

DOCTORADO EN CIENCIA Y TECNOLOGIA

Evaluado y acreditado por la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU).
Resolución N° 1178/11. Calificación "B".

*Integración de metodologías de bajo costo para la caracterización de la
calidad del aire*

Trabajo de tesis para optar por el título de Doctor en Ciencia y Tecnología de la Universidad
Nacional de General Sarmiento

Autor: **Ariel Scagliotti**

Director: Guillermo Jorge
Co-Director: Sebastián Díez

DOCTORADO EN CIENCIA Y TECNOLOGIA

Evaluado y acreditado por la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU).
Resolución N° 1178/11. Calificación "B".

Fecha: 30 de noviembre de 2021

DOCTORADO EN CIENCIA Y TECNOLOGIA

Evaluado y acreditado por la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU).
Resolución N° 1178/11. Calificación "B".

m. Temas tratados en la Tesis (palabras claves): calidad del aire, integración de metodologías, sensores de bajo costo, redes neuronales, desarrollo experimental.

n. Resumen en español:

Las partículas en el aire son un indicador importante de la calidad del aire (CA), pues tienen diferentes impactos sobre la salud y sobre el clima. El monitoreo de la CA de forma regulatoria se realiza a través de estaciones de medición en superficie. El costo hace que este tipo de mediciones suelen ser dispersas o inexistentes. Una alternativa viene dada por los sensores de bajo costo, aunque poseen numerosas limitaciones. Una segunda alternativa es el modelado numérico de la CA, que pueden requerir de información no siempre disponible. Esta Tesis se propone explorar estos dos enfoques, desarrollando e integrando metodologías nuevas y adaptadas con el objetivo de caracterizar el impacto de los aerosoles en la CA. Finalmente, estos métodos se integran utilizando datos propios y de otras fuentes. Se eligió como caso de estudio al Centro-Norte del Área Metropolitana de Buenos Aires. Los productos y resultados conseguidos aportan nueva información sobre las cantidades, distribuciones y tipos de partículas en el aire. Los desarrollos y los métodos presentados pueden ser replicados y adaptados a otros sitios con las mismas problemáticas.

o. Resumen en portugués:

As partículas no ar são um importante indicador da qualidade do ar (CA), pois têm diferentes impactos na saúde e no clima. O monitoramento regulatório de CA é realizado por meio de estações de medição de superfície. O custo significa que esses tipos de medidas são geralmente dispersos ou inexistentes. Uma alternativa é dada por sensores de baixo custo, embora tenham inúmeras limitações. Uma segunda alternativa é a modelagem numérica do CA, que pode requerer informações que nem sempre estão disponíveis. Esta Tese tem como objetivo explorar essas duas abordagens, desenvolvendo e integrando metodologias novas e adaptadas de forma a caracterizar o impacto dos aerossóis na CA. Por fim, esses métodos são integrados a partir de dados próprios e de outras fontes. O Centro-Norte da Área Metropolitana de Buenos Aires foi escolhido como estudo de caso. Os produtos e resultados alcançados fornecem novas informações sobre as quantidades, distribuições e tipos de partículas no ar. Os desenvolvimentos e os métodos apresentados podem ser replicados e adaptados para outros sites com os mesmos problemas.

p. Resumen en inglés:

Particles in the air are an important indicator of air quality (AQ), as they have different impacts on health and climate. Regulatory AQ monitoring is carried out through surface measurement stations. The cost means that these types of measurements are usually scattered or non-existent. An alternative is given by low-cost sensors, although they have numerous limitations. A second alternative is the numerical modeling of the AQ, which may require information that is not always available. This thesis aims to explore these two approaches, developing and integrating new and adapted methodologies in order to characterize the impact of aerosols in AQ. Finally, these methods are integrated using own data and from other sources. The Center-North of the Metropolitan Area of Buenos Aires was chosen as a case

DOCTORADO EN CIENCIA Y TECNOLOGIA

Evaluado y acreditado por la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU).
Resolución N° 1178/11. Calificación "B".

study. The products and results achieved provide new information on the amounts, distributions, and types of particles in the air. The developments and the methods presented can be replicated and adapted to other sites with the same problems.

q. Aprobado por (Apellidos y Nombres del Jurado):

Érica Yanina Sanchez, Pablo Eduardo Levy y Rodolfo Manuel Echarri

Firma y aclaración de la firma del Presidente del Jurado:

Érica Yanina Sanchez

Firma del autor de la tesis:

Ariel Fabricio Scagliotti

DOCTORADO EN CIENCIA Y TECNOLOGIA

Evaluado y acreditado por la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU).
Resolución N° 1178/11. Calificación "B".

Integración de metodologías de bajo costo para la caracterización de la calidad del aire

Publicaciones:

Scagliotti, A. F., & Jorge, G. A. (2021). Caracterización de sensor de material particulado de bajo costo. *Tecnología y Ciencia UTN*, 42(1), 96-111.

Scagliotti, A. F., & Jorge, G. A. (2020, July). Análisis de un año de mediciones con fotómetro solar en el noroeste del conurbano bonaerense. In *ANALES AFA* (Vol. 31, No. 2, pp. 46-50).

Scagliotti, A. F., & Jorge, G. A. (2020). Inter-comparison of environmental low-cost sensors on Arduino platform. *Journal of the IEST*, 63(1), 35-45.

Aportes Originales:

En primer lugar, se propone una metodología propia para el diseño de Redes Neuronales Perceptrón Multicapa con una estrategia fuera del modelo para elegir predictores en base a vínculos no lineales con la variable objetivo. El modelo desarrollado, RED3, demostró buenos indicadores de desempeño y puede ser utilizado para completar la serie histórica de PM2.5 medida por la estación EMC I de ACUMAR en Dock Sud. Otro modelo desarrollado RED4, explora los alcances de estas redes en la caracterización de la calidad del aire.

En segundo lugar, se presenta y estudian los resultados de las primeras mediciones sistemáticas de propiedades ópticas de aerosoles en el Centro-Norte del conurbano bonaerense, a partir de un fotómetro solar calibrado de bajo costo. Este estudio, integrado con análisis de trayectorias de masas de aire y modelado con Redes Neuronales de tipo Mapas Auto-Organizados, determina las concentraciones verticales de aerosoles y tipos de partículas típicas presentes en la zona, además de caracterizar escenarios en función de las condiciones meteorológicas. No se encontraron antecedentes de uso de este tipo de modelos con mediciones fotométricas y la información dada es nueva y única para la región.

DOCTORADO EN CIENCIA Y TECNOLOGIA

Evaluado y acreditado por la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU).
Resolución N° 1178/11. Calificación "B".

En tercer lugar, se llevaron a cabo diferentes desarrollos tecnológicos de bajo costo y pruebas de sensores para realizar mediciones ambientales en el Centro-Norte del conurbano. Se desarrolló un banco de pruebas de sensores para medir en el exterior, y una cámara para pruebas de laboratorio con diseño propio. Luego, se realizó una caracterización de sensores meteorológicos, lo cual brinda información útil para seleccionar el modelo adecuado para un proyecto específico. También se evaluó el desempeño de un sensor óptico de material particulado a partir de pruebas de laboratorio, comparación con referencia y simulaciones, originando nueva información sobre las fortalezas y limitaciones del mismo. Finalmente, se muestra el desarrollo del prototipo de estación de medición de calidad del aire ECAUNGS y un estudio de las primeras campañas de mediciones sistemáticas en la zona.

Finalmente, se presenta una integración de los métodos con un modelado de Red Neuronal Perceptrón Multicapa alimentada con las mediciones de bajo costo. Este modelo ofrece nueva evidencia sobre las condiciones de la calidad del aire en la región, en base a predicción de material particulado y su dinámica en función de la meteorología y las fuentes locales. También senta un precedente metodológico replicable donde sea necesario.

Integración de metodologías de bajo costo para la caracterización de la calidad del aire

Ariel Fabricio Scagliotti

Integración de metodologías de bajo costo para la caracterización de la calidad del aire

Ariel Scagliotti

Tesis presentada para optar al título de:
Doctor en Ciencia y Tecnología

Director:
Guillermo Jorge

Codirector:
Sebastián Diez

Universidad Nacional de General Sarmiento
Año 2021

Agradecimientos

Se agradece por ceder datos y herramientas que fueron utilizados en esta Tesis a:

- La Agencia Nacional de Protección Ambiental de Buenos Aires (APrA)
- El Servicio Meteorológico Nacional (SMN)
- La Autoridad de Cuenca Matanza Riachuelo (ACUMAR)
- La Oficina Nacional de Administración Oceánica y Atmosférica, “National Oceanic and Atmospheric Administration” (NOAA), de Estados Unidos
- Al Centro de Archivos Activos Distribuidos de Procesos en Superficie de la Administración Nacional de Aeronáutica y el Espacio, “NASA’s Land Processes Distributed Active Archive Center (LP DAAC), de Estados Unidos.

Se agradece a Lidia Otero, Eduardo Quel y Brent Holben por sus esfuerzos en establecer y mantener el sitio de medición CEILAP BA de la Red Robótica de Aerosoles de la NASA, “NASAs Aerosol Robotic Network (AERONET)”, cuyos datos han sido utilizados en esta Tesis.

Se agradece al Ministerio de Ambiente y Desarrollo Sostenible de la Nación, especialmente a Norberto Vidal por su acompañamiento y colaboración.

Se agradece a la iniciativa “Cambike” del proyecto “open-seneca” de la Universidad de Cambridge en Reino Unido por compartir sus recursos y desarrollos.

A mis viejos, por todo.

A mis directores, por la paciencia y por, junto con muchas otras, buscar acercarme al conocimiento.

A David y Marce, por bancarme la parada más de una vez.

A Lili Romanelli, que me acompañó desde mi primer día en la UNGS, y a Alejandro Fendrik. Ambos me abrieron la cabeza y me dieron ejemplos gigantes a seguir.

A la gente loca y linda.

A la Pepe

“La simplicidad es la máxima sofisticación”
Leonardo Da Vinci

Aportes originales

En primer lugar, se propone una metodología propia para el diseño de Redes Neuronales Perceptrón Multicapa con una estrategia fuera del modelo para elegir predictores en base a vínculos no lineales con la variable objetivo. El modelo desarrollado, RED3, demostró buenos indicadores de desempeño y puede ser utilizado para completar la serie histórica de $PM_{2,5}$ medida por la estación EMC I de ACUMAR en Dock Sud. Otro modelo desarrollado RED4, explora los alcances de estas redes en la caracterización de la calidad del aire.

En segundo lugar, se presenta y estudian los resultados de las primeras mediciones sistemáticas de propiedades ópticas de aerosoles en el Centro-Norte del conurbano bonaerense, a partir de un fotómetro solar calibrado de bajo costo. Este estudio, integrado con análisis de trayectorias de masas de aire y modelado con Redes Neuronales de tipo Mapas Auto-Organizados, determina las concentraciones verticales de aerosoles y tipos de partículas típicas presentes en la zona, además de caracterizar escenarios en función de las condiciones meteorológicas. No se encontraron antecedentes de uso de este tipo de modelos con mediciones fotométricas y la información dada es nueva y única para la región. En tercer lugar, se llevaron a cabo diferentes desarrollos tecnológicos de bajo costo y pruebas de sensores para realizar mediciones ambientales en el Centro-Norte del conurbano. Se desarrolló un banco de pruebas de sensores para medir en el exterior, y una cámara para pruebas de laboratorio con diseño propio. Luego, se realizó una caracterización de sensores meteorológicos, lo cual brinda información útil para seleccionar el modelo adecuado para un proyecto específico. También se evaluó el desempeño de un sensor óptico de material particulado a partir de pruebas de laboratorio, comparación con referencia y simulaciones, originando nueva información sobre las fortalezas y limitaciones del mismo. Finalmente, se muestra el desarrollo del prototipo de estación de medición de calidad del aire ECAUNGS y un estudio de las primeras campañas de mediciones sistemáticas en la zona.

Finalmente, se presenta una integración de los métodos con un modelado de Red Neuronal Perceptrón Multicapa alimentada con las mediciones de bajo costo. Este modelo ofrece nueva evidencia sobre las condiciones de la calidad del aire en la región, en base a predicción de material particulado y su dinámica en función de la meteorología y las fuentes locales. También senta un precedente metodológico replicable donde sea necesario.

Publicaciones en revistas

- Scagliotti, A. F., & Jorge, G. A. (2021). Caracterización de sensor de material particulado de bajo costo. *Tecnología y Ciencia UTN*, 42(1), 96-111.
- Scagliotti, A. F., & Jorge, G. A. (2020, July). Análisis de un año de mediciones con fotómetro solar en el noroeste del conurbano bonaerense. In *ANALES AFA* (Vol. 31, No. 2, pp. 46-50).

- Scagliotti, A. F., & Jorge, G. A. (2020). Inter-comparison of environmental low-cost sensors on Arduino platform. *Journal of the IEST*, 63(1), 35-45.

Conferencias y ponencias

- Scagliotti, A. F.; Urquiza, J.; Diez, S. C. *Uncertainties propagated from optical properties in aerosol classification schemes*. International Global Atmospheric Chemistry, United Kingdom (2021).
- Scagliotti, A. F.; Margarit, D. H.; Reale, M. V. *Environmental data imputation with hybrid artificial neural networks techniques*. XVIII Taller Regional de Física Estadística y Aplicaciones a la Materia Condensada, Argentina (2021).
- Scagliotti, A. F. *Enfoques de bajo costo para la exploración de la calidad del aire*. Conferencia en: Foro Internacional de Vigilancia Atmosférica: “Experiencias en Latinoamérica y el Caribe”. SENAMHI, Perú (2021).
- Scagliotti, A. F. *Desarrollo de estación de calidad del aire y plan de monitoreo ambiental*. Ponencia en: Jornadas “El trabajo experimental en la Universidad”. IDEI, UNGS (2019).
- Scagliotti, A. F.; Jorge, G. A. *Diseño y desarrollo de sistema de monitoreo de calidad del aire*. Ponencia en: “1ra Jornada de Vinculación Tecnológica y Social”. SDTyS, UNGS (2018).
- Scagliotti, A. F.; Llera, M. A.; Jorge, G. A. *Plan de diseño, desarrollo e instalación de estaciones remotas de medición de calidad de aire de bajo costo en el noroeste del conurbano bonaerense*. Ponencia en: XIII Congreso Argentino de Meteorología. CAM (2018).

Resumen

La calidad del aire (CA) es una creciente preocupación en todo el mundo, principalmente en zonas urbanas, debido a sus bien documentados efectos adversos sobre la salud. Las partículas en el aire (líquidas y/o sólidas) son un indicador importante al respecto pues, dependiendo de su tamaño, tienen diferentes impactos sobre el bienestar de las personas y sobre el clima. El monitoreo de la CA de forma regulatoria se realiza a través de estaciones de medición en superficie, pero los equipos que se requieren son extremadamente costosos. Es por ello que este tipo de mediciones son dispersas o inexistentes, especialmente en países como Argentina con bajos recursos económicos. Una alternativa viene dada por los denominados sensores de bajo costo, los cuales en principio permiten realizar mediciones en regiones sin antecedentes, aunque poseen numerosas limitaciones y se requiere una exhaustiva caracterización previo a su uso. Una segunda alternativa es el modelado numérico de la CA, que, dependiendo del enfoque, puede requerir de datos, información y/o capacidad de procesamiento no siempre disponibles. Esta Tesis se propone explorar estos dos enfoques, desarrollando e integrando metodologías nuevas y adaptadas con el objetivo de caracterizar el impacto de los aerosoles en la CA, intentando minimizar simultáneamente los costos y las incertidumbres. Por un lado, se muestra el desarrollo y la caracterización de equipos de medición y la implementación de instrumentos de bajo costo. Por otro, se propone la modelización de material particulado y de propiedades ópticas de aerosoles a través de la optimización de Redes Neuronales Artificiales. Finalmente, estos métodos se integran utilizando datos propios y de otras fuentes. Se eligió como caso de estudio al Centro-Norte del Área Metropolitana de Buenos Aires, donde hay desconocimiento acerca de la CA y características que hacen a una población vulnerable a los efectos de la contaminación ambiental. Los productos y resultados conseguidos aportan nueva información sobre las cantidades, distribuciones y tipos de partículas en el aire, apta como insumo para la toma de decisiones y planificación urbana. Los desarrollos, en cuanto a modelos y equipos de medición, y los métodos presentados pueden ser replicados y adaptados a otros sitios con las mismas problemáticas.

Nomenclatura

18b20	Sensor de temperatura DS18b20
α	Coefficiente de Ångstrom
ACUMAR	Autoridad de Cuenca Matanza Riachuelo (referido a la estación EMC I de Dock Sud)
<i>alphaCEI</i>	Coefficiente de Ångstrom para sitio CEILAP
<i>alphaUNGS</i>	Coefficiente de Ångstrom para sitio UNGS
ANN	Red neuronal artificial
AOD	Espesor óptico de aerosoles
<i>AODcei</i>	AOD para sitio CEILAP
<i>AODungs</i>	AOD para sitio UNGS
<i>BLHcaba</i>	Altura de capa límite en CABA
<i>BLHungs</i>	Altura de capa límite en UNGS
BME	Sensor de humedad relativa, presión atmosférica y temperatura BME280
BMP	Sensor de temperatura y presión atmosférica BMP180
CA	Calidad del aire
CABA	Ciudad Autónoma de Buenos Aires
CEILAP	Sitio de AERONET CEILAP BA
<i>classCEI</i>	Clase de aerosoles en sitio CEILAP
<i>classUNGS</i>	Clase de aerosoles en sitio UNGS
DER	Desviación estándar relativa a instrumento de referencia
DHT	Sensor de temperatura y humedad relativa DHT11
<i>diasem</i>	Día de la semana
ECAUNGS	Estación de medición de calidad del aire desarrollada en esta Tesis
<i>estab</i>	Estabilidad atmosférica
<i>f(RH)</i>	Factor de crecimiento higroscópico de aerosoles
<i>HCT</i>	Hidrocarburos totales
<i>HR</i>	Humedad relativa
<i>HRacu</i>	Humedad relativa medida en sitio ACUMAR
<i>HRcm</i>	Humedad relativa medida en estación SMN de Campo de Mayo
<i>HReca</i>	Humedad relativa medida por ECAUNGS
<i>invDp</i>	Diferencia de temperatura de rocío en una capa de inversión
<i>invH</i>	Altura de capa de inversión
<i>invT</i>	Diferencia de temperatura ambiente en una capa de inversión
LCS	Sensor de bajo costo

MLP	Red neuronal perceptrón multicapa
MLP-FF-BP	Red neuronal perceptrón multicapa con flujo de avance y propagación inversa
NOAA	Administración Oceánica y Atmosférica Nacional de EUA (NOAA READY)
<i>nub</i>	Nubosidad
<i>Pcm</i>	Presión atmosférica medida en estación SMN de Campo de Mayo
<i>Peca</i>	Presión atmosférica medida por estación ECAUNGS
<i>pla</i>	Altura de capa de nubes (plafond)
PM	Material particulado
<i>PMeca</i>	Material particulado medido por ECAUNGS
<i>Pacu</i>	Presión atmosférica medida en sitio ACUMAR
<i>PPTacu</i>	Precipitación medida en sitio ACUMAR
PM_1	Material particulado de tamaño menor a $1 \mu\text{m}$
<i>PM1eca</i>	PM_1 medido por ECAUNGS
$PM_{2,5}$	Material particulado de tamaño menor a $2,5 \mu\text{m}$
<i>PM25acu</i>	$PM_{2,5}$ medido en sitio ACUMAR
<i>PM25eca</i>	$PM_{2,5}$ medido por ECAUNGS
PM_{10}	Material particulado de tamaño menor a $2,5 \mu\text{m}$
<i>PM10acu</i>	PM_{10} medido en sitio ACUMAR
<i>PM10apra</i>	PM_{10} medido en sitio La Boca de APrA
<i>PM10eca</i>	PM_{10} medido por ECAUNGS
<i>PM10/PM1</i>	Razón de las concentraciones de PM_{10} y PM_1
PMS5003	Sensor de material particulado de ECAUNGS
<i>PPTungs</i>	Precipitación acumulada para la zona de la UNGS
PS	Sonda de temperatura de respuesta rápida PS-2135 (usada como referencia en laboratorio)
q_e	Error de cuantización de SOM
<i>RADacu</i>	Radiación solar medida en sitio ACUMAR
RED1	Red MLP-FF-BP de prueba entrenada con todas las variables disponibles en CABA para estimar $PM_{2,5}$
RED2	Red MLP-FF-BP con los mismos hiperparámetros que RED1, pero con los 21 predictores elegidos para estimar $PM_{2,5}$ en CABA
RED3	Red MLP-FF-BP dinámica recurrente con mismos hiperparámetros que RED2, pero con realimentación de la salida a la capa oculta, para estimar $PM_{2,5}$ en CABA
RED4	Red MLP-FF-BP alimentada sólo con variables meteorológicas y temporales, para estimar $PM_{2,5}$ en CABA

REDP	Red MLP-FF-BP de prueba con los hiperparámetros de la RED1, pero con solo una neurona de entrada para probar las variables una por una y estimar $PM_{2,5}$ en CABA
SMN	Servicio Meteorológico Nacional
SOM	Mapa auto-organizado
SOM3	Modelo SOM final utilizado con las propiedades ópticas de aerosoles y meteorología en UNGS
<i>Tacu</i>	Temperatura medida en sitio ACUMAR
<i>Tcm</i>	Temperatura medida en estación SMN de Campo de Mayo
t_e	Error topográfico de SOM
<i>Teca</i>	Temperatura medida por ECAUNGS
<i>Troc</i>	Temperatura de rocío
<i>u850</i>	Componente zonal del viento a altura referida a 850 hPa
<i>Uacu</i>	Componente zonal para sitio ACUMAR
<i>Ucm</i>	Componente zonal para estación SMN de Campo de Mayo
UNGS	Universidad Nacional de General Sarmiento
UNGS WS	estación meteorológica WT1081 instalada en UNGS (usada como referencia en exterior)
<i>v850</i>	Componente meridional del viento a altura referida a 850 hPa
<i>Vacu</i>	Componente meridional del viento para sitio ACUMAR
<i>Vcm</i>	Componente meridional del viento para estación SMN de Campo de Mayo
<i>vis</i>	Visibilidad
<i>wd850</i>	Dirección del viento en altura referida a 850 hPa
<i>WDacu</i>	Dirección del viento medida en sitio ACUMAR
<i>WDcm</i>	Dirección del viento medida en estación SMN de Campo de Mayo
<i>WSacu</i>	Velocidad del viento medida en sitio ACUMAR
<i>ws850</i>	Velocidad del viento en altura referida a 850 hPa
ws	Velocidad del viento
<i>WScm</i>	Velocidad del viento medida en estación SMN de Campo de Mayo

Índice general

Agradecimientos	5
Aportes originales	7
Resumen	9
Nomenclatura	10
1 Introducción general	15
2 Marco teórico y consideraciones generales	23
2.1 Redes Neuronales para la caracterización y modelado de datos ambientales	23
2.1.1 Bases de las Redes Neuronales MLP-FF-BP	27
2.1.2 Bases de las ANN tipo SOM	38
2.2 Fotometría solar y propiedades ópticas de aerosoles	41
2.2.1 Bases de las mediciones con fotómetro solar	45
2.3 Mediciones ambientales de bajo costo	46
2.3.1 Respecto al funcionamiento de los sensores de bajo costo	48
3 Metodología	57
3.1 Potencialidad de las Redes Neuronales en calidad del aire	57
3.1.1 Exploración de datos meteorológicos y de contaminantes	58
3.1.2 Modelado de PM	62
3.2 Caracterización de aerosoles	65
3.2.1 Exploración de datos meteorológicos	66
3.2.2 Mediciones fotométricas de bajo costo	69
3.2.3 Potenciales escenarios	72
3.3 Desarrollo e implementación de equipos de bajo costo	74
3.3.1 Evaluación de sensores meteorológicos	75
3.3.2 Caracterización de sensor de PM	81
3.3.3 Integración de metodologías de bajo costo	88
4 Potencialidad de las Redes Neuronales en calidad del aire	95
4.1 Objetivos e interés	95
4.2 Resultados y discusiones	96
4.2.1 Exploración de datos meteorológicos y de contaminantes	96
4.2.2 Modelado de PM	109
4.3 Conclusiones preliminares	119

5	Caracterización de aerosoles	121
5.1	Objetivos e interés	121
5.2	Resultados y discusiones	121
5.2.1	Exploración de datos meteorológicos	121
5.2.2	Mediciones fotométricas de bajo costo	125
5.2.3	Potenciales escenarios	132
5.3	Conclusiones preliminares	144
6	Desarrollo e implementación de equipos de bajo costo	149
6.1	Objetivos e intereses	149
6.2	Resultados y discusiones	150
6.2.1	Evaluación de sensores meteorológicos	150
6.2.2	Caracterización de sensor de PM	154
6.2.3	Integración de metodologías de bajo costo	168
6.3	Conclusiones preliminares	183
7	Conclusiones generales	187
8	ANEXOS	193
8.1	Potencialidad de las redes neuronales en calidad del aire	193
8.1.1	Conversión de datos de plafond	193
8.1.2	Método para el cálculo de estabilidad atmosférica	193
8.1.3	Tendencias de las variables de entrada del modelo RED3 no incluidas en el Capítulo 4	196
8.1.4	Matriz de correlación entre todas las variables consideradas como potenciales entradas del modelo RED3	198
8.2	Caracterización de aerosoles	199
8.2.1	Tendencias de variables no incluidas en el análisis del Capítulo 5	199
8.2.2	Calibración de fotómetro solar CALITOO	200
8.3	Desarrollo e implementación de equipos de bajo costo	202
8.3.1	Informe de calibración de Open-Seneca para equipo de Cambike	202
8.3.2	Correlaciones lineales entre todas las potenciales variables de entrada para el modelo REDungs	205
	Bibliografía	206

1 Introducción general

La calidad del aire (CA) es uno de los mayores problemas mundiales de la actualidad, pues su empobrecimiento está ligado a mortalidad prematura, especialmente en niños, adultos mayores y personas con problemas respiratorios y cardiovasculares [1, 2, 3, 4]. Según la Organización Mundial de la Salud (OMS), la contaminación del aire representa el mayor riesgo ambiental, ligado a cerca de siete millones de muertes cada año [5]. Además, la mayoría de las grandes ciudades no cumplen con los estándares de CA de la OMS [6]. Es por ello que resulta imprescindible avanzar en la comprensión de los procesos de contaminación del aire y en el desarrollo de modelos que los caractericen, pues el conocimiento al respecto es una herramienta indispensable en aplicaciones regulatorias, de investigación y forenses [7].

Uno de los principales indicadores de la CA son los aerosoles, partículas sólidas o líquidas (polvo, polen, hollín, metales, sales, gotas de ácido sulfúrico, dioxinas, pesticidas, etc.) que permanecen suspendidas en la atmósfera desde varias horas hasta algunas semanas. Pueden ser clasificados como primarios emitidos desde la superficie del planeta por eventos naturales (erupciones volcánicas, tormentas de polvo, etc.) o actividades antrópicas (combustión de combustibles fósiles en vehículos, plantas generadoras de energía, procesos industriales, etc.), o secundarios formados en la atmósfera por procesos químicos. Cercanos a la superficie, los aerosoles suelen ser caracterizados como concentraciones de material particulado (PM, por “particulate matter”) utilizando sistemas de monitoreo continuo/discontinuo en todo el mundo (como gravimétrico, atenuación beta, nefelometría, etc.) que brindan información sobre los impactos locales. Por otro lado, la carga atmosférica de aerosoles se representa comúnmente como espesor óptico de aerosoles (AOD, por “aerosol optical depth”) utilizando instrumentación óptica (por ejemplo, fotómetros solares, lidars, etc.) desplegados en la superficie de la tierra y/o en plataformas satelitales.

A causa de su irregularidad en forma, densidad, composición y estructura, el PM suele ser caracterizado por su tamaño. De acuerdo al diámetro aerodinámico equivalente, recurriendo a la simplificación de partículas esféricas con densidad igual a la unidad, la

1 Introducción general

concentración másica de partículas por unidad de volumen suele clasificarse en PM_{10} (partículas de tamaños menores a $10 \mu\text{m}$), $PM_{2,5}$ (partículas de tamaños menores a $2,5 \mu\text{m}$) y PM_1 (partículas de tamaños menores a $1 \mu\text{m}$). La mayor concentración de partículas se encuentra en la tropósfera, especialmente en la capa límite atmosférica donde se ubica la mayoría de las fuentes de emisiones [8] y donde los aerosoles se ven dispersados y mezclados por procesos de turbulencia y convección.

La medición y comprensión de la dinámica de las partículas en el aire es crucial dados sus impactos directos (CA) e indirectos (clima) en la salud humana [9, 10, 11, 12]. Algunas partículas pueden ser transportadas a través del tracto respiratorio superior hacia los bronquiolos y los pulmones produciendo peligros directos para la salud, mientras que las partículas más finas pueden penetrar más profundamente en el sistema respiratorio y depositarse en los pulmones aumentando la resistencia al flujo de aire [13]. La Agencia Internacional para la Investigación del Cáncer (IARC) determinó en 2013 que el PM es cancerígeno para los seres humanos [14], y la Agencia Europea de Medio Ambiente (EEA) atribuyó alrededor de 400.000 muertes prematuras en 41 países europeos a $PM_{2,5}$ en el aire [15]. Los aerosoles también modifican la transmisión de radiación tanto de onda corta como de onda larga mediante procesos de dispersión y absorción que producen efectos de enfriamiento y calentamiento. Esto puede alterar la microfísica de las nubes, el ciclo hidrológico y la estabilidad atmosférica, por nombrar algunas consecuencias bien documentadas [16, 17]. Sin embargo, los efectos de los aerosoles sobre el clima son complejos y existe una gran incertidumbre sobre su comportamiento y sus impactos, dada su variabilidad física y química [18]. Cabe aclarar que el interés actual de investigación acerca de PM se está centrando en las partículas de menor tamaño por su mayor incidencia sobre la salud humana, particularmente en $PM_{2,5}$ [19, 20].

El comportamiento del PM en la atmósfera depende tanto de las características de la zona en estudio (niveles de industrialización, densidad poblacional, tráfico vehicular, topografía) como de las variables meteorológicas a diferentes escalas espaciales y temporales [21]. En las últimas décadas se han desarrollado varios tipos de modelos de dispersión de contaminantes, estos han sido aplicados en diferentes ciudades del planeta con el fin de estimar los niveles de concentración en una distribución espacial y temporal en función de variables meteorológicas de superficie y propias de capa límite atmosférica. Los modelos más utilizados son los Gaussianos, que simulan el comportamiento de plumas de contaminantes en condiciones estacionarias, como el AERMOD [22]; los Lagrangianos, que estudian la trayectoria de una parcela de aire en el tiempo, como el NAME [23]; los Eulerianos, que utilizan sistemas de grilla para monitorear las

propiedades atmosféricas, como el modelo unificado EMEP [24]. Otros modelos simulan la emisión, transporte, mezcla y transformaciones químicas de contaminantes bajo diferentes condiciones meteorológicas en una escala regional, como el WRF-chem [25]. Todos estos modelos determinísticos pueden ser utilizados sin la necesidad de una gran cantidad de datos históricos (que no siempre están disponibles, si es que siquiera existen), pero requieren de conocimiento experto e información con la que no siempre se cuenta (acerca de las fuentes de contaminantes, las tasas de emisiones en tiempo real y las reacciones químicas involucradas en la transformación de gases), además de un alto poder de procesamiento computacional [26].

