

# Modelo de Análisis para el Dominio de Tránsito Ciclista

Leo Ordinez<sup>1</sup>, Damián Barry<sup>1</sup>, Alex Torrico<sup>1</sup>, Emanuel Mallon<sup>1</sup>, José Luis Devia<sup>1</sup>, Lucas Abella<sup>1</sup>, Nahuel Alamo<sup>1</sup>

<sup>1</sup>LINVI, Departamento de Informática, Facultad de Ingeniería, UNPSJB, 9120  
Puerto Madryn, Argentina,  
leo.ordinez@gmail.com, damian\_barry@unpata.edu.ar,  
<http://madryn.unp.edu.ar>

**Resumen** En este trabajo se presenta un modelo del dominio de aplicación enmarcado en el análisis del tránsito ciclista en una ciudad mediana, como la de Puerto Madryn. El objetivo del modelo es construir conocimiento que permita determinar los trayectos más adecuados para emplazar en esa ciudad ciclovías. El núcleo de la propuesta es el análisis de información de *recorridos*, que permita sintetizar *trayectos* con sus respectivas longitudes, frecuencias, densidades y ponderaciones. Además del modelo teórico, se presenta una herramienta de análisis de datos capturados (recorridos realizados por los ciclistas) y un generador de recorridos sintéticos.

**Keywords:** big data analytics, datos espacio-temporales, ciclovías, urbanismo, sistemas colaborativos

## 1. Introducción

Actualmente, muchos gobiernos municipales no cuentan con información de calidad, ni mecanismos de adquisición continua de datos, que permitan un monitoreo permanente del estado de los sistemas que hacen al funcionamiento global de la ciudad. La propuesta aquí presentada apunta a estudiar, experimentar y optimizar una infraestructura de Big Data enfocada en información heterogénea fuertemente ligada a datos espaciales y temporales, de un dominio particular como es el tránsito ciclista.

En línea con muchas de las principales ciudades del país (y del mundo), la localidad de Puerto Madryn tiene en carpeta la instalación de ciclovías que garanticen tanto el orden del tráfico en la ciudad, como la seguridad de los ciclistas. Este proyecto también persigue otro objetivo implícito que es el promover el uso de la bicicleta como medio de transporte, mejorando la calidad de vida de las personas y aportando a la ciudad en aspectos tales como: el tráfico, la seguridad, el medio ambiente, el turismo, etc.

El propósito del presente trabajo es la determinación y visualización de las alternativas más convenientes para la instalación de ciclovías en la ciudad de

Puerto Madryn, considerando los lugares con mayor afluencia de personas (centralidades), las características de las calles y los recorridos actuales con mayor tránsito ciclista.

La propuesta incluye la generación artificial de datos. En particular, aquellos referidos a recorridos realizados por ciclistas en la ciudad, serán generados en forma sintética. Asimismo, se incluye el diseño de las estrategias de elección y ponderación de los trayectos más convenientes para la instalación de ciclovías. Por otro lado, se contempla la visualización de los trayectos más convenientes en un mapa y con una escala de ponderación clara.

Aunque la problemática no es nueva y sobre todo ha cobrado gran relevancia en los últimos años, en general los trabajos científicos se enfocan en la aplicación de técnicas de optimización matemática y se concentran en grandes ciudades con sistemas de transporte multimodal. En [1] se presenta una metodología para modelar simultáneamente los modos de transporte de automóviles, autobuses y bicicletas, considerando las interacciones entre los tres modos a través del modelado de la división modal y la asignación de red de los diferentes recorridos de cada modo. Posteriormente, este modelo se utiliza para optimizar el diseño de la red de caminos ciclistas que obtengan un sistema de transporte eficiente y sostenible. En [2], aunque el objetivo es el análisis de estaciones para un sistema de bicicletas compartidas, se expone un modelo detallado útil para el modelado de dominio de tránsito ciclista. Por otro lado, en [3] se presenta un mapa que ilustra, por primera vez, el flujo real de ciclistas casuales y mensajeros de bicicletas en la ciudad de Madrid. Este mapa fue obtenido a partir del desarrollo y los resultados de la iniciativa Madrid Cycle Track, una plataforma en línea lanzada con el objetivo de recoger rutas ciclistas y otra información de voluntarios. Esta iniciativa es complementaria al trabajo aquí propuesto y se está en fase de desarrollo para lanzarla en la ciudad de Puerto Madryn. En relación a esto, en [4] se realiza un análisis para determinar si los datos de crowdsourcing de aplicaciones deportivas pueden usarse para cuantificar y mapear la variación espacial y temporal del número de ciclistas. Por otro lado, diversos autores han desarrollado modelos para describir el comportamiento de los ciclistas, ante la escasez de datos de los mismos [5], [6], [7], [8].

