

Desarrollo de un sistema de apoyo para la toma de decisiones en la gestión de la calidad del aire:

Implementación operativa de un producto espacial de PM_{10}

Descarga y pre-procesamiento de datos para el entrenamiento del modelo de predicción de PM_{10}

En este documento se describe el procedimiento para la descarga y pre-procesamiento de los datos que fueron utilizados en el entrenamiento del modelo de predicción de PM_{10} .

Preparado por: Lara Della Ceca

Revisado por: Soledad Represa y Fernanda García Ferreyra

Fecha: 13 de agosto 2021

1. ESPESOR ÓPTICO DE AEROSOLES: sensor MODIS (a bordo de los satélites Terra y Aqua/NASA)

El espesor óptico de aerosoles (AOD, siglas en inglés) se obtiene del sensor MODIS, que se encuentra a bordo de los satélites TERRA y AQUA (NASA). La hora de pasada por el ecuador es aproximadamente a las 10:30 hs (TERRA) y 13:30 hs UTC (AQUA). Este sensor brinda un producto satelital con frecuencia diaria de espesor óptico de aerosoles con $1 \times 1 \text{ km}^2$ de resolución espacial mediante un algoritmo genérico denominado 'Implementación de corrección atmosférica multi-ángulo' (MAIAC, siglas en inglés).

Para este proyecto se utiliza el producto diario de espesor óptico de aerosoles obtenido con el algoritmo MAIAC (denominado MCD19A2 Versión 006, <https://lpdaac.usgs.gov/products/mcd19a2v006/>).

Para información más detallada del producto MCD19A2 puede consultarse la Guía del usuario del producto (https://lpdaac.usgs.gov/documents/110/MCD19_User_Guide_V6.pdf).

1.1 Descarga del producto satelital AOD-MAIAC

El producto MCD19A2 es brindado en cuadrículas sinusoidales de $1200 \times 1200 \text{ km}^2$, con una resolución espacial de $1 \times 1 \text{ km}^2$ en archivos diarios con formato HDF (*Hierarchical Data Format*, siglas en inglés). El archivo diario del producto MCD19A2 contiene los valores de AOD de las pasadas (órbitas) de los satélites TERRA y AQUA sobre Argentina para ese día.

Para el entrenamiento del modelo se realizó la descarga del mosaico h13v12 que comprende la ciudad de Buenos Aires, Gran Buenos Aires y Bahía Blanca, donde se encuentran localizadas las estaciones superficiales de material particulado, para el período 2010-2019 (**Figura 2**). La descarga se realizó desde el servidor MOTA (<https://e4ftl01.cr.usgs.gov/MOTA/>) del Centro de distribución de archivos de procesos terrestres de la NASA (LP DAAC, siglas en inglés).

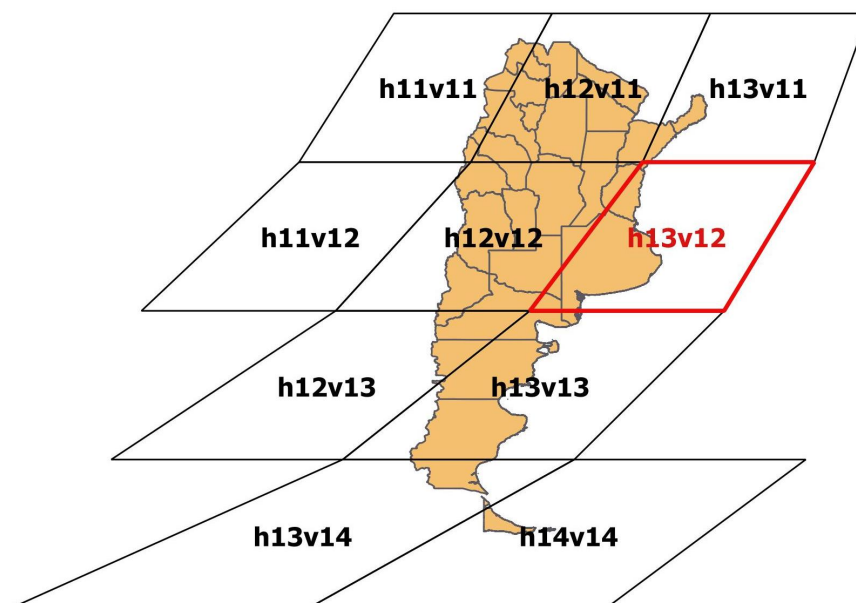


Figura 2. Superficie del territorio argentino cubierta por los mosaicos sinusoidales de MODIS.