Un enfoque distinto de modelado lo proporciona la minería de datos a partir de la inteligencia artificial, en el cual se intenta extraer la mayor cantidad de información posible con los datos disponibles. Las Redes Neuronales Artificiales (ANN, por Artificial Neural Networks”) son modelos capaces de interpretar la no linealidad entre variables, tienen un procesamiento computacional optimizado, son robustas y tolerantes al ruido y cuentan con amplias capacidades de aprendizaje y generalización [27]. Estos modelos han sido utilizado en numerosos campos de conocimiento y requieren un diseño e ingeniería específica para cada problema en cuestión.

Por otro lado, las mediciones de contaminantes a nivel de superficie requieren instrumentos generalmente muy costosos y equipos de profesionales para mantenerlos y administrarlos. Estas condiciones y requerimientos derivan en que se cuente en muchas ciudades del mundo, especialmente en aquellas con bajos recursos, con una cobertura limitada y una distribución irregular de estaciones de monitoreo [28]. Además, fuera de las grandes urbes, en zonas conurbanas o interiores de los países no se cuenta generalmente con ningún tipo de medición.

La sede central de la Universidad Nacional de General Sarmiento (UNGS) [29], donde se realizaron los desarrollos y mediciones para esta Tesis, se encuentra en una de estas zonas sin cobertura, emplazada cerca de la intersección entre los partidos de Malvinas Argentinas, San Miguel y J.C.Paz (con una población total cercana al millón de habitantes en una superficie de aproximadamente 200 km²), pertenecientes a la región Centro-Norte del conurbano bonaerense (parte de la también llamada Área o Región Metropolitana de Buenos Aires), a poco más de 30 km del centro de la Ciudad Autónoma de Buenos Aires (CABA). Esta zona pertenece a lo que se denomina segundo cordón del conurbano, el cual linda con el primer cordón que limita con CABA (con mayor densificación urbana) por un lado, y con zonas rurales y semi-rurales por el otro. En la **Figura 1.1** se puede observar la ubicación de la UNGS con sus alrededores y respecto

1 Introducción general

a CABA. El gran espacio verde en sus cercanías es la guarnición militar de Campo de Mayo.

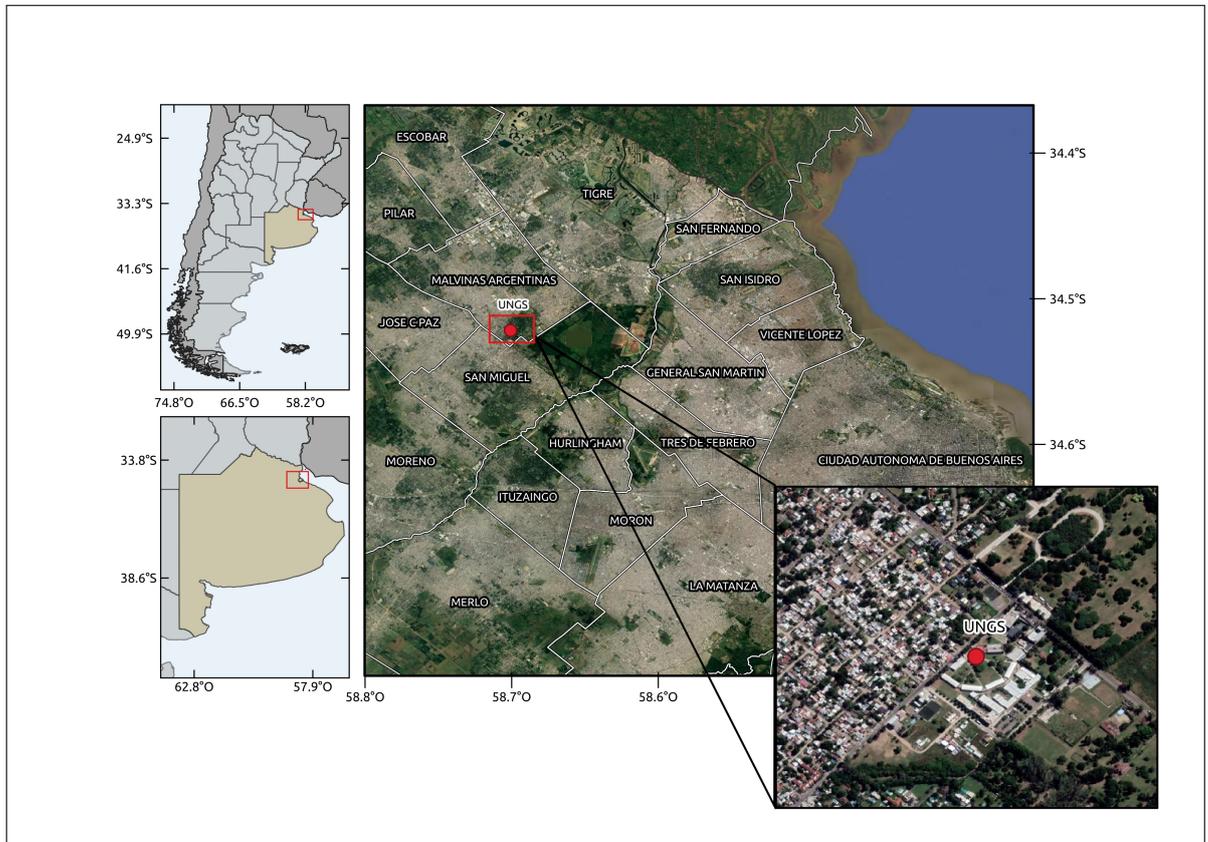


Figura 1.1: Ubicación de la UNGS. Figura desarrollada con software QGIS

El conurbano bonaerense no posee un conglomerado urbano continuo, sino que se extiende de forma “tentacular” en paralelo a las principales vías de transporte. Esto produce una heterogeneidad en la densidad poblacional, infraestructura y dotación de servicios esenciales que delinea las diferentes realidades socio-económicas de la región. Muchos sectores de la población se encuentran particularmente vulnerables a las problemáticas ambientales centradas en la calidad del agua y del aire[30]. Respecto a las preocupaciones vinculadas a la contaminación aérea, hay tres fuentes importantes a destacar en esta parte del conurbano[30]:

- El parque automotor ha crecido aceleradamente junto con la mancha urbana y de forma no planificada. Debido a la distribución particular de las ciudades, hay sectores con casi constantes atascos de tráfico sumados al resto de las actividades en lugares confinados por edificios. Todo esto concentrado en una superficie pequeña.

- La gestión de residuos ha sido una problemática casi constante en la zona. Si bien el relleno sanitario CEAMSE (ubicado a unos 9 km de la UNGS) se gestiona bajo regulaciones, sus efectos sobre la CA se hacen notar durante horas nocturnas, especialmente en verano. Además, hay dispersos varios basurales a cielo abierto improvisados por comercios, industrias y los mismos vecinos. Esta situación empeora con las quemas de basura y de biomasa en ciertas épocas del año.
- La mayoría de las industrias se ubican en Parques Industriales con estrictas normas de gestión ambiental. Pero existen también muchas otras emplazadas de manera dispersa, particularmente en zonas relegadas y de bajos niveles socioeconómicos, sin un control adecuado. De hecho, la información pública disponible sobre la actividad industrial de la zona se limita a los parques y es imprecisa o inexistente acerca del resto.

Luego de un primer relevamiento en la zona, no se encontraron datos sobre indicadores de la CA, ni la presencia de estaciones de monitoreo, de muestreo o cualquier tipo de iniciativa que se oriente a un control sistemático. Fagundez et al. (2001) [31] diseñaron instrumentos de muestreo pasivos que implementaron durante una campaña en el partido de San Miguel con el fin de medir material particulado total, NO_2 y SO_2 . Encontraron que las mayores concentraciones de contaminantes se hayan en las zonas céntricas con mayor tránsito vehicular. Este es el único antecedente encontrado.

En CABA y alrededores, se encuentran registradas nueve estaciones de medición de CA (gases y material particulado). Entre ellas, se encuentran actualmente en funcionamiento tres correspondientes a la Agencia de Protección Ambiental de Buenos Aires (APrA) [32] y cuatro a la Autoridad de Cuenca Matanza-Riachuelo (ACUMAR) [33]. Existen dos locaciones (CEILAP-BA y CEILAP-UTN) pertenecientes a la red AERONET (por “Aerosol Robotic Network”) [34] dedicada a mediciones de propiedades ópticas de aerosoles. En la **Figura 1.2** se muestra un mapa de la Red Federal de Monitoreo Ambiental (FEMA) [35] donde figuran las estaciones de superficie de medición de calidad de aire y de AERONET más cercanas a UNGS.

1 Introducción general

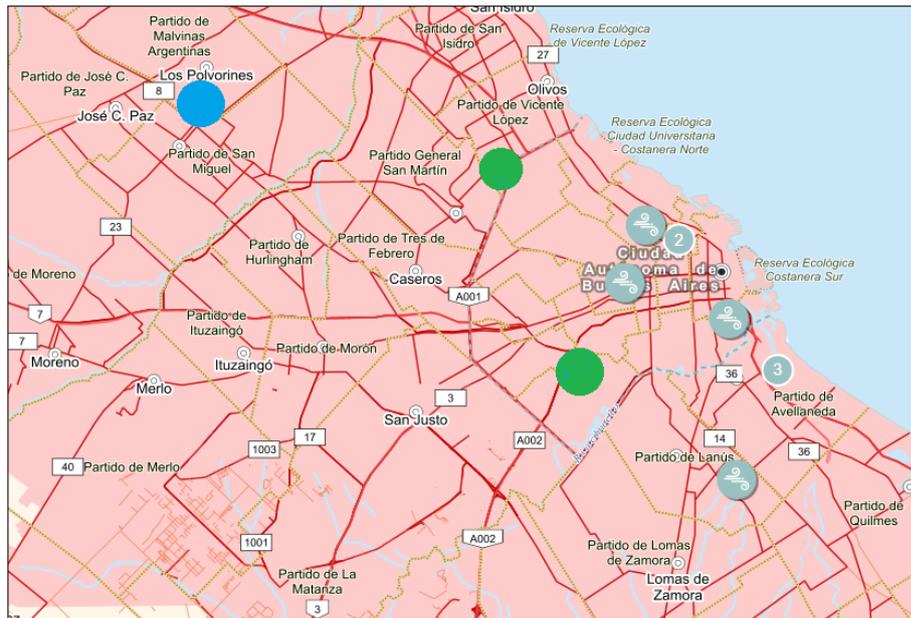


Figura 1.2: Mapa de la red FEMA. Se muestran con los símbolos grises las estaciones de calidad de aire más cercanas a la UNGS, que se encuentran en CABA y alrededores, y se agregaron las estaciones de AERONET con círculos verdes y la ubicación de la UNGS con un círculo celeste. Extraída y modificada de red FEMA [35]

Objetivos y organización de la Tesis

El objetivo principal de esta Tesis es integrar diferentes enfoques de bajo costo con el fin de determinar la CA, en cuanto a concentraciones y tipos de partículas, en una región sin antecedentes. Aquí se desprenden las siguientes preguntas de investigación enfocadas geográficamente en el Centro-Norte del conurbano bonaerense:

- ¿Cuál es el potencial de las Redes Neuronales Artificiales para la caracterización de la calidad del aire? ¿Cómo depende este potencial de la disponibilidad de mediciones?
- ¿Cuáles son las fuentes principales de aerosoles a nivel regional y local?
- ¿Cómo influye la meteorología en la dinámica y comportamiento de los aerosoles?
- ¿Cuál es el estado general de la CA en cuanto a partículas en la zona?

- ¿Qué desempeño se puede esperar de los sensores de bajo costo a la hora de entregar información ambiental?
- ¿Es posible determinar la CA en una zona sin antecedentes a partir de métodos de bajo costo?

Está probado que las ANN son modelos potentes y que ya han sido utilizados con éxito en CA, pero fallan ante una limitada disponibilidad de datos. La metodología propuesta aquí busca desarrollar modelos de bajo costo (computacional y en disponibilidad de datos) optimizando su desempeño. Respecto a las fuentes de partículas, se puede suponer que el transporte motorizado sea la principal a nivel local, y se espera identificar diferentes tipos a nivel regional. Si bien no hay muchos antecedentes al respecto, las características orográficas y meteorológicas de toda el Área Metropolitana de Buenos Aires (CABA y conurbano) deberían propiciar buenas condiciones de dispersión de los contaminantes. De todas formas, se prevé identificar condiciones vinculadas a eventos de mayor contaminación del aire. En base a la literatura y antecedentes relevados, se espera que los resultados de esta Tesis permitan demostrar que es posible estudiar la CA en base a desarrollos tecnológicos de bajo costo y modelización optimizada, estableciendo las limitaciones y definiendo precauciones adecuadas.

El enfoque propuesto se divide en dos categorías principales, uno de modelado computacional y otro experimental (y de desarrollo). Por un lado, se propone el diseño y uso de ANN a modo de modelos predictivos y de clasificación con un bajo costo computacional y optimizando su desempeño en base a las variables disponibles. Por otro lado, la categoría experimental se centra en el uso y desarrollo de tecnología de bajo costo económico que provea datos e información de manera sistemática sobre indicadores de CA en la región. Finalmente, se integran las mediciones realizadas a partir de equipos de bajo costo con modelos ANN para clasificar diferentes escenarios de aerosoles y para predecir PM.

Este capítulo introduce el trabajo y la estructura. El **Capítulo 2** desarrolla el marco teórico con el estado del arte actual, antecedentes, conceptos y teorías referido a lo trabajado en esta Tesis. El **Capítulo 3** describe la metodología utilizada para cada parte. El **Capítulo 4** detalla los resultados del diseño e implementación de modelos de predicción de PM, desarrollados con diferentes datos disponibles. El **Capítulo 5** incluye el análisis de datos meteorológicos del Centro-Norte del conurbano, el uso de tecnología óptica de bajo costo con mediciones de fotometría solar y el modelado de escenarios de espesor óptico de aerosoles con ANN de aprendizaje no supervisado. En

1 Introducción general

el **Capítulo 6** se presenta la caracterización de sensores ambientales y el desarrollo de la mini-estación de calidad de aire ECAUNGS, junto con el estudio de sus mediciones preliminares. Estas mediciones y el conocimiento generado en el **Capítulo 4** se utilizan para implementar un modelo ANN de predicción de PM para la zona. Finalmente, se presentan las conclusiones generales.

Los productos de esta tesis, en cuanto a desarrollo tecnológico y modelos (diseños, códigos, etc..) son de libre acceso en un repositorio digital creado para tal fin [36], con el espíritu de aportar un conocimiento compartido de manera abierta que pueda ser replicado en zonas con las mismas problemáticas y limitaciones.

2 Marco teórico y consideraciones generales

2.1. Redes Neuronales para la caracterización y modelado de datos ambientales

Los sistemas ambientales suelen ser complejos al involucrar numerosos procesos y variables interrelacionadas. Su estudio implica desafíos que atañen a la multidisciplinariedad, a estructuras complejas y dominios no lineales, a la alta dimensionalidad y multiescalaridad, a datos heterogéneos, a incertezas e imprecisión de la información y a comportamientos cíclicos, entre otros aspectos [37]. Todas estas características en el estudio de los sistemas ambientales se acentúan a partir de la inclusión de la actividad humana en su dinámica.

Diferentes técnicas basadas en minería de datos han lidiado con esta complejidad y facilitado herramientas que han demostrado ser superadoras a los enfoques estadísticos tradicionales [38]. Estas técnicas se pueden definir como un conjunto de procesos orientados a descubrir conocimiento, como información implícita con potencial de ser útil, de bases de datos extensas [39]. Un ejemplo de ello son las Redes Neuronales Artificiales (ANN), árboles de decisión, algoritmos genéticos, método del vecino más cercano y reglas de inducción, entre otras. Las tareas que permiten llevar a cabo abarcan agrupamiento, clasificación, extracción de asociaciones, predicción y regresión [37].

A partir de una de estas tareas o una combinación de ellas se puede predecir, es decir estimar datos de una variable con valores medidos de otra/s magnitud/es ambiental/es. Esta técnica puede utilizarse para, por ejemplo, imputar datos de una serie de contaminantes atmosféricos [40]. Cabe mencionar que en una serie temporal pueden encontrarse huecos de diferentes longitudes (cantidad de datos faltantes consecutivos) y distribuidos de diferentes maneras, y las técnicas de imputación pueden ser simples o múltiples (se usa un resultado o se combinan varios para completar un valor) y univariados o multiva-

riados (si se utiliza una o varias variables en la técnica). Entre los métodos tradicionales de imputación se encuentran: ignorar los datos faltantes (lo cual puede modificar la estructura inherente de la serie), completar con valores medios (lo cual produce un sesgo) e implementar interpolaciones lineales, spline y de vecinos cercanos, todos métodos univariados. En general, las técnicas de ANN muestran mejores resultados en predicción que los métodos tradicionales, principalmente debido a su capacidad de trabajar con relaciones no lineales entre las variables [41, 42].

Las ANN, pertenecientes al enfoque de aprendizaje automático y también conocidas como modelos conexionistas o de procesamiento de distribución paralela, consisten en un conjunto de unidades simples de procesamiento que se comunican a través de múltiples conexiones. El interés en su estudio surgió a partir de la presentación de neuronas simples por parte de Macculloch & Pitts (1943) [43]. Estas unidades o nodos como parte conceptual de circuitos que pueden llevar a cabo procesamientos computacionales, ingresándoles una entrada y obteniendo una salida, fueron llamados neuronas y al sistema completo red neuronal, debido al paralelo con el sistema nervioso en redes biológicas, aunque se trate de una sobre-simplificación. El clásico modelo perceptrón fue propuesto por Rosenblatt (1959) [44], quien incluyó pesos numéricos representando las conexiones entre neuronas. Este modelo fue perfeccionado y caracterizado por Minsky & Papert (1969) [45]. El modelo de Perceptrón Multicapa (MLP, por su denominación en inglés "Multi Layer Perceptron") es la forma más utilizada de ANN [46].

Todas las Redes Neuronales comparten un conjunto de aspectos principales [47], siendo los más relevantes: un conjunto de unidades de procesamiento (neuronas, nodos, celdas o unidades); un estado de activación y_k para cada neurona, que es equivalente a su salida; conexiones entre neuronas, cada una definida con un peso w_{jk} que determina el efecto que la señal de una neurona j tiene en una neurona k ; distribución jerárquica de las neuronas en la red, por lo que debe existir una norma o regla que indique la forma en que se propaga la información; una función de activación F_k y una entrada externa (sesgo o "bias") b_k en cada neurona; una regla de aprendizaje que indique de qué manera la información se genera y almacena en la red.

Estos sistemas "aprenden" en el sentido del procesamiento de actualizaciones de su estructura interna en respuesta a señales externas con el fin de desarrollar una tarea particular. Específicamente, el aprendizaje ocurre a través de un algoritmo numérico que actualiza los pesos entre neuronas. Diferentes modelos de ANN difieren principalmente en su finalidad (agrupamiento, clasificación, regresión), los patrones de interconexión (parcial o completo), la dirección del flujo de la información (recurrentes y no recurren-

tes), el tipo de algoritmo de aprendizaje, la regla de aprendizaje y el grado de supervisión necesitada durante el entrenamiento [27].

Las ANN han sido utilizadas en diferentes aplicaciones a sistemas ambientales a lo largo de los años, tales como deforestación [48], hidrología [49], calidad del agua [50], series temporales ecológicas [51], procesos de tratamientos de aguas residuales [52], predicción de magnitudes en sismos [53], entre otros.

También se encuentran antecedentes en las ciencias meteorológicas, en donde las ANN se han utilizado para predecir la ocurrencia de tornados [54], tormentas eléctricas [55], monzones [56], radiación solar diaria [57], para clasificación de nubes [58] y de patrones de circulación atmosférica [59], por mencionar algunos ejemplos.

En el ámbito de la calidad del aire, específicamente en PM (material particulado), las ANN han sido ampliamente utilizadas para pronóstico (estimar valores a tiempo futuro) y predicción (estimación) en diferentes partes del mundo con gran éxito en contraste con enfoques tradicionales como regresiones lineales, series temporales y promedio móvil integrado autoregresivo (ARIMA, por su denominación en inglés) [7]. Voukantsis et al. (2011) [60] han propuesto un enfoque combinando análisis de componentes principales con ANN para predecir concentraciones de PM_{10} y $PM_{2,5}$ en Thessaloniki y Helsinki, obteniendo coeficientes de correlación entre 0,65 y 0,8; Franceschi et al. (2018) [21] utilizaron ANN, análisis de componentes principales y agrupación “k-means” para pronosticar y descubrir relaciones entre PM_{10} y $PM_{2,5}$ en Bogotá, consiguiendo una raíz del error cuadrático medio RMSE (por “root mean square error”) de $14,01 \mu g m^{-3}$ para PM_{10} y de $9,34 \mu g m^{-3}$ para $PM_{2,5}$; Feng et al. (2015) [26] implementaron un modelo geográfico basado en ANN con transformación wavelet para pronosticar $PM_{2.5}$ utilizando trayectorias de masas de aire en diferentes regiones de China logrando un RMSE entre $15,65 \mu g m^{-3}$ y $19,75 \mu g m^{-3}$ (para predicción con 24h de anticipación); Ceylan & Bulkan (2018) [61] se valieron de ANN y regresión lineal múltiple para predecir niveles de PM_{10} en Sakarya con un coeficiente de correlación en la etapa de prueba de la red cercano a 0,93; Asghari & Nematzadeh (2016) [62] aplicaron un método híbrido de algoritmo genético y Red Neuronal Perceptrón Multicapa para predecir concentraciones de PM_{10} en Tehran con un coeficiente de correlación final de aproximadamente 0,74. Cabe aclarar que en general pronosticar resulta más difícil que predecir, porque modelizar a tiempo futuro se enfrenta a más imponderables e incertezas.

Enfoques de aprendizaje automático aplicados a calidad del aire se han utilizado mayormente en Norteamérica y Eurasia, mientras que Sudamérica y Oceanía son regiones menos estudiadas al respecto y con trabajos orientados a un solo contaminante [7]. Cabe

aclarar, de todas formas, que a nivel mundial la mayor parte de las aplicaciones de ANN a pronóstico o predicción de niveles de contaminantes se han centrado en un solo contaminante, siendo los más recurrentes PM_{10} , $PM_{2,5}$, los óxidos de nitrógeno y el ozono [40].

Los modelos orientados a la calidad del aire deben desarrollarse comprendiendo las interacciones entre fuentes de emisión, deposición y sumideros, concentraciones atmosféricas y procesos meteorológicos, entre otros [63]. Es por ello que actualmente se exploran las relaciones entre los niveles de los contaminantes en estudio y otras variables como las meteorológicas (generalmente dirección y velocidad del viento, humedad relativa y turbulencia atmosférica), emisiones de otros contaminantes, indicadores de tráfico (densidad de flujo, velocidad, grado de ocupación de espacios, longitud de filas de vehículos motorizados y tiempos de viaje) y, en menor medida aquellas como datos satelitales, información de usos del suelo y variables económicas [40].

A la hora de trabajar con una base de datos multidimensional de gran tamaño, existen diferentes enfoques que permiten organizarla para extraer información de ella y encontrar vínculos entre variables. Numerosas herramientas de minería de datos han sido más útiles también en este sentido que otros métodos estadísticos convencionales. Una de ellas, también dentro del campo del aprendizaje automático (ANN), son los Mapas Auto-Organizados (SOM, por “self organizing maps”), también llamados redes Kohonen [64], que se basan en el aprendizaje no supervisado y ayudan a visualizar similitudes y relaciones en un conjunto de datos. Estas ANN se utilizan ampliamente para explorar vínculos entre los parámetros de entrada y la importancia de cada uno en ese espacio [65, 66]. Esto se puede hacer mediante la capacidad de los SOM para realizar proyecciones no lineales desde un espacio de entrada multidimensional a un espacio de salida bidimensional. En comparación con las técnicas tradicionales de análisis y agrupación de datos, los SOM tienden a ser más robustos en el manejo de señales ruidosas y datos faltantes [67] y permiten que los nodos (neuronas) se visualicen en la forma original de los datos (es decir, vectores que representan el espacio de entrada), lo que amplía las posibilidades de análisis [65].

Los SOM se han utilizado para agrupamiento, clasificación, visualización y modelado en diferentes campos disciplinares y con distintos fines. En particular, en el análisis de datos meteorológicos [65, 68, 69, 70], en pronósticos y estudios de calidad del aire [71][72], en clasificación de estaciones meteorológicas [73], en identificación de regiones climáticas [74], entre otros. Los antecedentes indican que los SOM pueden ser útiles para extracción de características de un espacio de entrada con parámetros meteorológicos [75].

2.1.1. Bases de las Redes Neuronales MLP-FF-BP

En este apartado se detallan las bases de los tipos de redes que fueron utilizadas para la predicción de $PM_{2,5}$ en esta Tesis, cuyos resultados son descritos en los **Capítulos 4 y 6**, comenzando por la forma más general de Perceptrón Multicapa (MLP, por “Multilayer Perceptron”), de Propagación Inversa (BP, por “backpropagation”) y Flujo de Avance (FF, por “feedforward”). Estas ANN poseen una capa de entrada (con n neuronas de entrada), una o varias capas ocultas (con h neuronas ocultas) y una capa de salida (con una o varias neuronas de salida o).

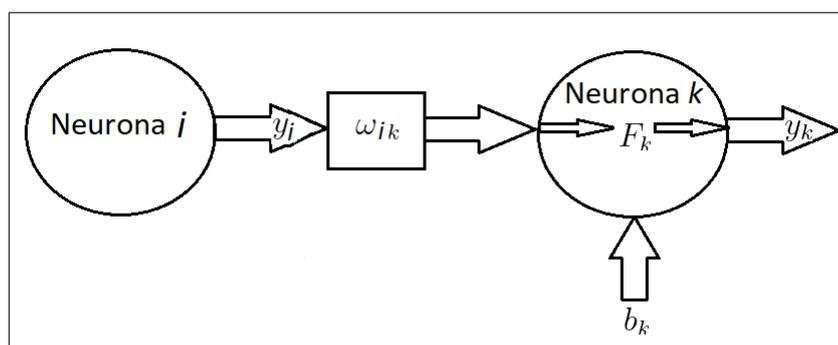


Figura 2.1: Componentes básicos de una ANN con una única neurona de entrada i y una única neurona k en una capa oculta y una salida y_k

La información proveniente por la neurona i de entrada y_i se multiplica por el peso $\omega_{i,k}$ y se le suma la compensación de sesgo b_k antes de ser procesada en la neurona k a partir de la función de activación o transferencia F_k para conseguir la salida y_k (**Figura 2.1**).

Luego de definir la arquitectura básica de la red, esta opera en tres etapas consecutivas: entrenamiento, validación y prueba. Por ello, los conjuntos de datos de las variables de entrada o predictores y del objetivo o “target” (que se desea predecir o estimar) se dividen en tres partes que se utilizan correspondientemente en cada etapa.

Respecto a las tres etapas

Durante el entrenamiento se ajustan los pesos correspondientes a las conexiones de las neuronas de entrada con las ocultas hasta que la red “aprende” a predecir la variable objetivo en función del conjunto de entrenamiento. La red es alimentada con los datos correspondientes a las variables seleccionadas como predictores (que se encuentran representados como neuronas de entrada). Los datos de entrada están organizados en

series temporales de cada predictor, donde a cada línea conformada por los datos de los predictores en un mismo tiempo se la denomina patrón. El flujo de los datos ocurre desde la capa de entrada hacia la/s oculta/s y luego hacia la salida (flujo de avance, FF). Luego de que el flujo de información alcanza a la capa de salida (donde se puede encontrar una o más neuronas de salidas que representan al o a las variables de salida), el valor estimado por la red se compara con el valor esperado objetivo (serie coincidente temporalmente con las de los predictores) para calcular el error. Este error se propaga hacia atrás en la entrada de la red (propagación inversa, BP) con el fin de ajustar los pesos en las conexiones de las neuronas de la capa oculta y comenzar el ciclo nuevamente. Este proceso ocurre iterativamente hasta que el valor de salida coincide con el valor objetivo dentro de un margen de error preestablecido. Este procedimiento es el más aceptado para entrenar las redes ANN-FF-BP [27].

La etapa de validación responde a evitar cualidades problemáticas típicas que puede poseer un modelo o red entrenada. Una de ellas es el sobre-ajuste, donde los resultados indican que la red posee una mayor cantidad de pesos y sesgos de los que necesita, y otra es el sub-ajuste, donde se da la situación contraria. En ambos casos, la red no generalizará correctamente en el sentido de que no tendrá un buen desempeño al simularse con datos nuevos. Los resultados de la validación permiten retrotraer los valores de la red a la instancia previa a la ocurrencia de estos problemas o a elegir entre distintos modelos posibles con diferentes parámetros, tales como cantidad de neuronas o capas ocultas. En otras palabras, la etapa de validación ofrece una oportunidad para reajustar los parámetros de la red en caso de ser necesario. Finalmente, la etapa de prueba es un ensayo imparcial de la red final con nuevos datos con el fin de evaluar su desempeño global.

Para entrenar la red se pueden utilizar tres modos: en línea (“online training”), por lote (“batch training”) o por mini-lote (“mini-batch training”). Estos se diferencian en la forma de presentación de los patrones de entrenamiento a la red.

Respecto a los modos de entrenamiento

En el modo en línea los pesos se actualizan después de cada presentación de patrón a la red. Este procedimiento se repite la cantidad de veces necesario hasta alcanzar un umbral de error predefinido, y luego se presenta el siguiente patrón para repetir el procedimiento. Sus ventajas respecto al modo por lote radica en que requiere menor capacidad de almacenamiento para los pesos y se corre un menor riesgo de caer en

un mínimo local de error. Su principal desventaja radica en que un patrón inadecuado (debido a la presencia de valores atípicos o faltantes) de entrada puede hacer que el aprendizaje caiga en un lazo cerrado sin poder avanzar, debido a que no se alcanza el umbral de error deseado.

En el modo por lote, la actualización de pesos ocurre luego de que todos los patrones son presentados a la red en un primer ciclo. Este proceso se repite iterativamente con nuevas presentaciones de los patrones hasta alcanzar el umbral de error requerido. Este modo posee las ventajas de estimar con mayor precisión el error y una medición del cambio necesario de pesos a cada iteración (denominada época, en el sentido de parte temporal) más representativo (al efectuarse con una cantidad de patrones igual al tamaño de cada vector de entrada en lugar de uno solo).

En el modo por mini lote, la red actualiza los pesos luego de la presentación de distintas porciones del conjunto de entrenamiento. Este enfoque representa un punto intermedio entre las otras dos opciones, por lo que puede compartir sus ventajas y desventajas[76].

Respecto al mecanismo de actualización de pesos y sesgos

Considerando n neuronas de entrada, la información total que recibe una neurona oculta k es la suma de los pesos de las salidas de cada neurona conectada a ella más el sesgo (**Ecuación 2.1**).

$$s_k(t) = \sum_{i=1}^n \omega_{ik}(t)y_i(t) + b_k(t) \quad (2.1)$$

La función $F_k(s_k)$ es una función, generalmente no decreciente, de la entrada total de la neurona (**Ecuación 2.2**).

$$y_k(t+1) = F_k(s_k(t)) = F_k\left(\sum_{i=1}^n \omega_{ik}(t)y_i(t) + b_k(t)\right) \quad (2.2)$$

En la **Figura 2.2** se muestran las funciones de transferencia más utilizadas en ANN. Dentro de la familia de funciones sigmoides, las más utilizadas son la logarítmica (“log-sig”) y la tangente hiperbólica (“tansig”). En aplicaciones de calidad del aire, una función de transferencia sigmoide es comúnmente utilizada al ser no lineal y fácilmente diferenciable [77]. Por otro lado, una función lineal o de identidad se suele utilizar en las neuronas de salida con el propósito de evitar un sesgo en los valores estimados [78].

