



Metodología para la vinculación de datos de tráfico vehicular con datos meteorológicos utilizando la fórmula de Haversine

Blanca Arenas-Ramírez¹, Edinalva Gomes-Bastos², Orlando Vaca³

¹ ETSI Industriales - Instituto Universitario de Investigación del Automóvil Francisco Aparicio Izquierdo (INSIA), Universidad Politécnica de Madrid, blanca.arenas@upm.es

² Instituto Universitario de Investigación del Automóvil Francisco Aparicio Izquierdo (INSIA), Universidad Politécnica de Madrid, edinalva.gbastos@upm.es

³ Facultad de Ingeniería, Universidad Nacional de Salta (UNSa) - Argentina, vacawo@gmail.com

Este trabajo describe la metodología desarrollada para vincular los datos de velocidad registrados en puntos de aforo ubicados en diversas carreteras, con los datos meteorológicos captados por estaciones meteorológicas cercanas. La metodología de obtención del indicador de velocidad de flujo libre sigue los lineamientos del proyecto europeo Trendline (2022-2025) en el que participa España entre otros países europeos, requiere para su determinación la selección de las mediciones de velocidad en condiciones meteorológicas favorables.

Este trabajo se inscribe en el proyecto de obtención del indicador y otros estadísticos de la velocidad de flujo libre realizado para la Dirección General de Tráfico en 2024.

1. Introducción

Este trabajo describe la metodología desarrollada para vincular los datos de medición de la velocidad (y otras variables como tiempo de paso entre vehículo, longitud de automóviles, etc.) en puntos de aforo ubicados en diversas carreteras, con los datos meteorológicos captados por estaciones meteorológicas en España. Este desarrollo se inscribe en el proyecto de obtención del indicador de velocidad de flujo libre KPI-VFL24 y otros estadísticos de la velocidad para la Dirección General de Tráfico. La metodología de obtención del indicador de velocidad de flujo libre sigue los lineamientos del proyecto europeo *Trendline* en el que participa España entre otros países europeos y requiere para su determinación la selección de las mediciones de velocidad en condiciones meteorológicas favorables, entre otros criterios. Este proyecto establece un procedimiento para que los valores de los indicadores informados por los países europeos sean homologables y comparables.

Las mediciones relativas al tráfico vehicular se realizaron en España durante el año 2023, en una muestra de 100 puntos de la red de carreteras españolas (en adelante se refiere como Datos 1).

Los datos meteorológicos fueron proporcionados por el director del contrato de la Dirección General de Tráfico, como resultado de una consulta de los datos meteorológicos del satélite *Copernicus*, para España en el periodo de noviembre y diciembre de 2023, coincidente con el calendario de mediciones de velocidad de la campaña. Estos registros se refieren como Datos 3.

Las informaciones de ambas fuentes de datos se han almacenado en diferentes entidades o tablas definidas en el diseño de una base de datos MySQL, construida para la creación de las bases de datos necesarias para la obtención del indicador de velocidad de flujo libre (KPI-VFL24) y otros estadísticos de la velocidad para España.

La metodología desarrollada vincula los datos de las dos fuentes de diferente procedencia y métodos de obtención de la información y permite seleccionar las lecturas que cumplen con las condiciones meteorológicas favorables de *Trendline*. La vinculación espacial de sendas estaciones se realizó con la fórmula de Haversine, que es un proceso simple desde el punto de vista computacional, recomendable cuando se trabaja con coordenadas geográficas y distancias grandes y no se necesita optimizar el proceso computacional como es el caso de esta aplicación. Además, fue utilizada previamente por la DGT en un proyecto similar en el año 2022.

La metodología aplicada permitió identificar las condiciones meteorológicas de todas las lecturas de la campaña de medición de la velocidad, y crear la base de datos para el cálculo posterior del indicador KPI-VFL24 y estadísticos en régimen de velocidad de flujo libre para España y el proyecto europeo *Trendline*.

2. El contexto: determinación del indicador KPI-VFL24 y estadísticos de velocidad de flujo libre en España

Durante los meses de junio a octubre de 2024, se realizaron los trabajos para obtener el indicador de velocidad de flujo libre (KPI-VFL24) y otros estadísticos de la velocidad con las mediciones de una campaña realizada en el año 2023, en el marco de un proyecto realizado para la DGT.