En particular, el presente trabajo propone un modelo de dominio específico y una aplicación prototipo que implementa dicho modelo. El modelo considera diferentes fuentes de información (recorridos, centralidades y calles) de una ciudad, para realizar un análisis integrado de las mismas. El objetivo principal del análisis es la determinación de los segmentos de calles (trayectos) más convenientes para instalar ciclovías, lo cual implica responder a otras preguntas previamente como son los trayectos con mayor tránsito ciclista, los que tienen menor, entre otras.

El resto del trabajo se organiza de la siguiente manera: en la sección 2 se presenta el modelo de dominio para el tránsito ciclista; en la sección 3 se expone el núcleo de análisis, el cual es la ponderación de trayectos; en la sección 4 muestra cómo la abstracción planteada en las secciones anteriores es utilizada

en la construcción de conocimiento útil para la toma de decisiones; finalmente, la sección 5 expone las conclusiones y trabajos futuros.

## 2. Modelado del Dominio

Se parte de que la ciudad se encuentra representada por una malla de puntos  $p$ , donde cada punto es una intersección de dos calles (esquina). Aquellos cruces de más de dos calles, se representan considerando las combinaciones en forma de pares. La unión entre dos puntos consecutivos se denomina *cuadra*  $b_{ab} = p_a \rightarrow p_b$ . Nótese que de lo anterior se desprende que  $b_{ab} \neq b_{ba}$ , lo cual implica indicar que las *cuadras* tienen sentido de circulación.

Se supone que un *ciclista*  $c_i$  realiza  $n$  recorridos  $r_{ij}$ , con  $j = 1 \dots n$ . A la vez, cada recorrido  $r_{ij}$  se *normaliza*, para adecuarlo a la representación de puntos, y cuadras en la malla de la ciudad (*i.e.*, el recorrido más corto que podría realizar un ciclista es de una cuadra, el cual se representa por dos puntos consecutivos). El recorrido  $r_{ij}$  normalizado se denota  $R_{ij}$ . En esta misma línea, el conjunto<sup>1</sup> de recorridos  $\mathcal{R}_i$  que realiza un ciclista a lo largo de su vida, se expresa como

$$\mathcal{R}_i = \sum_{j=1}^n R_{ij}$$

De igual modo, se construye el conjunto universal de recorridos  $\mathbf{R}$  de todos los ciclistas

$$\mathbf{R} = \prod_{i=1}^n \mathcal{R}_i$$

Por otro lado, se supone que cada recorrido  $R_{ij}$  se compone de *tramos*, que son conjuntos ordenados de cuadras, siendo  $t_{ij}^k$  el  $k$ -ésimo *tramo* del  $j$ -ésimo *recorrido* del  $i$ -ésimo *ciclista*. De este modo, cada recorrido se puede expresar como

$$R_{ij} = \sum_{k=1}^n t_{ij}^k$$

Así, cada tramo es un conjunto ordenado de cuadras consecutivas

$$t_{ij}^k = \sum_{b=1}^m b_{ab}, \quad m \leq n$$

De lo anterior, surge que cada tramo tiene un sentido de circulación.

A partir de los tramos que componen los recorridos normalizados de los ciclistas, se construye la entidad *trayecto*, que se representa con  $\tau$ , donde

$$\tau_{uv} = t_u \cap t_v, \quad \forall t \in \mathbf{R}$$

<sup>1</sup> En adelante se asume la siguiente convención para conjuntos  $\sum$  representa un conjunto ordenado temporalmente (secuencia); y  $\prod$  representa un conjunto que no necesariamente respeta un orden.

De la definición de trayecto surge que éstos son las intersecciones de dos tramos (*i.e.*, las cuadras compartidas por ambos). Los trayectos podrían tener un sentido de circulación, sin embargo para los fines de este trabajo no se considera esa característica.

### 2.1. Medidas de Análisis de Trayectos

A continuación se definen una serie de características de los trayectos, las cuales se utilizarán como base para la elaboración de los algoritmos de búsqueda y análisis de los recorridos, que permitan determinar los lugares más adecuados para instalar ciclovías.

