de una base de datos local. Entre los datos que se guardan se tienen la fecha y hora en que se realizó cada medición, así como el valor de ruido mínimo y máximo registrado. Una vez que el usuario accede a Internet, las mediciones almacenadas en el dispositivo son enviadas al servidor principal, donde se almacenan en una base de datos PostGIS. Cuando el servidor principal le responde al dispositivo del usuario que ha recibido y almacenado sus mediciones, éstas son eliminadas del almacenamiento local de su dispositivo.

Etapa de análisis y procesamiento de datos

En la etapa de análisis y procesamiento de datos, se predicen un conjunto de niveles de ruido en diversos puntos geográficos, entre los cuales existen algunos en los que es probable que no existan mediciones aportadas por los usuarios voluntarios, utilizando información contextual de esos puntos geográficos. Asimismo, en esta etapa se cubren varias actividades; primero, la creación de un modelo de aprendizaje automático; en este caso, una máquina de soporte vectorial. Posteriormente, el entrenamiento y evaluación del modelo, que incluye la separación aleatoria de mediciones en un conjunto de entrenamiento y otro de pruebas. Finalmente, la predicción del ruido ambiental en el área de interés, para lo cual es necesaria la creación de un conjunto de predicción compuesta por la unión de una *lattice* de puntos geográficos y los centroides geométricos de las colonias, además de la predicción del ruido en puntos del conjunto de predicción.

La predicción se basa en las mediciones georreferenciadas. Se utiliza la información contextual de los puntos donde se realizó cada medición, así como técnicas de aprendizaje automático para regresión; en este caso, máquinas de soporte vectorial y particularmente regresión de soporte vectorial (SVR). Estos métodos son aplicados

sobre las mediciones para predecir los valores de los niveles de ruido en puntos geográficos determinados dentro del área de interés.

Predicción a nivel de vecindad geográfica

Algunos factores alrededor de una medición de ruido se han considerado como relevantes para el modelo de predicción; es decir, factores en la vecindad geográfica de cada medición. Por lo tanto, se toma en cuenta lo siguiente:

- *Densidad de negocios.* Suponemos que es posible que una mayor concentración de negocios alrededor de un punto geográfico con mayor cantidad de personas en ellos, equivalga a mayor movimiento de personas, y por ende a mayores niveles de ruido.
- *Mediciones de ruido previas.* Suponemos que es posible que las mediciones de ruido tomadas previamente en la misma área geográfica en intervalos de tiempo similares, sean en cierta medida similares a la medición actual.
- *Tipo de colonias.* Suponemos que es posible que tanto el número, como el tipo de colonias alrededor de un área geográfica, determinen parcialmente el nivel de ruido existente en dicha región.
- *Horario de la medición.* Suponemos que es posible que el día y la hora en que se efectuó la medición determina de cierta forma el nivel de ruido registrado en el área geográfica de dicha medición.

La Figura 3 muestra un esquema que ejemplifica los factores antes mencionados. Los radios concéntricos representan el área de influencia de los factores respectivos alrededor del punto en donde se desea predecir el nivel de ruido; es decir, la vecindad geográfica de la medición.

Entonces, dado que la intensidad del ruido decae con la distancia, todos estos factores de carácter numérico son calculados considerando la distancia desde el punto geográfico de la medición hasta el punto geográfico que representa cada factor.

En el algoritmo presentado en la Figura 4, se muestra la forma en que son calculados los diversos factores alrededor de cada punto geográfico. Como se ha mencionado, se ha considerado la densidad de algunas características geográficas alrededor de cada medición, tales como la densidad de negocios, el tipo de colonias alrededor, y otras mediciones previamente capturadas. Una vez conocidas estas densidades, la influencia de cada factor se calcula numéricamente, considerando la distancia geográfica de la medición respectiva. Todos estos valores numéricos serán utilizados después por el algoritmo de aprendizaje automático para predecir valores de ruido en cualquier otro punto geográfico.

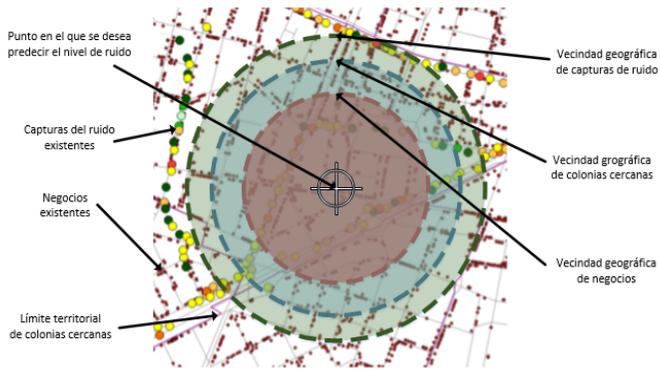


Figura 3. Factores considerados en las vecindades de las mediciones como influyentes para predecir el nivel de ruido.

```

Entrada:
• El conjunto universo c.universo, obtenido en el algoritmo de la Figura 4.5.
• Un conjunto finito de radios de las vecindades geográficas: {radio.negocios, radio.colonias, radio.mediciones}.
Salida:
• El conjunto universo c.universo con el cálculo de la afectación producida por diversos factores.
Datos:
• Un conjunto finito de negocios registrados en el DENEUE del INEGI: c.deneue = {d1, d2, ..., dn}.
1 para cada punto punto en c.universo hacer
2     vecindad.negocios ← vecindad geográfica circular con centro en punto y radio radio.negocios /* inicializar */
3     vecindad.colonias ← vecindad geográfica circular con centro en punto y radio radio.colonias
4     vecindad.mediciones ← vecindad geográfica circular con centro en punto y radio radio.mediciones
5     factor1 ← 0
6     factor2 ← 0
7     factor3 ← 0
8     factor4 ← 0
9                                     /* calcular el factor de negocios cercanos */
9     n ← 0
10    para cada negocio negocio en c.deneue hacer
11        si negocio en vecindad.negocios entonces
12            punto.negocio ← punto con las coordenadas de negocio
13            personas ← promedio de personas registradas en negocio
14            distancia ← distancia euclidiana entre punto y punto.negocio
15            factor1 ← factor1 + personas / (distancia * distancia)
16            n ← n + 1
17        fin
18    factor1 ← factor1 / n
19                                     /* calcular el factor de colonias aledañas */
20    n ← 0
21    para cada colonia colonia en vecindad.colonias hacer
22        punto.colonia ← punto con las coordenadas del centroide de colonia
23        tipo ← tipo de colonia registrado para colonia
24        distancia ← distancia euclidiana entre punto y punto.colonia
25        factor2 ← factor2 + tipo / (distancia * distancia)
26        n ← n + 1
27    fin
28    factor2 ← factor2 / n
29                                     /* calcular el factor de mediciones de ruido cercanas */
30    n ← 0
31    para cada medición medición en vecindad.mediciones hacer
32        punto.medición ← punto con las coordenadas de medición
33        ruido ← promedio de ruido registrado en medición
34        distancia ← distancia euclidiana entre punto y punto.medición
35        factor3 ← factor3 + ruido / (distancia * distancia)
36        n ← n + 1
37    fin
38    factor3 ← factor3 / n
39    factor4 ← n
40                                     /* actualizar los factores en los puntos del conjunto universo */
39    punto[factor1] ← factor1
40    punto[factor2] ← factor2
41    punto[factor3] ← factor3
42    punto[factor4] ← factor4
43 fin
44                                     /* retornar el conjunto universo actualizado */
44 devolver c.universo
    
```

Figura 4. Algoritmo para la actualización de los factores de predicción en el conjunto universo.