En la Figura 1, se muestra el diagrama de flujo del proyecto donde se resaltan los 3 productos finales obtenidos: las bases de datos finales entregados a la DGT denominadas: BBDDC, BBDDC_f y BBDDC_f. La base de datos denominada “BBDD preliminar” permitió al equipo de investigación realizar tareas previas y necesarias.

En el flujograma del proyecto se muestra el número de registros de cada una de las bases de datos una vez aplicados los procesos, filtros y criterios de selección de registros.

La aportación metodológica que se presenta en este trabajo permitió la aplicación del filtro identificado en el flujograma con el número 1 (en rojo). “Filtrado Criterios validación *Trendline*, en relación con las condiciones atmosféricas” para seleccionar las mediciones de velocidad válidas de la base de datos final para la obtención posterior del indicador KPI-VFL24 y otros estadísticos de velocidad de flujo libre.

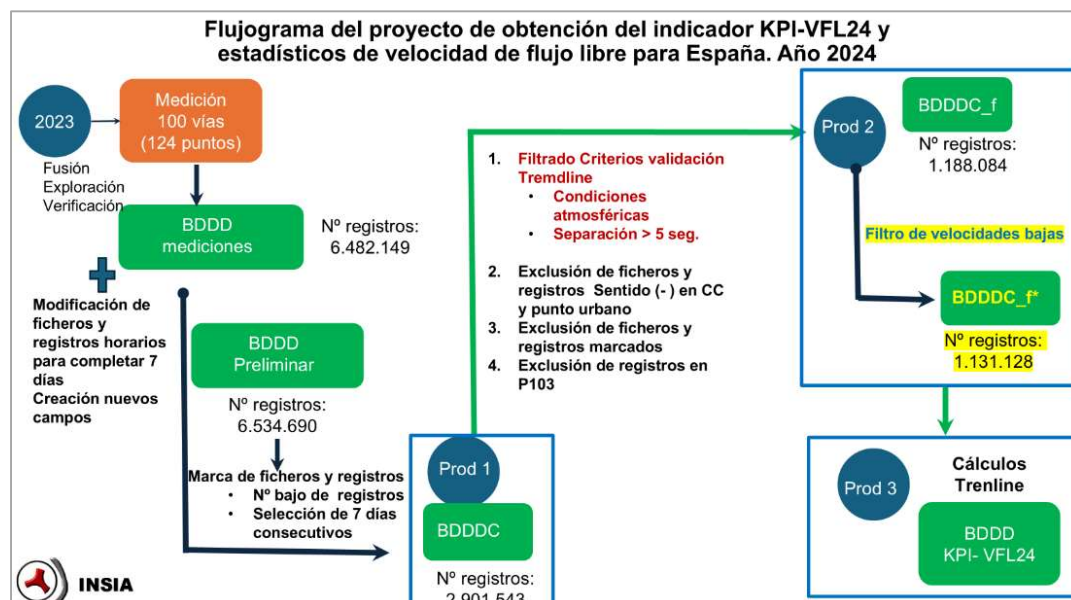


Figura 1: Flujograma del proyecto de obtención del indicador KPI-VFL24 y estadísticos de velocidad de flujo libre para España. Año 2024.

3. Datos relevantes para la metodología

La metodología desarrollada vincula dos fuentes de datos diferentes (Datos 1 y Datos 3) y permite seleccionar las lecturas que cumplen con las condiciones meteorológicas favorables de *Trendline* para la creación de la base de datos final para la determinación del indicador KPI-VFL24 y estadísticos en régimen de velocidad de flujo libre.

A continuación, se distinguen las fuentes de datos y las tablas asociadas utilizadas en la metodología descrita en este trabajo:

1. Datos 1: Contiene la información recogida en la campaña de medición de la velocidad de la campaña en los 100 puntos en las vías de la muestra seleccionada en el territorio español, y otros campos calculados según los requerimientos del PPT del contrato.