**Longitud:** Se define la *longitud* de un trayecto  $\tau_{uv}$  como  $long(\tau_{uv}) = |\tau_{uv}| = |t_u \cap t_v|$ .

**Frecuencia:** Se define la *frecuencia* de un trayecto  $\tau_{uv}$ , simbolizada como  $frec(\tau_{uv})$ , como la cantidad veces que se ha realizado ese trayecto, independientemente de si lo ha realizado uno o varios ciclistas. Esto es,  $frec(\tau_{uv}) = \omega$

**Densidad:** Se define la *densidad* de un trayecto  $\tau_{uv}$ , simbolizada como  $dens(\tau_{uv})$ , como la cantidad de ciclistas diferentes que realizaron ese trayecto. Esto es,  $dens(\tau_{uv}) = \gamma$

**Similitud:** Se dice que dos recorridos  $R_i$  y  $R_j$  son *similares* si existe un trayecto construido a partir de ambos, cuya longitud es mayor a un cierto porcentaje del menor. Esto es,  $\exists \tau_{uv} : \tau_{uv} = R_i \cap R_j \wedge long(\tau_{uv}) \geq \mu \cdot \min\{|R_i|, |R_j|\} \wedge \mu \in (0, 1)$

## 3. Análisis de Trayectos

Con el objetivo de determinar los lugares más aptos para la instalación de ciclovías, se busca obtener el conjunto de trayectos más transitados. A continuación se describe el algoritmo desarrollado. El resultado de los trayectos más transitados es fundamental para cualquier análisis que se haga. No obstante, éste no es único ni determinante, ya que se deben considerar otros factores, como por ejemplo las características de las calles (pendiente, ancho, material, etc.).

### Trayectos más frecuentes

Dado un conjunto de recorridos normalizados, ya sea obtenido desde la base de datos o de un conjunto preseleccionado, este algoritmo se encarga de realizar el análisis correspondiente para determinar los trayectos similares entre todo el conjunto y ponderar cada uno de los mismos, asignándole un peso mediante una fórmula determinada.

**Factores de peso** Para determinar el peso de cada trayecto, se tiene en cuenta la *frecuencia* y la *densidad*.

A partir de estos datos se determinan otros dos factores que influyen en el peso del trayecto ( $P_t$ ), como lo son:

1 - *Factor de frecuencia del trayecto* ( $F_{frec()}$ ): se calcula teniendo en cuenta:

- La frecuencia actual ( $F_{act}$ ): la cantidad de veces que se realizó el trayecto
- La frecuencia mínima ( $F_{mín}$ ): la frecuencia mínima del conjunto de recorridos analizado
- La frecuencia máxima ( $F_{máx}$ ): la frecuencia máxima del conjunto de recorridos analizado
- El rango de frecuencia ( $\Delta(F)$ ): la diferencia entre la frecuencia máxima y mínima ( $F_{máx} - F_{mín}$ )

$$F_{frec()} = \frac{(F_{act} - F_{mín})}{\Delta(F)}$$

con  $F_{act} - F_{mín} > 0$ , dado que los trayectos analizados son realizados por lo menos una vez.

2 - *Factor de densidad del trayecto* ( $F_{dens()}$ ): se obtiene a partir de:

- La densidad del trayecto ( $dens()$ ): el número de usuarios distintos que realizan el trayecto.
- La densidad total ( $dens_t()$ ): la cantidad de usuarios distintos del conjunto de recorridos analizado.

$$F_{dens()} = \frac{dens()}{dens_t()}$$

con  $dens_t() > 0$ , dado que un trayecto pertenece como mínimo a un usuario.

3 - *Factores de Importancia*: A fin de ponderar el trayecto, se establecen dos factores que pesan las medidas anteriores, ya que éstos no tienen la misma importancia. Se denominan  $I_{frec}$  e  $I_{dens}$ , haciendo referencia al Factor de frecuencia  $F_{frec()}$  y al Factor de densidad  $F_{dens()}$ , respectivamente. En general,

$$I_{frec} + I_{dens} = 1$$

**Fórmula de ponderación** De acuerdo a los factores descritos anteriormente, a continuación se presenta la descripción de la fórmula de ponderación utilizada:

$$P_t = (I_{frec} \cdot F_{frec()}) + (I_{dens} \cdot F_{dens()})$$

con  $0,1 \geq P_t \geq 1$ , donde  $I_{frec} = 0,7$  e  $I_{dens} = 0,3$ , puesto que se considera que  $F_{frec()}$  es más determinante que  $F_{dens()}$ .