Intervalos de tiempo

De acuerdo con diversas observaciones, se ha determinado que el cuerpo humano no llega a adaptarse a niveles altos constantes de ruido y que no todas las personas son igualmente sensibles al mismo y sus efectos; y que la sensibilidad cambia de acuerdo con factores como la hora, siendo más molesto el ruido durante la noche que durante el día (Ising y Kruppa, 2004; Liu et al., 2014). Tomando en consideración lo anterior, es conveniente analizar el ruido usando ciertos intervalos de tiempo (análisis espacio-temporal).

Los intervalos de tiempo sobre los que se realizan los análisis, han sido prefijados tomando en cuenta las actividades cotidianas, que en general, la mayor parte de la población realiza en la Ciudad de México. Cada intervalo de tiempo sirve para obtener un mapa de ruido distinto. En la presente propuesta se consideran intervalos de tiempo acordes entre los días laborables (lunes a viernes) y no laborables (sábado y domingo), obteniendo los siguientes periodos de tiempo.

- Horario H-00: por la madrugada, de 00:00 a 06:00 horas
- Horario H-06: por la mañana, de 06:00 a 12:00 horas
- Horario H-12: por la tarde, de 12:00 a 18:00 horas
- Horario H-18: por la noche, de 18:00 a 24:00 horas

Aprendizaje automático para la predicción de ruido ambiental

Para desarrollar un modelo de aprendizaje, es necesario crear, entrenar, evaluar y utilizar el modelo de aprendizaje. Para la construcción del modelo, es necesario obtener dos conjuntos de puntos de muestreo (datos de ruido): el conjunto de entrenamiento (C_A) y el conjunto de predicción (C_P). La unión de ambos conjuntos forma el conjunto universo (C_U), tal como se muestra en la Figura 5.

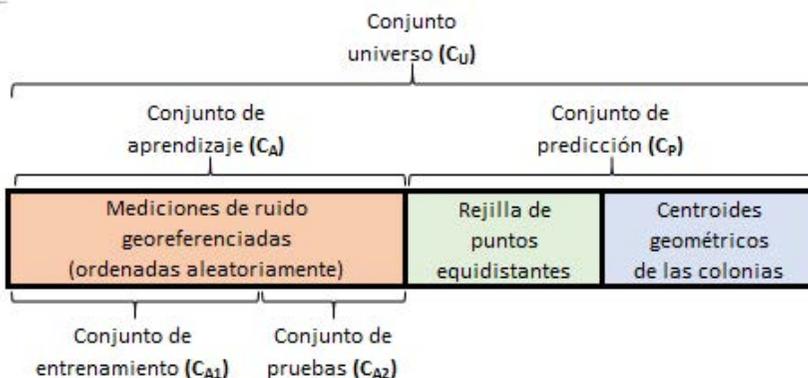


Figura 5. Conjuntos de puntos geográficos utilizados para la modelo de aprendizaje.

Los puntos de las mediciones georreferenciadas corresponden a los lugares en donde los usuarios hayan realizado mediciones de ruido, utilizando sus dispositivos móviles, éstos constituyen el conjunto CA. El conjunto CP se compone de los centroides de las colonias (ya que interesa predecir el nivel de ruido en cada colonia, independientemente de si existen en ella o no mediciones de ruido), así como una lattice de puntos dentro del área de interés; para cubrir la predicción también en las áreas donde no existan colonias o existan pocas de ellas. La distancia que separa a los puntos de la lattice es arbitraria y puede ser configurada (aumentada o disminuida) según se prefiera.

El conjunto CA se divide aleatoriamente en dos subconjuntos: el conjunto de entrenamiento (CA1) y el conjunto de pruebas (CA2). A su vez, se utiliza k-fold cross validation con $k=10$. Es posible definir la proporción de CA1 como el 90% de CA, y CA2 como el 10% restante. Con CA es posible entrenar y evaluar al modelo de aprendizaje automático; en este caso, una máquina de soporte vectorial.

Etapa de visualización de la información

La última etapa de la metodología corresponde a la visualización de la información. En esta etapa se integran, con ayuda de un servidor de datos geoespaciales, como GeoServer, las mediciones y los valores predichos, con los datos de un servicio de mapas, como es el caso de OpenStreetMap. Posteriormente, empleando el código de colores estándar para los niveles de ruido, así como el algoritmo Barnes de interpolación sobre los valores de los puntos del CU, se generan y visualizan los mapas de ruido resultantes, los cuales describen el comportamiento del ruido en el área mostrada, así como datos estadísticos relevantes. El método de interpolación empleado (Barnes) es un algoritmo de interpolación geoespacial que opera sobre una rejilla regular de celdas que cubren un área específica en la que existen mediciones en puntos distribuidos irregularmente (Weymouth *et al.*, 1999; Xia *et al.*, 1999).

Esta etapa incluye la elaboración de mapas de ruido que pueden visualizarse tanto en equipos de escritorio como en equipos móviles, así como la presentación de reportes y estadísticas, tanto del ruido por colonias de municipios (o delegaciones) de interés, así como del ruido por colonias para los diferentes horarios. La unidad básica de medida del ruido ambiental, recomendado por algunos estándares internacionales, tales como la norma ISO 1996-1:2003, es el dB(A). Su código de colores está definido en la norma ISO 1996-2:2007 y está representado junto con su código hexadecimal (RGB) en la Figura 6.

Debido a que la mayoría de los sonidos existentes son de naturaleza fluctuante (y en algunos casos efímera), se emplean medidas como Leq,T y SEL, que permiten describir sonidos fluctuantes; es decir, trasladar la medida fluctuante del sonido a un valor constante. La norma ISO 1996-1:2003 también recomienda utilizar a Leq,T y a SEL como descriptores básicos de los niveles de ruido. Con respecto a las categorías

en que se agrupan las mediciones de ruido según su intensidad, se ha empleado la clasificación establecida en la norma ISO 1996-2, que considera el rango de sonidos menores a 35 decibeles, entre 35 y 80 decibeles en intervalos de 5 decibeles, y mayores a 80 decibeles.

dB (A)	color (HEX)
> 80	#000066
75 - 80	#0000FF
70 - 75	#AC9AD5
65 - 70	#990033
60 - 65	#FF3333
55 - 60	#FF6600
50 - 55	#FFC649
45 - 50	#FFFF00
40 - 45	#005000
35 - 40	#00CC00
< 35	#C0FFC0

Figura 6. Código de colores estandarizado para representar el nivel de ruido en dB(A).