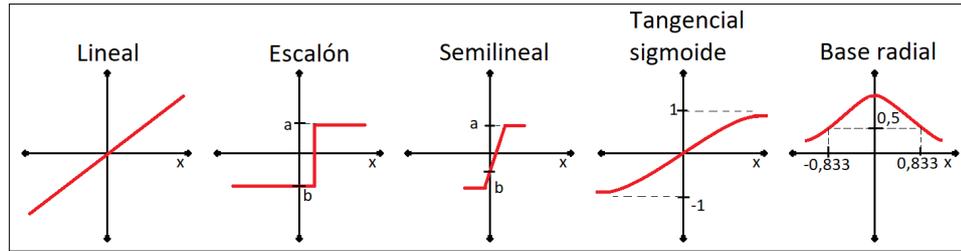


Figura 2.2: Funciones de transferencia típicas en neuronas

La salida y_k debe compararse con el valor deseado para estimar un error y corregir los pesos de las conexiones en la siguiente época. Esto se logra a partir de lo que se denomina regla de aprendizaje.

Una sola capa oculta con el suficiente número de neuronas ocultas con su función de activación no lineal es suficiente para lograr la propiedad universal de aproximación de las redes ANN MLP-FF-BP [79, 80]. Además, más de una capa oculta aumenta la probabilidad de caer en un mínimo local y que el cálculo del error se vuelva inestable [81]. En base a estos antecedentes, las redes MLP descritas en esta Tesis poseen una única capa oculta. El algoritmo básico para MLP utiliza una regla de aprendizaje de gradiente descendente, como caso especial del algoritmo BP, denominado regla delta.

El conjunto de datos de entrada (patrones) está conformado por una matriz \mathbf{X} de “ p x n ” dimensiones, donde p es la cantidad de patrones igual al tamaño de las series en las n variables que offician como predictores durante el entrenamiento (igual a la cantidad de neuronas de entrada). La dinámica de esta regla precisa que se alimente a la red con un conjunto de patrones de entrada \mathbf{X} con sus respectivas salidas \mathbf{d} (vector de tamaño p) a modo de ejemplos para que la red “aprenda” de ellos, lo que se conoce como aprendizaje supervisado y se esquematiza en la **Figura 2.3**.

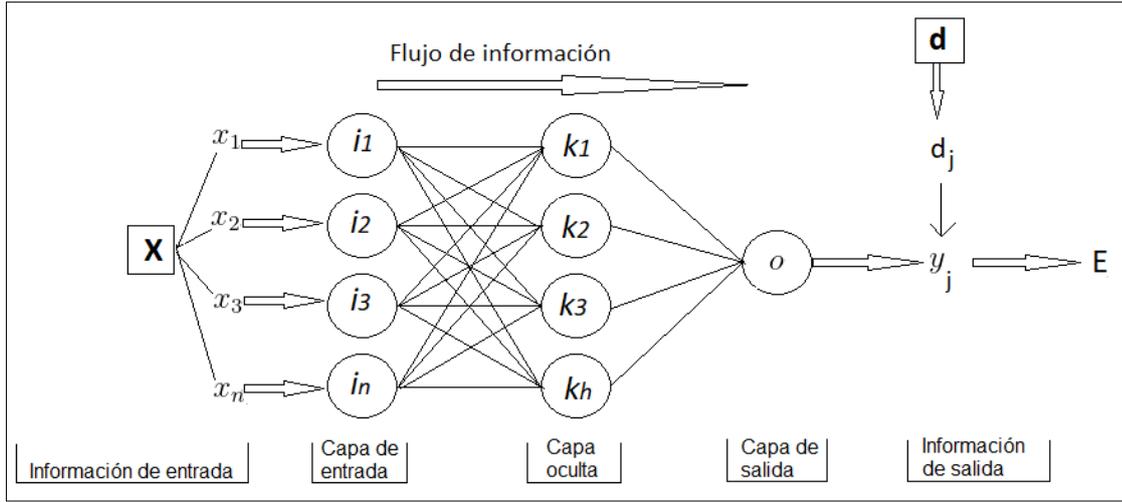


Figura 2.3: Representación de una red neuronal MLP con i_n neuronas de entrada, k_h neuronas ocultas y una neurona o de salida. La información fluye de izquierda a derecha, con excepción del error E que realimenta la red y se calcula comparando cada salida y_j con su respectivo valor deseado d_j . En el caso de entrenamiento por lote, una época involucra j comparaciones igual en número a la cantidad de patrones p

El objetivo es ajustar la red lo mejor posible a un conjunto de muestras de entrenamiento consistentes en la matriz de entrada \mathbf{X} y valores deseados de salida \mathbf{d} . Para cada entrada, la salida de la red difiere del valor esperado por $d - y$. La regla delta usa una función de error o costo basada en estas diferencias para ajustar los pesos y que se calcula por cuadrados mínimos (la **Ecuación 2.3** muestra el cálculo del error E para un modo de entrenamiento por lote).

$$E = \frac{1}{2} \sum_{j=1}^p (d_j - y_j)^2 \quad (2.3)$$

El procedimiento de cuadrados mínimos encuentra los valores de todos los pesos que minimizan la función de error por el método de gradiente descendente. Se cambian los pesos de manera proporcional a la derivada negativa del error (**Ecuación 2.4**).

$$\Delta\omega_{i,k} = -\gamma \frac{\partial E}{\partial \omega_{i,k}} \quad (2.4)$$

Donde γ es una constante de proporcionalidad denominada tasa de aprendizaje y los subíndices i y k representan a las neuronas de entrada y ocultas, respectivamente. La actualización de los sesgos sigue un procedimiento equivalente (**Ecuación 2.5**).

$$\Delta b_k = -\gamma \frac{\partial E}{\partial b_k} \quad (2.5)$$

El gradiente descendente requiere que se den pasos infinitesimales y la constante de proporcionalidad es la tasa de aprendizaje γ . Por cuestiones prácticas se elige la mayor tasa de aprendizaje sin que se llegue a una oscilación en torno al valor deseado. Una forma de evitar la oscilación a grandes valores de γ es hacer el cambio de pesos dependiente del cambio de peso anterior añadiendo un término de momento (**Ecuación 2.6**).

$$\Delta \omega_{i,k}(t+1) = -\gamma \frac{\partial E}{\partial \omega_{i,k}} + \beta \Delta \omega_{i,k}(t) \quad (2.6)$$

t se refiere a cada paso, iteración o época y β es una constante de proporcionalidad que determina la importancia del término de momento en el ajuste de los pesos. En la variación de aprendizaje por gradiente conjugado, β no es una constante sino una función que ajusta la dirección de búsqueda (signo y amplitud del cambio del peso y/o del sesgo) de acuerdo a la variación de Fletcher-Reeves[82]. La **Ecuación 2.7** muestra cómo se determina la función β .

$$\beta = \frac{\left\| \frac{\partial E}{\partial \omega_{i,k}}(t+1) \right\|^2}{\left\| \frac{\partial E}{\partial \omega_{i,k}}(t) \right\|^2} \quad (2.7)$$

El algoritmo de aprendizaje regla delta de BP, generalizado a funciones no lineales, ha sido un avance sustancioso en la investigación de ANN. Aunque resulta muy lento para diversas aplicaciones prácticas debido a la cantidad de cálculos y uso de memoria necesario para cada época [83]. A lo largo de los años, se han propuesto técnicas de optimización del algoritmo BP que se pueden categorizar en heurísticas (modificación de las tasas de aprendizaje y momento, y cambio de variables) y optimización numérica. La presente descripción se centrará en el algoritmo Levenberg-Marquardt (LM) perteneciente a la última categoría, ya que es utilizado en esta Tesis para algunas de las ANN desarrolladas.

El algoritmo LM se deriva del método de Newton que fue diseñado con el fin de encontrar el mínimo de funciones que son suma de cuadrados de funciones no lineales, como el error cuadrático medio. Específicamente, este método aproxima una velocidad de entrenamiento de segundo orden sin necesidad de calcular la matriz Hessiana (matriz de derivadas parciales segundas). Con los fines de generalizar, se define un vector de desempeño $\mathbf{e}_{(o)}$ de o dimensiones (cantidad de neuronas de salida) con el error calculado E a cada época para cada salida de acuerdo a la **Ecuación 2.3**, y un vector de pesos

y sesgos de z dimensiones (donde z es igual al producto entre cantidad de neuronas de entrada n y cantidad de neuronas ocultas h), $\mathbf{w}_{(n,h)}$. El método de Newton para optimizar el vector de desempeño se muestra en la **Ecuación 2.8**.

$$\mathbf{A}_{(i,k)}(t+1) = \mathbf{A}_{(i,k)}(t) - \mathbf{H}(t)^{-1} \quad (2.8)$$

$$\mathbf{g}(t) = \mathbf{J}^T \mathbf{e}_{(o)} \quad (2.9)$$

Donde t identifica a una época particular, \mathbf{H} es la matriz Hessiana (derivadas parciales segundas de cada componente de $\mathbf{e}_{(o)}$ respecto a cada componente de $\mathbf{w}_{(n,h)}$) y \mathbf{g} (**Ecuación 2.9**) es un vector gradiente igual al producto entre la traspuesta de la matriz Jacobiana \mathbf{J} (derivadas parciales de cada componente de $\mathbf{e}_{(o)}$ respecto a cada componente de $\mathbf{w}_{(n,h)}$) y el vector $\mathbf{e}_{(o)}$. Luego, $\mathbf{a}_{(n,h)}$ es inicialmente un vector de pesos y sesgos predefinidos y se actualiza en cada época con sus mismos valores restados a los del vector resultante del producto entre la inversa de la matriz Hessiana con el vector \mathbf{g} . El algoritmo LM utiliza la aproximación mostrada en la **Ecuación 2.10** para la matriz Hessiana, a partir del producto de la matriz Jacobiana con su traspuesta.

$$\mathbf{H} = \mathbf{J}^T \mathbf{J} \quad (2.10)$$

Finalmente, el algoritmo de aprendizaje LM es el mostrado en la **Ecuación 2.11**.

$$\mathbf{a}_{n,h}(t+1) = \mathbf{a}_{n,h}(t) - [\mathbf{J}^T \mathbf{J} + \mu \mathbf{I}]^{-1} \mathbf{J}^T \mathbf{e}_{(o)} \quad (2.11)$$

Donde \mathbf{I} es la matriz identidad y μ es un escalar denominado parámetro de amortiguación que cambia a cada época (decrece luego de un ciclo exitoso donde disminuya el error y aumenta en los ciclos donde ocurre lo contrario).

Cabe aclarar que existen diferentes formas de inicializar los pesos y sesgos. Se pueden iniciar al azar (“rands”), al azar con pequeños valores (“randsmall”), calculando un valor medio entre los valores de entrada (“midpoint”) o con el algoritmo de Nguyen-Widrow (“initnw”) que elige valores para distribuir la región activa de cada neurona en la capa oculta, aproximadamente de manera uniforme, a través de su espacio de entrada con una cierta componente de azar.

Respecto a la convergencia y a la generalización

La convergencia indica que la red finalizó sus etapas de procesamiento (entrenamiento, validación y prueba). Las razones por la que no suceda esto pueden ser varias, tales como que el modelo es demasiado complejo y el costo computacional demasiado elevado o que la selección de parámetros de la red no es la adecuada. Una vez lograda la convergencia, se debe evaluar si la red converge en un mínimo local o global de lo que se suele denominar superficie de errores, en función de los parámetros de desempeño.

En el apartado anterior se describió matemáticamente el proceso del cálculo de los errores y su vector gradiente. El vector gradiente de errores es un vector multidimensional (de dimensión z igual a la cantidad de pesos, ya que cada error calculado corresponde al cálculo de un peso). Con el fin de simplificar, si una red posee una neurona de entrada y dos ocultas se le asigna al vector gradiente dos elementos (dos errores del cálculo de dos pesos w_1 y w_2) que se actualiza a cada época (que vale recordar en este contexto significa iteración). Con esta información se puede representar una superficie de errores en tres dimensiones (un eje representa a los valores de error de w_1 a cada época, otro a los valores de error de w_2 a cada época, y un tercero a los valores de los errores). Esta superficie posee un mínimo global (que es el que se desea alcanzar) y varios mínimos locales (que son los que se quieren evitar). Esta misma idea se traslada a un espacio de más dimensiones.

La generalización de un modelo se refiere a su capacidad de demostrar el mismo desempeño al operar con datos nuevos que al operar con los datos que se utilizaron para su entrenamiento. Esto se busca lograr con el modelo más simple posible, ya que al aumentar su complejidad aumenta la probabilidad de ocurrencia de errores. En términos de ANN, la simplicidad radica en encontrar el modelo con menor cantidad de parámetros libres (pesos y sesgos o cantidad de neuronas) que exhiba un buen desempeño (se aproxime lo mejor posible al mínimo global en la superficie de errores). Existen diferentes enfoques que persiguen la generalización de ANN, tales como crecimiento (implica implementar el modelo sin neuronas y luego ir aumentando su cantidad hasta alcanzar el desempeño deseado con el menor número posible de pesos y sesgos), poda (técnica opuesta al crecimiento, se comienza con una gran cantidad de neuronas que aseguren un sobreajuste y luego se va disminuyendo su cantidad), búsquedas globales (generan un espacio con todas las posibles arquitecturas para luego encontrar el modelo más simple con buena generalización), detención temprana (se monitorean los errores de entrenamiento y validación con el propósito de detener la actualización de pesos y sesgos si estos errores aumentan luego de una cierta cantidad de épocas), regularización (modifica el cálculo de

errores agregando un término que penaliza el nivel de complejidad creciente del modelo a cada época, lo cual implica utilizar un modelo con una cantidad de neuronas que logre converger pese a este criterio).

Sobre ANN dinámicas

Las ANN descritas en los apartados anteriores pertenecen a la categoría de redes estáticas. Otro grupo de redes, denominadas dinámicas, se caracterizan por poseer elementos de realimentación y retrasos temporales en su arquitectura, y su respuesta depende de sus entradas y del registro y secuencia de las mismas (**Figura 2.4**). En lugar de aproximar funciones, las redes dinámicas aproximan sistemas (enfocados a control, predicción, reconocimiento, entre otros).

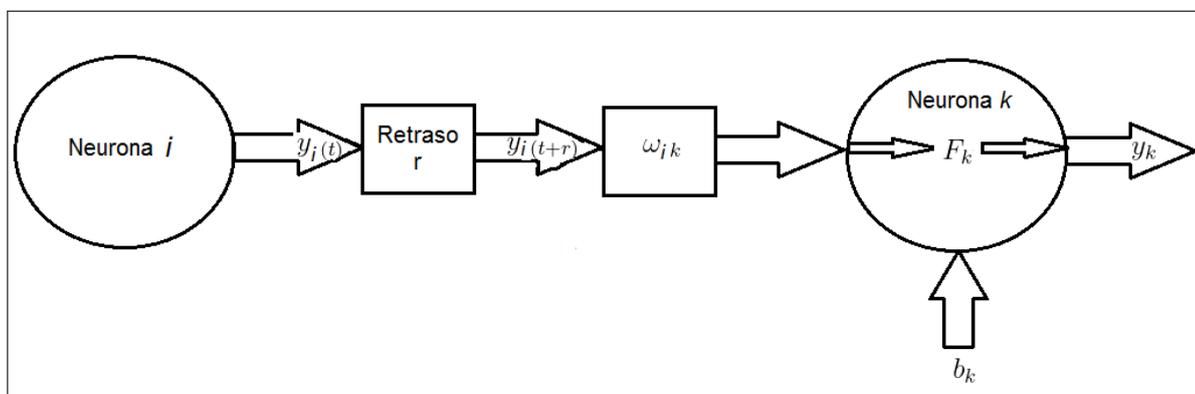


Figura 2.4: Transferencia de información de una neurona de entrada i a otra neurona oculta k con un bloque de retraso que cambia la presentación de la información de i , saltando r posiciones en la serie temporal

Un tipo de ANN dinámicas puede poseer una línea de retraso temporal en su capa de entrada (retraso temporal focalizado, útil para la predicción de series temporales), o en todas sus capas (retraso temporal distribuido, normalmente utilizado en modelos de reconocimiento). Los modelos NARX (por “nonlinear autoregressive network with exogenous inputs”, usadas en sistemas de control) poseen conexiones de realimentación en varias de sus capas. Las redes recurrentes por capa, utilizadas en procesos de filtrado, poseen un lazo de realimentación desde la salida hacia la entrada de cada una de sus capas, con excepción de la última. La **Figura 2.5** muestra un ejemplo de ANN recurrente con un lazo de realimentación de la salida a la capa oculta con un retraso temporal.

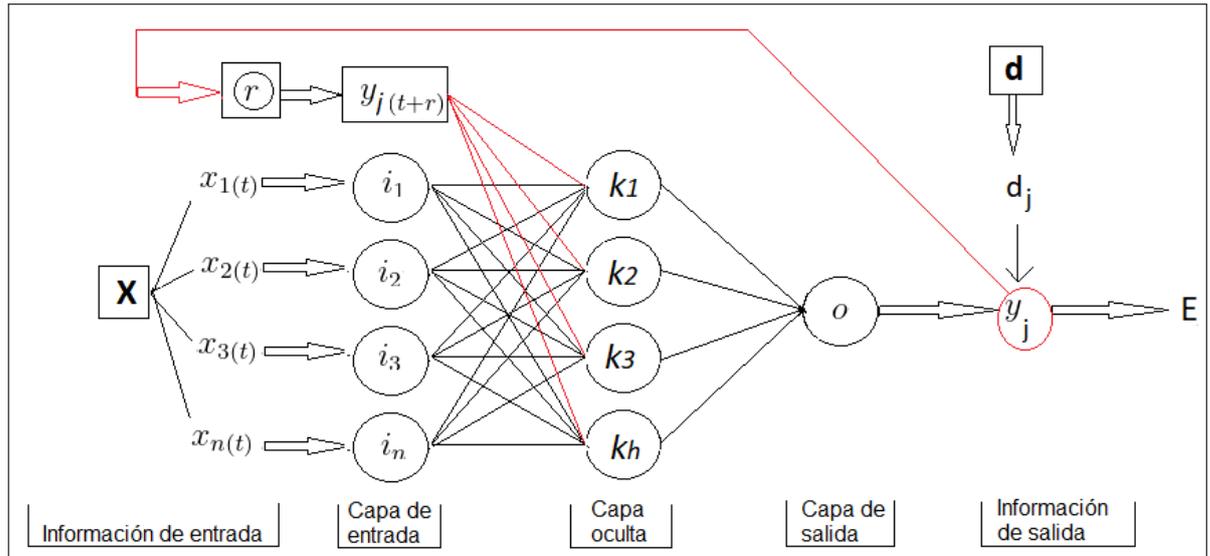


Figura 2.5: Ejemplo de ANN recurrente. Cada salida entra en un lazo de realimentación, pasando a través de un retraso temporal r antes de alcanzar la capa oculta

El entrenamiento de ANN dinámicas puede ser más complicado que en redes estáticas, debido a la complejidad de la superficie de error y a una mayor posibilidad de caer en un mínimo local durante el entrenamiento [83]. Los procedimientos de optimización de aprendizaje utilizados en estas redes son los ya descritos para redes estáticas, con la diferencia del cálculo del gradiente $\mathbf{g}(t)$ para lo que se puede utilizar BP en el tiempo (comenzar por el último paso temporal y operar hasta el primero) o aprendizaje recurrente en tiempo real (se comienza en el primer paso temporal y el cálculo se actualiza en cada época).

Respecto al diseño y entrenamiento de una ANN

El diseño de una ANN presenta varios desafíos ligados a su complejidad intrínseca. Con el objetivo del desarrollo de un modelo adecuado al problema de estudio, el usuario debe seleccionar diversos parámetros, variables y técnicas a implementar que no son independientes entre sí y son implementadas en un esquema con abundancia de sensibilidades cruzadas. Un procedimiento general se describe en la **Figura 2.6**. Cabe mencionar, que dicho procedimiento es el normalmente utilizado a la hora de diseñar modelos basados en ANN, pero aún no existen protocolos generales y consistentes al respecto [40].

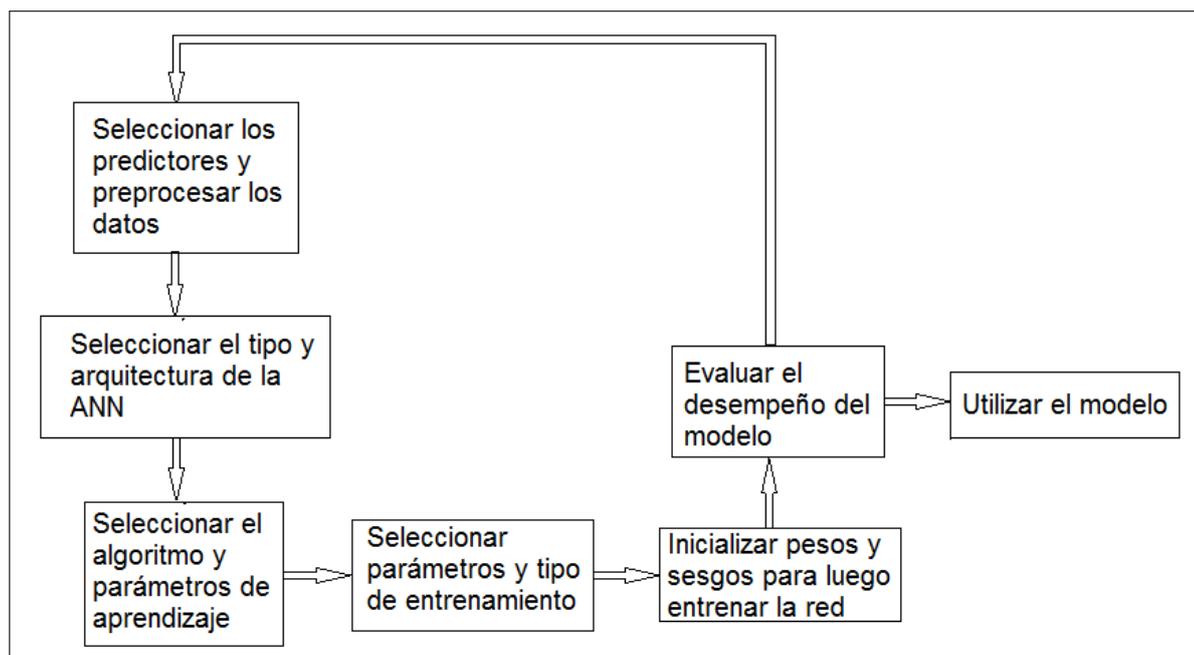


Figura 2.6: Esquema general de diseño de una ANN. Luego de evaluar el desempeño, si resulta satisfactorio el modelo estará listo para ser utilizado, caso contrario se deberá reiniciar el proceso con el fin de realizar las modificaciones necesarias

Como se mencionó anteriormente, el uso de numerosos predictores posee el riesgo de linealizar la respuesta general de la red [84]. Pero en general, resulta dificultoso seleccionar predictores con un conocimiento experto acerca del fenómeno. Es por ello que los investigadores dedican esfuerzos en encontrar los predictores que mejor explican el comportamiento de la variable objetivo, y utilizar la menor cantidad posible que asegure una buena generalización. Los enfoques para seleccionar a los predictores pueden clasificarse en: basados (o dentro de) en el modelo y libres (o fuera de) del modelo. En el primer caso, se analiza la influencia de los predictores en el desempeño de la ANN, aunque esto posee la incongruencia de que cada vez que se agrega o quita un predictor cambia la red, por lo que no se está realmente comparando en base al mismo modelo. En el segundo caso, se analizan las relaciones entre los predictores y la variable objetivo con diferentes técnicas. En general, se suelen utilizar métodos analíticos lineales de correlación cruzada entre los predictores y el objetivo, pero resulta contradictorio utilizar un método de relación lineal para seleccionar variables que luego se utilizarán en un modelo no lineal, además del hecho de que se puede omitir variables que tengan una correlación no lineal con el objetivo [85]. Otro método ampliamente utilizado es el de análisis de componentes principales (PCA por “principal component analysis”) que describe el espacio de datos en términos de nuevas variables no correlacionadas, aunque el método no focaliza per

se en la variable objetivo. Resulta necesario avanzar en enfoques no lineales para la selección de predictores [40].

Una vez seleccionados los predictores, existen diferentes pasos en el pre-procesamiento de los datos con el fin de optimizar el entrenamiento de la red. Entre estos, se encuentran la normalización y la extracción de características, entre otros. La normalización busca que los datos de cada predictor posea una influencia similar en el cálculo de los pesos y sesgos y puede darse como re-escalado de los datos en un rango determinado o como una estandarización (para que el conjunto de datos tenga valor medio igual a cero y desviación estándar igual a la unidad). La extracción de características busca generalmente reducir las dimensiones del espacio de entrada, un ejemplo de ello es PCA.

Como parte de un pre-procesamiento, cabe mencionar también que diferentes métodos de eliminación de ruido o tratamiento de periodicidades en datos de predictores y/u objetivos han sido reportados como exitosos por algunos autores [26, 60, 86].

Un problema usual de los modelos ANN es que no son eficientes extrapolando datos, en cuanto a que exhiben un pobre rendimiento al implementarse con datos que no se encuentran dentro del rango para los que fueron entrenados. Es por ello que si el error durante el test es alto, mientras que los de entrenamiento y validación son bajos, es posible que la red esté extrapolando. Es por esta razón que los modelos basados en ANN presentan una reducida precisión en la predicción de concentraciones extremas de contaminantes atmosféricos [7], ya que estos casos son escasos en comparación con los valores habituales. Una forma de evitar la extrapolación o reducirla, es cambiar la técnica de división de datos en los tres grupos que serán utilizados durante el entrenamiento, validación y prueba. Esto se puede realizar de manera azarosa, por índices seleccionados por el usuario, por grupos de datos consecutivos para cada categoría, etc..., y en diferentes porcentajes del conjunto de datos de entrada original. Respecto a la arquitectura de la red, tampoco existen criterios unificados y universales. Pero se puede indicar que para el problema de predicción o estimación se suelen utilizar MLP con funciones de transferencia no lineales en las neuronas ocultas y lineal en la/las neuronas de salida, y una cantidad de neuronas ocultas seleccionadas por prueba y error.

2.1.2. Bases de las ANN tipo SOM

Una red SOM básica está formada por una capa de entrada y una capa de Kohonen que representa las neuronas de salida en una grilla o arreglo bidimensional, lo que es también llamado mapa de salida [64]. Las conexiones entre estas dos capas se encargan

de propagar la información desde la entrada hacia la salida. Cada neurona de entrada está conectada a todas las neuronas de salida mediante un peso. Por lo tanto, cada neurona de salida tiene asociado un vector de pesos o de referencia (se trata de un vector prototipo o promedio de la categoría representada por la neurona de salida).

El proceso de auto-organización de los SOM se basa en adaptación no supervisada que genera un orden global en los datos presentados a través de competencia entre las neuronas del mapa de salida preservando el espacio de entrada. La unidad de mejor ajuste o “ganadora” b (BMU, por sus siglas en inglés “Best Matching unit”), resulta ser la neurona cuyo vector de pesos se aproximó mejor a la entrada.

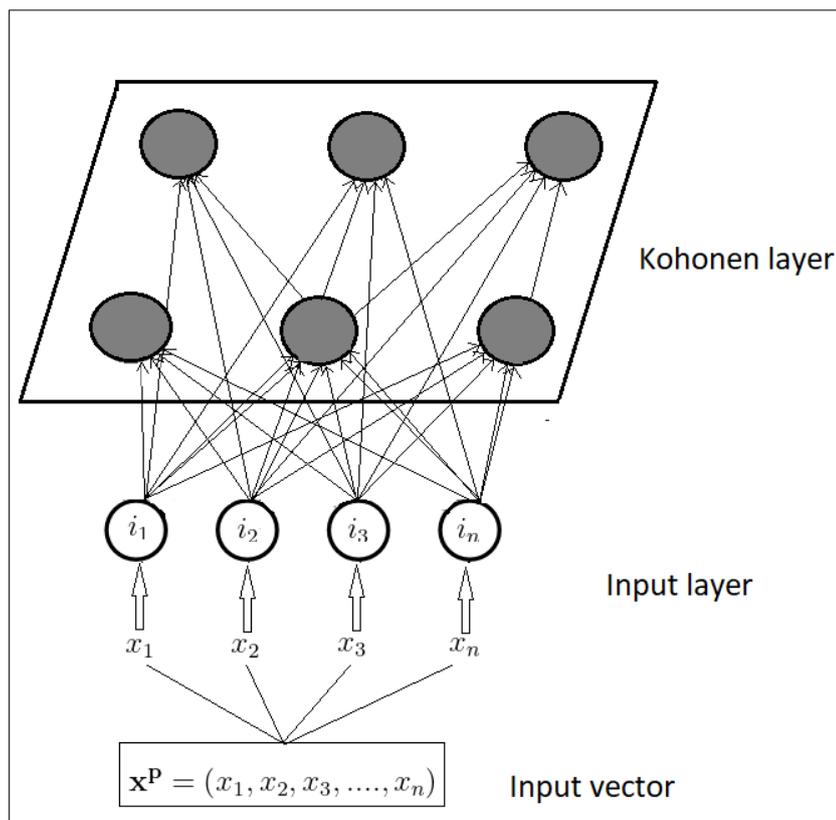


Figura 2.7: Representación de mapa autoorganizado

Cada neurona k en el mapa de salida está representada por un vector de pesos de n dimensiones $\mathbf{w}_k = [w_1, \dots, w_n]$, donde n coincide con la cantidad de neuronas de entrada.

Cuando un vector de entrada $\mathbf{x} \in R^n$ es presentado a la red, las neuronas en el mapa compiten entre ellas para ser la “ganadora”, que resultará ser la más cercana al vector de entrada en términos de distancia euclidiana (**Ecuación 2.12**), pues el mapa se proyecta sobre un plano euclidiano.

$$\|x - w_b\| = \min \|x - w_k\| \quad (2.12)$$

Durante el entrenamiento, los pesos de las neuronas que están topográficamente cerca en el mapa son desplazados hacia la entrada x usando la regla de aprendizaje de auto-organización (**ecuación 2.13**).

$$\mathbf{w}_k(t+1) = \mathbf{w}_k(t) + \eta h_{bk}(t) [\mathbf{x}(t) - \mathbf{w}_k(t)], \forall k \in N_b(t) \quad (2.13)$$

Donde t es el retraso temporal, η es la tasa de aprendizaje (que define la magnitud del ajuste a cada vector de pesos) y $h_{bk}(t)$ es la función de vecindad Kernel alrededor de la unidad b al tiempo t [87]. $N_b(t)$ es una función que denota a los nodos que pertenecen al vecindario de la unidad “ganadora” al tiempo t . Mientras se produce la actualización de los pesos, la tasa de aprendizaje se reduce gradualmente hasta cero (y junto con ella el radio de vecindad).