## 4. Construcción de Conocimiento

La construcción de conocimiento está ligada a la interpretación de hechos de un dominio específico. Así, el análisis de los modelos presentados tiene un correlato en las metas de conocimiento esperadas. En este sentido, el dominio específico se orienta a facilitar la construcción de conocimiento buscando alcanzar las siguientes métras de negocio:

- Ordenamiento urbano, especialmente en lo que se refiere a la movilidad de los habitantes.
- Mejoramiento de la calidad de vida de la población, especialmente en la promoción de mecanismos de vida saludable, fomentando actividades que permitan interactuar con los planes de medicina preventiva.
- Mejorar mecanismos que permitan optimizar el presupuesto de infraestructura urbana relacionada con la movilidad.

A fin de alcanzar estas metas se plantean una serie no exhaustiva de interrogantes, los cuales se pretende responder a través del modelo propuesto. A continuación se enuncian algunos de ellos:

- ¿Cuáles son los trayectos mínimos que tienen mayor cantidad de tráfico ciclista?
- ¿Cuáles son los trayectos máximos que tienen mayor cantidad de tráfico ciclista?
- ¿Cuáles son los trayectos mínimos que tienen menor cantidad de tráfico ciclista?
- ¿Cuáles son los trayectos mínimos que tienen menor cantidad de tráfico ciclista?
- ¿En qué mes del año se producen los recorridos de los trayectos mínimos y máximos?
- ¿En qué días de la semana se realizan los mayores desplazamientos?
- ¿En qué franjas horarias se realizan los mayores desplazamientos?
- ¿Qué zonas de la ciudad (barrios) son los menos transitados por los ciclistas?
- ¿Qué centralidades se encuentran cercanas a los trayectos con mayor cantidad de tráfico?
- ¿Cuáles son los radios mínimos y máximos entre las centralidades halladas y los trayectos de mayor tráfico?

En base al modelo propuesto, se desarrolló un prototipo de aplicación de simulación y análisis de recorridos ciclistas. El simulador genera conjuntos sintéticos de recorridos y el analizador aplica el modelo planteado anteriormente, a fin de que los tomadores de decisión puedan contar con la información suficiente para determinar los lugares por los cuales instalar las ciclovías.

### 4.1. Arquitectura

En la Figura 1 se muestra la arquitectura completa de la aplicación desarrollada. El diseño se basó en capas. En particular, se dividió la arquitectura

en cuatro capas. Esto se debió a cuestiones de diseño y a imposiciones de la tecnología utilizada en algunas capas, como Laravel.

1. Capa de Presentación
2. Capa de Servicio
3. Capa de Datos
4. Capa de Base de Datos.

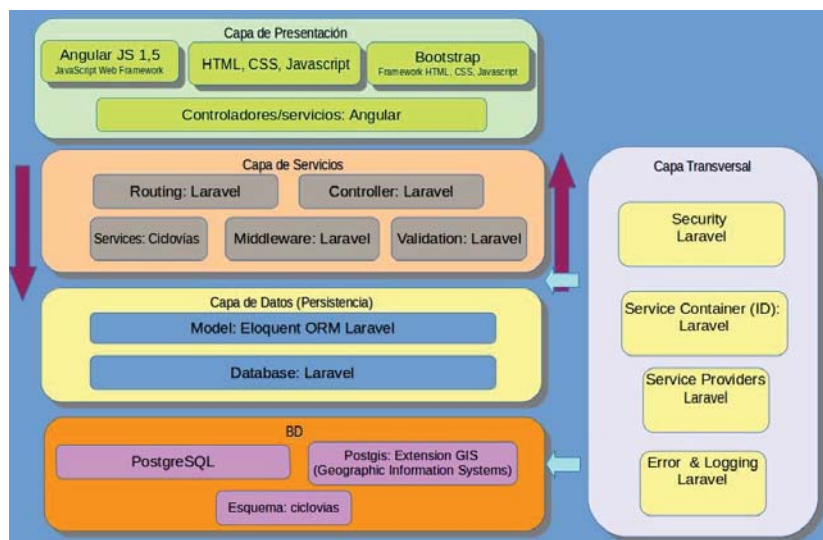


Figura 1. Arquitectura completa de la aplicación.