Experimentos y resultados

Caso de estudio

Para verificar la efectividad de la metodología, se ha elegido como caso de estudio el monitoreo de ruido ambiental en la Delegación Cuauhtémoc de la Ciudad de México, que puede visualizarse como un área a nivel macro, y en la que es posible enfocarse, a nivel micro, en el ruido ambiental de las calles alrededor del Centro Histórico de la Ciudad de México, en la colonia Centro.

Área de estudio a nivel macro

El horario H-12 fue el elegido para monitorear el ruido en la Delegación Cuauhtémoc. Las razones para elegir esta área y este horario como caso de estudio son las siguientes:

- *Ubicación:* esta área incluye la capital de la Ciudad de México.
- *Actividad:* Es un área de las más activas en términos de movilidad poblacional.
- *Negocios:* contiene una gran concentración de negocios, por lo que se espera que las mediciones resulten más útiles para el modelo de aprendizaje.

- *Seguridad*: este lugar contiene en general, varias de las colonias en las que hay mayor vigilancia y mejor infraestructura, por lo que representa menor riesgo durante los horarios del recorrido para realizar mediciones.

Área de estudio a nivel micro

Los horarios H-06 y H-12 fueron los elegidos para monitorear el ruido en el Centro Histórico de la Ciudad de México. Las razones para elegir esta área y sus alrededores en estos horarios como caso de estudio, son las siguientes:

- *Ruido*: se determinó que esta es una de las regiones que presentan la mayor concentración de ruido por su amplia actividad mercantil y tránsito de personas.
- *Actividad*: es una de las áreas más activas considerando la afluencia y flujo de personas.
- *Seguridad*: al ser uno de los principales destinos turísticos de esta ciudad, este lugar cuenta con vigilancia constante, por lo que también representa menor riesgo durante los recorridos para realizar las mediciones.

Implementación del sistema de información geográfica

El modelo funcional del sistema de información geográfica (GIS) se muestra en la Figura 7. Este sistema consiste en una aplicación para dispositivos móviles, que permite a los usuarios voluntarios efectuar con el sensor del micrófono, mediciones georeferenciadas de los niveles de ruido en los lugares en que se encuentre; las mediciones son enviadas a un servidor, éste las recibe utilizando una interfaz de servicios de sincronización y las almacena en una base de datos PostGIS; estos módulos cubren la fase de adquisición de datos de la metodología. Las mediciones son usadas por un servicio de análisis para predecir los valores de ruido; este módulo cubre la etapa de análisis y procesamiento de los datos de la metodología. Finalmente, un servidor de datos geoespaciales, genera los mapas de ruido y los presenta a través de una interfaz web; este módulo cubre la etapa de visualización de la información, conforme al diagrama presentado en la metodología.

Aplicación para dispositivos móviles

Se desarrolló una aplicación, llamada NoiseMonitor, para monitorear los niveles de ruido utilizando los dispositivos móviles, con el fin de cubrir las necesidades de la etapa de adquisición de datos. Esta aplicación está basada en Android, y es compatible con la versión 2.3.3 (API nivel 10) o superior; algunas pantallas de la aplicación se muestran en la Figura 8. En esta figura se puede apreciar una pantalla principal, desde la que el usuario puede navegar entre tres secciones diferentes: la pantalla para capturar los niveles de ruido, la pantalla para ver el mapa de ruido, y la pantalla para realizar ajustes en la configuración del usuario y del dispositivo.

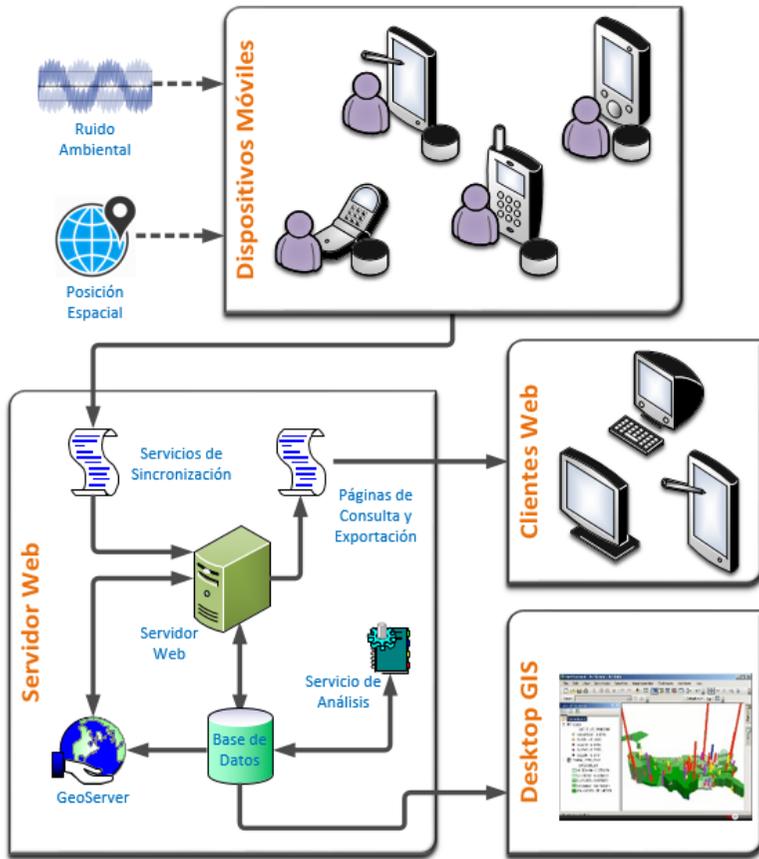


Figura 7. Modelo funcional y arquitectura del sistema.

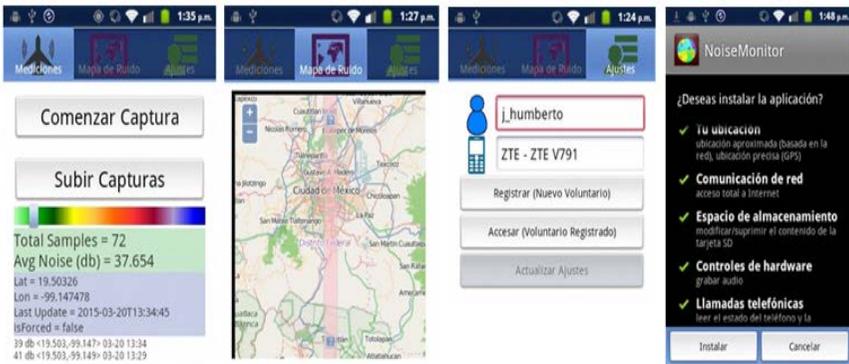


Figura 8. Aplicación móvil NoiseMonitor desarrollada para el monitoreo de ruido ambiental.