La forma generalmente usada para la función de vecindad es la gaussiana (**Ecuación 2.14**), aunque también pueden utilizarse otros tipos como “burbuja”, “gaussiana truncada” o una función “Epanechnikov”).

$$h_{bk}(t) = \exp \left[\frac{-sqdist(b,k)}{2\sigma^2(t)} \right] \quad (2.14)$$

Donde $sqdist(b,k)$ es el cuadrado de la distancia geométrica entre los nodos b y k en el arreglo, y $\sigma(t)$ es una función decreciente, generalmente lineal, que representa el radio del vecindario. Tanto $h_{bk}(t)$ como $N_j(t)$ decrecen mientras avanza el entrenamiento, dividiéndolo en una fase de entrenamiento “grueso” (donde ocurre el ordenamiento gradual de las neuronas) y otra fase “fina” (donde el vector de pesos se ajusta finamente hasta sus valores óptimos). El achicamiento de los vecindarios promueve el orden emergente, con las neuronas aprendiendo a responder de acuerdo a cada vector de entrada. De esta manera, los vectores de entrada similares son agrupados en una sola neurona o neuronas vecinas en el mapa una vez logrado el aprendizaje. La matriz de distancias unificada (U-Matrix) constituye una forma estándar de visualizar las distancias promedios entre los vectores de pesos de las neuronas vecinas y particionar el mapa en grupos [88]. De esta manera, un paisaje de grupos se forma sobre el SOM para visualizar el grado de clasificación.

Sobre indicadores de desempeño

Como criterios de calidad del mapa se utilizan la precisión de la proyección (que describe cómo se adaptan las neuronas a los datos) y la preservación de la topología. Como medida de la precisión se utiliza el error medio de cuantización q_e sobre el total de datos, determinado según la **Ecuación 2.15**.

$$q_e = \frac{1}{p} \sum_{i=1}^p \|x_i - w_b\| \quad (2.15)$$

Donde p es la cantidad de vectores de entrada.

El error topográfico t_e (**Ecuación 2.16**) mide la preservación de la topología del conjunto de datos considerando la estructura del mapa.

$$t_e = \frac{1}{N} \sum_{k=1}^N u(x_k) \quad (2.16)$$

Donde x_k representa a cada neurona en el mapa de salida y $u(x_k)$ es igual a 1 si la primer y segunda neuronas ganadoras más cercanas a x_k no están próximas la una a la otra. Caso contrario es igual a 0.

La distorsión del SOM es una función de costo que puede ser utilizada para cuantificar un valor de error para todo el mapa. Se calcula como una medición pesada de las distancias [89], como se indica en la **Ecuación 2.17**.

$$Dis = \sum_{i=1}^p \sum_{k=1}^m h_{bk} \|x_i - w_k\| \quad (2.17)$$

En general se utiliza una función de vecindad h_{bk} con un radio fijo.

El método más comúnmente usado para definir el tamaño del mapa de salida es por prueba y error, ya que no se puede producir un valor que sirva como criterio absoluto para ello. Sin embargo, los errores de cuantización, topográfico y la distorsión final del mapa pueden ayudar a deducir un rango adecuado.

2.2. Fotometría solar y propiedades ópticas de aerosoles

Los aerosoles dispersan y absorben la radiación solar en sus diferentes longitudes de onda, y los fotómetros solares hacen uso de este fenómeno registrando un valor proporcional a la intensidad de luz en determinadas componentes espectrales incidentes en

superficie. A partir de ese valor, se pueden determinar diferentes parámetros relacionados con el tipo y cantidad de aerosoles presentes en la columna atmosférica, integrada desde la superficie hasta su tope con una sección horizontal de 10km x 10km, aproximadamente [12]. Uno de estos parámetros es el espesor óptico de aerosoles (AOD, por sus siglas en inglés “Aerosol Optical Depth”), que es el coeficiente integrado de extinción de radiación en una determinada longitud de onda en una columna atmosférica desde el tope hasta el punto de medición [90]. Se utiliza como indicador de propiedades ópticas de aerosoles [9, 91, 92, 93, 94, 95, 96]. Dado que AOD es una propiedad columnar de la atmósfera, ciertamente está influenciado por procesos locales y de mesoescala, y muchos investigadores lo han utilizado para caracterizar la contaminación regional [13, 16, 17]. La Organización Meteorológica Mundial (WMO) incluyó al AOD en el listado básico de medidas del Programa de Vigilancia Atmosférica Mundial [97]. Localmente, el AOD puede estar influido por la calidad del aire, condiciones meteorológicas o por fenómenos naturales y antrópicos como quema de biomasa, actividad agrícola o transporte de polvo [98].

El AOD puede ser derivado a partir de mediciones satelitales de radiación reflejada por la Tierra [11, 99, 100], aunque existe una incerteza asociada a las aproximaciones en cuanto a los cambios de reflectividad [101], o puede ser medido en superficie con fotómetros solares. El proyecto AERONET (por “Aerosol Robotic Network”) [34] coordinado por la NASA (por “National Aeronautics and Space Administration”) de Estados Unidos nuclea mediciones en superficie con fotómetros solares de diferentes propiedades ópticas de aerosoles en la atmósfera (entre ellos AOD) en 1527 sitios a nivel mundial, 10 de ellos en Argentina. AERONET, considerado como la referencia estándar, ofrece valores de AOD en tres niveles: el nivel 1.0 son los datos sin filtro o post-procesamiento, el nivel 1.5 corresponde a los resultados de someter al nivel 1.0 a ajustes por nubosidad y a procesos automáticos de control de calidad, y el nivel 2.0 resulta de someter al nivel 1.5 a un control manual de calidad.

En cuanto a los datos satelitales, el espectroradiómetro de imágenes de media resolución (MODIS, por sus siglas en inglés) es un instrumento que, puesto en órbita, es capaz de medir en 36 franjas espectrales, temperatura de la superficie, color del océano, vegetación global, nubosidad y AOD. Los satélites Aqua y Terra cuentan con este instrumento a bordo, e históricamente se han creado diferentes algoritmos para mejorar la calidad y resolución espacial de sus productos [102, 103, 104]. Uno de los más recientes y utilizados es la Implementación Multiangular de Corrección Atmosférica (MAIAC, por sus siglas en inglés) para procesar la colección 6 de MODIS usando un modelo físico de

atmósfera-superficie con una resolución de 1 km, cuyos parámetros se definen a partir de mediciones[105]. Con el fin de contar con mediciones de AOD absolutas, los productos del algoritmo deben ser calibrados/validados con mediciones de referencia provenientes de la red AERONET en superficie, lo cual es una limitante considerando la alta dispersión con la que se encuentran instaladas sus estaciones. Como se mencionó en el **Capítulo 1**, la estación AERONET más cercana a la UNGS se encuentra a unos 16 km, por lo que una calibración de productos MAIAC realizada con esta estación para píxeles cercanos a la UNGS tendrá una incerteza real desconocida.

Las mediciones con fotómetros solares en el mundo aportan información valiosa acerca de las propiedades ópticas de aerosoles, pero aún existen muchas regiones sin cobertura. Es por ello que, con el fin de profundizar el conocimiento sobre los efectos de los aerosoles, resulta necesario aumentar el número de sitios de mediciones de AOD, especialmente en áreas metropolitanas donde su dinámica es muy cambiante [106].

El AOD cambia en respuesta a diferentes procesos, como las emisiones de aerosoles primarios, el transporte y su eliminación en la atmósfera [107]. Asimismo, tiene tendencias y cambios estacionales y diarios en respuesta a la relación entre las propiedades de los aerosoles con la temperatura, el contenido de humedad del aire y la velocidad del viento. Los vínculos entre los parámetros meteorológicos y el AOD fueron estudiados por diferentes autores. Usando la plataforma CALIPSO, Niu & Lee (2012) [16] identificaron dos correlaciones distintas de aerosol con nubes y precipitación. Ali et al. (2020) [107] utilizaron datos satelitales de MODIS a largo plazo sobre las regiones de Pakistán y encontraron tendencias crecientes de AOD relacionadas con la disminución de las precipitaciones y las tendencias crecientes de temperatura. Estos autores también encontraron valores altos/bajos de AOD correlacionados con anomalías en la humedad relativa y los campos de viento de la troposfera media. Chabane et al. (2020) [108] investigaron las relaciones entre el AOD y la humedad relativa, la temperatura y la presión bajo masas de aire en la ciudad de Biskra, Argelia; encontraron que, en condiciones claras, los aerosoles son los principales responsables de los efectos directos sobre la radiación solar. Gu et al. (2018) [109] construyeron un modelo utilizando aerosoles terrestres y mediciones meteorológicas en Beijing; sus resultados muestran algunas asociaciones entre AOD, viento e inversión de temperatura. Khalid et al. (2021) [110] encontraron variaciones estacionales y mensuales de AOD sobre las rutas este y oeste del Corredor Económico China-Pakistán y sus relaciones con los parámetros meteorológicos utilizando datos satelitales y análisis de trayectorias de masas de aire. Todos estos antecedentes muestran fuertes vínculos entre AOD y la meteorología, pero su fuerza depende en gran medida

de las condiciones locales [109]. Es por eso que es importante identificar las conexiones subyacentes y sus intensidades para llegar a una mejor comprensión del comportamiento y los impactos de los aerosoles en un área determinada.

Los aerosoles pueden clasificarse según el tamaño de las partículas y sus características de absorción de luz. La clasificación nos da una idea sobre los principales tipos de aerosoles presentes en la atmósfera y podemos rastrear su origen y comportamiento a lo largo del tiempo, pero existen diferentes enfoques para hacerlo. Chen et al. (2019) [111] compararon cuatro esquemas de clasificación diferentes utilizando observaciones terrestres y satelitales, y descubrieron que cada uno de ellos era coherente con las referencias y entre sí. Un tipo de esquema ampliamente utilizado (utilizado también en esta Tesis) recurre al coeficiente de Ångström, que es sensible a la distribución del tamaño de los aerosoles, y al AOD en una longitud de onda determinada [112, 113, 114, 115, 116].

Algunas veces también es posible rastrear la fuente de los principales tipos de aerosoles presentes en la atmósfera en un lugar particular. El análisis de la trayectoria hacia atrás en el tiempo de las masas de aire es una herramienta útil para hacerlo, y puede complementarse con los resultados de la clasificación de aerosoles. Varios investigadores han utilizado el modelo Híbrido de Trayectoria Integrada Lagrangiana de Partícula Individual (HYSPLIT, por sus siglas en inglés) de la Administración Nacional Oceánica y Atmosférica (NOAA) [117][118] en trabajos relacionados con aerosoles [110, 119, 120, 121]. El modelo HYSPLIT utiliza datos cuadrículados para calcular las trayectorias de las parcelas de aire y simular la dispersión y la deposición de contaminantes atmosféricos. Calcula las trayectorias hacia adelante y hacia atrás en diferentes ubicaciones y altitudes utilizando información meteorológica del Centro Nacional de Predicción Ambiental / Centro Nacional de Investigación Atmosférica (NCEP / NCAR) y reanálisis global de datos meteorológicos.

En Argentina hay algunos estudios basados en AOD y su clasificación utilizando datos de AERONET. Otero et al. (2013) [94] analizaron datos de AOD en un sitio urbano en la provincia de Córdoba y encontraron tipo de aerosoles “Continental Limpio” sin humo ni influencia industrial y urbana. Otero et al. (2019) [115] presentaron un trabajo basado en propiedades de aerosoles en el sur, en la provincia de Chubut, y encontraron que la dirección del viento está altamente relacionada con la presencia de aerosoles de tipo “Marítimo” y “Continental” en esa región. Casasola et al. (2020) [116] realizaron un trabajo similar en una estación urbana de AERONET en Tucumán, y encontraron una fuerte presencia de aerosoles de tipo “Quema de biomasa” y “Continental Contaminado” relacionados con industrias y quemas después de la cosecha. CEILAP BA

(34.555S, 58.506 W) es la estación de AERONET más cercana al emplazamiento de la UNGS (unos 16 km) y es una zona muy poblada e industrializada con aerosoles clasificados principalmente como “Continental limpio”, “Continental medio” y “Continental contaminado” [91, 93, 113].

2.2.1. Bases de las mediciones con fotómetro solar

Los fotómetros solares regulatorios utilizan filtros interferenciales y foto-detectores, pero es una tecnología no asequible para instituciones con recursos limitados. Es por ello que los fotómetros solares basados en tecnología LED (por “light emitting diode”) representan una alternativa de bajo costo que permite adquirir mediciones válidas de AOD, especialmente en las regiones sin cobertura de los programas oficiales.

Las mediciones directas de un fotómetro LED se utilizan para calcular el AOD en cada longitud de onda en base a la ley de Lambert-Beer aplicada a la atmósfera, considerando que el espesor óptico atmosférico total depende únicamente de la dispersión de la luz causada por las moléculas conocida como dispersión de Rayleigh, por las moléculas de ozono (O_3) y por los aerosoles (**Ecuación 2.18**).

$$I(\lambda) = I_0(\lambda) \exp(-m(\tau_a + \tau_{O_3} + \tau_r)) \quad (2.18)$$

Siendo $I(\lambda)$ la intensidad de la luz solar recibida en superficie para la longitud de onda dada, $I_0(\lambda)$ la intensidad de la luz solar fuera de la atmósfera para la longitud de onda dada, m es el coeficiente de la masa de aire que la luz atraviesa hasta llegar a la superficie, calculada como la inversa del seno del ángulo de elevación solar [122], τ_a es el coeficiente de transparencia de aerosoles también conocido como AOD, τ_{O_3} es el coeficiente de transparencia de ozono y τ_r (**Ecuación 2.19**) es el coeficiente de transparencia de difusión de Rayleigh, proporcional a la relación entre la presión atmosférica P medida en el punto de observación y la presión al nivel del mar p_0 [123].

$$\tau_r = a_r \frac{P}{p_0} \quad (2.19)$$

a_r es la contribución al espesor óptico de la dispersión molecular de la luz de Rayleigh en la atmósfera y es informado con la calibración de cada fotómetro.

El coeficiente de transparencia de aerosoles puede calcularse a partir de un valor N directamente proporcional a la intensidad de la luz que registra el fotómetro, como indica la **Ecuación 2.20**.

$$\tau_a = \frac{\left[\ln \left(N_0 \left[\frac{r_0}{r} \right]^2 \right) - \ln(N) \right]}{m} - a_r \frac{P}{p_0} - \tau_{O_3} \quad (2.20)$$

N_0 es el valor que registraría el fotómetro para una medida de la intensidad luminosa fuera de la atmósfera a una distancia al Sol r_0 igual a 1 UA (unidad astronómica), y es un parámetro de calibración del instrumento [98][106]. A su vez, el cálculo del AOD requiere de la distancia al Sol r en UA para el momento en que se tomó la medición (que se calcula en función de las coordenadas geográficas del sitio de medición y del día del año).

El coeficiente de Ångstrom [124] α (**Ecuación 2.21**) es una magnitud sensible a la distribución de tamaños de los aerosoles, y se calcula a partir del cociente entre la pendiente del logaritmo de AOD medido en dos longitudes de onda, λ_1 y λ_2 , y el logaritmo de la proporción entre las longitudes de onda [9, 96, 125]. Este coeficiente resulta ser inversamente proporcional al radio cuadrático medio de las partículas, variando entre 4 (para partículas pequeñas) y 0 (para partículas grandes) [124, 126, 127].

$$\alpha = \frac{\ln \left(\frac{\tau_{a1}}{\tau_{a2}} \right)}{\ln \left(\frac{\lambda_1}{\lambda_2} \right)} \quad (2.21)$$

Mayor información puede aportar la relación entre el coeficiente de Ångstrom y el AOD. Una gráfica entre estas dos variables permite identificar los grupos principales de aerosoles presentes y sus características de acuerdo a la clasificación climática [96, 128, 129, 130].

2.3. Mediciones ambientales de bajo costo

El monitoreo y muestreo, junto con los modelos de predicción y los inventarios de emisiones, son parte integral de la gestión de la CA. La medición de contaminantes atmosféricos se puede realizar a partir de diferentes métodos, tales como muestreo pasivo, muestreo con bioindicadores, muestreo activo, método automático y método óptico de percepción remota. El enfoque en el que se centra el desarrollo tecnológico llevado a cabo para esta Tesis es el de monitoreo automático, a partir del cual se realizan mediciones periódicas, continuas y sistemáticas, sin intervención del usuario a partir de un sistema analógico (cada vez menos utilizado) o electrónico. Este método posee múltiples ventajas respecto al resto, pues permite ajustar la resolución temporal de las mediciones y se puede complementar con un sistema de comunicación para generar una base de datos

que se actualice en tiempo real.

Un equipo de medición tiene como finalidad la cuantificación de un parámetro ambiental y, cuando se complementan varios equipos de medición, se habla de una estación de monitoreo. Si una estación se combina con sistemas de transmisión de datos, un centro de control, oficinas, laboratorios y talleres, da lugar a un sistema de medición. Dependiendo del sitio de instalación y sus características (topografía, entorno, meteorología, fuentes de contaminación), un equipo, estación o sistema domina una escala de representatividad [131, 132, 133]. La microescala y las escalas regional, nacional y global, a nivel de mediciones, no son abarcadas en esta Tesis, por lo que se brindan detalles de las escalas media, local y urbana.

La escala media posee dimensiones entre los 100 m² y los 0,5 km², lo cual representa varias manzanas dentro de una ciudad. Las estaciones con esta representatividad se ubican distanciadas de las fuentes de emisión, aunque se ven influenciadas por las mismas. Los datos de estas estaciones pueden utilizarse en estrategias de control de emisiones y para caracterizar eventos de contaminación ambiental.

La escala local abarca entre 0,5 km² y 4 km² y las estaciones con este alcance no se ven afectadas por una fuente de emisión en particular. La homogeneidad en las parcelas de aire bajo su detección supone tasas bajas de cambios de concentraciones, por lo que suele ubicarse en las periferias de los centros urbanos. Los datos obtenidos con esta representatividad pueden usarse para caracterizar niveles de fondo y en estudios sobre los efectos de la CA en la salud humana.

La escala urbana se representa entre 4 km y 100 km y abarca áreas metropolitanas completas. Cubrir esta escala requiere de varios sitios de medición y los datos obtenidos pueden utilizarse para evaluar las tendencias de la calidad del aire en una ciudad.

La parte esencial del enfoque automático de medición son los sensores, que cumplen una transducción de un fenómeno físico-químico a valores cuantitativos del parámetro que se desea medir. La salida de los sensores puede ser analógica (un valor de tensión eléctrica que se traduce de acuerdo a un algoritmo apropiado) o digital (señal de pulsos eléctricos que un lenguaje traduce en valores numéricos).

Los sensores o equipos calibrados y enmarcados en protocolos oficiales poseen un coste elevado de adquisición y operación, pero el aumento sostenido en la disponibilidad comercial de micro sensores, placas de desarrollo y programación en código abierto está contribuyendo a la rápida adopción de sensores de bajo costo (LCS, por su denominación en inglés “low cost sensor”) para el monitoreo de la calidad del aire por iniciativas de ciencia ciudadana, investigación científica y autoridades públicas [134]. De acuerdo a

la Agencia de Protección Ambiental de EUA (EPA), estos dispositivos son reconocidos como LCS cuando tienen un costo menor a USD 1000 y están equipados con electrónica miniaturizada[135].

Los LCS pueden proporcionar mediciones en tiempo real a un costo menor al de un equipo de referencia, lo que facilita una mayor cobertura espacial que los métodos de referencia actuales para mediciones de contaminantes en el aire. Por otra parte, su inclusión en proyectos de desarrollo personalizados permite modificar la frecuencia de muestreo abarcando en su espectro efectos debidos a eventos de corta duración.

2.3.1. Respecto al funcionamiento de los sensores de bajo costo

Los LCS pueden clasificarse de acuerdo al tipo de tecnología que utilizan para la transducción. Pueden ser capacitivos (p. ej. presión atmosférica), electroquímicos (p. ej. gases), de óxido metálico (también aptos para medir diferentes tipos de gases) u ópticos autónomos que incluyen celdas de absorción ultravioleta (generalmente utilizados para mediciones de ozono) o sensor infrarrojo no dispersivo (p. ej. dióxido de carbono) o de dispersión de luz (p. ej. PM) [136]. Al margen de estas diferencias intrínsecas en el elemento básico de detección, todos comparten en mayor o menor medida algunos componentes en común, tales como hardware para amplificación de señal, conversión analógica digital, procesamiento de señal, controles ambientales, manejo de energía, fuente de alimentación, encapsulado físico y componentes de software (para procesamiento y almacenamiento de datos, telecomunicaciones y visualización).

Algunos aspectos importantes a considerar sobre los LCS, que hacen a su desempeño, son su linealidad, precisión, exactitud, límites de detección, estabilidad, equivalentes de interferencia y tiempos de respuesta. Estos aspectos deben determinarse para contar con un conocimiento cabal del sensor y ajustar su utilidad a un proyecto particular. En base a esto, un punto definitivamente en contra de esta tecnología es que aún no cuenta con protocolos de calibración, validación y evaluación de desempeño exhaustivos que sean universalmente aceptados. Es por ello que su validez formal y acabada para uso institucional y científico aún tiene un camino por recorrer, aunque este es un campo de investigación que se amplía continuamente. Se debe tener en cuenta que el campo técnico está evolucionando rápidamente reemplazando modelos de sensores individuales de manera superadora a sus versiones previas, sin dar tiempo suficiente para establecer métodos de evaluación revisados por pares o, en el caso de que los haya, actualizándolos.

Los equipos integrados por LCS (como todo sistema de medición) son susceptibles a malfuncionamiento o a ser vulnerados por condiciones en las que su desempeño se

ve comprometido, por lo que es fundamental contar con un sistema que dimensione la calidad de los datos. En este sentido hay dos enfoques complementarios, el de gestión automática de calidad (QA) y el de control de calidad (QC). QA está conformado por procesos o etapas que se aseguran de minimizar las inexactitudes de los datos y la necesidad de medidas correctivas posteriores, y se llevan a cabo de forma automática mientras se están registrando los datos. QC es un proceso que se implementa con los datos ya registrados y comprueba que cumplan con los requisitos de calidad determinados por el usuario[137].

En los últimos años, numerosos trabajos se han dedicado a calibrar, validar y estimar el desempeño y limitaciones de sensores de calidad de aire de bajo costo. Kamionka et. al (2006)[138] calibraron un multisensor de gases en laboratorio y con mediciones de campo; Wang et. al (2015)[139] compararon el desempeño y calibraron tres sensores diferentes de PM; De Vito et. al (2008)[140] presentaron una calibración basada en ANN para un sensor electrónico de estimación de concentraciones de benceno; Spinelle et. al (2014 y 2017) [141, 142] calibraron en campo diferentes sensores de gases con un enfoque multimetódico; Shi et. al (2017)[143] validaron un sensor de PM de dispersión de luz haciendo uso de un equipo de medición gravimétrica de largo plazo y calibraron sensores de PM bajo condiciones de estado estacionario en cuanto a concentración de partículas; Lin et. al (2015)[144] evaluaron y calibraron los sensores de gases de una estación comercial de bajo costo; Piedrahita et. al (2014)[145] y Fuertes et. al (2015) [146] desarrollaron estaciones de monitoreo de bajo costo con diferentes sensores de gases en plataforma ARDUINO. A estos antecedentes, se pueden sumar la existencia de varios proyectos de investigación que recolectan datos de calidad de aire a partir de equipos y plataformas de bajo costo. Algunos ejemplos de ello son OpenSense [147], Citi-Sense [148] y Everyaware [149].

En general y a nivel mundial, se ha despertado un interés de las autoridades públicas en aumentar la densidad de monitoreo y mediciones de CA con LCS, pues no pueden permitirse costear suficientes estaciones de monitoreo de CA regulatorias [150]. El campo es prometedor y la literatura respalda que los enfoques de bajo costo pueden ser útiles para la evaluación cualitativa de las concentraciones de partículas en un ambiente moderadamente contaminado y que el despliegue de muchos sensores a escala de vecindario puede proporcionar información sobre los patrones espaciales y/o temporales y la distribución de una fuente [136]. Estas particularidades de los LCS ayudan a refinar los enfoques de modelado, a analizar tendencias a largo plazo y a evaluar las exposiciones humanas a diferentes contaminantes desde una perspectiva más amplia.

Respecto a los LCS de humedad, presión y temperatura

La información meteorológica es fundamental para establecer un diagnóstico de la calidad del aire en una localidad determinada, pues influye en diversos mecanismos que afectan la formación, distribución, dispersión y deposición de contaminantes [134]. Los LCS también permiten medir este tipo de parámetros, los cuales son útiles para diferentes aplicaciones. En los últimos años ha habido un creciente interés en el desarrollo de sistemas autónomos para medir y controlar la temperatura, la presión atmosférica y la humedad relativa en áreas muy sensibles, como los laboratorios biomédicos [151] y biológicos [152], áreas de almacenamiento de sustancias peligrosas y áreas de trabajo controladas con sensibilidad crítica [153]. Estos parámetros también son de particular importancia en el monitoreo ambiental orientado (p. ej. para estudios de isla urbana de calor) dada la tendencia a aumentar tanto la densidad espacial de los sitios de medición como la resolución temporal en cada uno de ellos, para mejorar la precisión espacial y temporal de los modelos predictivos [154].

En general, los sensores de temperatura son resistivos, mientras que los de humedad relativa y presión son capacitivos. Sin embargo, los LCS dedicados a mediciones de estos parámetros son muy diversos y aún no hay disponible una cantidad de información técnica suficiente para seleccionar el sensor que mejor se adapte a una aplicación específica. Por lo tanto, existe una necesidad urgente de caracterizar este tipo de sensores, sus incertidumbres y rendimiento para comprender mejor sus potenciales y limitaciones [155].

Respecto a LCS de PM

Las técnicas de medición de PM de bajo costo se basan más comúnmente en mediciones ópticas (basadas en la dispersión de la luz), haciendo uso de una fuente de luz de baja potencia, ya sea un LED o un láser, donde las partículas que se recogen dispersan la luz (en un fenómeno conjunto de reflexión, refracción y difracción) que es medida por un fotodetector (generalmente un fototransistor) en el dispositivo. La concentración de partículas es proporcional a la intensidad de luz dispersa y se supone una distribución de densidad y tamaño de partículas determinados. La mayoría de estos sensores suele contar con un mini-ventilador para forzar la circulación de aire a través de la cámara de muestreo.

Las aplicaciones de LCS para la medición de PM emplean dos técnicas. Por un lado, la nefelometría (que mide la dispersión de la luz de partículas de un conjunto de aerosoles)

y, por el otro, el recuento óptico de partículas (que mide el tamaño de las partículas y el número de partículas individuales). El límite de detección real de tamaño de partículas de estos sensores suele ser mayor a $3 \mu\text{m}$ y no hay sensores disponibles que detecten partículas ultrafinas (PM_1).

Aunque poseen ventajas como tiempos de respuesta despreciables (dependen de la transferencia de luz dentro de la recámara y de la velocidad del transporte de electrones en el circuito) y no poseer equivalentes de interferencia, su rendimiento reportado en la literatura es un tanto heterogéneo [156, 157, 158] y puede depender sobre el tipo de sensor de partículas empleado (de los cuales hay gran variedad en el mercado). La variable interferente más significativa observada con los LCS de PM es la humedad relativa, debido al crecimiento higroscópico que experimentan las partículas en el aire y a verse afectado el detector. Estudios recientes sugieren una degradación en el rendimiento de los LCS de PM cuando la humedad relativa supera el 80-85 % [159]. Por otro lado, la dispersión de la luz se ve fuertemente afectada también por la densidad, índice de refracción y composición de partículas. Todos estos factores varían de un sitio a otro y con la estacionalidad, por lo que en la actualidad no existe un protocolo común para probar LCS de PM contra un instrumento de referencia.

Los sensores de la serie PMS de “Plantower” (en esta Tesis se trabajó con el PMS5003) tienen una buena calidad de fabricación y proveen un buen desempeño, pese a su bajo costo [20]. Sayahi et. al (2019) evaluaron los sensores PMS1003 y PMS5003 en el exterior durante varias temporadas y bajo diferentes episodios de elevadas concentraciones de PM, consiguiendo buenas correlaciones con el monitor de referencia especialmente en invierno[160]. Bulot et. al (2019) instalaron en el exterior diferentes sensores de PM (entre ellos el PMS5003) durante un año y encontraron de moderadas a buenas correlaciones con una estación de referencia en las cercanías, aunque remarcaron que el desempeño se ve afectado por la temperatura, humedad relativa y eventos de concentraciones altas de partículas[161]. Kuula et. al (2020) realizaron pruebas de laboratorio con el sensor PMS5003, entre otros, variando las distribuciones de tamaños de partículas y, al comparar con un instrumento de referencia, encontraron inconsistencias en todos los sensores[162]. Todos los LCS de PM poseen sus problemas a la hora de distribuir los tamaños de partículas en bins y, en general, los trabajos de calibración/validación utilizan como referencia las concentraciones de $PM_{2,5}$ por ser las más confiables en estos instrumentos [20]. Esto se basa en que los LCS solo capturan una fracción muy pequeña de las partículas de aerosol y, por lo tanto, dependen en gran medida de las estadísticas y la extrapolación. La densidad de recuento de números de la fracción PM_{10} de

aerosoles típicos es extremadamente baja; por lo tanto, esa fracción no se puede medir directamente con estos sensores.

Resulta necesario profundizar en algunos conceptos y teorías que se utilizan a la hora de caracterizar la presencia de partículas en el aire en base a su interacción con la luz, para contar con más herramientas de evaluación de los LCS de PM. La intensidad de la luz dispersada por aerosoles es una función de la longitud de onda (λ), el ángulo de dispersión (θ) medido respecto a la dirección de la luz incidente, el diámetro de partículas (D_p) y el índice de refracción (n) relativo entre la partícula y el medio. La relación entre D_p y λ indica el régimen de dispersión en el cual nos encontramos:

- Si el tamaño de partículas es similar a la longitud de onda de la luz incidente, estamos en el régimen de Mie para la dispersión (caso más común para los aerosoles atmosféricos).
- Si el tamaño es mucho menor que la longitud de onda, la dispersión se explica a través del régimen de Rayleigh. Esta componente es la que se resta en la **Ecuación 2.20** para determinar la extinción de la luz en una medición fotométrica.
- Si el tamaño de partícula es mucho mayor que la longitud de onda de la luz, entonces se da una dispersión geométrica simplificada.

La teoría de Mie[163] predice un patrón de dispersión en base a la relación entre la intensidad de la luz y el ángulo de detección, con el cual puede determinarse una distribución de tamaños de partículas. Para ello, una serie de suposiciones se deben asumir[20]:

- La luz es monocromática y se compone de ondas planas
- Las partículas son esféricas e isotrópicas
- Se consideran tanto la absorción como la dispersión de la luz
- La luz dispersada de una partícula a otra es despreciable
- Las características de dispersión son independientes del movimiento de las partículas
- No se consideran efectos cuánticos

Un espectrómetro láser para medición de aerosoles (mecanismo que aprovechan los LCS de PM) funciona colimando un haz de luz perpendicular al flujo de aire que ingresa a la cámara de detección. Cuando el láser ilumina una partícula la luz se dispersa y se colecta una fracción de la misma a un ángulo dado por un detector. De acuerdo con la teoría de Mie, la amplitud de cada pulso medido se correlaciona con el tamaño de la partícula. Cada pulso se cuenta y asigna a alguno de los bins (rangos) de tamaños estandarizados[20].

Con el fin de modelar la distribución de aerosoles presentes en una muestra de aire, se la puede considerar como compuesta por una suma de n modos lognormales, como se indica en la **Ecuación 2.22**[164].

$$\frac{dN}{d\log(D_p)} = \sum_{i=1}^n \frac{N_i}{\sqrt{2\pi}\log(\sigma_i)} \exp\left(-\frac{(\log(D_p) - \log(\bar{D}_{p,i}))^2}{2\log^2(\sigma_i)}\right) \quad (2.22)$$

Donde N es el número de partículas (luego N_i se denomina número de concentración de cada modo i , es decir número de partículas por cm^3), σ_i la desviación estándar geométrica (variedad del tamaño de partícula) y $\bar{D}_{p,i}$ es la media geométrica del diámetro de partícula (la mediana de la distribución de tamaños para cada modo).