1.2 Procesamiento del producto MCD19A2

El procesamiento del producto MCD19A2 incluyó el resampleado al sistema de referencia WGS84 (EPSG: 4326) y la extracción de las variables *Optical Depth 047* (espesor óptico de aerosoles a 470 nm) y *Optical Depth 055* (espesor óptico de aerosoles a 550 nm) para cada órbita satelital en archivos con formato GeoTIFF. Previo a su extracción, los valores de dichas variables fueron filtrados según la banda AOD QA que determina el status de calidad de los mismos, de manera tal de eliminar los valores de AOD con posibles errores de detección.

Mediante el filtro de calidad de los datos se desestimaron los píxeles donde se detectaron nubes (máscara de nubes), sombras de nubes, focos de calor de incendios y sedimentos de agua. Se preservan los valores de los píxeles considerados libres de nubes, los que presentan hasta 1 píxel vecino con nubes, con un fondo para todo tipo de superficies (tierra, agua, nieve y hielo), los considerados de mejor calidad para el cálculo del AOD, sin el reflejo intenso de la luz solar (sun glint), y que son procesados mediante el modelo de base (no el modelo de polvo o humo). Por lo tanto, y según la definición de la banda QA que presenta el producto MCD19A2.006, los datos que se publican en esta implementación son aquellos que valen 1, 2, 9, 10, 17, 18, 25, 26, 97, 105, 113, 121, 98, 106, 114, 122 en la banda QA. Para ello se utilizó el script de R denominado **MAIAC_01_filterQA_save_orbits_TIF.R**

Una vez filtrados los valores, los archivos GeoTIFF de cada órbita fueron recortados para la región 59.2°–57.7° O 35°–34° S que cubre la superficie de la Ciudad Autónoma de Buenos Aires y Gran Buenos Aires, donde se

encuentran localizadas las estaciones superficiales de material particulado (MP). Posteriormente, a las imágenes con más de un 10% de valores con datos válidos se les aplicó el modelo de interpolación IDW para completar los valores de AOD en píxeles sin dato válido de AOD. Para ello se utilizó el script de R denominado **MAIAC_02_crop_and_fill_gaps.R**

1.3 Extracción de valores de AOD para el pixel de las estaciones de MP

De las imágenes GeoTIFF recortadas e interpoladas se obtuvieron los valores de AOD correspondientes a los píxeles de cada estación de MP superficial considerando su localización y se completó la tabla **MAIAC_pixel_data_pm_stations_2010-2019_INTERPOLADO.csv** con información sobre: Fecha y hora de la órbita de MODIS, satélite (Terra o Aqua), longitud de onda de AOD (470 o 550 nm), valor de AOD, estación de MP. Para ello se utilizó el script de R denominado **MAIAC_03_get_pixel_data_pm_stations.R**

A la tabla del archivo **MAIAC_pixel_data_pm_stations_2010-2019_INTERPOLADO.csv** se le incorporaron las siguientes columnas: **DEM_asnm**, **VIIRS_night_lights**, **ALBEDO**, **BCCMASS**, **CLDHGH**, **CLDLOW**, **DMSSMASS**, **DUSMASS**, **OCSMASS**, **PBLH**, **PRECTOT**, **PS**, **RH**, **SO2SMASS**, **SO4SMASS**, **SPEED**, **SPEEDMAX**, **SSSMASS**, **T**, **U**, **USTAR**, **V**.

1.4 Dominio del modelo: grilla de 1x1 km² de resolución espacial

A partir de los mosaicos de MODIS que cubren el territorio argentino (**Figura 2**), resampleados al sistema de referencia WGS84, se elaboró un archivo raster en formato GeoTIFF con 1x1 km² de resolución espacial que se utilizó luego para el resampleo de las demás bases de datos (denominado **MOSAICO_MAIAC_dominio_modelo.tif**). El mismo fue cortado para ajustarse al territorio argentino (**Figura 3**).