De la fuente de Datos 1, se destacan dos tablas:

- **Tabla puntos_aforo:** Los atributos de esta tabla son: dirección del punto, nomenclatura, punto kilométrico de instalación del radar, total de días aforados, número de registros, latitud y longitud, comunidad, provincia, ciudad, etc.
 - **Tabla Datos:** contiene la información de las mediciones de la velocidad, la longitud de los vehículos y otros parámetros relevantes del flujo vehicular registrados en los puntos de aforo de la campaña de mediciones de velocidad del proyecto KPI-VFL 2024
2. Datos 3: contiene la información meteorológica del satélite europeo *Copernicus*, en estaciones situadas en el territorio español, estas son:
 - a. u10: Velocidad horizontal del viento a 10m sobre el suelo (dirección Este), en m/s.
 - b. v10: velocidad vertical del viento a 10m sobre el suelo (dirección norte), en m/s.
 - c. d2m: Punto de rocío a 2 metros de la superficie, en K.
 - d. t2m: Temperatura a 2 metros de la superficie, en K.
 - e. ptype: Tipo de precipitación: 0, sin precipitación; 1, lluvia; 3, lluvia gélida; 5, nieve; 6, aguanieve; 7, mezcla de lluvia y nieve; 8, "Ice Pellets".
 - f. sp: Presión atmosférica en la superficie en Pa.
 - g. tcc: Proporción del cielo cubierto de nubes.
 - h. tp: Precipitación acumulada, en m.

La información de esta fuente se organizó en dos tablas:

- **Tabla Estaciones:** Esta tabla tiene 1.632 filas de pares de latitud-longitud de las estaciones consideradas para el proyecto KPI-VFL 2024.
- **Tabla datos_meteoro:** contiene las variables de clima; u10, v10, etc.

4. Metodología para la vinculación de datos de tráfico con datos meteorológicos

El objetivo de la metodología es vincular las tablas principales de dos fuentes de datos (Datos 1 y Datos 3) y analizar la influencia de las condiciones meteorológicas en el comportamiento del tráfico en régimen de velocidad de flujo libre. Datos 1 contiene los registros de velocidad y otras variables del tráfico y Datos 3 es la fuente externa de mediciones meteorológicas. Los dos conjuntos de informaciones se volcaron en una base de datos MySQL de cuyo tratamiento se obtuvo la base de datos que fue utilizada para la obtención del KPI-VFL24 y estadísticos de velocidad de flujo libre.

En la infografía de la Figura 2 se muestra el proceso seguido para la obtención de la base de datos final BBDDC_f* para la obtención de indicador y estadísticos de velocidad de flujo libre 2024.



Figura 2: Infografía de la Metodología para la vinculación de datos de tráfico vehicular con datos meteorológicos del proyecto de obtención de indicador y estadísticos de velocidad de flujo libre 2024.

En este trabajo se describen las tareas realizadas bajo el Hito “Vinculación de datos” y las relacionadas con la creación de las bases de datos finales.

4.1. Procedimiento de vinculación de estaciones y puntos de aforo

Los pasos de la metodología se resumen en los siguientes puntos:

1. Selección de variables relevantes: Latitud y longitud, Fechas y horas de lecturas, d2m y t2m y ptype

Ambas fuentes recogen la fecha y hora de las mediciones y se utilizaron para la búsqueda de proximidad de fechas y horas (y así determinar la proximidad horaria en una ventana temporal de una hora), para asociar la información meteorológica de Datos 3 con la medición de velocidad de Datos 1.

Consideraciones:

- Precisión de Ubicación: La precisión de los datos de latitud y longitud es importante para la exactitud del cálculo de distancias con la fórmula de Haversine, un proceso simple desde el punto de vista computacional, recomendado cuando se trabaja con coordenadas geográficas y distancias grandes y cuando no es necesario optimizar el proceso computacional como es el caso de esta aplicación. Además, fue utilizada previamente por la DGT en un proyecto similar en el año 2022 y permitiría hacer análisis comparativos.
- Su aplicación requiere la conversión de coordenadas geográficas de grados decimales a radianes.
- Tolerancia Temporal: La elección de la tolerancia temporal para la coincidencia de los datos es importante y debe basarse en la frecuencia de los registros y la dinámica de los fenómenos meteorológicos.