A su vez, se deben determinar algunas características que afectan a las propiedades ópticas de aerosoles para terminar de definir un modo de distribución de partículas. Estas son: el factor higroscópico de crecimiento ($f(RH)$), la densidad de partículas (ρ) y el índice complejo de refracción (m_i) (cuya parte real indica la razón entre la velocidad de la luz en el medio y en el vacío, y cuya parte imaginaria corresponde al coeficiente de extinción de la luz).

La función de distribución de probabilidades está dada por la suma final de todos los modos en la **Ecuación 2.22**, y se denomina también función de número de concentración ya que el área bajo la curva de la gráfica de $dN/d\log(D_p)$ vs D_p es igual al número de concentración total. La distribución de números de partículas en función del diámetro (en escala logarítmica de base 10) suele simbolizarse como $n_N^o(\log(D_p))$, o $n_N(D_p)$ para el diámetro en una escala lineal.

Al multiplicarse la distribución de la **Ecuación 2.22** por el volumen de partícula simplificado a una esfera se obtiene la distribución volumétrica. Si se integra esta distribución entre dos tamaños de partículas se consigue la función de distribución acumulada volumétrica, y su producto con la densidad permite estimar la carga másica de aerosoles por unidad de volumen, que llamamos material particulado (PM en $\mu g/m^3$), entre los diámetros elegidos (**Ecuación 2.23**). Por ejemplo, $PM_{2,5}$ surge de integrar la distribu-

ción volumétrica entre los diámetros 0 y 2,5 μm y multiplicar por la densidad ρ .

$$PM = \rho \frac{\pi}{6} \int_{D_{min}}^{D_{max}} D_p^3 n_N(D_p) dD_p \quad (2.23)$$

Respecto a los protocolos de comunicación

Como se mencionó anteriormente, los sensores realizan el trabajo de transducción desde un fenómeno físico-químico a una respuesta que puede traducirse en valores del parámetro que se desea medir. Con el fin de que el usuario pueda acceder a esta información, se requiere que el sensor se vincule a un microcontrolador que provea linealización integrada, corrección de errores y una salida que pueda ser volcada en una base de datos. La interfaz entre el sensor y el microcontrolador debe contar con un despliegue inteligible, de bajo consumo y de alta fiabilidad [165]. Una estructura serial “bus” es un elemento eficiente que provee dichas características y se presenta en diferentes formas generales, siendo la UART (por su denominación en inglés “Universal Asynchronous Reception and Transmission”), la I²C (por su denominación en inglés “Inter-Integrated Circuit”) y la SPI (por su denominación en inglés “Serial Peripheral Interface”) las más comunes y cubiertas por este capítulo.

UART sostiene transmisión de datos de manera bidireccional, asincrónica (sin pulso de reloj) y serial con dos líneas para datos, una para transmitir (TX) y otra para recibir (RX), que vinculan dos dispositivos (**Figura 2.8**). Puede operar en modo “simplex” (transmisión de datos en una dirección), “half dúplex” (transmisión de datos en dos direcciones, alternadamente) o “full dúplex” (transmisión de datos en dos direcciones, simultáneamente).

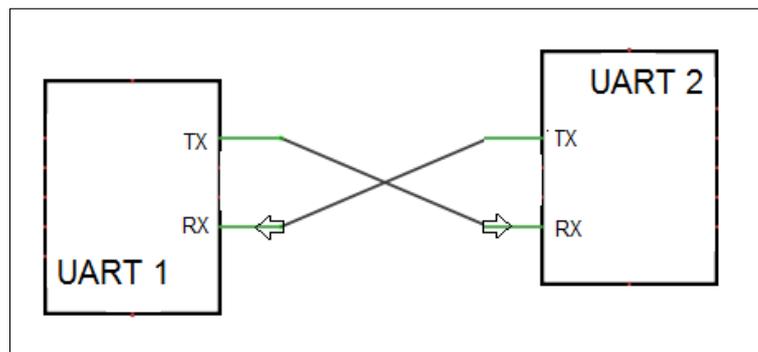


Figura 2.8: Conexión UART entre dos dispositivos

I²C es un “bus” de dos líneas, una de datos (SDA) bidireccional (alternado) y otro de reloj (SCK), que pueden conectar varios dispositivos (cada uno con una identifica-

ción única) que actúan como “maestros” (generalmente microcontroladores) o “esclavos” (sensores) (**Figura 2.9**). Los dispositivos “maestro” inician la transferencia de datos y generan los pulsos de reloj, mientras que los que funcionan como “esclavo” pueden recibir o enviar datos.

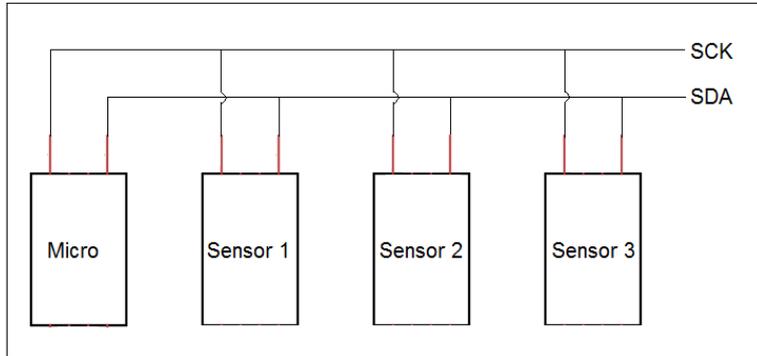


Figura 2.9: Conexión I²C entre cuatro dispositivos, un “maestro” microcontrolador y tres sensores “esclavos”

SPI es un “bus” de tres líneas de comunicación donde los dispositivos también actúan como “maestros” o “esclavos”. En la **Figura 2.10** se puede observar que dos líneas corresponden a transferencia de datos de entrada (desde los “esclavos” al “maestro”) y salida (del “maestro” a los “esclavos”) y una línea al pulso de reloj controlado por el “maestro”. Cada sensor se encuentra conectado al microcontrolador por una línea selectora que alterna la transferencia de datos entre los dispositivos.

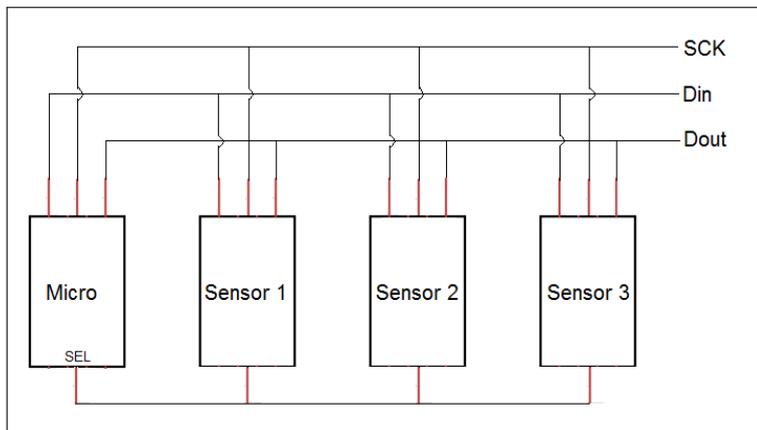


Figura 2.10: Configuración típica de conexión SPI entre tres sensores y un microcontrolador.

3 Metodología

3.1. Potencialidad de las Redes Neuronales en calidad del aire

El desarrollo e implementación de los modelos se centró en $PM_{2,5}$ como variable objetivo. El fin es explorar y dimensionar las capacidades de las ANN tipo MLP con un sitio de estudio que posea suficiente cantidad de datos disponibles para su entrenamiento (CABA), así como proponer una metodología para su diseño y desarrollo. En este sentido, se propone una metodología para llegar a un modelo de alto costo (computacional, disponibilidad de datos, complejidad) en contraste con otro de bajo costo (variables más sencillas de conseguir). El diagrama de la **Figura 3.1** esquematiza esto de forma general.

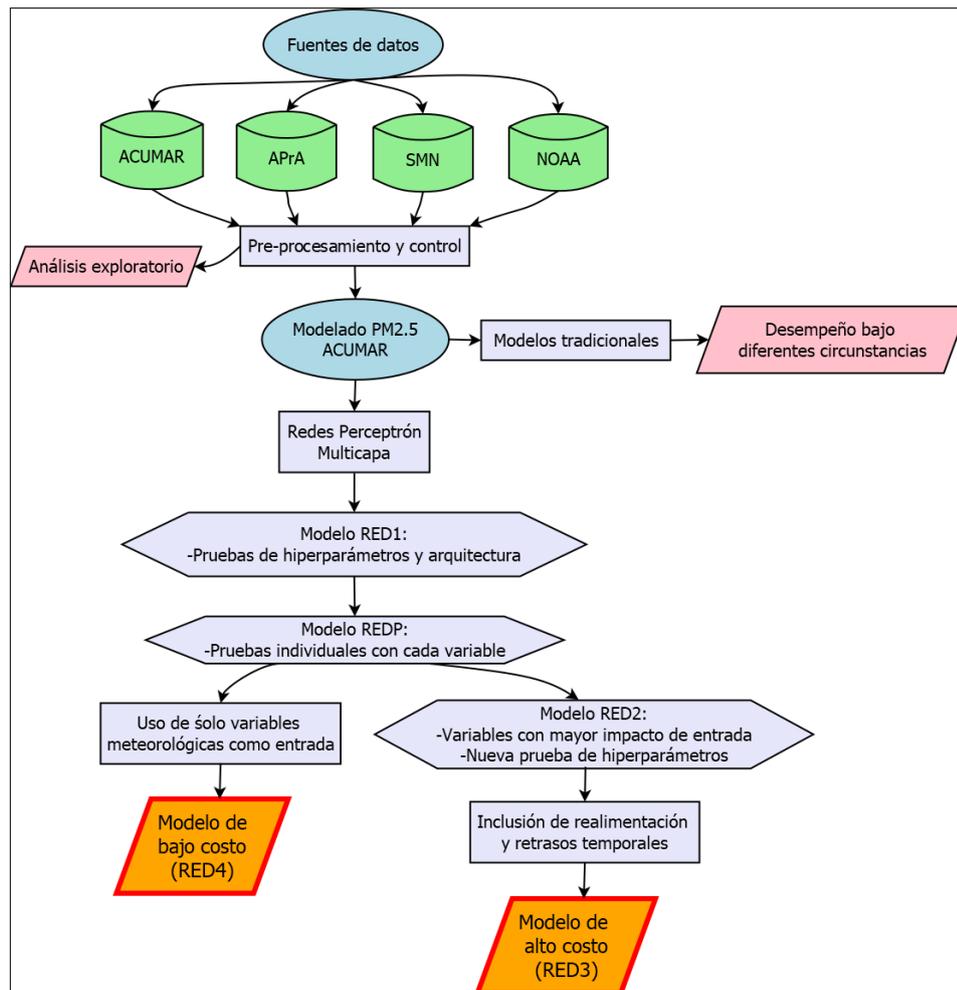


Figura 3.1: Diagrama de flujo explicativo de la metodología implementada para el desarrollo de los modelos ANN

3.1.1. Exploración de datos meteorológicos y de contaminantes

Se utilizaron datos de contaminantes medidos en las estaciones de CA de APrA [32] en La Boca (34.62S, 58.37W) y EMC I en Dock Sud de ACUMAR [33] (34.66S, 58.32W). La estación de La Boca se encuentra emplazada dentro de la Cuenca Matanza-Riachuelo, en una zona residencial/industrial con incidencia de fuentes fijas (**Figura 3.2**). La estación EMC I se encuentra ubicada en el límite de la Cuenca Matanza Riachuelo, en un predio con pocas edificaciones, alejado del tráfico, lindante con el Polo Petroquímico Dock Sud (**Figura 3.3**).

3.1 Potencialidad de las Redes Neuronales en calidad del aire



Figura 3.2: Visualización de la ubicación de la estación de monitoreo en La Boca de APRA. Imagen satelital adaptada de Google Earth

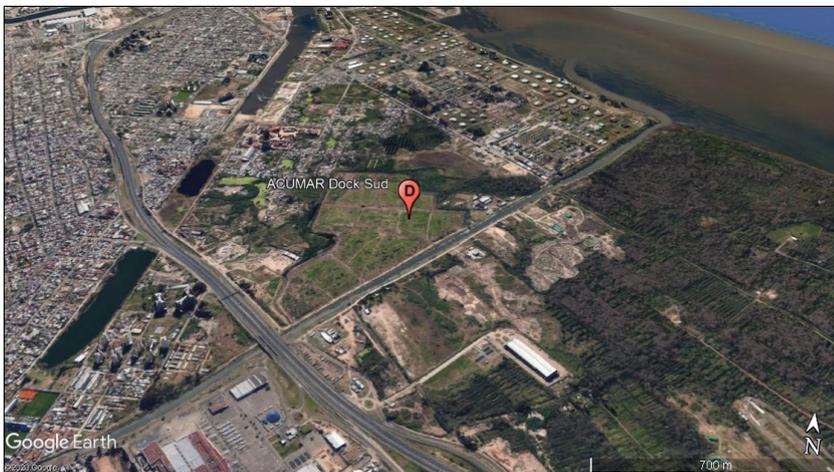


Figura 3.3: Visualización con la ubicación de la estación de monitoreo en Dock Sud de ACUMAR. Imagen satelital adaptada de Google Earth

Respecto a la información meteorológica, se utilizaron datos de la estación del Aeroparque Jorge Newbery (WMO 87582, 34.55S, 58.41W) del Servicio Meteorológico Nacional (SMN) [166]. También se descargaron datos de asimilación (observaciones incorporadas a un modelo) del Laboratorio de Recursos del Aire del NOAA ("National Oceanic and Atmospheric Administration" de Estados Unidos) [167, 118], en una grilla de resolución de 1° centrada en las coordenadas de la estación EMC I de ACUMAR.

En la **Figura 3.4** se pueden observar las estaciones de monitoreo de contaminantes de APRA en La Boca y de ACUMAR en Dock Sud, la estación meteorológica del SMN en Aeroparque y la ubicación de la UNGS, en un mismo mapa. La estación del SMN en

3 Metodología

Aeroparque se encuentra en CABA, limitando con el Río de La Plata. Si bien se encuentra a una distancia considerable de las estaciones de calidad de aire, corresponde a la fuente de variables meteorológicas específicas necesarias para este estudio más cercana de la que se dispone hasta el momento.

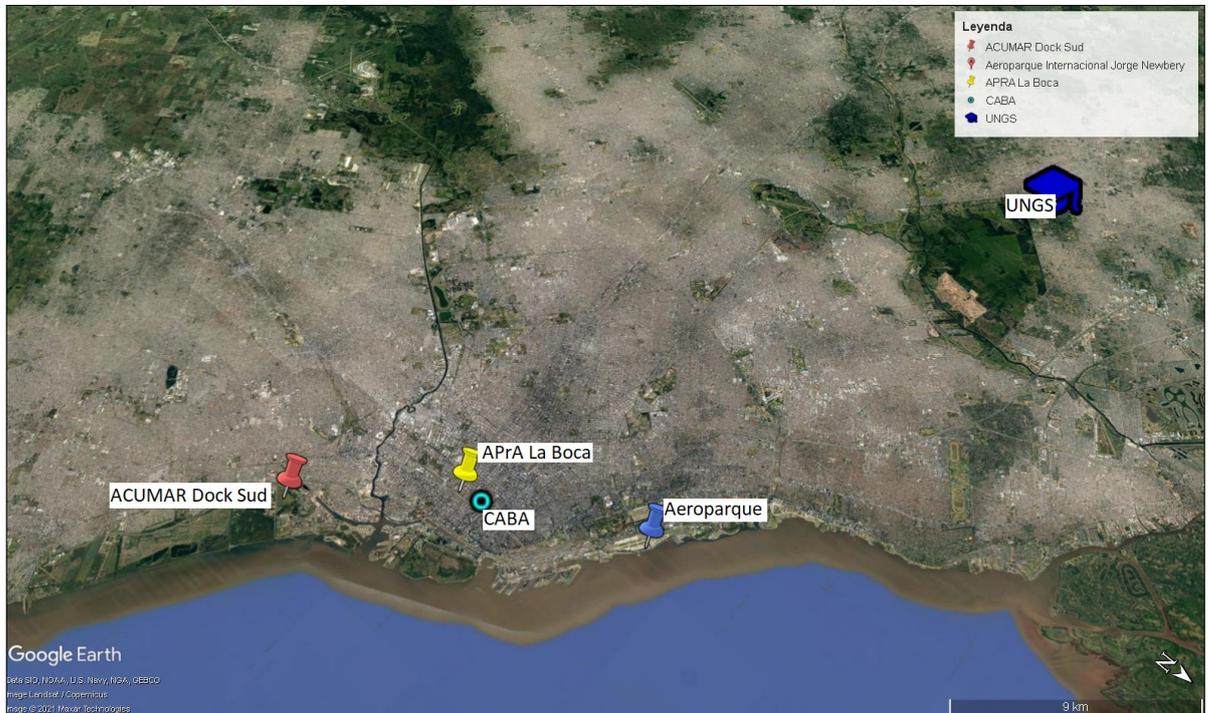


Figura 3.4: Visualización de CABA con las estaciones de monitoreo de contaminantes y meteorología. Aeroparque se encuentra a unos 7Km de la estación de APrA y a unos 14Km de la estación de ACUMAR. Imagen satelital adaptada de Google Earth

La **Tabla 3.1** lista las variables utilizadas con sus unidades y fuentes, además de la nomenclatura que se les asignó. En ella se presentan distintos gases, PM, variables meteorológicas y de capa límite atmosférica. Se utilizaron datos horarios correspondientes al período entre el 14/10/2015 y el 31/12/2017 (19440 líneas de datos).

Los datos de algunas variables meteorológicas generalmente son presentados con un código denominado Synop, el cual es utilizado en esta Tesis para convertir los datos de plafond a unidades de metros (m) (ver **Anexo 8.1.1**).

Variable	Nomenclatura	Unidades	Fuente
PM (<10 μm)	<i>PM10apra</i>	$\mu\text{g}/\text{m}^3$	APrA
Dióxido de azufre	<i>SO2</i>	$\mu\text{g}/\text{m}^3$	ACUMAR
Sulfuro de hidrógeno	<i>SH2</i>	$\mu\text{g}/\text{m}^3$	ACUMAR
Monóxido de nitrógeno	<i>NO</i>	$\mu\text{g}/\text{m}^3$	ACUMAR
Dióxido de nitrógeno	<i>NO2</i>	$\mu\text{g}/\text{m}^3$	ACUMAR
Óxidos de nitrógeno totales	<i>NOx</i>	$\mu\text{g}/\text{m}^3$	ACUMAR
Ozono	<i>O3</i>	$\mu\text{g}/\text{m}^3$	ACUMAR
Benceno	<i>ben</i>	$\mu\text{g}/\text{m}^3$	ACUMAR
Etilbenceno	<i>etilben</i>	$\mu\text{g}/\text{m}^3$	ACUMAR
MP-xileno	<i>MPxi</i>	$\mu\text{g}/\text{m}^3$	ACUMAR
O-xileno	<i>Oxi</i>	$\mu\text{g}/\text{m}^3$	ACUMAR
Tolueno	<i>tol</i>	$\mu\text{g}/\text{m}^3$	ACUMAR
PM (<10 μm)	<i>PM10acu</i>	$\mu\text{g}/\text{m}^3$	ACUMAR
PM (<2.5 μm)	<i>PM25acu</i>	$\mu\text{g}/\text{m}^3$	ACUMAR
Monóxido de carbono	<i>CO</i>	mg/m^3	ACUMAR
Hidrocarburos totales	<i>HCT</i>	mg/m^3	ACUMAR
Hidrocarburos no metánicos	<i>HCNM</i>	mg/m^3	ACUMAR
Hidrocarburos metánicos	<i>HCM</i>	mg/m^3	ACUMAR
Temperatura	<i>Tacu</i>	$^{\circ}\text{C}$	ACUMAR
Humedad Relativa	<i>HRacu</i>	%	ACUMAR
Velocidad del viento	<i>WSacu</i>	m/s	ACUMAR
Dirección del viento	<i>WDacu</i>	$^{\circ}$ sex.	ACUMAR
Presión	<i>Pacu</i>	hPa	ACUMAR
Precipitación	<i>PPTacu</i>	mm	ACUMAR
Radiación	<i>RADacu</i>	W/m^2	ACUMAR
Temperatura de Rocío	<i>Troc</i>	$^{\circ}\text{C}$	ACUMAR
Plafond	<i>pla</i>	ad.	AEROPARQUE
Visibilidad	<i>vis</i>	m	AEROPARQUE
Nubosidad	<i>nub</i>	octavas	AEROPARQUE
Altura de capa límite	<i>BLHcaba</i>	m	NOAA

Tabla 3.1: Variables disponibles con unidades, fuente y nomenclatura utilizada. Se separa con doble línea los contaminantes de las variables meteorológicas

Pre-procesamiento y análisis

En primera instancia, se realizó una exploración básica de los datos de contaminantes y meteorológicos. Los datos de altura de capa límite de asimilación se proveen con una resolución temporal de 3 h, por lo que se implementó una interpolación lineal a dicha serie para completar los datos horarios faltantes.

3 Metodología

Todas las series de datos son sometidas a un proceso de análisis de calidad antes de ser publicados o brindados por la institución administradora correspondiente, por lo que se procedió a realizar un control de los mismos limitado a la búsqueda de valores fuera de rango (límites de detección informados por las instituciones) ausentes o faltantes clasificándolos como valores no numéricos (NaN).

Generalmente, no se cuenta con observaciones de estabilidad atmosférica y, por ello, se han establecido métodos alternativos que utilizan datos de mayor disponibilidad y que continúen representando a los forzantes físicos de la turbulencia. El método que se utilizó aquí consiste en emplear la clasificación propuesta por Turner (1964) [168] (se adjunta el detalle en el **Anexo 8.1.2**). De esta manera se obtiene una clasificación de 7 clases, siendo: 1. Extremadamente inestable; 2. Moderadamente inestable; 3. Ligeramente inestable; 4. Neutral; 5. Ligeramente estable; 6. Moderadamente estable; 7. Extremadamente estable.

Dado el comportamiento cíclico de la dirección del viento, que puede introducir un sesgo en los modelos, los datos de viento se transformaron en componentes zonal u (este-oeste) y meridional v (norte-sur), como se indica en las **Ecuaciones 3.1** y **3.2**.

$$u = ws * \sin (wd) \quad (3.1)$$

$$v = ws * \cos (wd) \quad (3.2)$$

Donde ws y wd son la velocidad y dirección del viento, respectivamente. Para los datos de viento de ACUMAR, las componentes zonal y meridional se nombraron $Uacu$ y $Vacu$, respectivamente.

Se procedió realizar gráficas de tendencias diarias, semanales y anuales de valores medios con intervalos de confianza al 95 %, rosas de los vientos y rosas de contaminantes. Se realizó este análisis con los datos de las variables de probada incidencia en el PM, lo cual tendrá una incidencia en los modelos desarrollados.

3.1.2. Modelado de PM

Implementación de modelos univariados y lineales

La serie original de PM25acu tiene datos faltantes, como generalmente sucede con series temporales de magnitudes ambientales. En este sentido, el modelado de la mag-

nitud es una herramienta útil no sólo para entender relaciones con otras variables y patrones, sino también para imputar los datos faltantes de una serie medida. Previo a la implementación de los modelos ANN, se emplearon métodos univariados como interpolaciones lineal, spline (segmentada cúbica) y vecinos cercanos (interpolación proximal) para completar la serie de $PM_{2,5}$ de ACUMAR ($PM25acu$). Asimismo, se generaron subseries a completar con diferentes tamaños y distribuciones de huecos con el fin de probar la sensibilidad de estos métodos tradicionales a esas variaciones y comparar los resultados con los de los modelos de tipo ANN. Primero, se generaron 2900 datos faltantes artificialmente distribuidos al azar (aproximadamente un 15 % de la serie total de $PM25acu$). Luego, en una subserie de $PM25acu$ sin faltantes de 768 datos consecutivos se generaron artificialmente huecos de longitud 2 (dos faltantes consecutivos) hasta 10 (diez faltantes consecutivos) distribuidos al azar en un 30 % de la subserie. Todos estos casos fueron imputados con los métodos univariados con el fin de estudiar el desempeño de estos modelos a la hora de modelar $PM_{2,5}$.

Modelado con Redes Perceptrón Multicapa

Posteriormente, se procedió a diseñar un primer modelo de ANN Perceptrón Multicapa (MLP), de flujo directo (FF) y propagación inversa (BP) de estimación de $PM_{2,5}$, llamada RED1. Este tipo de modelos no tienen una versión acabada y completa de protocolo para su diseño e implementación, dado que son dependientes del problema en cuestión. Aquí se presenta una metodología propia partiendo de un MLP de tres capas y se utilizaron como variables de entrada a las presentadas en la **Tabla 3.1**, la de estabilidad (*estab*) construida, y otras variables auxiliares que son proxy a las actividades humanas (como día de la semana “*diasem*”, hora “*hora*” y mes del año “*mes*”). La estructura básica de la RED1 consiste en una capa de entrada con todas las variables descritas anteriormente (33 neuronas de entrada), una capa oculta y una capa de salida con una neurona ($PM25acu$, usado como variable objetivo). Si bien no es aconsejable utilizar una gran cantidad de neuronas de entrada, dado que esto aumenta las probabilidades sobre-entrenar el modelo, la RED1 se tomó como banco de pruebas con el espacio de entrada completo para determinar luego los hiper-parámetros con mejor rendimiento.

Se comenzó con los parámetros por defecto descritos en la **Tabla 3.2**. La función “mapminmax” normaliza las series en un rango entre -1 y 1, la función de transferencia “tansig” es la tangente hiperbólica sigmoide y “purelin” es lineal. “trainlm” se refiere a la función de entrenamiento Levenberg-Marquardt con método por lote y la función “dividerand” divide los patrones de entrada al azar (con proporción de 70 %

3 Metodología

para entrenamiento, 15 % para validación y 15 % para prueba). Los valores de épocas, amortiguación y fallas corresponden al control de detención temprana que busca disminuir la probabilidad de que la red converja en un mínimo local y detiene el entrenamiento para lograr una convergencia y generalización adecuadas. Para evaluar el desempeño del modelo se utilizaron la raíz del error cuadrático medio (RMSE) y el coeficiente de correlación lineal (R) durante la etapa de validación. Se evaluó el desempeño de la RED1 con diferente cantidad de neuronas ocultas utilizando la técnica de crecimiento (debido a la presencia de componentes azarosas en los parámetros de las redes, en cada prueba se entrenaron los modelos 10 veces con el fin de determinar el mejor desempeño).

Descripción	Función o parámetro
Normalización de variables y objetivo	mapminmax
Función para neuronas ocultas	tansig
Función para neurona de salida	purelin
Inicialización de pesos y sesgos	Nguyen-Widrow
Entrenamiento	trainlm
División de datos	dividerand
Número máximo de épocas	1000
Parámetro de amortiguación	1×10^{-10}
Gradiente mínimo	1×10^{-7}
Número máximo de fallas	6

Tabla 3.2: Funciones y parámetros utilizados por defecto en el modelo RED1

Con el fin de suavizar la señal del objetivo *PM25acu*, se le aplicó un filtro de medianas de tercer orden y se entrenó nuevamente la RED1 con los mismos parámetros descritos y diferente cantidad de neuronas ocultas. Con esto se buscó reducir el efecto de posibles “outliers” (valores de concentración atípicos) y ruidos en la serie temporal, como fue también descrito en otros antecedentes (ver **Apartado 2.1.1**). Al mejorar el desempeño del modelo, se continuaron las pruebas con la serie filtrada.

Luego, se procedió a reducir la dimensionalidad del espacio de entrada con el fin de reducir el costo computacional y la linealización del modelo. Para ello, se propuso e implementó un método fuera del modelo entrenando otra red de prueba (REDP) alternadamente con cada variable de entrada para evaluar el R de validación en la estimación de la variable objetivo. Esta información, permitió determinar los mejores predictores para RED1 con un método no lineal. También se practicó una correlación lineal cruzada entre todos los predictores con el fin de detectar variables que entreguen información similar a la red. Adicionalmente, se entrenó la RED1 con un método directo

gradual que consistió en agregar uno por uno los predictores (en el orden de mayor a menor R de validación obtenido) y determinar la cantidad óptima de acuerdo a los indicadores de desempeño de la red. De esta manera, se estableció un modelo RED2 con los mismos parámetros dictados para RED1, pero con 15 predictores de entrada y 20 neuronas ocultas.

A continuación, se probó la sensibilidad de RED2 con diferentes hiper-parámetros tales como funciones de pre-procesamiento, funciones de entrenamiento, funciones de transferencia de neuronas ocultas y funciones de inicialización de pesos y sesgos.

En algunas ocasiones, alimentar una ANN MLP con su salida y un delay puede mejorar su desempeño. Por ello, se diseñó una ANN dinámica recurrente con retraso de un paso temporal (RED3) con las mismas propiedades y predictores de la RED2. Luego, se procedió a evaluar su desempeño bajo distintas cantidades de neuronas ocultas. El modelo resultante puede considerarse como de “alto costo” (computacional y de requerimiento de mediciones).

Finalmente, una red (RED4) se construyó utilizando como entrada sólo parámetros meteorológicos, para contar con una referencia de modelo de “bajo costo”. El desempeño de este modelo es comparado con el de la RED3 con el fin de demostrar, a la hora de entender los mecanismos de contaminación atmosférica, la importancia de contar con mediciones de contaminantes in situ. Pero también resulta útil para comparar con un modelo que pueda implementarse en el Centro-Norte del conurbano con los datos disponibles (ya que no se cuenta en esa zona con estaciones de medición de CA).

Los resultados obtenidos a partir de la metodología descrita en esta sección se pueden consultar en el **Capítulo 4**.

3.2. Caracterización de aerosoles

Los métodos presentados aquí tienen el fin de estudiar las propiedades ópticas de aerosoles en el Centro-Norte del conurbano a partir de mediciones fotométricas propias y la implementación de ANN tipo SOM (sin antecedentes en este tipo de estudios). El diagrama de la **Figura 3.5** esquematiza esta metodología de forma general.