Aplicaciones para análisis y procesamiento de datos

De acuerdo con la etapa 2 de la metodología, se utilizó el lenguaje R para realizar el modelo de aprendizaje automático, específicamente se implementó una Máquina de Soporte Vectorial (SVM). Un programa diseñado para funcionar como un servicio que pueda ejecutarse periódicamente fue escrito en Python; éste primero actualiza los factores de predicción en el conjunto universo, después recupera los datos necesarios desde la base de datos, luego invoca al programa escrito en lenguaje R que se encarga de predecir los valores de nivel de ruido en los puntos de interés del conjunto de predicción, el cual está compuesto por los centroides de cada colonia de los municipios (o delegaciones) de interés, más una lattice de puntos que cubren toda el área de interés. Después de obtener las predicciones, el programa escrito en Python actualiza en la base de datos los valores de las predicciones correspondientes.

Aplicaciones para la visualización de la información

La aplicación Web-GIS para la etapa de visualización de la información, que se muestra en la Figura 9, fue dividida en cuatro secciones o pestañas: la primera muestra el mapa de ruido, la segunda permite al usuario realizar búsquedas de lugares geográficos, la tercera muestra algunas estadísticas sobre los niveles de ruido en distintas regiones, y la cuarta permite descargar recursos asociados con la aplicación. Esta aplicación fue desarrollada utilizando PHP, HTML5, CSS, JSON, JavaScript y AJAX. Para mostrar el mapa se empleó el framework de JavaScript OpenLayers, el cual lee datos en formato WMS desde el servidor GeoServer, y éste a su vez, de la base de datos PostgreSQL con la extensión PostGIS.

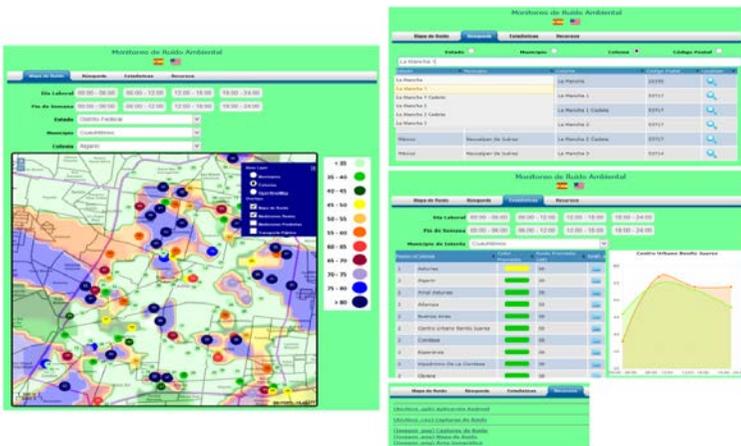


Figura 9. Aplicación Web-GIS desarrollada mostrando mapa de ruido (sintético) de la Delegación Cuauhtémoc.

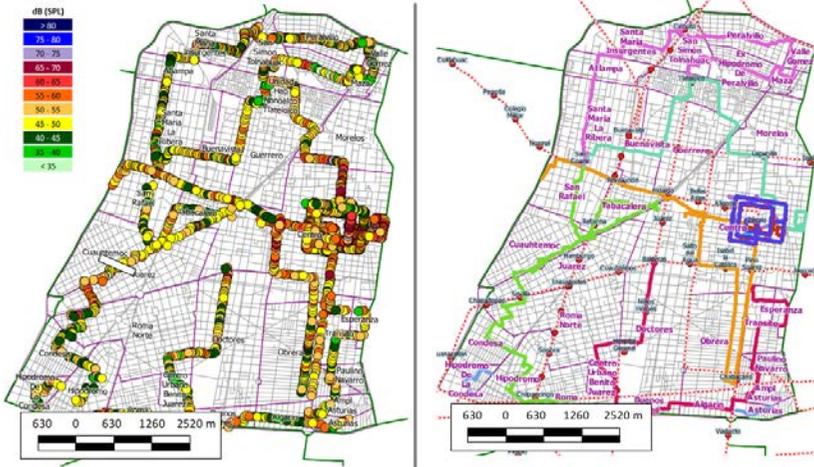


Figura 10. Capturas de ruido ambiental en la Delegación Cuauhtémoc.

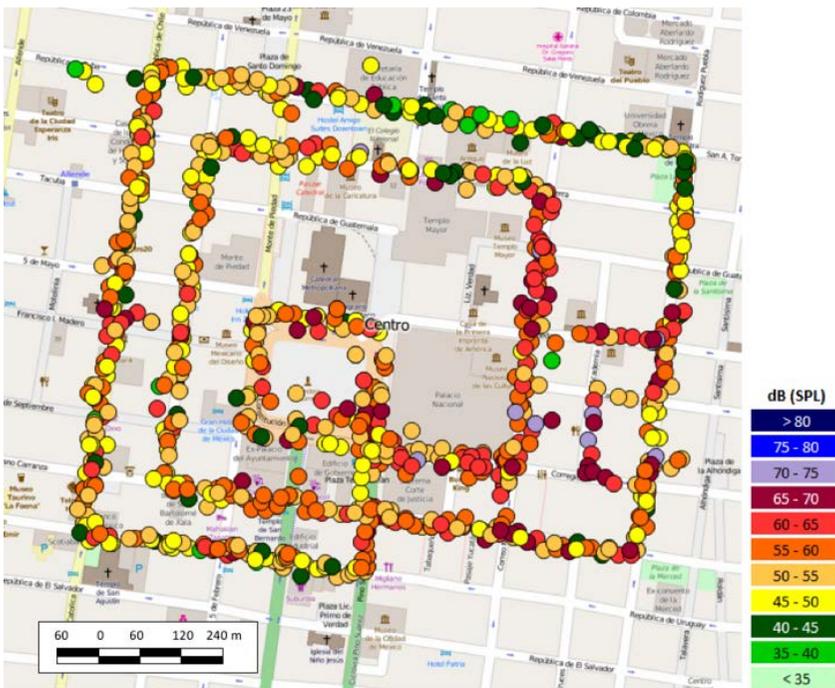


Figura 11. Capturas de ruido ambiental en el Centro Histórico.

Muestreo de ruido en el área de interés

El primer grupo de mediciones se realizó entre el 6 y el 10 de julio de 2015, para hacer el monitoreo se definieron rutas que buscaban recorrer al menos una parte de cada colonia de la Delegación Cuauhtémoc. La Figura 10 muestra las rutas (derecha) y las mediciones registradas (izquierda). El segundo grupo de mediciones se realizó en las inmediaciones del Centro Histórico (véase Figura 11).

Así, en total se monitorearon y coleccionaron con el dispositivo móvil, 2 355 mediciones georeferenciadas de ruido ambiental en el área de interés, habiéndose recorrido poco más de 83 km. Algunas estadísticas sobre la distribución de las mediciones de ruido ambiental capturadas se muestran en la Figura 12, en forma de un boxplot (izquierda), una tabla de estadísticas (centro) y un histograma (derecha). En el boxplot se aprecia en escala original (unidades de dB(SPL)) y normalizada (en un rango entre 0 y 1). Por tanto, utilizando como unidades dB(SPL), aunque las mediciones registradas se encuentran en un rango entre 32 y 76, el 50% de ellas caen en el rango entre 48 y 57, siendo la media de 52.45. El histograma permite confirmar visualmente las observaciones anteriores, así como presentar un comportamiento similar a una distribución normal o gaussiana en los datos.