2. La vinculación espacial se realizó con:

- **Latitud y longitud** de los puntos de aforo y cada estación meteorológica son básicas para la metodología de vinculación desarrollada en el proyecto, mediante:
 - **Cálculo de distancias geográficas** mediante una función SQL (*“store procedure”*). Para cada punto de aforo y cada estación meteorológica, se calculó la distancia geodésica entre sus ubicaciones utilizando la fórmula de Haversine:

$$a = \sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos(\varphi_1) * \cos \varphi_2 * \sin^2\left(\frac{\Delta\lambda}{2}\right)$$

$$c = 2 * \operatorname{atan2}(\sqrt{a}, \sqrt{1-a})$$

$$d = R * c$$

Donde:

- φ_1 y φ_2 son las latitudes de los puntos en radianes
- λ_1 y λ_2 son las longitudes de los puntos en radianes
- $\Delta\varphi = \varphi_2 - \varphi_1$ es la diferencia de latitud
- $\Delta\lambda = \lambda_2 - \lambda_1$ es la diferencia de longitud
- R es el radio de la Tierra (aproximadamente 6,371 km)
- d es la distancia entre los dos puntos
- **Identificación de la estación meteorológica más cercana** a cada punto de aforo, y registro de dicha relación en una tabla intermedia para su posterior análisis. Con la función SQL se evalúa cada par de coordenadas de la tabla “estaciones” y su cercanía al par de coordenadas de la tabla “puntos_aforo”, que da como resultado la estación meteorológica más cercana al punto de medición de velocidades.

Una vez determinada la proximidad espacial se busca la proximidad temporal (días y horas de las lecturas) entre los datos meteorológicos y los datos de medición de velocidades.

3. La vinculación temporal requirió establecer una ventana temporal, mediante:

- **Fechas y horas de lecturas** de ambas fuentes para asociar los registros de puntos de aforo y estaciones meteorológicas en fechas y horas cercanas o coincidentes, con los siguientes resultados:
 - **Integración temporal de los datos**, asociando a cada registro de tráfico vehicular los datos meteorológicos de la estación más cercana correspondientes al mismo intervalo temporal.
 - **Creación de Datos Vinculados**. Finalmente, se generó una vista con los datos vinculados que combina los datos del punto de aforo (velocidad, longitud promedio, fecha, hora, etc.) con los

datos meteorológicos correspondientes de la estación más cercana (precipitación, temperatura, velocidad del viento, etc.) para la misma o similar marca de tiempo.

4.2. Procedimiento de validación de registros de la base de datos final para la obtención del indicador y estadísticos de velocidad de flujo libre

En este apartado se describe la validación de condiciones meteorológicas del proyecto europeo *Trendline* para generar una base de datos con registros válidos para la obtención del indicador y estadísticos de velocidad de flujo libre en España en el año 2024.

Una vez realizada la vinculación de estaciones y puntos de aforo, con la creación de la tabla de datos vinculados se generó una etiqueta de condición favorable o desfavorable basada en las variables meteorológicas seleccionadas **d2m** y **t2m** (necesarias para calcular el valor de corte que determina la presencia de niebla) y **p_{type}** para presencia de lluvia, con los siguientes resultados:

- **Creación de la etiqueta para filtro de registros válidos según las condiciones meteorológicas.** La creación de las etiquetas (S si la condición es favorable) / (D si la condición es desfavorable) para cada registro de velocidad realizada en España es el filtro necesario para la obtención de los registros válidos para la obtención del indicador y estadísticos del proyecto KPI-VFL 2024.
- **Adopción del Criterio de selección / descarte de lecturas general.** Con la etiqueta “d_clima” D si la condición es desfavorable, se descartaron los registros **cuando se verifican** al menos las condiciones meteorológicas adversas siguientes:
 - El valor de la variable relacionada con la precipitación **p_{type}** es distinto de 0.
 - La presencia de niebla definida como: $t2m - d2m < 2,5\text{ }^{\circ}\text{C}$
 - La presencia de a) o b) o las 2 condiciones.

Además, la metodología *Trendline* prevé la verificación de una separación entre vehículos mayor o igual a 5 segundos.

5. Creación de Bases de Datos

El proceso de creación de las bases de datos BBDDC y BBDDC_f (y la final BBDDC_f*) se muestra de forma resumida en las tablas siguientes.