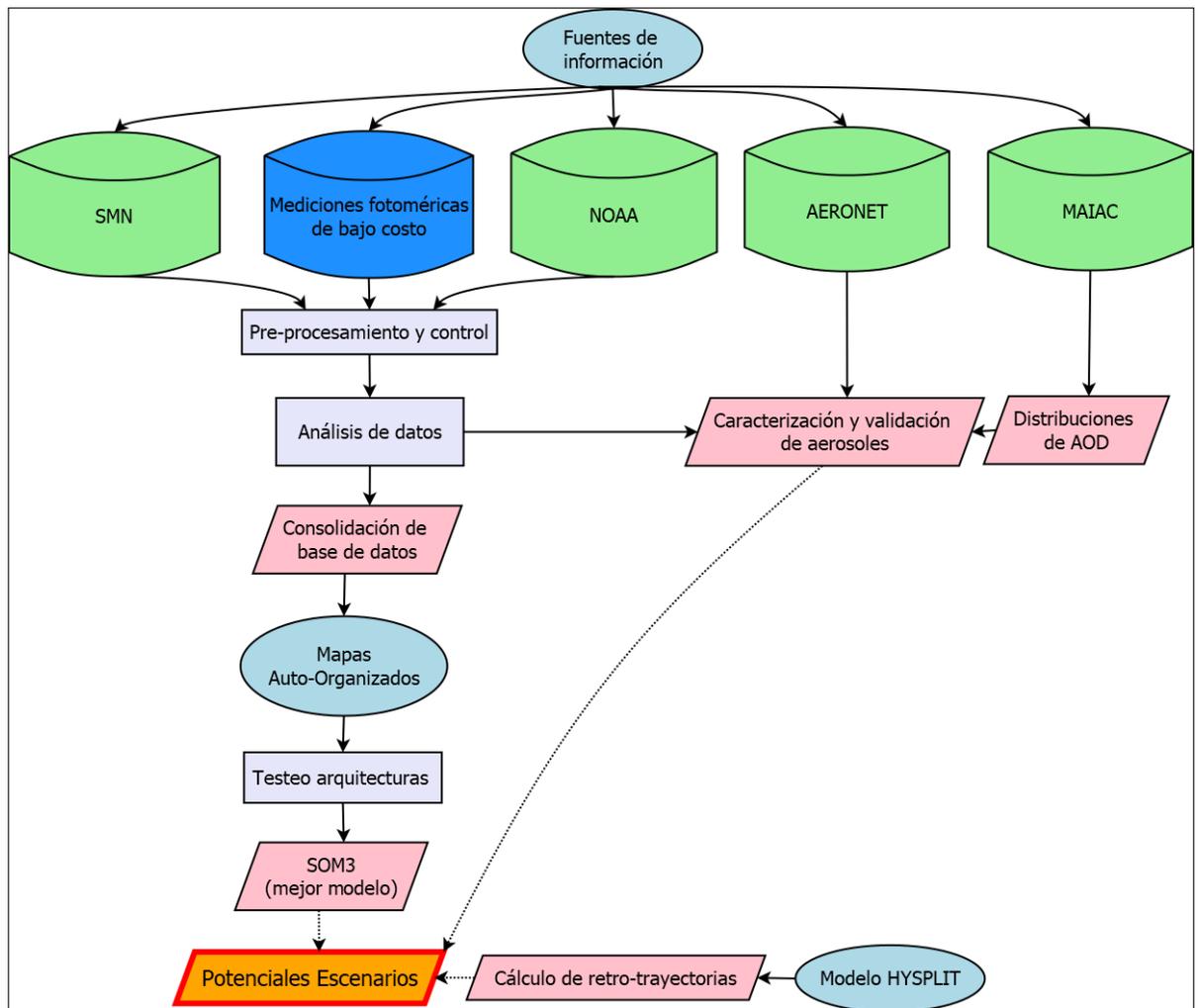


Figura 3.5: Diagrama de flujo para la metodología utilizada en esta sección

3.2.1. Exploración de datos meteorológicos

Se realizó una búsqueda y exploración de la meteorología de la zona en torno a la UNGS con el fin de comprender sus principales características y usarla en el posterior modelado de los escenarios con las propiedades ópticas de aerosoles.

Los parámetros meteorológicos de superficie medidos en la zona (temperatura (T_{cm}), humedad relativa (HR_{cm}), presión (P_{cm}), velocidad del viento ($WScm$) y dirección (WD_{cm})) fueron proporcionados por el Servicio Meteorológico Nacional (SMN). Nótese que se utilizó una nomenclatura diferente para las mismas variables registradas en la estación de ACUMAR utilizadas en el desarrollo de los modelos ANN MLP. Estos datos representativos para la zona de la UNGS se registraron en la estación Campo de Mayo

(WMO 87570, 34.53S, 58.67W), ubicada aproximadamente a 3,8 km de la universidad. Los datos del SMN tienen una resolución horaria, por lo que se utilizó el registro más cercano en el tiempo con cada medición de AOD en UNGS. Los datos de viento se utilizaron para calcular las componentes zonal y meridional siguiendo las **Ecuaciones 3.1 y 3.2**. Con el fin de diferenciar su nomenclatura respecto a las mismas variables para ACUMAR, se las nombró U_{cm} y V_{cm} , respectivamente.

Más allá de la velocidad del viento en las componentes de la dirección, es importante representar los efectos del módulo de la velocidad del viento en los valores de AOD. Por lo tanto, U_{cm} , V_{cm} y W_{Scm} fueron consideradas como variables de interés.

Dado que el efecto principal de la humedad sobre los aerosoles es el crecimiento higroscópico, se calculó el factor de influencia higroscópico $f(HR)$ [169, 170, 144] (**Ecuación 3.3**) y se incluyó como variable de entrada en lugar de HR_{cm} . El crecimiento higroscópico de aerosoles juega un papel importante en la caracterización de aerosoles atmosféricos, ya que las propiedades físicas, químicas y ópticas de los aerosoles dependen de la humedad relativa, así como de su tamaño y composición.

$$f(HR) = \frac{1}{1 - HR} \quad (3.3)$$

Con un rango de HR (humedad relativa) entre 0 y 1 (aplicado en este caso a HR_{cm}).

Se utilizó también meteorología archivada del Sistema Global de Asimilación de Datos (GDAS) del NOAA [167]. Estos datos tienen 1° de resolución espacial y la ventana seleccionada está centrada en las coordenadas de UNGS. Como la resolución temporal es de 3 horas, se utilizó una interpolación lineal para obtener los datos horarios del modelo y seleccionar los datos más cercanos en tiempo a las mediciones de AOD en UNGS. Las variables seleccionadas fueron: precipitación acumulada de lluvia (PPT_{ungs}), altura de la capa límite (BLH_{ungs}), velocidad (ws_{850}) y dirección (wd_{850}) del viento a 850 hPa.

La selección de la altura de referencia a 850 hPa de presión para los datos de viento responde a observaciones que indican que el entorno atmosférico a esa altura puede representar las condiciones reales de la atmósfera en superficie sin verse influenciado por edificios y la topografía [109]. Bajo el mismo criterio establecido para datos de viento con otras fuentes utilizadas, se calcularon las componentes zonal (u_{850}) y meridional (v_{850}) utilizando las **Ecuaciones 3.1 y 3.2**. Dado que el AOD responde a la integración a lo largo de una columna vertical desde el suelo hasta la parte superior de la atmósfera, los datos a 850 hPa se utilizan como entrada para el modelo de agrupamiento SOM en

3 Metodología

lugar de los datos del viento en superficie del SMN. Para validar el uso de viento a 850 hPa como representativo para la zona de la UNGS, se realizó una comparación entre los datos de viento del SMN en Campo de Mayo y a 850 hPa del NOAA a través de una regresión lineal entre los componentes zonal y meridional de ambas fuentes.

A su vez, se utilizaron datos de sondeos también proporcionados por NOAA [167], que permitieron identificar capas de inversión que juegan un papel importante en la dispersión de contaminantes, actuando como barreras para el transporte vertical. La **Figura 3.6** muestra un sondeo del 7 de enero de 2019, a modo de ejemplo. En el gráfico de la izquierda, la curva roja corresponde a la temperatura del aire y la verde al punto de rocío a diferentes alturas, y en el gráfico de la derecha se representa la curva de temperatura potencial. En el último, las condiciones de inversión se identifican en amarillo y se corresponden con un aumento de temperatura y una disminución del punto de rocío con la altura. Como variables, la ubicación de la capa de inversión se estableció como su altura media ($invH$) y la intensidad de la inversión como la diferencia de temperatura ($invT$) y punto de rocío ($invDP$) dentro de la capa. Un enfoque similar fue presentado por Gu et. al (2018) [109].

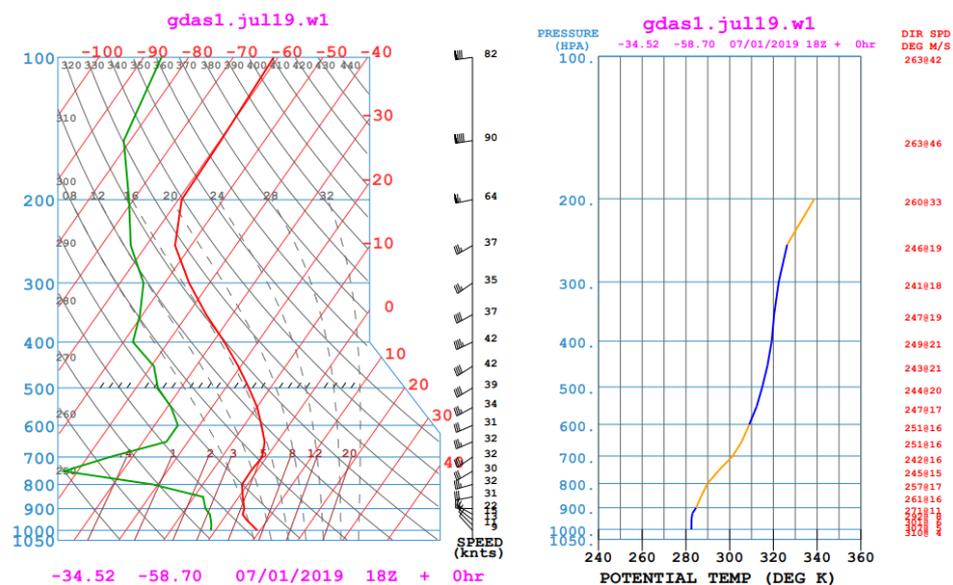


Figura 3.6: Ejemplo de gráficos de sondeo del 7 de enero de 2019. Un diagrama de sesgo-T a la izquierda y un perfil de temperatura potencial a la derecha. El dato corresponde a las 15 h (hora local) y presenta una fuerte capa de inversión entre 800 hPa y 700 hPa. Fuente: NOAA READY [167]

A modo de exploración, se realizó una descripción básica de los datos de las variables

de asimilación y sondeos del NOAA ($PPTungs$, $BLHungs$, $invH$, $invT$, $invDp$) como caracterización de meteorología regional (para el período de estudio entre 20/11/2018 al 29/02/2020). También se realizó una descripción de la meteorología de superficie, en este caso utilizando las series históricas disponibles del SMN en la estación de Campo de Mayo (entre marzo de 2018 y marzo de 2020) de temperatura (T_{cm}), humedad relativa (HR_{cm}), velocidad y dirección del viento ($WScm$ y $WDcm$).

3.2.2. Mediciones fotométricas de bajo costo

Se llevaron a cabo mediciones manuales de AOD en el sitio de la UNGS con un fotómetro solar CALITOO (identificación #1703-0304) del programa “Global Learning and Observation to Benefit the Environment” (GLOBE). El instrumento utiliza un diodo emisor de luz (LED) RGB (por su especialización de emisión en los colores rojo, verde y azul) como fotodetector selectivo, entregando un valor proporcional a la intensidad de luz recibida en las longitudes de onda de 465 nm (canal azul), 540 nm (canal verde) y 619 nm (canal rojo), que luego son convertidos en valores de AOD en base a la ley de Lambert-Beer aplicada a la atmósfera [171]. Cada valor de AOD es registrado junto con la fecha, hora, temperatura, presión atmosférica, altitud, latitud, longitud y ángulo de elevación solar. Los valores de AOD que calcula el fotómetro a partir de las mediciones realizadas están calibrados y son de nivel 1.0, según la denominación de AERONET, con un error inferior a 0,02 unidades de AOD [98].

El fotómetro solar CALITOO incluye, además de un sensor de luz, un módulo GPS, sensores de presión y temperatura, un módulo de memoria y una pantalla LCD. Se alimenta con cuatro baterías tipo AA (1,5V) y puede operar a temperaturas entre -20°C y 55°C .

La **Figura 3.7** muestra una imagen frontal del fotómetro. Se puede observar que, justo arriba de la pantalla, el dispositivo posee una mira donde se concentra la luz solar directa recibida con el fin de facilitar la orientación del equipo hacia el Sol. En la **Figura 3.8** se observa la parte superior del equipo, con la entrada USB para la extracción de datos, el orificio de alfiler para la “mira” que permite orientar el equipo y la entrada de luz para el sensor que funciona como un colimador que permite el ingreso de luz directa y atenúa el ingreso de luz difusa. Su campo visual, cono de luz recogido por el fotómetro, es de $2,5^{\circ}$.



Figura 3.7: Fotómetro solar CALITOO utilizado en las mediciones realizadas en UNGS



Figura 3.8: Vista de la parte superior del fotómetro. De izquierda a derecha se puede observar el puerto USB, el orificio del sensor y el orificio de la “mira”

Este instrumento fue calibrado, validado y calificado por el Laboratorio de Óptica Atmosférica en Lille, Francia [172]. Por lo que su adquisición incluyó un boletín de calibración con los parámetros correspondientes (ver el **Anexo 8.2.2**).

Cada medición en el sitio de la UNGS (34.52S, 58.70W) con el fotómetro se efectuó en ausencia de nubosidad en torno a una circunferencia de 30° de diámetro angular centrada en el disco solar, sin reportes de bancos de niebla en ese momento y con un ángulo de elevación solar mayor a 30°, con el fin de que la masa relativa de aire sea inferior a 2 unidades, según los procedimientos sugeridos por la literatura [98, 94, 95, 106, 112]. A su vez, se tomaron imágenes desde una cámara estacionaria con filtro rojo (absorción selectiva para observar vapor de agua) cada vez que se realizó una medición con el fin de contar con un control adicional del estado del cielo en torno al disco solar.

Pequeños movimientos del fotómetro pueden provocar que la luz directa no incida completamente en el campo visual de 2,5° del instrumento, alterando las mediciones.

Incluso, fluctuaciones propias de la atmósfera pueden tener el mismo efecto. Por ello, cada medición constó de 1 minuto de orientación del colimador al disco solar para que el fotómetro registre el máximo valor de intensidad de luz detectado en ese tiempo.

Caracterización y validación de aerosoles

Para el análisis, se utilizaron los datos de AOD calculados a partir de las mediciones con el fotómetro manual en el campus de la UNGS desde el 20/11/2018 al 29/02/2020 (226 registros). El análisis básico se realizó con los datos correspondientes al canal azul, de 465nm de longitud de onda, por ser el canal más cercano a los 440nm utilizados como referencia en la literatura [128].

También se calcularon los coeficientes de Ångstrom (α) y se graficaron en función de los valores de AOD para clasificar los tipos de aerosoles presentes durante las mediciones. Se utilizó la **Tabla 3.3** de clasificación climática de aerosoles, desarrollada por Otero et al. (2006)[93], ya que se trata de la única referencia local para tal fin. Además, se trata de un esquema que utiliza mediciones de radiación directa como las que registra el fotómetro CALITOO.

Clase de aerosoles	AOD	α
Continental Limpio	0,05-0,15	0,8-1,5
Continental Promedio	0,1-0,3	1-1,7
Continental Contaminado	0,08-0,3	1,5-2,5
Urbanos, Industriales	>0,25	>1,2
Desértico	>0,2	<0,8
Marítimo	<0,15	<0,3
Quema de Biomasa	>0,2	1,5-2,6
Antártico	<0,1	<0,8

Tabla 3.3: Categorías de aerosoles en función de los valores de AOD (a 440nm o cercano) y el coeficiente de Ångstrom α

Esta metodología posee varias incertidumbres, pero proporciona pistas sobre el origen de los aerosoles presentes en la atmósfera. Contempla ocho clases de aerosoles diferentes: “Continental Limpio”, “Continental Promedio”, “Continental Contaminado”, “Urbano/Industrial”, “Desértico”, “Marítimo”, “Quema de biomasa” y “Antártico”. Debe tenerse en cuenta que los límites establecidos para los grupos de clasificación climática son difusos y varían para diferentes autores [128, 93, 96, 129, 130]. Es por ello que las observaciones en base a estos resultados deben realizarse considerando las características propias de cada zona de estudio.

En paralelo, se utilizaron datos de AOD a 440nm del sitio de AERONET más cercano a la UNGS, CEILAP BA (**Figura 1.2**)[34], con el fin de comparar las distribuciones horarias y estacionales de AOD, α y clasificación de este sitio con las de UNGS. Si bien se esperan similitudes regionales entre los aerosoles en UNGS y CEILAP, el objetivo es encontrar características específicas de aerosoles en el sitio de estudio. Con el fin de evitar confusiones en la nomenclatura, se diferenciaron los espesores ópticos de UNGS y CEILAP como AOD_{ungs} y AOD_{cei} , los coeficientes de Ångström como $alpha_{UNGS}$ y $alpha_{CEI}$, y las clasificaciones como $class_{UNGS}$ y $class_{CEI}$, respectivamente.

Con el propósito de contar con una visión más regional y comparar las mediciones en UNGS con otra fuente, se utilizaron datos satelitales de AOD a 470 nm procesados por el algoritmo MAIAC descargados de la pagina EARTHDATA de NASA[173]. Se utilizó una capa de calidad que asegura cielo limpio sin presencia de nubes, y no se implementaron calibraciones posteriores. Luego, se volcaron valores medios por píxel y estación del año en mapas centrados en la UNGS y mostrando también la estación de CEILAP a través del programa QGIS.

3.2.3. Potenciales escenarios

Cálculo de retro-trayectorias

Como complemento para el estudio de escenarios de propiedades ópticas, se caracterizó el origen de masas de aire. Para ello, se modelaron las trayectorias inversas usando el modelo NOAA HYSPLIT disponible en línea [167]. Las retro-trayectorias se calcularon hasta 7 días antes de cada tiempo de medición de AOD en UNGS, y con destino final en esas coordenadas. Como método de movimiento vertical se estableció un campo de velocidad vertical, y se calcularon las trayectorias a tres diferentes alturas (500 m, 1500 m, 2500 m sobre el nivel del mar) por cada tiempo de medición. Se obtuvo también un campo meteorológico a lo largo de cada trazado de trayectoria para la temperatura ambiente, la altura de la capa límite y la humedad relativa. En total, se calcularon 678 retro-trayectorias para rastrear fuentes de aerosoles y relacionarlas con las clases de aerosoles estimadas y el comportamiento de las variables meteorológicas.

Clasificación con Mapas Auto-Organizados

Para trabajar con los modelos SOM, se buscaron los valores más cercanos en tiempo de cada variable meteorológica descrita anteriormente con las mediciones de AOD realizadas. Como paso previo a la modelización, se realizó una regresión lineal entre

todas las variables (excepto para $u850$ y $v850$ debido a su naturaleza cíclica). Para reducir la dimensionalidad del espacio de entrada de los modelos SOM, de las variables meteorológicas que tuvieron una correlación significativa entre ellas, se descartaron aquellas con menor correlación con AOD_{ungs} . Finalmente, se construyó la base de datos de entrada para el modelo SOM con el resto de las variables (**Tabla 3.4**).

Variable	Nomenclatura	Unidades	Fuente
AOD fotométrico	AOD_{ungs}	ad.	Mediciones en UNGS
Coefficiente de Ångström	α_{UNGS}	ad.	Mediciones en UNGS
Clasificación de aerosoles	$class_{UNGS}$	ad.	Mediciones en UNGS
Temperatura	T_{cm}	°C	SMN
Factor de crec. higroscópico	$f(HR)$	ad.	SMN
Precipitación acumulada	PPT_{ungs}	mm	NOAA
Altura de capa mezclada	BLH_{ungs}	m	NOAA
Dif. temp. de rocío de sondeo	$invDp$	°C	NOAA
Altura de inv. de sondeo	H_{inv}	m	NOAA
Velocidad del viento (850 hPa)	$ws850$	m/s	NOAA
Comp. meridional (850 hPa)	$v850$	m/s	NOAA
Comp. zonal (850 hPa)	$u850$	m/s	NOAA

Tabla 3.4: Listado de variables seleccionadas como entradas del modelo SOM. “ad” significa adimensional

Se construyó un SOM básico con una capa de entrada y una capa de Kohonen que representa las neuronas de salida dispuestas en una cuadrícula bidimensional [87]. Las conexiones entre estas dos capas son las encargadas de propagar la información de la entrada a la salida. Cada neurona de entrada está conectada a todas las neuronas de salida por un peso. Por lo tanto, cada neurona de salida está asociada con un vector de referencia o de peso. Los modelos SOM se diseñaron y entrenaron utilizando el paquete de Mapas Auto-Organizados [174] en lenguaje MATLAB. La topología se estableció como una forma de mapa rectangular y enrejado hexagonal, mientras que la función de vecindad seleccionada fue gaussiana, ya que estos ajustes son los recomendados en la literatura [174, 175] y una exploración preliminar realizada mostró que otras combinaciones resultan con el mismo o mayor error. Para minimizar la influencia de cada rango de magnitud variable, los datos de entrada se estandarizaron utilizando una normalización de varianza unitaria y media cero.

Como criterio de calidad del mapa final se utilizaron la precisión de la proyección (que describe cómo se adaptan las neuronas a los datos) y la preservación de la topología.

Los indicadores utilizados para este propósito son el error de cuantificación medio q_e y el error topográfico t_e , respectivamente (**Ecuaciones 2.15** y **2.16**). A su vez, se controló también el indicador de distorsión del mapa (**Ecuación 2.17**). El método más utilizado para definir el tamaño del mapa de salida (número de neuronas) es por ensayo y error, ya que no es posible obtener un índice de criterio absoluto. Sin embargo, los errores de cuantificación, los errores topográficos y la distorsión del mapa pueden ayudar a deducir un tamaño de mapa casi óptimo. Se diseñaron y probaron 32 configuraciones diferentes de Mapas Auto-Organizados (número de neuronas y distribución) con la base de datos de la **Tabla 3.4** como entrada.

Los resultados del modelado se utilizaron para identificar escenarios típicos de cantidad, tamaño y tipo de aerosoles en función de condiciones meteorológicas y el rastreo de fuentes a partir de las retro-trayectorias de masas de aire. Con la integración de las herramientas y métodos de bajo costo implementados, se busca aportar nueva información sobre las propiedades ópticas de aerosoles en la región.

En el **Capítulo 5** se encuentran los resultados obtenidos a partir de la metodología descrita en esta sección.

3.3. Desarrollo e implementación de equipos de bajo costo

Aquí se describen los diseños experimentales desarrollados e implementados para pruebas de los LCS seleccionados y el desarrollo de un sistema de control personalizado. Los métodos se resumen en el diagrama de la **Figura 3.9**.

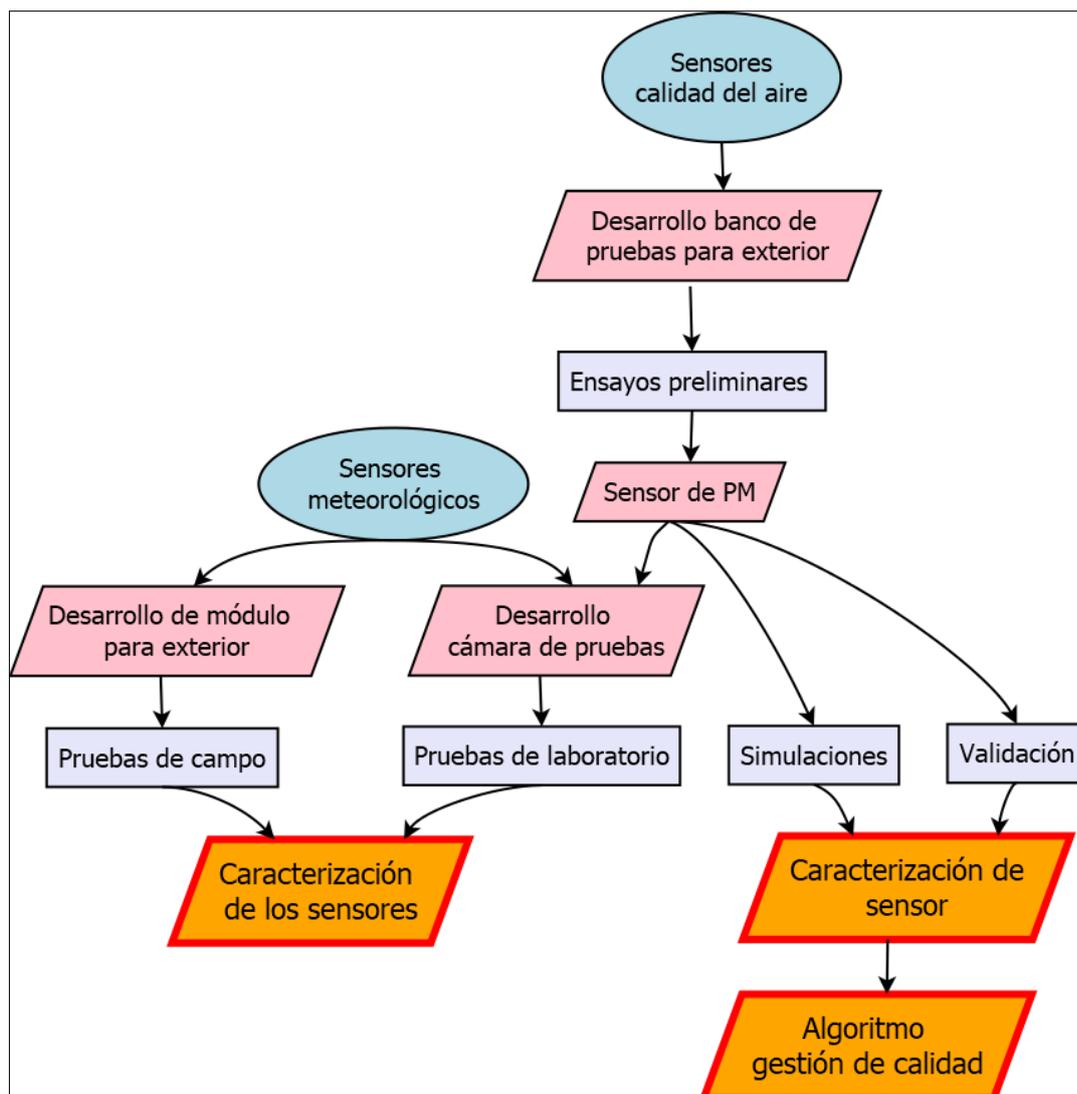


Figura 3.9: Diagrama de flujo explicativo sobre la metodología utilizada para la caracterización de los LCS y el desarrollo del sistema de control

3.3.1. Evaluación de sensores meteorológicos

Se realizó un estudio comparativo de las características de diferentes sensores de humedad, presión y temperatura de bajo costo [176] con el fin de seleccionar el más adecuado para el desarrollo de la estación de medición de calidad del aire ECAUNGS. Los LCS seleccionados para el análisis cumplen las siguientes condiciones [155, 177]: se encuentran disponibles comercialmente en todo el mundo, tienen un diseño liviano y pequeño, su tasa de adquisición sólo está limitada por la placa de procesamiento que se utilice. Teniendo en cuenta estas características, se seleccionaron los sensores de temperatura

3 Metodología

DS18B20 (18b20), de humedad y temperatura DHT11 (DHT), de presión y temperatura BMP180 (BMP) y de humedad, presión y temperatura BME280 (BME), que se utilizan comúnmente en diferentes proyectos de código abierto [178, 179].

El sensor 18b20 es un termómetro digital que proporciona entre 9 y 12 bits para mediciones de temperatura. Se comunica a través de un bus que, por definición, requiere solo una línea de datos (y tierra) para la comunicación con un microprocesador central. Además, puede derivar energía directamente de la línea de datos, eliminando la necesidad de una fuente externa [180]. También tiene un código de serie único de 64 bits que permite que múltiples sensores del mismo tipo funcionen con el mismo bus [181].

El sensor de humedad y temperatura DHT tiene una señal de salida digital calibrada. Incluye un componente resistivo para medir la humedad relativa y un semiconductor NTC para medir la temperatura, ambos conectados a un microcontrolador de 8 bits. Tiene una interfaz de comunicación en serie de 1 línea [182].

El sensor de presión y temperatura BMP tiene una interfaz de comunicación I^2C y está basado en tecnología piezo-resistiva [183].

El BME es un sensor digital combinado de humedad, presión y temperatura. Admite protocolos de comunicación SPI e I^2C y se puede operar en tres modos de consumo: reposo, normal y forzado [184].

La **Tabla 3.5** muestra que todos los sensores tienen bajo consumo de energía. DHT tiene los rangos de medición más restringidos, 18b20 informa la mejor precisión y resolución para medir la temperatura y BME acusa la mejor precisión y resolución que BMP para medir la presión.

Parámetro	18b20	DHT	BMP	BME
Tensión de entrada	[3;5,5]V	[3;5,5]V	[1,6;3,6]V	[1,7;3,6]V
Corriente de consumo	[1;1,5]mA	[0,5;2,5]mA	[3;32] μ A	[1,8;3,6] μ A
Rango de medición	[-55;125] $^{\circ}$ C	[0;50] $^{\circ}$ C; [20;90] %	[-40;85] $^{\circ}$ C; [300;1100]hPa;	[-40;85] $^{\circ}$; [0;100] %; [300;1100]hPa
Precisión	[0,5;2] $^{\circ}$ C	2 $^{\circ}$ C;5 %	2 $^{\circ}$ C; [2;4,5]hPa	1 $^{\circ}$ C; 3 %;1hPa
Resolución	0,0625 $^{\circ}$ C	1 $^{\circ}$ C;1 %	0,1 $^{\circ}$ C; 0,01hPa	0,01 $^{\circ}$ C;0,008 %; 0,8hPa

Tabla 3.5: Características generales de los sensores extraídas de las hojas de datos

Se evaluaron diferentes características de los sensores seleccionados en condiciones de

laboratorio conocidas y se compararon con instrumentos de referencia. Para ello se utilizó una estación meteorológica WT1081 (UNGS WS) (para las pruebas en el exterior) y una sonda de temperatura de respuesta rápida PS-2135 (PS) conectada a un multímetro digital DM3068 (para las pruebas de laboratorio). La estación meteorológica UNGS WS mide la temperatura entre $-40\text{ }^{\circ}\text{C}$ y $+65\text{ }^{\circ}\text{C}$ (resolución de $0,1\text{ }^{\circ}\text{C}$), humedad relativa en un rango de 10-99% (resolución del 1%) y presión entre 700hPa a 1100hPa (resolución de $0,1\text{hPa}$)[185]. Antes de realizar las comparaciones, se la validó comparándola con la estación del SMN CAMPO DE MAYO AERO, N ° 87570. La sonda de temperatura de respuesta rápida PS es un sensor de tipo termistor que mide la temperatura entre $-10\text{ }^{\circ}\text{C}$ y $+70\text{ }^{\circ}\text{C}$, con una precisión de $0,5\text{ }^{\circ}\text{C}$ y resolución de $0,01\text{ }^{\circ}\text{C}$ [186]. Estas características son limitantes para la salida del multímetro digital DM3068[187].

Diseño experimental

Se evaluaron los sensores en base a los siguientes indicadores: (i) linealidad, (ii) exactitud, (iii) precisión, (iv) resolución, (v) tiempo de respuesta y (vi) respuesta a la pérdida de energía.

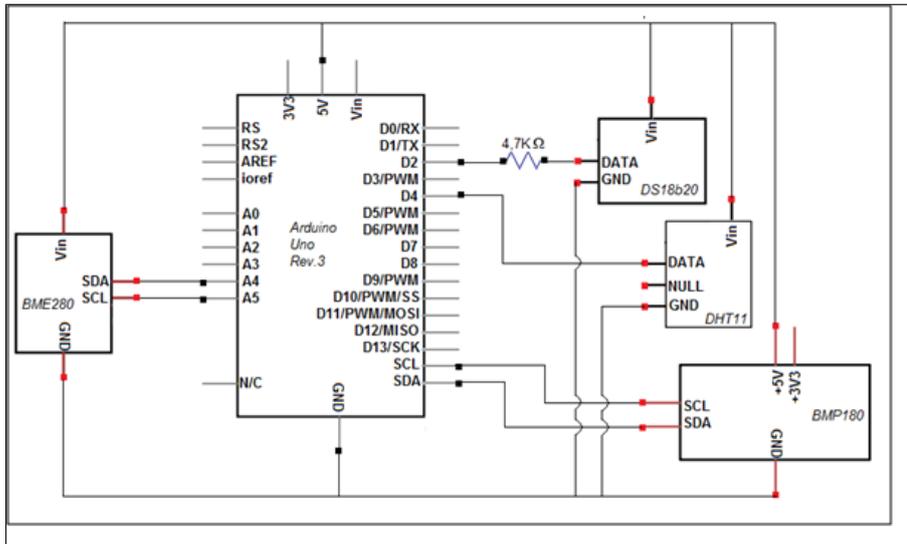


Figura 3.10: Circuito básico ensamblado con la placa Arduino Uno y los sensores BME280, DS18B20, DHT11 y BMP180. La fuente de alimentación difiere según el tipo de experimento y los datos fueron almacenados en una tarjeta SD o visualizados en una computadora o una pantalla LCD.