Predicción de ruido ambiental utilizando aprendizaje automático

Para encontrar los valores adecuados a emplear en el modelo de aprendizaje automático (*machine learning*) desarrollado en el lenguaje R, se implementó en este lenguaje un programa para evaluar el rendimiento de las máquinas de soporte vectorial bajo diversos parámetros, así como compararlo con otras técnicas de aprendizaje automático, específicamente, contra las redes neuronales artificiales (ANN). Por simplicidad, de aquí en adelante se utilizarán abreviaturas tales como

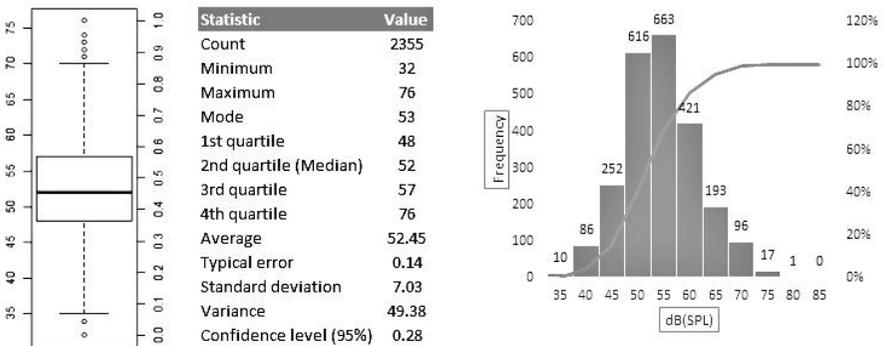


Figura 12. Mediciones de ruido capturadas: boxplot (izquierda), estadísticas (centro) e histograma (derecha).

ML para referirse a aprendizaje automático, SVM para máquinas de soporte vectorial, ANN para redes neurales artificiales, MAE para error absoluto promedio, CORR para correlación, SSE para suma de errores cuadráticos, entre otras.

Por tanto, del conjunto de las 2 355 mediciones de ruido, se tomaron aleatoriamente 1 871 (aprox. 80%) como conjunto de entrenamiento, y las 484 restantes (aprox. 20%) para el conjunto de pruebas. Estos conjuntos se mantuvieron constantes para evaluar a las SVM y a las ANN. Al finalizar el entrenamiento de cada modelo de ML, utilizando el conjunto de entrenamiento (y bajo cada combinación de los distintos parámetros) y después de predecir con el modelo entrenado y el conjunto de pruebas los niveles de ruido, fue posible comparar los niveles de ruido reales (previamente conocidos) contra los nuevos niveles de ruido (recién predichos por el modelo). Con las diferencias entre ambos valores se calculó el MAE y la CORR. Con estos resultados se determinó la exactitud de cada modelo, métrica necesaria para compararlos entre ellos, considerando como más exacto al modelo con menor MAE y mayor CORR. Para los radios de las vecindades geográficas se evaluaron diversas longitudes, y se utilizó al final un radio de 200 m para el de los negocios del DENUÉ, 1 000 m para el de las colonias aledañas, y 1 000 m para el de las mediciones de ruido cercanas.

Desempeño de las redes neuronales artificiales

Para evaluar las ANN se utilizó la función *neuralnet* de la librería *neuralnet* del lenguaje R. Con ello, se consideraron diferentes cantidades de nodos en las capas ocultas, y utilizando internamente el algoritmo por defecto, *rprop+*. Así, se construyó una ANN con 0 nodos ocultos (ANN 0), otras ANN con una capa oculta de 1, 2, 3, 4, 5 y 10 nodos, así como una ANN con dos capas ocultas de 5 nodos en cada una (ANN 5-5), otra ANN con dos capas ocultas de 10 nodos en cada una (ANN 10-10), y finalmente una ANN con tres capas ocultas de 5 nodos en cada una (ANN 5-5-5).

Utilizando una escala logarítmica, en la Figura 13 se muestra el desempeño de estas distintas configuraciones de ANN. En ellas se aprecia el aumento de CORR, respecto a otras ANN del mismo número de capas ocultas a medida que aumenta el número de nodos entre ellas. El MAE y el SSE se comportan de manera similar, ya que el SSE sólo amplifica el efecto del MAE. El tiempo para entrenar el modelo (en segundos) depende directamente de la cantidad de pasos efectuados por el algoritmo, y en general crece exponencialmente a medida que se aumentan los nodos en las capas ocultas. Los mejores resultados en cuanto a exactitud en la predicción se obtuvieron con una ANN 10-10, si bien su tiempo de entrenamiento es de aproximadamente 3.3 minutos, es considerablemente alto, en comparación por ejemplo, con una ANN 3 que tardó aproximadamente 1.2 minutos en entrenarse y presentó un desempeño promedio. En cuanto a tiempo de entrenamiento, los peores resulta-

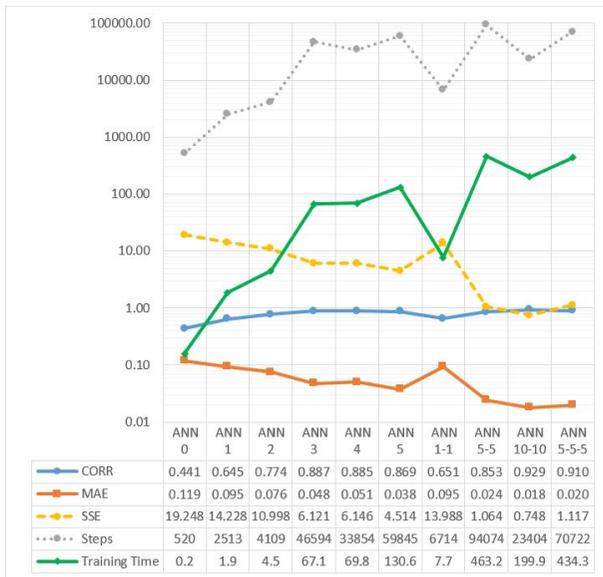


Figura 13. Desempeño de diferentes configuraciones de ANN (precisión, tiempo de entrenamiento, entre otros).

dos se obtuvieron con una ANN 5-5, con aproximadamente 7.7 minutos, si bien ésta fue una de las mejores ANN respecto a la exactitud en la predicción.