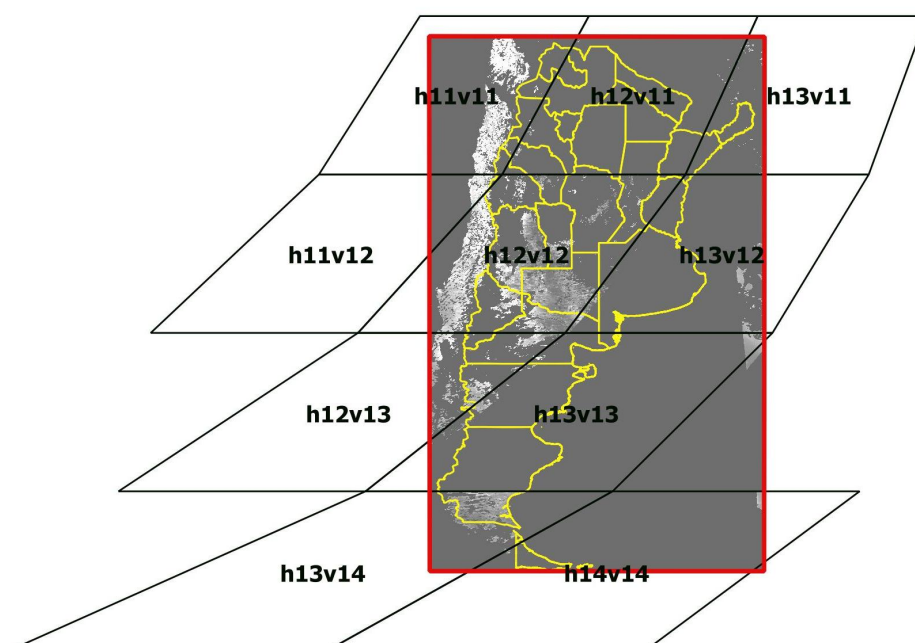


Figura 3. Archivo raster de 1 km de resolución espacial (con sistema de referencia WGS84) elaborado a partir del recorte de los mosaicos MODIS que cubren el territorio argentino (recuadros en negro) recortados según área cubierta por el dominio del modelo (recuadro rojo) y reproyectados a WGS84.

2. MERRA-2: Modern-Era Retrospective analysis for Research and Applications

El MERRA-2 es un reanálisis atmosférico utilizando el GEOS-5 con su ADAS, versión 5.12.4, a una resolución espacial de $0.5^\circ \times 0.625^\circ$ (Gelaro et al., 2017). El análisis se realiza a intervalos de 3 o 6 horas, con asimilación de distintos productos satelitales, incluyendo el AVHRR, MISR, MODIS sobre superficies brillantes y AERONET.

Para mayor detalle de la elaboración y distribución de estos productos puede consultarse <https://disc.gsfc.nasa.gov/daac-bin/FTPSubset2.pl>

2.1 Descarga de productos MERRA-2 y variables seleccionadas

Los productos de MERRA-2 son distribuidos en formato NETCDF y fueron descargados de la web <https://disc.gsfc.nasa.gov/> de NASA. En la siguiente tabla se detalla el nombre de cada producto descargado y las variables seleccionadas para el presente proyecto:

Nombre del producto	Variables seleccionadas	Descripción	Unidad	Resolución temporal
tavg1_2d_aer_Nx	BCSMAS	Concentración de masa superficial de carbono negro	kg.m^{-3}	3 horas
	DMSSMAS	Concentración de masa superficial de sulfuro de dimetilo	kg.m^{-3}	3 horas
	DUSMAS	Concentración de masa superficial de polvo	kg.m^{-3}	3 horas
	OCSMAS	Concentración de masa superficial de carbono orgánico	kg.m^{-3}	3 horas
	SO2SMAS	Concentración de masa superficial de SO_2	kg.m^{-3}	3 horas
	SO4SMAS	Concentración de masa superficial de SO_4	kg.m^{-3}	3 horas
	SSSMAS	Concentración de masa superficial de de sal marina	kg.m^{-3}	3 horas
tavg1_2d_flg_Nx	PBLH	Altura de la capa límite planetaria	m	3 horas
	PRECTOT	Precipitación total del modelo físico atm	$\text{kg.m}^{-2}.\text{s}^{-1}$	3 horas
	SPEED	Velocidad del viento superficial	m.s^{-1}	3 horas
	SPEEDMAX	Velocidad máxima de viento en la superficie	m.s^{-1}	3 horas
	USTAR	Escala de velocidad superficial	m.s^{-1}	3 horas
tavg1_2d_rad_Nx	ALBEDO	Albedo superficial	%	6 horas
	CLDHGH	Fracción de área de nubes para nubes altas	%	3 horas

	CLDLow	Fracción de área de nubes para nubes bajas	%	3 horas
inst3_3d_asm_Nv	PS	Presión en superficie	hPa	3 horas
	T	Temperatura del aire a 1000 hPa	K	3 horas
	RH	Humedad relativa a 1000 hPa	%	3 horas
	U	Viento del este	m.s ⁻¹	3 horas
	V	Viento del norte	m.s ⁻¹	3 horas

La descarga de dichos productos se realizó para el período 2010-2019, para la región 59.2°–57.7° O 35°–34° S que cubre la superficie de la Ciudad Autónoma de Buenos Aires y Gran Buenos Aires (donde se encuentran localizadas las estaciones superficiales de MP) y para la franja horaria 12-20 horas que corresponde a la hora de pasada de los satélites Terra y Aqua.

2.2 Procesamiento de los productos MERRA-2

El procesamiento de los productos MERRA-2 comprendió la generación de un archivo GeoTIFF por cada variable de interés y el resampleado a la grilla descrita en el ítem 1.4 (1x1 km² de resolución espacial y WGS84). Para ello se utilizó el script de R denominado **MERRA_01_resample_save_tif.R**.

2.3 Extracción de valores de variables MERRA-2 para el pixel de las estaciones de MP

De las imágenes GeoTIFF resampleadas de cada variable se obtuvieron los valores correspondientes a los píxeles de cada estación de MP superficial considerando su localización y la hora más cercana a la hora de pasada de cada órbita MODIS, completando la tabla de entrenamiento incorporando una columna por cada variable MERRA-2 considerada. Para ello se utilizó el script de R **TRAINING_TABLE_append_data.R**

3. Producto luces nocturnas del sensor VIIRS (SNPP-NASA)

El sensor VIIRS proporciona mediciones diarias globales de la luz nocturna visible e infrarroja cercana (NIR) con las cuales se generan un nuevo conjunto de productos de imagen de luces nocturnas que permiten estudiar fenómenos nocturnos como las fuentes antropogénicas de emisiones de luz.

Para este proyecto se utilizó el producto denominado VNP46A1 - VIIRS / NPP, que consiste en un producto con frecuencia diaria y resolución especial de 500 m, que contiene 26 conjuntos de datos científicos (SDS) que incluyen la luminosidad del sensor, ángulos cenital y azimutal (sensor at, solar y lunar), banderas de máscara de nube, tiempo, radiancia IR de onda corta, temperaturas de brillo, indicadores de calidad VIIRS, ángulo de fase lunar y fracción de iluminación lunar, entre otros.

Para mayor detalle sobre la elaboración de este producto puede consultarse la guía de usuario disponible en https://viirsland.gsfc.nasa.gov/PDF/VIIRS_BlackMarble_UserGuide.pdf

3.1 Descarga del producto VNP46A1 - VIIRS / NPP

El producto VNP46A1 - VIIRS / NPP es distribuido en formato HDF5 y fue descargado de la web <https://ladsweb.modaps.eosdis.nasa.gov/> de NASA. Se descargó el producto diario para el mosaico h12v12 que cubre parte de la provincia de Buenos Aires (incluyendo CABA y AMBA, **Figura 4**) para los meses Abril, Mayo y Junio de cada año del período 2012-2019. Se eligieron estos meses porque son los que presentan la menor frecuencia de incendios y de otros tipos de eventos luminosos en Argentina.