Los datos de los sensores se procesaron con la ayuda de una placa Arduino UNO, la cual tiene un microcontrolador ATmega 328P que representa su parte central y funciona con

3 Metodología

una frecuencia de 16 MHz. La placa del microcontrolador tiene 6 pines analógicos y 14 entradas/salidas digitales, de las cuales 6 se pueden usar como salidas PWM (modulación de ancho de pulso). También contiene pines de alimentación + 5 V y + 3,3 V y 3 pines GND (tierra). La placa se puede alimentar a través de USB o un adaptador de 7-16 V CA a CC[188]. La **Figura 3.10** muestra el circuito ensamblado básico.

La **Figura 3.11(a)** muestra el arreglo de uno de los experimentos que se llevaron a cabo con diferentes temperaturas, humedad y condiciones de presión en una cámara diseñada para este propósito. La temperatura fue modificada con compresas de hielo (para temperaturas inferiores a la temperatura ambiente) y la radiación de una lámpara infrarroja colocada a diferentes distancias de la cámara (temperaturas superiores a la temperatura ambiente). La humedad fue controlada por un humidificador ultrasónico y la presión se estableció a través de una bomba de aire.

Para comparar los datos de los LCS con la estación meteorológica UNGS WS, se diseñó una caja compacta que contiene los sensores según el circuito de la **Figura 3.10**, con la adición de un reloj de tiempo real DS3231 y un módulo de tarjeta SD, todo alimentado por una batería de 7,4 V y 1300 mA LiPo. Esta caja se colocó al lado de la estación meteorológica UNGS en diferentes etapas planificadas, como se puede ver en la **Figura 3.11(b)**.



Figura 3.11: a) Diseño experimental para ensayos con diferentes valores de temperatura, humedad relativa y presión en la cámara. La placa Arduino y los sensores fueron alimentados por una fuente de alimentación “switching” de 9V para estos experimentos. b) Estuche compacto con sensores de medición instalados junto a UNGS WS

En la comparación de la salida de los sensores con el PS, se cambió la temperatura (entre 18°C y 43°C) en la cámara del laboratorio y se recopilaban los puntos de datos cuando cada salida de los sensores haya alcanzado un estado estacionario. La recopilación de datos de los sensores para comparar con los datos de la UNGS WS de temperatura, humedad y presión se tomaron en diferentes campañas con una duración de 40 horas cada una (temperatura entre 3°C y 26°C, humedad entre 30 % y 93 % y presión entre 1017hPa y 1023hPa). La tasa de muestreo en todos los experimentos se estableció en 100 ms y se

calcularon promedios por hora para el análisis. De esta manera, se evaluó la linealidad trazando las salidas de los sensores contra los instrumentos de referencia, utilizando regresión de cuadrados mínimos. Se calcularon los coeficientes de determinación lineal (R^2), que expresan la fuerza de la relación lineal entre las mediciones promedio del sensor y los valores del instrumento de referencia. También se calculó para cada caso la raíz del error cuadrático medio (RMSE), como una medida de la dispersión en las comparaciones.

La exactitud se refiere a la cercanía entre las medidas y los valores reales. En ausencia de curvas de calibración universales para los sensores analizados y los sensores de referencia o instrumentos estándar, se utilizaron las correlaciones lineales descritas anteriormente para evaluar la exactitud de los sensores y se calcularon las desviaciones estándar relativas a los instrumentos de referencia (DER)[139], como se muestra en la **Ecuación 3.4**.

$$DER = 100 * \frac{DE}{\mathbf{v}} \quad (3.4)$$

Donde DE se refiere a la desviación estándar de la serie de datos del instrumento a evaluar, y \mathbf{v} el valor medio de la serie del instrumento de referencia. La DER entonces se informa como el porcentaje de desviación del instrumento a evaluar respecto a la referencia.

La precisión se evaluó alcanzando estados estacionarios (valores casi constantes de las variables) en experimentos con la cámara de laboratorio (con PS como referencia) y en el exterior (con UNGS WS como referencia). Para la temperatura, se usaron 11 minutos de muestreo con un valor medio de 28°C en el sensor PS, para la humedad relativa se usaron 10 minutos de muestreo con un valor medio del 46% en el UNGS WS y para la presión, se usaron 23 minutos de muestreo con un valor medio de 1015 hPa en el UNGS WS. Se consideró a la precisión como la variación alrededor de un valor promedio en mediciones repetidas de la misma variable en condiciones experimentales idénticas o similares[155] y se determinó a partir de la ecuación **Ecuación 3.5**.

$$P(\%) = 100 - \frac{\mathbf{E}_s}{\mathbf{x}} \quad (3.5)$$

Donde \mathbf{x} es el valor promedio y \mathbf{E}_s es el error estándar del valor promedio de los sensores durante los períodos considerados como estado estable. El error estándar se calculó de acuerdo con la **Ecuación 3.6**.

$$\mathbf{E}_s = \frac{\sqrt{\sum(x - \mathbf{x})^2}}{n} \quad (3.6)$$

3 Metodología

Donde n es el número de mediciones y x representa todos los datos medidos.

La resolución se puede interpretar como el cambio más pequeño en la señal de salida de un sensor que se puede detectar de manera confiable. Este término está determinado por el ruido del instrumento, entendido como desviaciones espontáneas de corta duración en la salida promedio que no son causadas por cambios en los niveles de la variable que se mide [177]. En este estudio, la resolución se informó como las desviaciones estándar de las señales de salida de los sensores calculadas durante los diferentes estados estacionarios logrados en las pruebas de cámara de laboratorio y en el exterior mencionadas anteriormente para evaluar la precisión. En este caso, se calcularon las desviaciones estándar de los conjuntos de datos medidos tomando una sub-muestra de 1 minuto de estado estacionario para cada variable (600 muestras en cada conjunto de datos).

Para evaluar el tiempo de respuesta de los sensores, se probaron dos parámetros[177]. Por un lado, el tiempo de retraso definido como el intervalo de tiempo entre un cambio en la variable y el primer cambio observable en la respuesta de medición. Ese cambio observable se considera como la tolerancia de precisión informada en las hojas de datos. Por el otro, el tiempo de subida, que es el intervalo de tiempo entre el tiempo de respuesta inicial y el 95 % de la respuesta final después de un aumento en la variable. Para ello, los sensores se sometieron a cambios instantáneos de temperatura (22°C a 28°C), humedad relativa (45 % a 58 %) y presión (1019 hPa a 1008 hPa) en la cámara del laboratorio y se midieron los tiempos de retraso y subida. La frecuencia de muestreo se estableció en 1 ms para estos experimentos, a excepción de las medidas de tiempo de aumento con cambios de temperatura que se establecieron en 100 ms (a causa del método utilizado para variarla).

La energía de los sensores se cortó y restableció varias veces y los conjuntos de datos registrados se graficaron y analizaron. Esto con el fin de evaluar la respuesta a la falta de energía, considerada como la cantidad de tiempo que un sensor tarda en entrar en régimen y en funcionar correctamente después de una pérdida de energía[155]. Cada conjunto de datos contó con 147 muestras (con una frecuencia de muestreo de 100 ms) con una interrupción de energía cada 20 muestras.

Los resultados se utilizaron luego para seleccionar el sensor meteorológico a incorporar en la mini-estación ECAUNGS.

3.3.2. Caracterización de sensor de PM

Ensayos preliminares

Luego de evaluar los sensores meteorológicos, se continuó con un relevamiento, adquisición y pruebas con diversos LCS de CA. Algunos de ellos electroquímicos de la familia MQ (MQ-2 para gases combustibles, MQ-7 para monóxido de carbono y MQ-135 sensible a diversos gases), uno de óxido metálico (MiCS-5524 sensible a monóxido de carbono, entre otros gases), uno de infrarrojo no dispersivo (MHZ-14 para dióxido de carbono) y diferentes sensores de dispersión de luz para medir PM (GP2Y1010AU0F, HK-A5, PMS1003 y PMS5003).

Se diseñó y desarrolló un banco de pruebas piloto compartimentado para realizar mediciones en el exterior y evaluar el desempeño de diferentes sensores. Al igual que lo detallado para las pruebas con los sensores meteorológicos, el control y adquisición de datos se realizó a través de una placa Arduino UNO R3 y se ensambló un circuito con un módulo SD OKY-3301 [189] con interfaz SPI (**Figura 2.10 del Apartado 2.3.1**) para almacenar los datos y, para incorporar una marca de tiempo a las líneas de datos, un módulo de reloj DS3231 [190] con “bus” I²C (**Figura 2.9 del Apartado 2.3.1**) y batería para transmitir pulsos y actualizar la fecha y hora cuando la alimentación desde la placa se detenga por cortes de energía.

Una primera exploración de las características principales de los sensores disponibles, así como una evaluación preliminar con datos recogidos por el banco de pruebas, permitió descartar varios sensores para su uso final. Todos los sensores MQ y el MiCS-5524 poseen características que no los hacen aptos para medir en el exterior (límites de detección, ruido en las señales). El sensor de CO_2 MHZ-14 puede medir en el exterior, pero requiere de calibraciones que excedieron a los tiempos y posibilidades de esta Tesis (especialmente considerando las limitaciones dadas por la pandemia de COVID-19). Finalmente, se decidió continuar trabajando con los sensores de PM que mostraron resultados consistentes y además trabajan con el contaminante de referencia transversal a esta Tesis.

De los 4 sensores de PM mencionados (GP2Y1010AU0F, HK-A5, PMS1003 y PMS5003), el PMS5003 es el que mejores características posee, por lo que se continuó trabajando con este sensor. Esta superioridad se vincula a límites de detección, sensibilidad e incertezas informadas [191, 192, 193, 194] que claramente diferencia a estos sensores y fueron corroboradas con las pruebas preliminares.

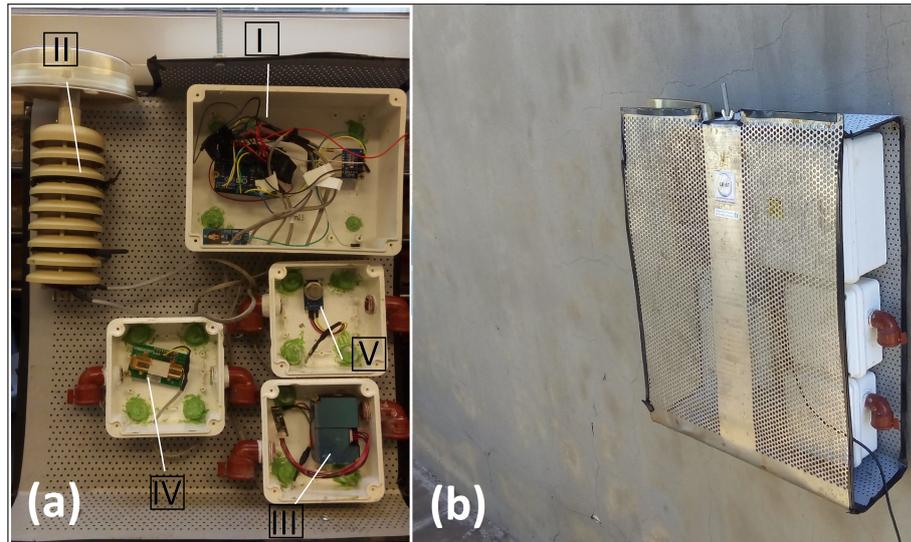


Figura 3.12: Prototipo de banco de pruebas para sensores. a) Compartimentos con electrónica y sensores (I. Placa de control con módulos de reloj y SD; II. Sensor BME280 con abrigo meteorológico; III. Sensor de PM; IV. Sensor MHZ-14; V. Sensor de la familia MQ), b) Equipo con cubierta protectora e instalado

Diseño experimental

Se evaluó el desempeño del sensor de PM elegido, a partir de pruebas de laboratorio, mediciones de campo y comparaciones con otros sensores calibrados. Si bien cada sensor trae consigo información de fábrica y una calibración informada, estas pruebas permiten determinar algunas limitaciones y fortalezas del sensor a la hora de ser utilizado en término de valores relativos de concentración entre tamaños de partículas y bajo diferentes condiciones. Para utilizar los valores dados por el sensor con fines de investigación, el mismo debería ser calibrado junto con instrumentos de referencia, pero no existen protocolos establecidos para cada sensor y no fue posible tampoco, durante los plazos de esta Tesis, acceder a instrumentos de referencia que cumplan con las regulaciones internacionales.

Se evaluaron para el sensor de PM (señales de PM_1 , $PM_{2,5}$ y PM_{10}) los mismos indicadores que los detallados para los sensores meteorológicos, además del sesgo durante comparación con equipo de referencia. Algunos de estos indicadores precisaron pruebas de laboratorio, para lo que se adaptó la cámara de pruebas de la **Figura 3.12(a)** de acuerdo al diseño mostrado en la **Figura 3.13**.

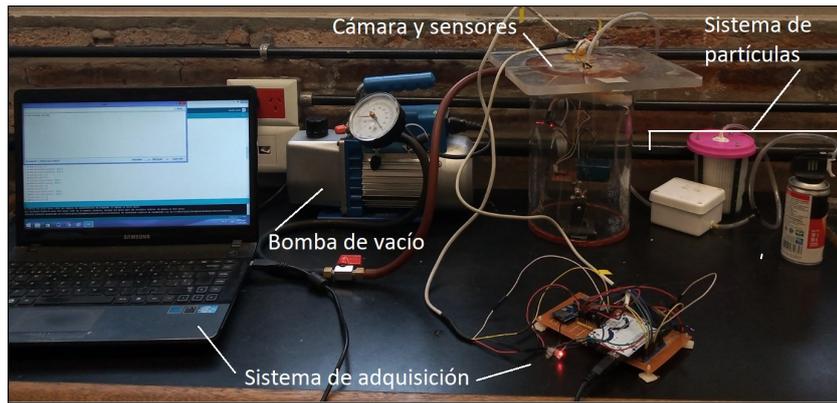


Figura 3.13: Diseño experimental con la cámara de prueba de sensores adaptada para material particulado. Dentro de la cámara se colocaron el sensor de PM PMS5003, un BME280 y un ventilador en la base. Se adaptaron dos entradas, una de ellas conectada a una bomba de vacío, la otra al suministro de partículas y aire limpio. Externamente, se diseñó un sistema de adquisición de datos en tiempo real

La bomba de vacío posee dos utilidades, la primera es realizar pruebas con bajas presiones internas con el fin de evaluar el límite inferior de detección del sensor de PM, y la segunda disminuir una fracción de la presión interna con el fin de facilitar el posterior ingreso de partículas a la cámara al abrir una válvula de recuperación.

Dentro de la cámara, se colocó un ventilador para facilitar la circulación de partículas y disminuir el efecto de deposición durante las pruebas, además del sensor de PM y un BME280 para medir humedad, presión y temperatura internas (**Figura 3.14(a)**). El sistema de partículas consta de un circuito que permite ingresar a la cámara la muestra de partículas, o aire limpio a través de un filtro HEPA o de un cilindro con aire a alta presión (**Figura 3.14(b)**).

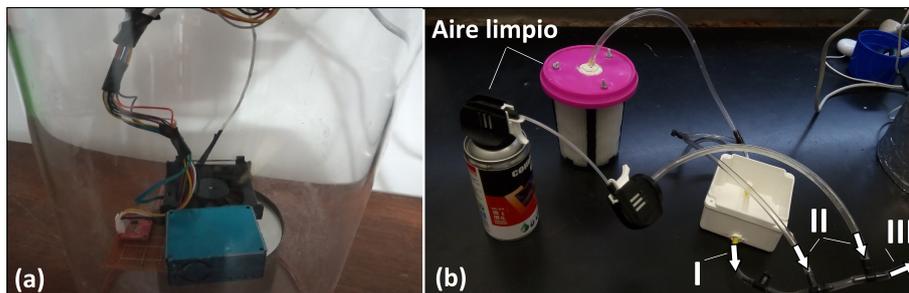


Figura 3.14: Imágenes más cercanas de la cámara (a) y del sistema de partículas (b). En el último se indican los circuitos para el ingreso de partículas (I), de aire limpio (II) y la dirección de flujo hacia la cámara (III)

3 Metodología

El procedimiento que se diseñó para ingresar partículas a la cámara consta de los siguientes pasos:

1. Cerrar las válvulas del sistema de partículas y colocar la muestra de partículas en el circuito del mismo.
2. Poner en funcionamiento la bomba de vacío con el fin de remover la mayor parte de las partículas presentes dentro de la cámara.
3. Apagar la bomba y cerrar la válvula que conecta la misma a la cámara.
4. Abrir las válvulas que permiten el paso de la muestra de partículas a la cámara junto con aire limpio (la diferencia de presión lleva las partículas dentro de la cámara).
5. Una siguiente prueba requiere de una limpieza del interior de la cámara, seguida de la repetición de los pasos anteriores.

Cabe mencionar que el procedimiento anterior y el diseño experimental permiten el ingreso de una cantidad cualitativa de partículas, pues cierta parte de la muestra se pierde indefectiblemente en el circuito del sistema. De todas formas, las pruebas realizadas no requirieron del ingreso de una masa conocida de muestra, sino más bien una solución de concentraciones determinadas o partículas de diferentes tamaños.

La precisión se calculó con las **Ecuaciones 3.5** y **3.6**. Se realizaron dos pruebas con aire sin filtrar (dos concentraciones de partículas diferentes) de medición continua del sensor dentro de la cámara sellada y con circulación interna de aire durante 15 minutos cada vez. Las mismas pruebas se utilizaron para determinar la resolución, calculando las desviaciones estándar de las señales del sensor.

Los tiempos de respuesta (retraso y subida) del sensor fueron evaluados en dos pruebas ingresando a la cámara una cantidad de partículas de polvo (tiza) en una de ellas, y humo (sahumerio) en la otra, y luego retirando el sensor de la cámara mientras está funcionando hacia un espacio con concentraciones de ambiente. Se midió también el tiempo de bajada, definido como el tiempo desde que el sensor se trasladó a concentraciones ambiente y el 95 % de la respuesta de las señales a concentraciones ambiente. Con el fin de evaluar esta característica del sensor en las mismas condiciones para las que será utilizado, se colocó en la cámara al sensor dentro de una caja estanco con una manguera conectada a una toma de aire, siendo la manguera y la toma las mismas que luego fueron utilizadas en el desarrollo de la mini-estación ECAUNGS final. Esto se realizó de esta manera porque

los tiempos de respuesta dependen del circuito por el cual ingresan las partículas a la cámara de muestreo del sensor.

La respuesta a la falta de energía se evaluó cortando la alimentación y reconectándola varias veces dentro de la cámara sellada con concentraciones de aire no filtrado.

El límite inferior de detección no ha sido evaluado para los sensores meteorológicos, pero es un indicador interesante para los sensores de PM. El método de Kaiser & Specker (1956)[195] fue utilizado, el cual se basa en la **Ecuación 3.7**. Para la prueba de laboratorio, se determinó el valor de k como 1 (al no depender de la lectura de un instrumento de referencia en este caso) y se disminuyó la presión dentro de la cámara hasta 400 hPa con el fin de conseguir una baja concentración estimada de partículas. Como límites superiores de detección se tomaron los indicados en la hoja de datos del sensor ($500 \mu\text{g}/\text{m}^3$ para $PM_{2,5}$)[194].

$$LLOD = \frac{3 * DE}{k} \quad (3.7)$$

La **Ecuación 3.7** describe la forma de calcular el límite inferior de detección del sensor considerando un período de medición de campo del sensor en conjunto con un instrumento de referencia (sensor calibrado)[195]. Se utiliza la desviación estándar (DE) de la señal del sensor cuando la referencia mide concentraciones menores a $1 \mu\text{g}/\text{m}^3$ y la pendiente k de la regresión lineal entre las mediciones del sensor a evaluar y la referencia.

Validación

A continuación, se procedió a realizar una comparación del sensor PMS5003 con un equipo calibrado Open-Seneca[196] del programa “Cambike” de la Universidad de Cambridge llevado a cabo en Buenos Aires, con el vínculo logrado a través del Ministerio de Ambiente y Desarrollo Sustentable de la Nación. Este equipo posee un sensor de PM SPS30 del fabricante “Sensirion” que es de bajo costo, pero que fue calibrado en varias campañas con un equipo regulatorio de la Embajada de Estados Unidos en CABA (en el **Anexo 8.3.1** se adjunta la calibración informada), por lo que se utilizó como referencia para evaluar el desempeño del PMS5003. Con las mediciones de campo de 11 días de duración de ECAUNGS (PM, humedad relativa, presión y temperatura) y el equipo de Cambike (PM), se realizó una descripción comparativa y se evaluaron las características del PMS5003 que se detallan a continuación.

La linealidad, con el mismo método descrito para los sensores meteorológicos, utilizando las regresiones lineales entre PMS5003 y el SPS30. Además, se practicaron análisis de

varianzas (ANOVA) y se verificaron las potencias de los ajustes. Las mismas regresiones lineales se utilizaron para evaluar la exactitud del sensor a partir de la DER (desviación estándar respecto a la referencia) como se definió previamente según la **Ecuación 3.4**.

El sesgo del sensor se determinó como la razón entre las concentraciones del PMS5003 y el SPS30 de Cambike en función del tiempo (**Ecuación 3.8**).

$$sesgo_t = \left(\frac{PMeca_t}{PMcam_t} - 1 \right) * 100 \quad (3.8)$$

Donde $PMeca_t$ es la concentración de PM en función del tiempo del sensor PMS5003 y $PMcam_t$ es lo equivalente para el sensor de Cambike. Esta ecuación indica el sesgo para cada medición (los residuales normalizados por la referencia, en porcentaje), luego se presenta la pendiente de todas las mediciones como sesgo o deriva general para cada señal. Esto se puede interpretar también como una cuantificación de la estabilidad del sensor[160]. Con el fin de mostrar una gráfica menos ruidosa, se practicó una media móvil horaria con los resultados de los sesgos en el período muestral original de 1 minuto.

Nuevamente se evaluó el límite inferior de detección, pero esta vez utilizando la **Ecuación 3.7** con el criterio de mediciones menores a $1 \mu g/m^3$ de la referencia y utilizando la pendiente de la regresión lineal entre las señales de ambos sensores.

También se buscaron anomalías y errores sistemáticos en las señales del sensor, a partir de estudiar los cuantiles condicionales y diagramas de Taylor[197] entre las mediciones del PMS5003 y del SPS30.

Simulaciones

Con el fin de contar con una referencia en torno a las diferentes posibles distribuciones de partículas que afectan a las señales del sensor de PM PMS5003, se utilizó la librería “opcsim”[198] para lenguaje Python que permite programar a medida sensores de PM (contadores de partículas y nephelómetros) desde un enfoque físico, así como construir distribuciones de partículas en base a diferentes parámetros. En primer lugar, se muestran distribuciones típicas construidas en base a la suma de modos lognormales (**Apartado 2.3.1**) con parámetros (números de concentración, medianas del diámetro de partículas y desviación estándar de tamaños) extraídos de Seinfeld & Pandis (2016) [164]. Se muestran las distribuciones volumétricas y las funciones de distribución acumuladas para cada tipo creado con el fin de determinar las posibles lecturas del sensor de PM en sus rangos de tamaños.

El sensor PMS5003 mide la luz láser, de $680nm \mp 10nm$, dispersada a un ángulo de 90°

[160]. Reporta seis rangos de tamaños como si se tratara de un contador de partículas, pero en realidad son salidas computadas a partir de mediciones nefelométricas [199]. Por ello, se construyó un nefelómetro con la librería “opcsim” basado en la longitud de onda y ángulo de dispersión mencionados con el fin de comparar luego con el desempeño del sensor real. El nefelómetro programado (llamado “neph1”) se calibró bajo condiciones lo más cercanas posibles a las usadas por el fabricante del PMS5003[194], con distribuciones del tipo urbana y bajo 50% de humedad relativa. Luego, se calibró el mismo sensor programado con una distribución de sulfato de amonio en solución y 0% de humedad (llamado “neph2”). Esta solución se estableció con número de concentración $N = 1000$, mediana de diámetros $\bar{D}_p = 0,08$, desviación estándar de tamaños $\sigma_i = 1,5$, índice de refracción $Ir = 1,521 + 0j$, densidad $\rho = 1,77 \text{ g/cm}^3$ y factor de crecimiento higroscópico $f(RH) = 0,53$ [198].

Luego, se calcularon los cambios en las distribuciones y carga másica de partículas medidas por los nefelómetros programados para la solución de sulfato de amonio bajo diferentes condiciones de humedad. Para comparar estos resultados con el comportamiento real del sensor PMS5003, se utilizó nuevamente la cámara de laboratorio para inyectar una solución de sulfato de amonio de concentración 10% m/V con el sensor funcionando dentro bajo diferentes condiciones de humedad (5 niveles controlados con un humidificador ultrasónico y partiendo de un nivel bajo en presencia de esferas de silicagel). Los valores de humedad conseguidos no son exactos por la variabilidad dentro de la cámara, pero sí están diferenciados entre sí y son de aproximadamente 20, 30, 50, 70 y 90%.

Gestión de calidad

Con la información y resultados obtenidos de las pruebas de laboratorio, comparaciones con el equipo de Cambike y simulaciones con librería “opcsim” se establecieron criterios de gestión de calidad de los datos de PM del PMS5003 para ser evaluados de forma automática con la estación en funcionamiento. Esto se logró diseñando un algoritmo adosado al control de la mini-estación ECAUNGS que permite incorporar un dato más a la serie de la estación con un código que facilite al usuario identificar valores sospechosos o eventos puntuales, con marcadores que en este contexto suelen llamarse “banderas”. Estos marcadores pueden advertir acerca de valores fuera de los límites de detección, presencia de curvas ascendentes o descendentes de concentración y valores de humedad con elevado efecto de crecimiento higroscópico en los aerosoles.

3.3.3. Integración de metodologías de bajo costo

En este apartado se muestra la integración de los diferentes métodos de bajo costo estudiados, adaptados e implementados. Por un lado, se muestra el desarrollo del prototipo de mini-estación de CA ECAUNGS y se propone una metodología para el estudio de las primeras mediciones automáticas de PM en la región. Por otro lado, se detalla la implementación de un modelo ANN tipo MLP con estos datos. Esta metodología se esquematiza en el diagrama de flujo de la **Figura 3.15**

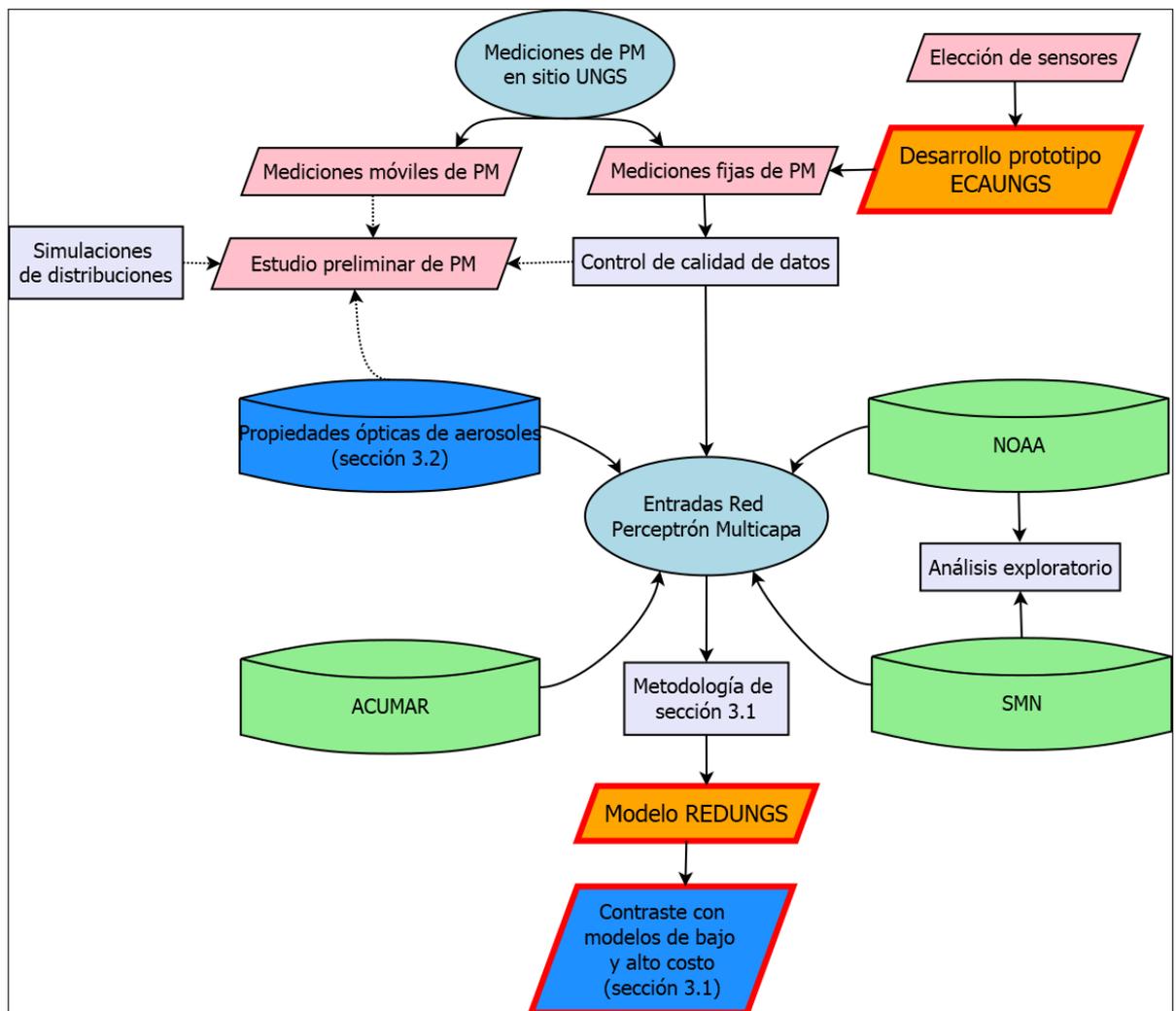


Figura 3.15: Diagrama de flujo que describe la metodología implementada para las mediciones de PM y modelización

Desarrollo de prototipo de bajo costo ECAUNGS

Una vez seleccionados y evaluados los sensores meteorológico (BME280) y de PM (PMS5003), se continuó con el diseño y desarrollo definitivo del prototipo de mini-estación de CA ECAUNGS. De la misma forma que el banco de pruebas ya descrito, la estación se controla a través de una placa Arduino UNO R3 y cuenta con módulos de reloj y SD para almacenamiento de los datos. Asimismo, cuenta con un abrigo meteorológico para el sensor de humedad, presión y temperatura. Posee una entrada de aire que se encuentra conectada a través de una manguera al sensor de PM, un LED indicador verde para indicar la frecuencia de muestreo y otro de color rojo para indicar alguna falla en el funcionamiento. Todo el equipo se alimenta con una fuente switching de 9 V y 1 A, pero con el fin de darle independencia energética se implementó un sistema de alimentación a través de un panel solar de 10 W (tensión de 18,2 V y corriente de 0,52 A, operando en potencia máxima) conectado a través de un regulador a una batería de gel de 12 V y 7 Ah. En un reservorio digital [36], creado con los productos de esta Tesis, se puede consultar el código de control desarrollado para ECAUNGS en lenguaje C++ y ejecutado en la “Interfaz De Usuario” de Arduino.