Desempeño de las máquinas de soporte vectorial

Para evaluar el desempeño de las SVM se utilizó la función `ksvm` de la librería `kernlab` del lenguaje R, considerando diferentes funciones kernel y parámetros. Los kernels evaluados fueron el de base radial o Gaussiano (`rbfdot`), el Laplaciano (`laplacedot`), el Bessel (`besseldot`), el Anova RBF (`anovadot`), el polinomial (`polydot`), el lineal (`vanilladot`) y el de tangente hiperbólica (`tanhdot`). Por cada kernel empleado, se evaluó el parámetro C (para el costo de violación de las restricciones), en un rango de 0.1 a 3.1 con intervalos de 0.5. Por cada kernel y cada valor de C, también se evaluó el parámetro ξ (utilizado en la función de regresión de insensibilidad a la pérdida), en un rango de 0.1 a 2.6 con intervalos de 0.5. Bajo estas configuraciones, se obtuvieron 294 resultados. El parámetro k (para realizar k-fold cross validation con el conjunto de entrenamiento mejorando la calidad del modelo) se mantuvo constante con el valor de 10, por ser el comúnmente utilizado.

Una comparación de desempeño de los diferentes kernels, considerando el MAE mínimo, la CORR máxima, y el tiempo de entrenamiento mínimo alcanzados (en segundos), se muestra en la Figura 14. Respecto al MAE y a la CORR, con el kernel `rbfdot` se obtuvieron los mejores resultados, seguido del kernel `laplacedot`,

mientras que el peor desempeño correspondió a los kernels polydot (de segundo grado), al kernel tanhdot y al kernel vanilladot. Con respecto al tiempo de entrenamiento mínimo, el mejor desempeño correspondió también al kernel rbfdot, con casi 6 segundos, seguido de los kernels vanilladot y laplacedot, con 8 segundos aproximadamente, finalmente el kernel polydot obtuvo el peor desempeño, con casi 3.8 minutos, incluso cuando estaba configurado con grado 2, mientras que grados mayores (no reportados aquí) para este mismo kernel rebasaban por mucho este tiempo.

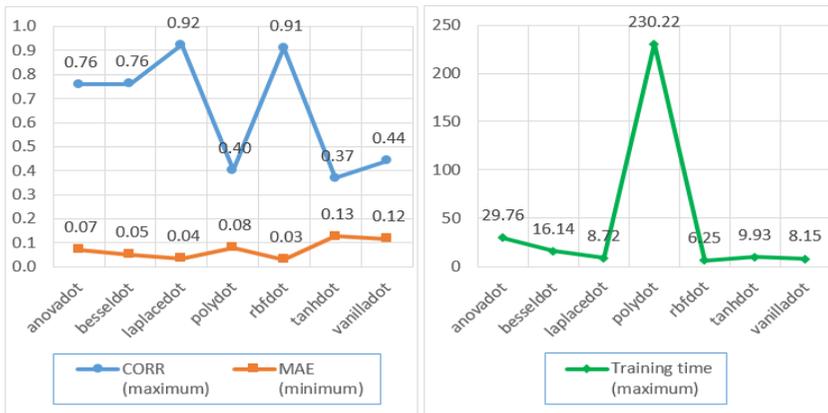


Figura 14. Desempeño de diferentes SVM: precisión (izquierda) y tiempo de entrenamiento (derecha).

Mapas de ruido ambiental de las áreas de interés

Como resultado, se desarrolló un mapa de ruido utilizando QGIS, basado en las mediciones obtenidas en la Delegación Cuauhtémoc, el cual se muestra en la Figura 15 (izquierda). El mapa se obtuvo utilizando el método de interpolación IDW con el parámetro $C=2$, y el código de colores definido en la norma ISO 1996-2:1987. De manera inmediata puede apreciarse que los niveles más altos de ruido registrados durante los recorridos se ubican en la zona comprendida por las inmediaciones del Centro Histórico. Con el mismo método de interpolación y el mismo código de colores, se obtuvieron otros mapas de ruido, usando las mediciones obtenidas en el Centro Histórico para los diferentes intervalos de tiempo, mismos que se muestran en la Figura 15 (derecha).

Con respecto a la Delegación Cuauhtémoc, la mayor concentración de ruido durante estos recorridos fue encontrada en la colonia Centro (casi en su totalidad) y la Colonia Morelos (en las inmediaciones del mercado de la Merced), tal como se

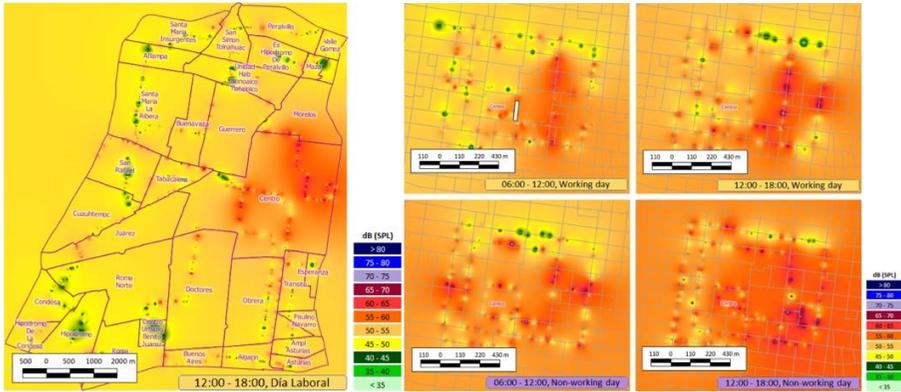


Figura 15. Mapa de ruido ambiental para las colonias de la Delegación Cuauhtémoc (izquierda) y para las calles alrededor del Centro Histórico (derecha).

aprecia en la Figura 15 (izquierda). Más aún, para la Colonia Centro particularmente, la Figura 15 (derecha) muestra que la mayor concentración de ruido se encontró en calles aledañas al Palacio Nacional (calles con alta densidad de vendedores ambulantes, tales como Corregidora, Correo Mayor, La Academia, Soledad y Venustiano Carranza), y también muestra que, en general, los mayores niveles de ruido registrados se encontraron en el horario H-12 para días no laborables, mientras que los niveles más bajos se encontraron en el horario H-06 para días laborables.

En general se determinó que los niveles más altos de ruido promedio eran provocados por el ruido constante de la gente (por ejemplo, los gritos de los vendedores ambulantes o inclusive la charla continua de las multitudes), seguidos por el ruido provocado por los automóviles. Las zonas industriales, como la colonia Atlampla, al contrario de lo que se esperaría, resultaron estar entre las más silenciosas y tranquilas, probablemente debido a que las industrias aíslan el ruido en su interior y en sus calles no transitan demasiados autos ni personas. Las zonas con los niveles de ruido más bajos fueron los interiores de grandes parques y en general, de zonas rodeadas de árboles, como la colonia Hipódromo y Tlatelolco. Algunos de los ruidos más intensos detectados con la aplicación, fueron sonidos impulsivos, tales como estornudos, el grito de un vendedor de gas, la sirena de una patrulla o el ladrido de los perros.