En la Figura 2 se muestra el modelo de datos de la aplicación, con las principales entidades involucradas.

- **Centralities** (Centralidades): muestra los datos referidos a Centralidades, que son puntos de interés para los ciclistas. En ellos se destacan establecimientos educativos, farmacias, etc. Los datos que se representan, son el nombre, dirección, ubicación geográfica.
- **Zones** (Zonas): muestra los datos referidos a las Zonas en las que está delimitada la ciudad. Los datos que se representan, son el nombre, dirección, puntos que delimitan la extensión de una zona.
- **Roads** (Calles): muestra los datos referidos a las calles de la ciudad. Los datos que se representan, son el nombre y los puntos de todas las esquinas referidas a esa calle.
- **Trips** (Recorridos): muestra los datos referidos a los Recorridos realizados por los ciclistas en la ciudad. Los datos que se representan, son el nombre, descripción, usuario, tiempo del recorrido, distancia, el recorrido que realiza (Linestring), y una referencia al datalog.

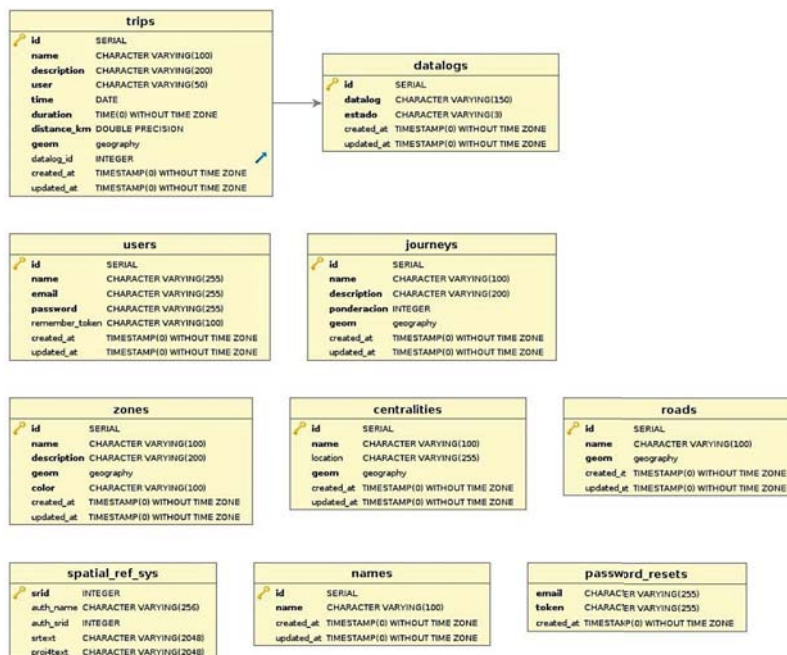


Figura 2. Modelo de Datos.

- **Journeys** (Trayectos): representación de los trayectos en línea recta que unidos forman un Recorrido. Los datos que se representan son el nombre, descripción, usuario, ponderación, la línea del trayecto realizado (Linestring).
- **Names** (Nombres): representación de ciclistas ficticios que realizan Recorrido generados en la ciudad. Representa el nombre.

#### 4.2. Funcionalidades

La aplicación cuenta con características que permiten visualizar los comportamientos presentados en el modelo de análisis, presentado en las secciones 2 y 3. El mismo cuenta con segmentos para:

- Visualizar y analizar recorridos.
- Visualizar y analizar trayectos.
- Visualizar y analizar densidades y ponderaciones:
  - Trayectos similares.
  - Recorridos por Zona.
  - Recorridos cercanos a un punto o centralidad (Figura 3).
  - Recorridos por rangos de distancia (Figura 4).
- Visualizar y analizar recorridos más frecuentes.
- Visualizar y analizar comportamiento de movilidad respecto de centralidades.