Dado que para los años 2010 y 2011 no hay disponibilidad de este producto, se les asignará el valor del año 2012.

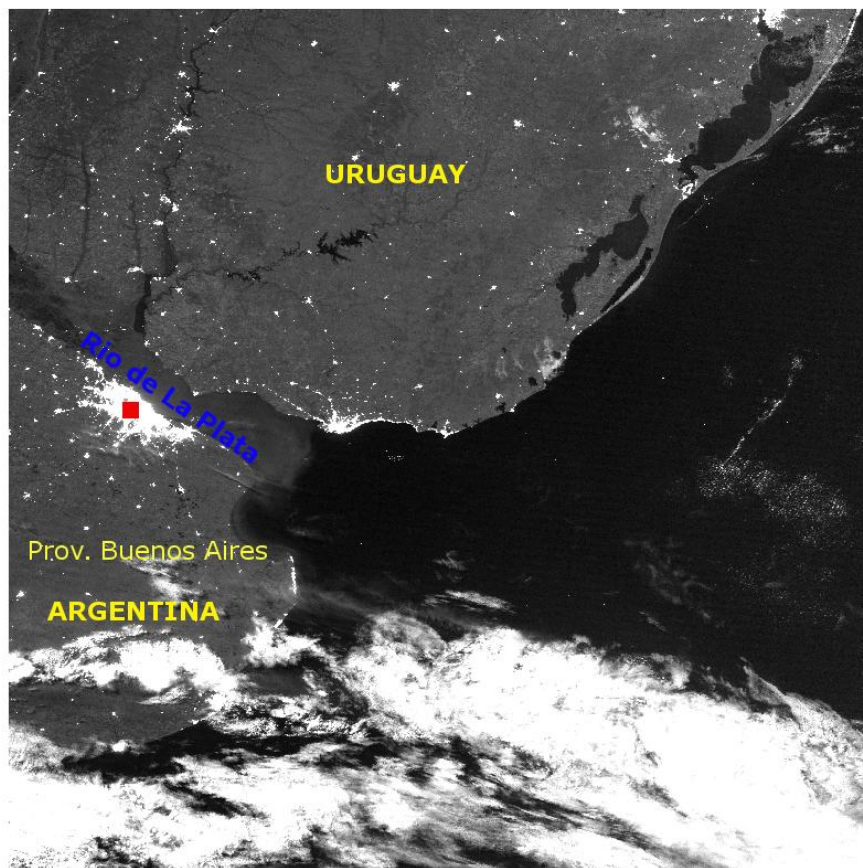


Figura 4. Imagen de luces de noche obtenida del sensor VIIRS (NPP) para el día 28/03/2015 del mosaico h12v12 que comprende parte de la provincia de Buenos Aires, incluyendo CABA y AMBA (punto rojo).

3.2 Procesamiento del producto del producto VNP46A1 - VIIRS / NPP

A partir del producto diario VNP46A1 se elaboró un archivo TIF con el promedio de los meses Abril a Junio para cada año con datos disponibles (2012-2019).

En primera instancia, se guardó en formato TIF el subdataset *DNB_At_Sensor_Radiance_500m* (contenido en el archivo VNP46A1) resampleado a la grilla descrita en el ítem 1.4 (1x1 km² de resolución espacial y WGS84). Para ello se utilizó el script de R denominado **VIIRS_01_save_daily_resampled.tif.R**

Luego, a partir de los archivos obtenidos se realizó el promedio del período Abril-Junio para cada año (2012 a 2019). Para ello se utilizó el script de R denominado **VIIRS_02_year_mean.R**.