Después de predecir los niveles de ruido en diversos puntos utilizando aprendizaje automático y luego de aplicar el método de interpolación con el algoritmo *Barnes* usando *GeoServer*, fue posible obtener el mapa de ruido, con mediciones reales y valores predichos, para la Delegación Cuauhtémoc. La Figura 16 muestra una comparación entre el mapa de ruido obtenido por interpolación con Quantum GIS (izquierda) sobre el que se encuentran los puntos predichos por el modelo de ML, contra el mapa

de ruido obtenido por interpolación con *GeoServer* (derecha) para el horario H-12 en días laborables. También es posible comparar los mapas de ruido ambiental previamente mostrados contra las observaciones de ruido obtenidas mediante el proyecto *NoiseWatch*, aunque estas últimas presentan una escala categórica subjetiva diferente, que incluyen categorías de nivel de ruido muy malo, malo, moderado, bueno y muy bueno. Las mediciones de *NoiseWatch* muestran buena correlación con las de los mapas de ruido ya mostrados.

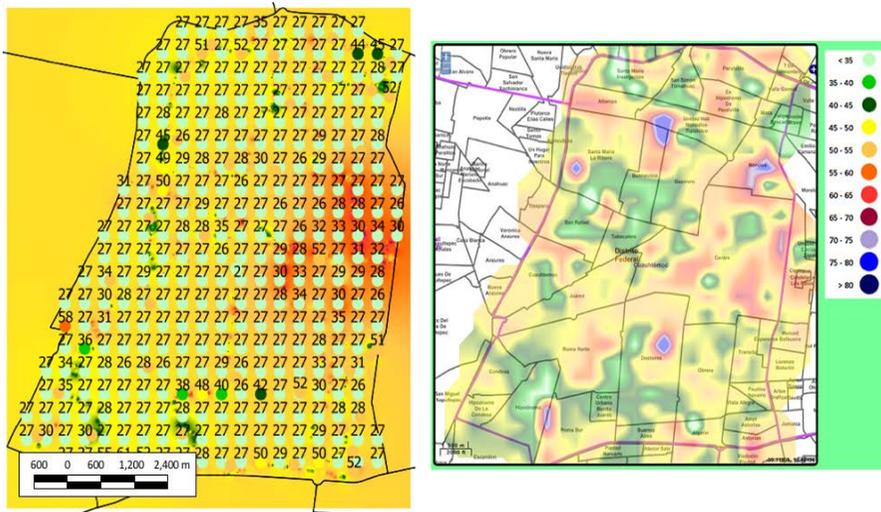


Figura 16. Mapa de ruido generado con mediciones reales sobre el que se sobreponen valores predichos (izquierda) comparado contra el mapa de ruido generado con valores reales y predichos (derecha).

Consideraciones acerca de los resultados obtenidos

Las mediciones de ruido se realizaron con un dispositivo móvil ZTE-V791. Las posiciones geográficas se obtuvieron principalmente utilizando el GPS del dispositivo, sólo una muy pequeña cantidad de ellas se obtuvo mediante posicionamiento por Wi-Fi. Dado que las mediciones corresponden a nivel de presión sonora, para obtenerlas el usuario debía permanecer inmóvil durante los 10 segundos que duraba cada captura. Se evitaba estar en movimiento durante la medición para evitar que el aire golpeando al micrófono provocara una mayor presión y en consecuencia una lectura más elevada del nivel de ruido que la realmente existente.

Además, se procuró ser imparcial al tomar las mediciones, intentando capturar el ruido a intervalos regulares (por ejemplo, cada que se hubiera recorrido caminando

media cuadra) independientemente del nivel de ruido que se percibía subjetivamente en ese lugar y buscando que las condiciones climáticas fueran favorables.

Conclusiones y trabajo a futuro

En el enfoque expuesto en este artículo, los ciudadanos juegan un rol importante al convertirse de simples usuarios a productores de información, representando una gran ventaja sobre enfoques tradicionales, especialmente en términos de velocidad, bajo costo y volumen de datos que pueden ser obtenidos. El principal beneficio de la creación de este tipo de herramientas y mecanismos, es que nos permiten ayudar a informar y concientizar a la población y a las autoridades acerca de los altos niveles de ruido a los que la sociedad se encuentra expuesta, en un esfuerzo para detectar y erradicar, o al menos disminuir, las consecuencias de este tipo de problemas. El trabajo aquí presentado ofrece una de esas posibilidades y tiene algunos beneficios, tales como el uso de la VGI. Asimismo, este trabajo representa un enfoque para construir big data desde una perspectiva colaborativa y soportar aplicaciones de Internet de las Cosas (IoT), así como de ciudades inteligentes (smart cities).

En este trabajo encontramos que, aprovechando la VGI, a partir de un conjunto de mediciones de ruido ambiental georeferenciadas, obtenidas voluntariamente por las personas con sus propios dispositivos móviles, se pudieron crear mapas de ruido ambiental a partir del análisis de dichas mediciones, el análisis geoespacial y la predicción con métodos de ML. El estudio de trabajos relacionados con el monitoreo de ruido ambiental nos permitió definir una metodología así como efectuar el diseño y desarrollo del software para cubrir cada fase de la misma, al tiempo que el análisis de un caso de estudio nos permitió experimentar con los datos y las aplicaciones desarrolladas. Los experimentos de comparación de las predicciones obtenidas con SVM contra las obtenidas con ANN mostraron que, a pesar de que ANN obtuvo mayor correlación y menor error absoluto promedio, su tiempo de entrenamiento fue mucho mayor que el de las SVM. Preferimos utilizar las SVM debido a que su correlación y error absoluto promedio son aceptables, y a que su tiempo de entrenamiento es mucho menor que el de las ANN, lo que permite efectuar predicciones más rápidas.

Como trabajos futuros se buscará probar esta metodología con otros modelos, así como con otros dispositivos móviles que cuenten con métodos de calibración del sensor del micrófono. Asimismo, es necesario incorporar un módulo de calibración que permita realizar ajustes sobre los sensores. De igual manera, sería deseable enfocar el trabajo a una red de sensores sonoros que permitan monitorear mayor cantidad de datos para refinar el modelo de predicción. Por otra parte, podría ser de utilidad incluir un método de ponderación multicriterio para asignar pesos a las variables involucradas, como es el caso de CRITIC (Diakoulaki *et al.*, 2015). Sin embargo, hay que considerar que en un enfoque de aprendizaje de máquina la asig-

nación de pesos es manejada por medio de los kernels para desambiguar los valores predichos.

Agradecimientos

Agradecemos el apoyo que este trabajo recibió del Instituto Politécnico Nacional (IPN), el Consejo Nacional de Ciencia y Tecnología (CONACYT) y a la Secretaría de Investigación y Posgrado (SIP) con los proyectos SIP 20171192 y 20171918.