En la Tabla 1 se muestran los valores resultantes de la creación de la base de datos preliminar del proyecto. Partiendo de las lecturas en 100 puntos de medición, el número final se incrementa hasta 6.534.690 registros completando los datos de 7 días consecutivos según lo indicado en el PPT del contrato del proyecto KPI-VFL 2024.

Tabla 1: BBDD preliminar: Número final de registros: 6.534.690.

		Nº Registros
BBDD preliminar	Registros de la campaña en 100 puntos de la red de carreteras (2023)	6.482.149
	Se completan 7 días consecutivos de los registros de la campaña en 100 vías con horas faltantes (excepto P106, que no cuenta con horas para el estrato LD)	52.541
	Número de registros final de BBDD preliminar	6.534.690

En la Tabla 2 se muestran los valores resultantes de la base de datos BBDDC del proyecto. La mayoría de los criterios de depuración y filtrado del listado (6 de 7) están relacionados con las mediciones de la campaña, siendo el criterio de condiciones climáticas favorables (ausencia de lluvia y niebla) el que elimina el mayor número de registros (39,9%) y es el resultado de la aplicación de la metodología de vinculación desarrollada en base a la distancia con la fórmula de Haversine. El número de registros de BBDDC es de 2.901.543 (44,4% de los totales iniciales de BBDD preliminar).

Tabla 2: BBDDC: Filtros aplicados. Número total de registros: 2.901.543.

BBDD preliminar		6.534.690	
	Razones de los registros depurados / eliminados	Nº Registros	Pérdida %
BBDDC	Datos incoherentes (velocidad 0 y 1)	-289	-0,004%
	Lecturas de 3 archivos eliminados porque no tienen lecturas suficientes, datos anómalos, incompletos	-28.088	-0,4%
	Sentido de circulación en vías CC100 (sólo quedan observaciones con signo +)	-650.382	-10,0%
	Sentido de circulación en el punto P12 urbano (sólo quedan observaciones con signo +)	-33.100	-0,5%
	Selección 7 días, se descartan lecturas excedentes	-927.303	-14,2%
	Se descartan registros diurnos porque no cuenta con horas para el estrato LD (d_LDdenidpunto 106='D')	-66.607	-1,0%
	No cumplen las condiciones climáticas	-2.606.167	-39,9%
	Combinaciones	678.789	
Nº de registros eliminados		-3.633.147	
Nº de registros final de BBDDC		2.901.543	

En la Tabla 3, se muestran los valores resultantes de la base de datos BBDDC_f del proyecto. A la base de datos previa (**BBDDC**) se aplicó el criterio de separación entre vehículos de más de 5 segundos de la metodología *Trendline* que da como resultado la base BBDDC_f con un total de registros de 1.188.084 registros (40,95% de los registros de la base de partida).

Uno de los estudios realizado por el equipo investigador, consistió en la exploración y valoración del efecto sobre los estadísticos de velocidad de valores de velocidad anormalmente bajas en cada tipo de vía, que determinó la aplicación de criterios de corte discriminando por tipo de vía y la velocidad límite correspondiente. Este último filtro (no especificado en la metodología *Trendline*), da como resultado final la base de datos BBDDC_f* con un número total de 1.131.128 registros (el número de observaciones eliminadas en todas las vías de la muestra representa el 4,8% de los registros).

Tabla 3: Prod 2. Filtros aplicados. (BBDDC_f. Número total de registros: 1.188.084 -- BBDDC_f*. Número final de registros: 1.131.128).

BBDDC		2.901.543		
	Criterio de filtrado	Nº de registros	% de registros de BBDDC	% de registros de BBDDC_f
BBDDC_f	Condiciones de flujo libre: separación mayor o igual a 5 segundos	1.188.084	40,95%	
BBDDC_f*	Filtro con umbrales para velocidades bajas:			
	En U30 = 20 km/h			
	En U50 = 25 km/h	-56.956		4,8%
	En vía convencional = 45 km/h			
	En AP y AV = 60 km/h (según normativa)			
Nº de registros final de BBDDC_f*		1.131.128		

La base de datos BBDDC_f* es la que se usó finalmente para el cálculo del indicador KPI-VFL24 y de los estadísticos de velocidad de flujo libre (velocidad media, percentil 85, etc.).