3.3 Extracción de valores VNP46A1 para el pixel de las estaciones de MP

De las imágenes GeoTIFF con el promedio de los meses Abril a Junio de cada año se obtuvieron los valores correspondientes a los pixeles de cada estación de MP superficial considerando su localización y el año del producto VNP46A1. Para ello se utilizó el script de R denominado **VIIRS_03_get_pixel_data_pm_stations.R**, del cual se obtiene la tabla **VIIRS_pixel_data_pm_stations_mean_Apr-Jun_2012-2019.csv** que luego se utiliza para completar la tabla de entrenamiento mediante el script de R denominado **TRAINING_TABLE_append_data.R**

4. Producto Modelo de Elevación Digital

La altura sobre el nivel del mar para cada estación de MP superficial se obtuvo del producto GMTED2010¹ (Global Multi-resolution Terrain Elevation Data, en inglés), de 30 arc-sec (~ 1 x 1 km) de resolución espacial, elaborado por el Servicio Geológico de los Estados Unidos (USGS, siglas en inglés) y la Agencia Nacional de Inteligencia Geoespacial (NGA, siglas en inglés).

Para mayor información sobre este producto puede consultarse el reporte técnico disponible en <https://pubs.usgs.gov/of/2011/1073/pdf/of2011-1073.pdf>.

4.1 Descarga del producto GMTED2010

El producto GMTED2010 es distribuido en formato TIF y fue descargado de la web https://topotools.cr.usgs.gov/gmted_viewer/viewer.htm de NASA. Se descargaron todos los mosaicos que cubren la superficie del territorio argentino.

4.2 Procesamiento del producto GMTED2010

Con los mosaicos del producto GMTED2010 se generó un único mosaico para el territorio argentino que fue resampleado y recortado a la grilla descrita en el ítem 1.4 (1x1 km² de resolución espacial y WGS84). Para ello se utilizó la primer parte del script de R denominado **DEM_01_USGS_resample_and_get_pm_st_pixel_data.R**

De esta manera, se generó el archivo **DEM_Arg_30arcsec_1km_USGS_dominio_modelo_resampled.tif** que será utilizado no sólo en el entrenamiento sino además en la operativización del modelo de este proyecto.

4.3 Extracción de valores del producto GMTED2010 para el pixel de las estaciones de MP

A partir del archivo **DEM_Arg_30arcsec_1km_USGS_dominio_modelo_resampled.tif** se obtuvieron los valores de altura sobre el nivel del mar (en metros) correspondientes a los pixeles de cada estación de MP superficial considerando su geolocalización. Para ello se utilizó la segunda parte del script de R denominado **DEM_01_USGS_resample_and_get_pm_st_pixel_data.R**, del cual se obtiene la tabla

¹ Global multi-resolution terrain elevation data 2010 (GMTED2010); 2011; OFR; 2011-1073; Danielson, Jeffrey J.; Gesch, Dean B.

DEM_pixel_data_pm_stations.csv que luego se utiliza para completar la tabla de entrenamiento mediante el script de R denominado **TRAINING_TABLE_append_data.R**

5. Material particulado

Para el entrenamiento del modelo se utilizaron las mediciones de PM_{10} de las estaciones de monitoreo continuo del Gobierno de la Ciudad Autónoma de Buenos Aires (C.A.B.A) y de la Autoridad de Cuenca Matanza Riachuelo (ACUMAR).

En la **Tabla 1** se presentan las características de cada estación y en la **Figura 5** su localización.

Tabla 1. Información relevante sobre las estaciones de monitoreo de calidad del aire consideradas en este proyecto.

Autoridad responsable	Estación	Coordenadas	Característica de la medición	Disponibilidad de mediciones	Descripción del área de emplazamiento
Gobierno de C.A.B.A	LA BOCA	58.37°O - 34.62°S	PM10 (ug/m3) - media móvil 24 hs	Desde 01/05/2009 - actualidad	Zona mixta con flujo vehicular medio-bajo e incidencia de fuentes fijas. Emplazada dentro del área de incidencia de la cuenca Matanza-Riachuelo
	CENTENARIO	58.43°O - 34.60°S		Desde 01/01/2005 - actualidad	Área residencial-comercial con flujo vehicular medio y muy escasa incidencia de fuentes fijas. Próxima a un espacio arbóreo ubicado en el centro geográfico
	CORDOBA	58.39°O - 34.59°S		Desde 01/05/2009 - actualidad	Área residencial-comercial con flujo vehicular alto y muy escasa incidencia de fuentes fijas. Representativa de un conjunto de zonas de similares características
ACUMAR	LA MATANZA	58.68°O - 34.88°S	PM10 ug/m3 - El límite de detección del analizador de Material Particulado PM10 es de 4,8 µg/m3 para períodos de tiempo de 1 hora.	Desde 07/2017 a 02/2019	
	DOCK SUD	58.68°O - 34.91°S		Desde 12/2010 a 09/2016	