Bibliografía

- Antón, D. y Antón, A., (2012). *Ecología y Medio Ambiente*, ST Editorial, México, pp. 99-115.
- Bilandzic, M.; Banholzer, M.; Peev, D.; Georgiev, V.; Balagtas-Fernandez, F. y De Luca, A., (2008). “Laermometer: a mobile noise mapping application”, *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges*, ACM, New York, United States of America, pp. 415-418.
- Connors, J.P.; Lei, S. y Kelly, M., (2012). “Citizen science in the age of neogeography: Utilizing volunteered geographic information for environmental monitoring”, *Annals of the Association of American Geographers*, 102 (6), Taylor & Francis, United Kingdom, pp. 1267-1289.
- Diakoulaki, D.; Mavrotas, G. y Papayannakis, L., (1995). “Determining objective weights in multiple criteria problems: The critic method”, *Computers & Operations Research*, 22(7), Elsevier, United Kingdom, pp. 763-770.
- Elwood, S.; Goodchild, M.F. and Sui, D.Z., (2012). “Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice”, *Annals of the Association of American Geographers*, 102 (3), Taylor & Francis, United Kingdom, pp. 571-590.
- Fonte, C.; Bastin, L.; Foody, G.; Kellenberger, T.; Kerle, N.; Mooney, P.; Olteanu-Raimond, A.M. y See, L., (2015). “VGI Quality Control”, *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1, Elsevier, United States of America, pp. 317-324.
- García-Martí, I.; Rodríguez-Pupo, L.; Díaz, L. y Huerta, J. (2013). “Noise Battle: A Gamified application for Environmental Noise Monitoring in Urban Areas”, *Proceedings of the 16th AGILE Conference on Geographic Information Science*, Leuven, Belgium, pp. 14-17.
- Goodchild, M.F., (2007). “Citizens as Sensors: The World of Volunteered Geography”, *Geo Journal*, 69(4), Springer, Netherlands, pp. 211-221.
- , (2009). “Geographic Information Systems and Science: Today and Tomorrow”, *Annals of GIS*, 15(1), Taylor & Francis, United Kingdom, pp. 3-9.

- Goodchild, M.F. y Glennon, J.A., (2010). "Crowdsourcing Geographic Information for Disaster Response: A Research Frontier, *International Journal of Digital Earth*, 3 (3), Taylor & Francis, United Kingdom, pp. 231-241.
- Halonen, J.I.; Hansell, A.L.; Gulliver, J.; Morley, D.; Blangiardo, M.; Fecht, D.; Toledano, M.; Beever, B.; Sean, D.; Anderson, H.; Hugh, R. y Ross, K.F., (2015). "Road traffic noise is associated with increased cardiovascular morbidity and mortality and all-cause mortality in London", *European Heart Journal*, 36(39), Eur Soc Cardiology, United Kingdom, pp. 2653-2661.
- Ising, H. y Kruppa, B., (2004). "Health effects caused by noise: evidence in the literature from the past 25 years", *Noise and Health*, 6 (22), Medknow Publications, India, p. 5.
- Jiang, B. y Thill, J.C., (2015). "Volunteered Geographic Information: Towards the establishment of a new paradigm", *Computers, Environment and Urban Systems*, 53, Elsevier, United Kingdom, pp. 1-3.
- Jirka, S.; Bröring, A.; Kjeld, P.; Maidens, J. y Wytzisk, A., (2012). "A lightweight approach for the sensor observation service to share environmental data across Europe", *Transactions in GIS*, 16 (3), Wiley Online Library, United Kingdom, pp. 293-312.
- Kaiser, C. y Pozdnoukhov, A., (2013). "Enabling Real-Time City Sensing with Kernel Stream Oracles and MapReduce", *Pervasive and Mobile Computing*, 9 (5), Elsevier, Netherlands, pp. 708-721.
- Kanjo, E., (2010). "Noisespy: A real-time mobile phone platform for urban noise monitoring and mapping", *Mobile Networks and Applications*, 15 (4), Springer, Netherlands, pp. 562-574.
- Karolus, J., (2013). "Creating Noise Pollution Maps Based on User-Generated Noise Measurements: Erstellung Von Lärmkarten Basierend Auf Nutzergenerierten Lautstärkemessungen", *Ph.D. Thesis*, Telecooperation Group, TU Darmstadt.
- Liu, T.; Zheng, Y.; Liu, L.; Liu, Y. y Zhu, Y., (2014). "Methods for Sensing Urban Noises", *Technical Report*, MSR-TR-2014-66, Shanghai, China.
- Maisonneuve, N.; Stevens, M.; Niessen, M.E. y Steels, L., (2009). "NoiseTube: Measuring and mapping noise pollution with mobile phones", *Information Technologies in Environmental Engineering, Proceedings of the 4th International ICSC Symposium*, Springer, Thessaloniki, Greece, pp. 215-228. Middleton, S.; Middleton, L. y Modafferi, S. (2014). "Real-Time Crisis Mapping of Natural Disasters Using Social Media", *IEEE Intelligent Systems*, 29 (2), IEEE, United States, pp. 9-17.
- Nakano, N., Tratz-Ryan, B. y Freeman, D. (2015). "Predicts 2016: IoT Technologies Optimize Smart City Operations", *Gartner Report*, United States, pp. 1-10.

- Parker, C.J.; May, A. and Mitchell, V. (2013). "The role of VGI and PGI in supporting outdoor activities", *Applied Ergonomics*, Elsevier, United Kingdom, 44 (6), pp. 886-894.
- Quintero, G., (2013). "Sistema móvil georeferenciado para la medición y análisis de ruido ambiental", Ph.D. Thesis, Centro de Investigación en Computación, IPN, México.
- Rana, R.K.; Chou, C.T.; Kanhere, S.S.; Bulusu, N. y Hu, W., (2010). "Ear-Phone: An End-to-End Participatory Urban Noise Mapping System", *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, ACM, Stockholm, Sweden, pp. 105-116.
- Rodríguez Manzo, F.E. y Garay Vargas, E., (2012). "El mapa de ruido de la Ciudad de México", *Reporte Técnico*, Universidad Autónoma Metropolitana, pp. 45-68.
- Schweizer, I.; Bärtl, R.; Schulz, A.; Probst, F. y Mühläuser, M. (2011). "NoiseMap-Real-Time Participatory Noise Maps", *Proceedings of the 2nd International Workshop on Sensing Applications on Mobile Phones*, ACM, United States, pp. 1-5.
- Weymouth, G.; Mills, G.; Jones, D.; Ebert, E. y Manton, M., (1999). "A continental-scale daily rainfall analysis system", *Australian Meteorological Magazine*, 48 (3), Sydney, Australia, pp. 169-179.
- Xia, Y.; Winterhalter, M. y Fabian, P., (1999). "A model to interpolate monthly mean climatological data at Bavarian forest climate stations", *Theoretical and applied climatology*, 64 (1-2), Springer, Germany, pp. 27-38.
- Zheng, Y.; Capra, L.; Wolfson, O. y Yang, H. (2014a). "Urban Computing: concepts, methodologies, and applications", *ACM Transaction on Intelligent Systems and Technology*, 5 (3), ACM, United States, pp. 1-55.
- Zheng, Y.; Liu, T.; Wang, Y.; Zhu, Y. y Chang, E. (2014b). "Diagnosing New York City's Noises with Ubiquitous Data", en *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, Seattle, United States, pp. 715-725.