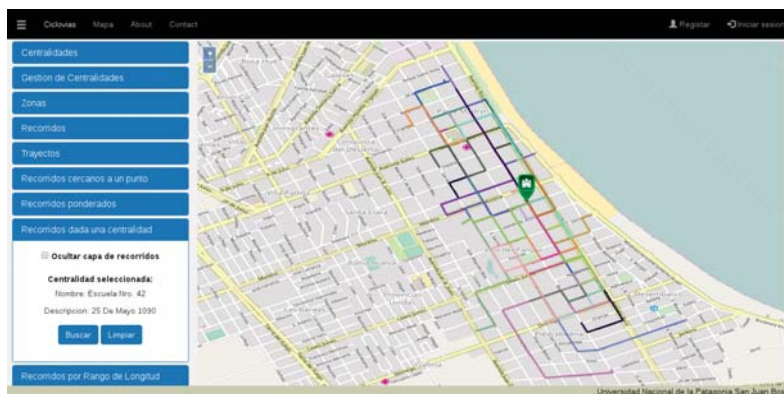


Figura 3. Recorridos dada una centralidad

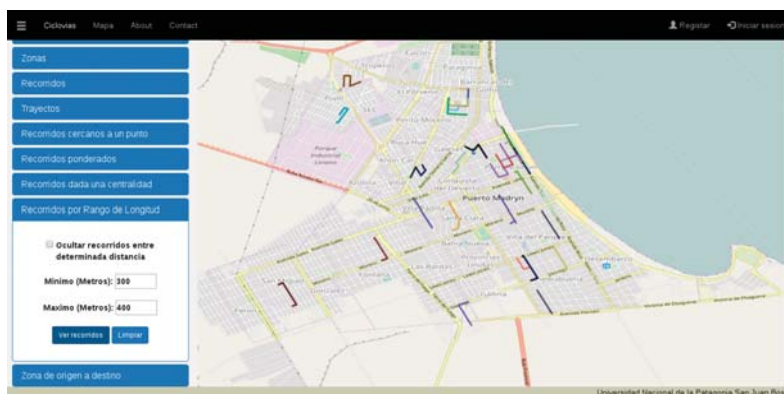


Figura 4. Recorridos por rango de longitud

- Visualizar y analizar las densidades poblacionales y los compromientos según la zona de residencia (Figura 5).

## 5. Conclusiones y Trabajos Futuros

En este trabajo se propuso un modelo para el análisis de un dominio específico, como es el tránsito ciclista en una ciudad pequeña como Puerto Madryn. A la vez, se expuso una aplicación prototipo que implementa dicho modelo. El modelo considera diferentes fuentes de información (recorridos, centralidades y calles) de una ciudad. Al mismo tiempo, en base a los datos obtenidos de esas fuentes construye conceptos nuevos, a fin de determinar los segmentos de calles (trayectos) más convenientes para instalar ciclovías. El principal concepto introducido es el de trayecto, el cual posee características como longitud, frecuencia, densidad y peso.



Figura 5. Zona de origen-destino

En cuanto a trabajos futuros, se continuará profundizando el modelo descrito, a partir de la utilización de grafos en la descripción de algunos aspectos y se avanzará en la construcción de una aplicación para la recolección de recorridos reales. Estos aspectos se utilizarán a la vez, para continuar mejorando la aplicación desarrollada.

## Referencias

1. Gaspar, I., Benavente, J., Bordagaray, M., Alonso, B., Moura, J.L., Ibeas, A.: A bilevel mathematical programming model to optimize the design of cycle paths. In: *Transportation Research Procedia*. (2015)
2. Kim, E.S.: Mathematical model for cost-efficient installation of public transportation system. PhD thesis, University of California at Berkeley (2014)
3. Romanillos, G., Zaltz Austwick, M.: Madrid cycle track: visualizing the cyclable city. *Journal of Maps* (10 2015) 1–9
4. Jestico, B., Nelson, T., Winters, M.: Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography* **52** (4 2016) 90–97
5. Gosse, C., Clarens, A.: Estimating Spatially and Temporally Continuous Bicycle Volumes by Using Sparse Data. *Transportation Research Record: Journal of the Transportation Research Board* **2443** (12 2014) 115–122
6. Broach, J., Dill, J., Gliebe, J.: Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice* **46**(10) (12 2012) 1730–1740
7. Casello, J., Usyukov, V.: Modeling Cyclists' Route Choice Based on GPS Data. *Transportation Research Record: Journal of the Transportation Research Board* **2430** (12 2014) 155–161
8. El Esawey, M.: Estimation of Annual Average Daily Bicycle Traffic with Adjustment Factors. *Transportation Research Record: Journal of the Transportation Research Board* **2443** (12 2014) 106–114