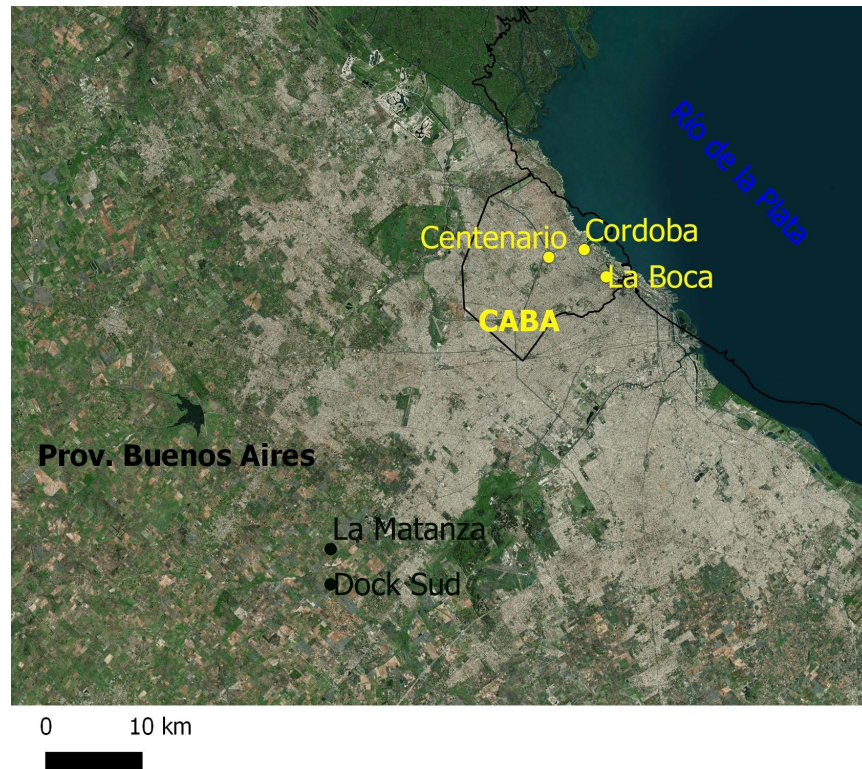


Figura 5. Localización de las estaciones de monitoreo de calidad del aire de ACUMAR (puntos negros) y de Ciudad Autónoma de Buenos Aires (CABA) (puntos amarillos).

5.1 Descarga de mediciones de PM_{10}

Las mediciones de PM_{10} se encuentran disponibles y fueron obtenidas en formato de archivos tipo tabla (Excel) de las páginas web de ambos organismos responsables:

- **Gobierno C.A.B.A** <https://data.buenosaires.gob.ar/dataset/calidad-aire>
- **ACUMAR** <http://jmb.acumar.gov.ar/calidad/reportes/archivosExcel.php?archivo=datosExcel.xlsx>

5.2 Procesamiento de mediciones de PM_{10}

Dado que las mediciones de la base de datos del Gobierno de C.A.B.A se brindan como media móvil de 24 hs, y la de ACUMAR como valor horario, se procedió a calcular el valor como media móvil de 24 hs para esta última base de datos con el objetivo de homogeneizar la información incorporada al modelo.

Las bases de datos descargadas fueron preprocesadas para obtener sólo las mediciones de MP (PM_{10}), y descartar las de otros contaminantes que no son de interés para este proyecto y reordenar la información para poder completar la tabla de entrenamiento. Para ello se utilizó el script de R denominado **PM_01_preprocessing_ACUMAR_and_CABA_tables.R** del cual se obtiene la tabla **PM_ACUMAR_y_APRA_PREPROCESADO.csv** que luego se utiliza para completar la tabla de entrenamiento mediante el script de R denominado **TRAINING_TABLE_append_data.R**