6. Resultados

Los resultados obtenidos en el proyecto se pueden resumir en:

- **Generación de Bases de datos**, partiendo de 2 fuentes de datos diferentes Datos 1 de una campaña de mediciones de velocidad para un proyecto de la DGT y Datos 3 de registros de climatología del satélite europeo *Copernicus*.
- **Selección de un gestor de bases de datos como MySQL** para el trabajo con datos masivos.
- **Desarrollo de una metodología para la vinculación de las fuentes de datos y selección de datos específicos**.
- **Adopción de la distancia geográfica empleando la fórmula de Haversine, como criterio principal para la vinculación**.
- **Definición de la vinculación temporal** para aproximar las fechas de ambas mediciones: tráfico y variables meteorológicas mediante la definición de una ventana de tiempo.
- **Implementación de la vinculación espacial y temporal, creación de tablas, vistas y bases de datos finales mediante comandos MySQL**.
- **Definición de los criterios para la selección de datos válidos según las condiciones meteorológicas**.
- **Análisis del impacto de las condiciones meteorológicas sobre el tráfico en régimen de velocidad de flujo libre** (criterios más restrictivos, porcentaje de registros inválidos, etc.).
- **Validación de condiciones de velocidad de flujo libre del proyecto Trendline** para el cálculo del indicador KPI-VFL24 y estadísticos de velocidad de flujo libre para España, año 2024.

7. Conclusiones

La Metodología para la vinculación de datos de tráfico vehicular con datos meteorológicos utilizando la fórmula de distancia de Haversine, permitió asociar los datos de dos fuentes de información: una fuente de datos externa del satélite *Copernicus* y la fuente de datos de mediciones de velocidad en la red de carreteras españolas.

La tabla de los datos vinculados combina las informaciones de los puntos de aforo (velocidad, longitud promedio, tipo de vehículo, fecha_hora, etc.) con los datos meteorológicos (precipitación, temperatura, velocidad del viento, etc.) correspondientes a la estación más cercana para la misma o similar marca de tiempo, identificada con la metodología desarrollada.

Se definieron las condiciones meteorológicas adversas a partir de las informaciones de **d2m** y **t2m** (con las que se obtuvo el valor de presencia de niebla) y **ptype** relacionada con la lluvia y se creó la etiqueta o campo “d_clima” de dos estados: S si la condición es favorable / D si la condición es desfavorable.

El campo “d_clima= D” permite filtrar los registros con condiciones meteorológicas adversas para el cálculo del KPI-VFL24 y estadísticos de velocidad de flujo libre en España en el año 2024, según las condiciones de la metodología del proyecto europeo *Tredline*. Este criterio elimina el 39,9% de registros, siendo la presencia de niebla la condición meteorológica que penaliza más.

Otro requisito de la metodología *Trendline* para la obtención del indicador de velocidad de flujo libre es verificar la separación entre vehículos mayor o igual a 5 segundos. La aplicación de este criterio retiene el 40,95% de registros de la base de datos vinculados.

Finalmente, se aplicó un filtro final de velocidades anormales discriminado por valores de corte de la velocidad en función del límite en cada tipo de carretera. Este criterio elimina el 4,8% de los registros de BBDDC_f, y la base final (BBDDC_f*) para el cálculo del KPI-VFL24 y estadísticos de velocidad de flujo libre en España en el año 2024 tiene 1.131.128 registros.

8. Agradecimientos

Este trabajo realizado por el INSIA, durante los meses de junio a octubre de 2024, se ha desarrollado dentro del Servicio para la obtención del indicador de velocidad de flujo libre. Expediente DGT: 3DGT6A000171 y se inscribe en el Proyecto *Trendline* de la Unión Europea para la “CONSTRUCCIÓN DEL INDICADOR DE

VELOCIDAD DE FLUJO LIBRE. ESPAÑA. 2024”. Los autores también quieren agradecer a la Comunidad de Madrid, la financiación parcial para la difusión, a través del programa SEGVAUTO 5G (TEC-2024- ECO-277).

9. Referencias

- [1] Trendline project. <https://trendlineproject.eu/>
- [2] trendline-speed-template2024_ver3. <https://trendlineproject.eu/>
- [3] Copernicus Climate Change Service, Climate Data Store, (2023): ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.adbb2d47 (Accessed on 19-JUN-2024)