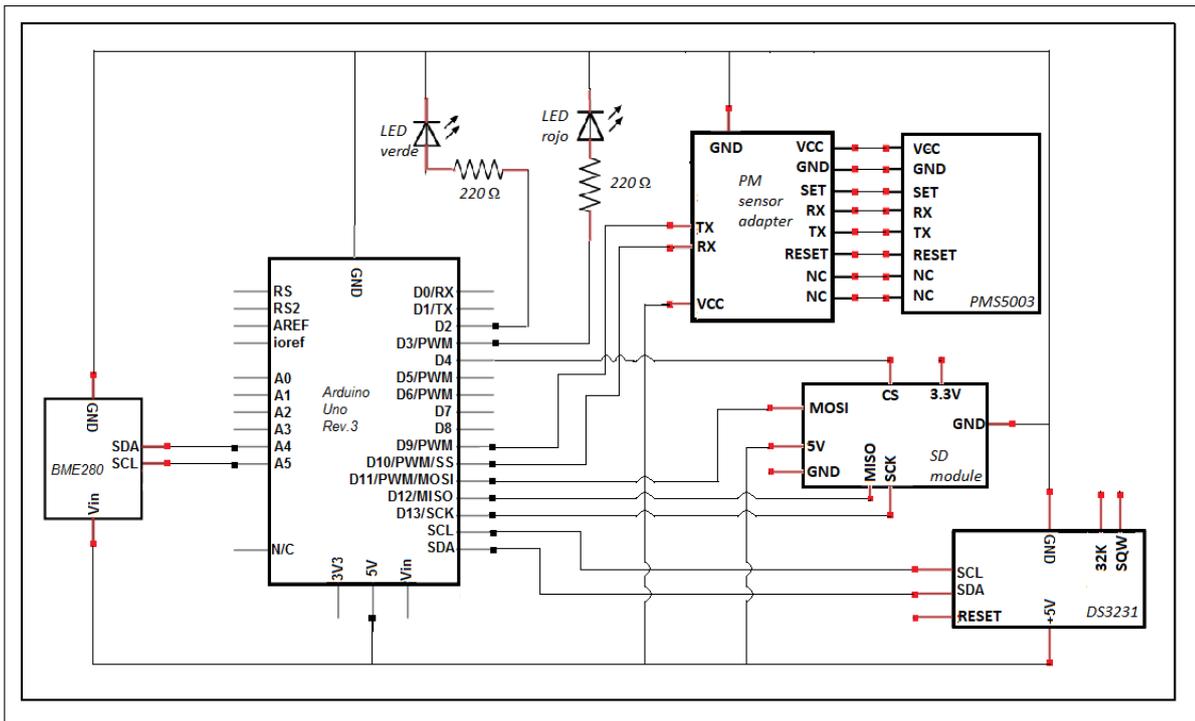


Figura 3.16: Circuito de conexiones de los sensores, actuadores y la placa de control

La **Figura 3.16** muestra el circuito diseñado para la mini-estación, mientras que

la **Figura 3.17** muestra los componentes físicos internos de la misma, así como las entradas en la parte inferior con pasacables de goma para evitar aberturas innecesarias. Una manguera de goma fue utilizada para conectar la toma de aire exterior con la entrada a la cámara de medición del sensor, mientras que la salida del aire circulado por el sensor se encuentra al lado y en contacto con el interior de la caja estanco que alberga los componentes de la central con el fin de no perturbar las mediciones.

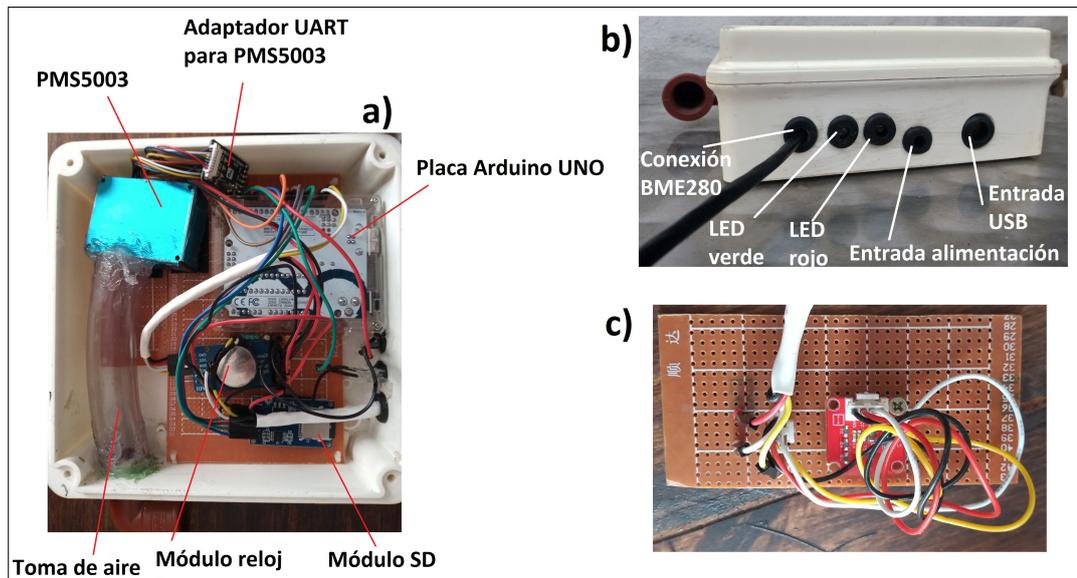


Figura 3.17: Disposición física de los componentes de ECAUNGS. a) Parte central de ECAUNGS, con sensor de PM, módulos de reloj y SD y placa de control; b) Vista inferior de la central ECAUNGS, con cable de conexión al abrigo meteorológico, LEDs indicadores y entradas de alimentación y USB para la placa; c) Placa de conexión para BME280 cuya ubicación final es dentro del abrigo meteorológico

La **Figura 3.18** muestra las partes del circuito de alimentación y la instalación de ECAUNGS. El regulador se encarga de, valga la redundancia, regular la entrada de tensión del panel solar y con ella cargar la batería de gel que, a su vez, alimenta el funcionamiento del regulador y de la estación. Finalmente, el equipo completo cuenta con una caja estanco para la central del equipo, otra caja para los componentes de alimentación solar, un abrigo meteorológico para las mediciones de humedad, presión y temperatura, y el panel solar orientado al norte. El panel solar posee dimensiones de 260mm x 350mm x 2mm (alto, ancho y grosor) y la caja central con el abrigo de 280mm x 160mm x 70mm, por lo que la disposición final ocupa poco espacio (420mm x 350mm x 250mm) y se puede instalar fácilmente cuidando únicamente la orientación del panel

solar.

Por otro lado, el consumo máximo de ECAUNGS es bajo y cercano a 1 W, por lo que el sistema de alimentación es más que suficiente para un funcionamiento autónomo y estable.

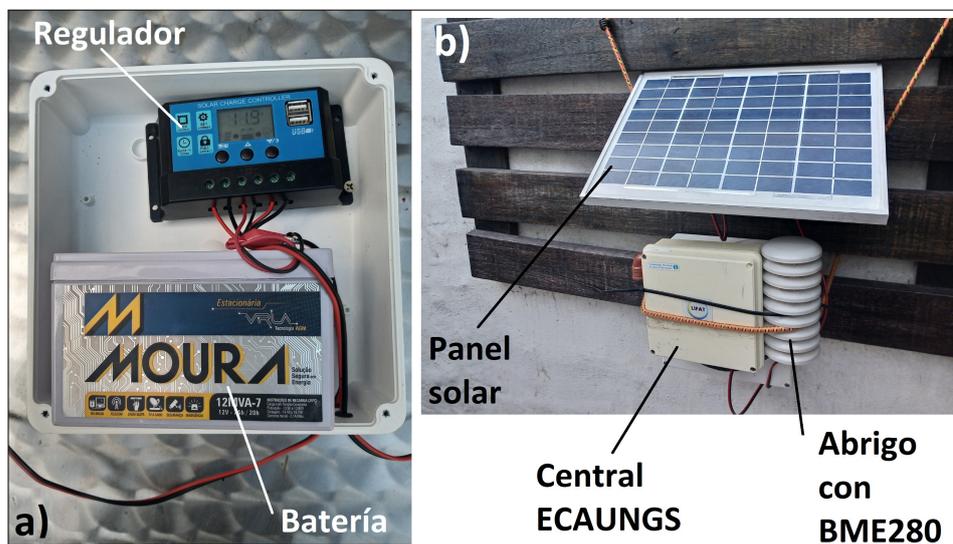


Figura 3.18: Componentes físicos del sistema de alimentación solar e instalación final del equipo. a) Regulador y batería de gel, ambos conectados junto con los bornes del panel solar; b) Presentación de la instalación final del prototipo ECAUNGS. El panel solar se ubica apuntando al Norte con unos 15° de inclinación respecto a la superficie, y por debajo se ubican la central y el abrigo

Mediciones de PM de bajo costo

Finalmente, se realizó un estudio con los valores de concentración de PM registrados por ECAUNGS instalada en un domicilio particular en el partido de San Miguel (con el fin de tener acceso al equipo en todo momento), con mediciones entre junio de 2019 y enero de 2021. Se había previsto instalar el equipo en el campus de la UNGS, pero no resultó conveniente dada la situación condicionada por la pandemia de COVID-19. De todas formas, se contó con una gran cantidad de datos medidos en el exterior, con períodos intermedios sin datos debido a mantenimiento y pruebas con el equipo. El período de muestreo de ECAUNGS se estableció en aproximadamente 1 segundo (que varía levemente por tiempos de comunicación entre los componentes), con el fin de que los datos generados puedan ser utilizados bajo distintas bases temporales de acuerdo a la duración de los eventos o períodos de las tendencias que se quisieron estudiar. En total, se

contó con 165 días de mediciones, con un total de 13711356 líneas de datos (cada línea con datos de fecha en formato DD/MM/AAAA, hora en formato HH:MM:SS, PM_1 , $PM_{2,5}$, PM_{10} , humedad, presión y temperatura).

Se practicó un control de calidad de datos partiendo de la frecuencia de muestreo original. En primera instancia se buscaron datos anómalos de acuerdo al límite superior del sensor ($500 \mu g/m^3$ para $PM_{2,5}$ [194]), los datos por debajo del límite de detección inferior estimado con las pruebas de laboratorio y los valores no numéricos. Luego, se buscaron en la serie temporal pendientes elevadas en búsqueda de datos atípicos (para ello se determinó como dato sospechoso a aquel que involucre una diferencia con su dato vecino mayor a 2 desviaciones estándar de la serie completa). Este método se implementó posteriormente en una base minutal (tras aplicar un promedio móvil a las series) y horaria, como se recomienda en Bulot et. al (2019)[161]. Todo el análisis de control de calidad se realizó a partir de un algoritmo no automatizado generado para tal fin, ya que el código de gestión de calidad creado se incorporó con posterioridad a estas mediciones. Finalmente, se estudiaron las señales de PM en función del tiempo y variables meteorológicas. Estos resultados se ponen en diálogo con los correspondientes a la metodología descrita en la **Sección 3.2**.

Luego, se llevó a cabo una campaña de mediciones de PM en la zona con el equipo de Cambike. Este equipo cuenta con un módulo de georeferenciación GPS, lo que permitió trazar un mapa con las concentraciones medidas en diferentes calles. Este mapa muestra una resolución espacial más fina y aporta nueva información al generado con datos de AOD del algoritmo MAIAC mencionado en el **Apartado 3.2.2**. Las mediciones se realizaron entre el 21 de noviembre de 2019 y el 21 de enero de 2020 montando el equipo sobre una bicicleta, para lo que fue diseñado. Los datos fueron subidos a una aplicación oficial de Open-Seneca que permite visualizarlos y trazar mapas[200].

Modelado de bajo costo de la calidad del aire

Con el fin de completar el estudio dando una aplicación específica a las mediciones, se implementó un modelo integrador de tipo ANN MLP-FF-BP para predecir $PM_{2,5}$ basado en 165 días de mediciones de ECAUNGS y se comparó con los modelos de alto y bajo costo desarrollados de acuerdo a lo mostrado en la **Sección 3.1** (RED3 y RED4, respectivamente), obtenidos con datos de CABA. Se utilizaron medias horarias (3396 horas) y las variables consideradas para tal fin son las listadas en la **Tabla 3.6**.

3.3 Desarrollo e implementación de equipos de bajo costo

Variable	Nomenclatura	Unidades	Fuente
Hora de medición	<i>hora</i>	h	ECAUNGS
Componente zonal del viento en CABA	<i>Uacu</i>	m/s	ACUMAR
Componente meridional del viento en CABA	<i>Vacu</i>	m/s	ACUMAR
Radiación en CABA	<i>RADacu</i>	W/m^2	ACUMAR
PM_{10} en CABA	<i>PM10acu</i>	$\mu g/m^3$	ACUMAR
$PM_{2,5}$ en CABA	<i>PM25acu</i>	$\mu g/m^3$	ACUMAR
Día de la semana de medición	<i>diasem</i>	ad.	ECAUNGS
Mes de medición ECAUNGS	<i>mes</i>	ad.	ECAUNGS
$PM_{2,5}$ medido en UNGS	<i>PM25eca</i>	$\mu g/m^3$	ECAUNGS
Temperatura en UNGS	<i>Teca</i>	$^{\circ}C$)	ECAUNGS
Humedad relativa en UNGS	<i>HReca</i>	%	ECAUNGS
Presión atmosférica en UNGS	<i>Peca</i>	hPa	ECAUNGS
Componente zonal del viento en Campo de Mayo	<i>Ucm</i>	m/s	SMN
Componente meridional del viento en Campo de Mayo	<i>Vcm</i>	m/s	SMN
Estabilidad atmosférica	<i>estab</i>	ad.	NOAA
Altura de capa de mezcla	<i>BLHungs</i>	m	NOAA
Precipitación acumulada de 6h	<i>PPTungs</i>	mm	NOAA
Espesor óptico de aerosoles en UNGS	<i>AODungs</i>	ad.	Fotómetro solar Calitoo
Coefficiente de Ångstrom en UNGS	<i>alphaUNGS</i>	ad.	Fotómetro solar Calitoo

Tabla 3.6: Variables consideradas para el desarrollo de la ANN de estimación de $PM_{2,5}$ en UNGS. Las magnitudes adimensionales se marcan como “ad.”

Se incorporaron al modelo los datos de PM medidos por la estación de ACUMAR en

3 Metodología

Dock Sud, dado que a escala regional puede haber algunas similitudes con las concentraciones dadas en el Centro-Norte del conurbano. Los datos de viento de esa estación se incorporaron también dado que la estación de Campo de Mayo del SMN posee datos faltantes durante la noche, al igual que los de radiación que, excepto en casos de nubosidades diferentes, serán prácticamente los mismos en ambos sitios.

Los datos de asimilación del NOAA se agregaron también dado que pueden aportar información importante a la red acerca de propiedades de la capa límite atmosférica, las cuales no son medidas en estaciones de superficie de la región. Finalmente, aunque son pocos datos, se agregaron las mediciones realizadas con el fotómetro solar Calitoo, ya que es factible encontrar alguna correlación con las mediciones en superficie de PM.

La metodología para la elección de predictores, arquitectura de la red e hiperparámetros es la misma descrita en la **Sección 3.1**. La variable “objetivo” es nuevamente $PM_{2,5}$, en este caso PM_{25eca} .

Los resultados obtenidos a partir de lo detallado de esta sección se pueden consultar en el **Capítulo 6**.

4 Potencialidad de las Redes Neuronales en calidad del aire

4.1. Objetivos e interés

En este capítulo se describe el desarrollo de Redes Neuronales Artificiales orientadas a predecir concentraciones de PM en un sitio de mediciones de ACUMAR en Dock Sud, lindando con CABA. La elección de esta estación de monitoreo se basa en el hecho de ser la más completa hasta el momento en cuanto a datos históricos y cantidad de magnitudes monitoreadas en la región. Por otro lado, como se explicó en el **Capítulo 1**, el interés actual de investigación de PM se centra principalmente en partículas de pequeño tamaño (menor a $2,5\mu\text{m}$), por ello se implementan los modelos con la categoría de $PM_{2,5}$ como objetivo.

Los modelos fueron desarrollados con métodos personalizados y las estimaciones se persiguen a través de modelos en función de datos de otros contaminantes y variables meteorológicas y cumplen dos objetivos principales. El primero de ellos es explorar las potencialidades de las ANN para predecir concentraciones de contaminantes bajo diferentes condiciones y disponibilidad de datos. El segundo es profundizar en la comprensión de los vínculos entre las variables en cuanto a la relación predictor-objetivo, alguna de las cuales pueden ser extrapoladas al Centro-Norte del conurbano al tratarse de vínculos generales y de escala sinóptica o regional. Se comparan también los resultados de los modelos desarrollados con otros implementados correspondientes a métodos tradicionales univariados y lineales.

4.2. Resultados y discusiones

4.2.1. Exploración de datos meteorológicos y de contaminantes

En la exploración inicial de datos no se encontraron valores fuera de los rangos normales. Por otro lado, el porcentaje de datos faltantes en las series de cada variable se encuentra dentro de los valores aceptables para ser utilizadas en los modelos (la mayor cantidad la posee el MP-xileno con un 20 % sobre el total).

A continuación, se detallan características generales de las series de datos de mayor interés para el trabajo con los modelos que se implementan en el presente capítulo. Las tendencias del resto de las variables finalmente utilizadas se pueden consultar en el **Anexo 8.1.4**.

Tanto las intensidades como direcciones del viento influyen en el transporte horizontal de distintos contaminantes y en la resuspensión de partículas en superficie. Durante el período de estudio, las intensidades del viento más frecuentes se encuentran entre 3 ms^{-1} y 5 ms^{-1} con valores máximos cerca del mediodía (**Figura 4.1**). Estacionalmente, las velocidades son más intensas durante los meses cálidos.

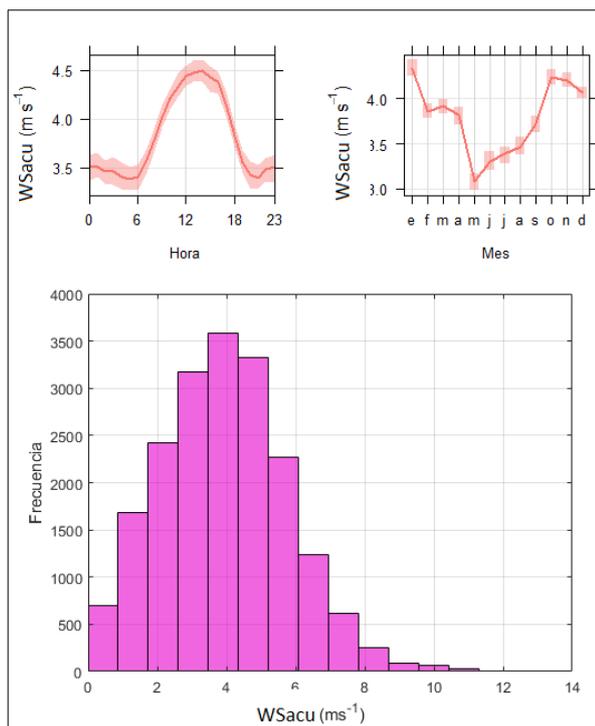


Figura 4.1: Histograma y tendencias mensual y horaria de datos de velocidad del viento

En horas diurnas la dirección del viento más frecuente es del Noreste y durante la noche

prevalecen los vientos del Sudeste, mientras que las mayores intensidades corresponden para ambas franjas horarias a las direcciones Noreste, Sudeste y Sudoeste (**Figura 4.2**). Estacionalmente, se puede observar que en primavera y otoño (hemisferio sur) prevalecen vientos del Sudeste, mientras que en verano e invierno son más frecuentes las direcciones Norte y Noreste.

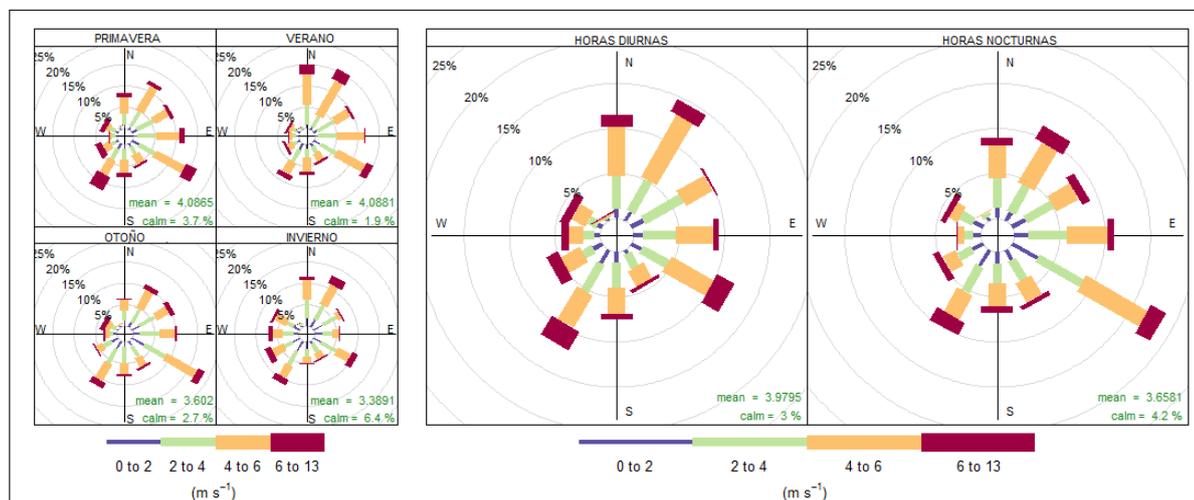


Figura 4.2: Rosas de los vientos estacionales y diferenciadas en horas diurnas y nocturnas

La humedad ambiente interviene en reacciones físico-químicas con el SO_2 , precursor de PM, y se vincula al crecimiento higroscópico de los aerosoles en la atmósfera. Sus valores más frecuentes son elevados (entre 80 % y 90 %), durante el día se observa un ciclo característico con valores máximos antes del amanecer y mínimos después del mediodía en horas de mayor calentamiento por radiación. Su ciclo anual se caracteriza por valores menores en los meses cálidos y mayores durante la época invernal (**Figura 4.3**), aún así los valores medios mensuales son todos mayores al 65 %.

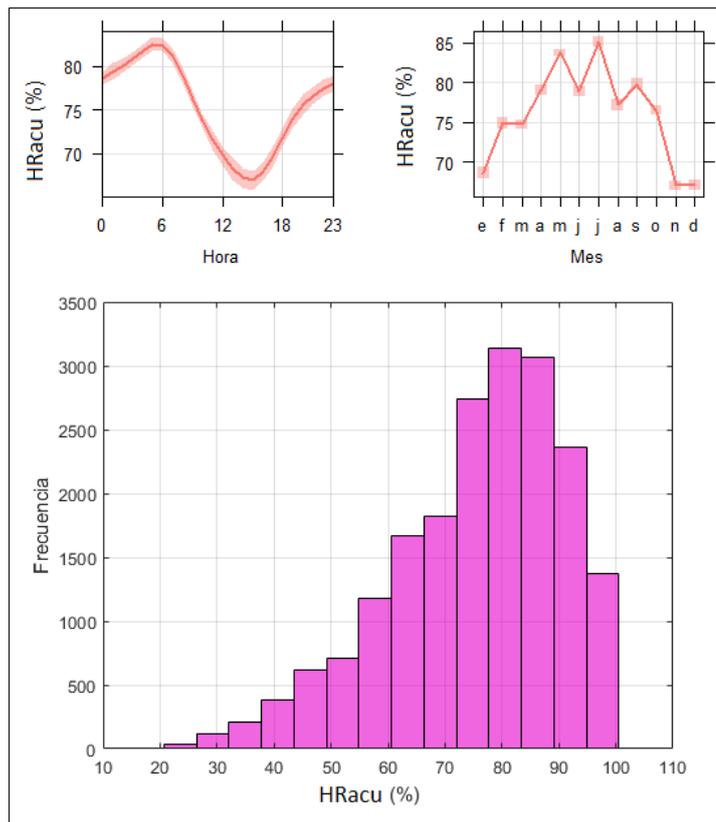


Figura 4.3: Histograma y tendencias diarias y anuales de la humedad relativa

La altura de capa de mezcla es un parámetro propio de la capa límite atmosférica que suele ser utilizado en modelos de calidad del aire. En esta parte de la atmósfera ocurre una intensa mezcla vertical en las propiedades del aire (calor, humedad, etc..) y la turbulencia es conducida principalmente por convección. En días libres de nubosidad, el crecimiento de la capa límite se encuentra vinculado al calentamiento en superficie por radiación, comenzando alrededor de media hora después de la salida del sol y alcanzando su máximo en las últimas horas de la tarde. Su límite superior se da con una capa estable que actúa como barrera que contiene a los contaminantes dentro de la capa límite [201].

En la **Figura 4.4** se observan las tendencias de *BLHacu* durante el período de estudio. El ciclo diurno coincide con lo descrito anteriormente, se observa una altura base de unos 200 m en promedio que crece luego del amanecer hasta alcanzar su valor máximo (con una media poco superior a los 1200 m) al atardecer. La altura decae rápidamente durante la noche hasta volver a su valor de base. Estacionalmente, durante los meses cálidos *BLHacu* alcanza sus mayores valores medios y en la época fría sus valores mínimos. Esto ocurre por la relación entre la altura de la capa límite y la radiación solar incidente en superficie, que en invierno alcanza sus valores medios más bajos sumado a una mayor

cantidad de días nublados.

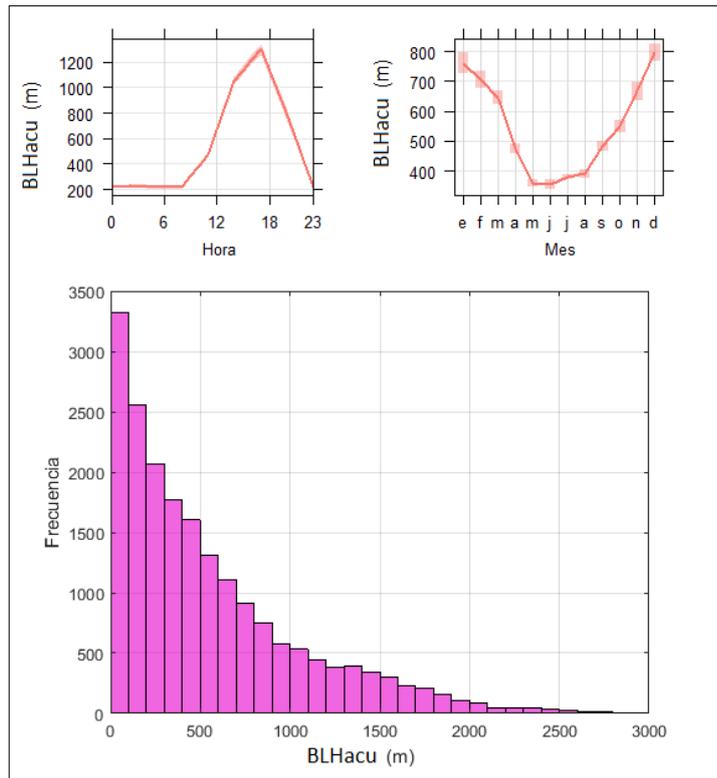


Figura 4.4: Histograma y tendencias diarias y anuales de la altura de capa límite

En la serie de estabilidad (*estab*) construida domina la clase 4 (neutral), aunque esta es una característica sesgada por el método utilizado (**Figura 4.5**). Durante el día se pueden observar las clases neutrales e inestables en horas de mayor radiación solar incidente y las estables en horas nocturnas, como es de esperar. Debido al mismo fenómeno radiativo, se observa que durante los meses fríos hay prevalencia de condiciones estables y durante el verano aumenta la cantidad de casos de inestabilidad. Las condiciones atmosféricas estables inhiben los movimientos verticales favoreciendo la permanencia de los contaminantes cerca de la superficie.

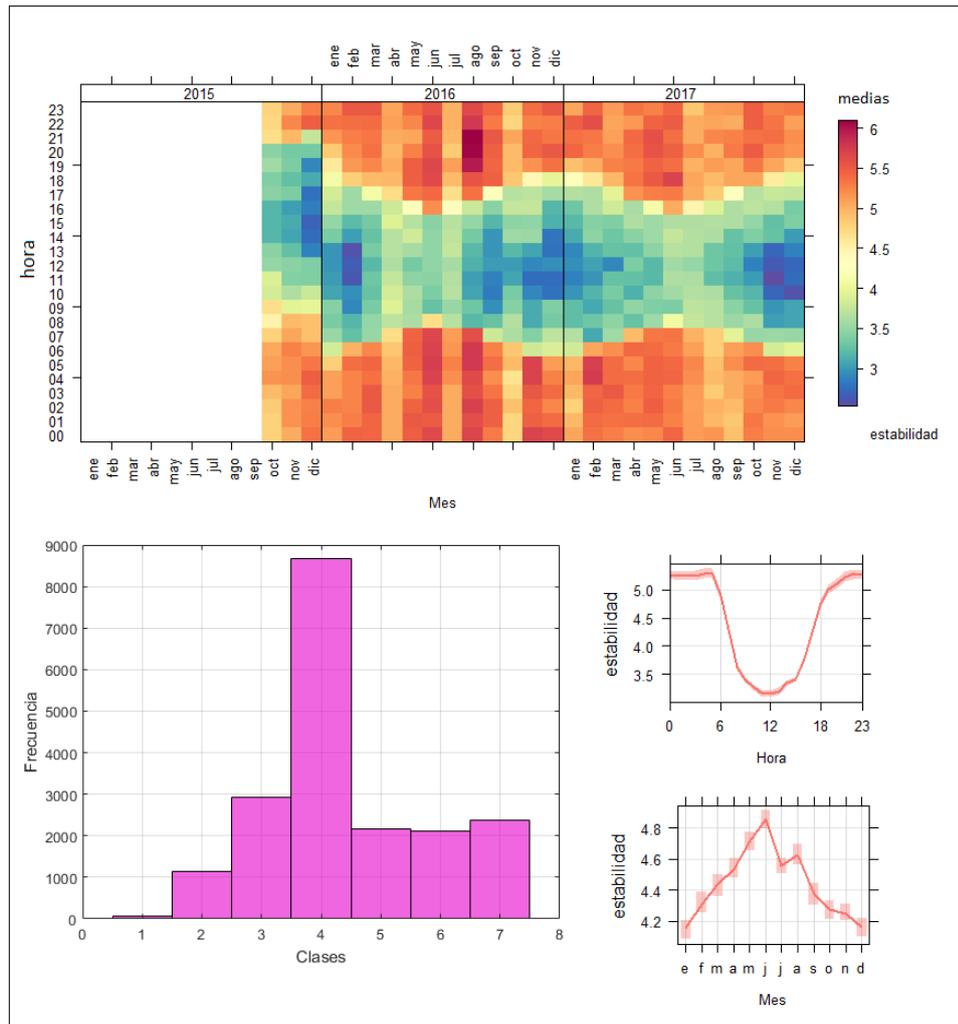


Figura 4.5: Tendencias de las clases de estabilidad atmosférica

En cuanto a los gases, se presenta un análisis de aquellos conocidos como precursores del PM: SO_2 , NO_x (como suma de NO y NO_2) y HCT (como suma de compuestos orgánicos volátiles).

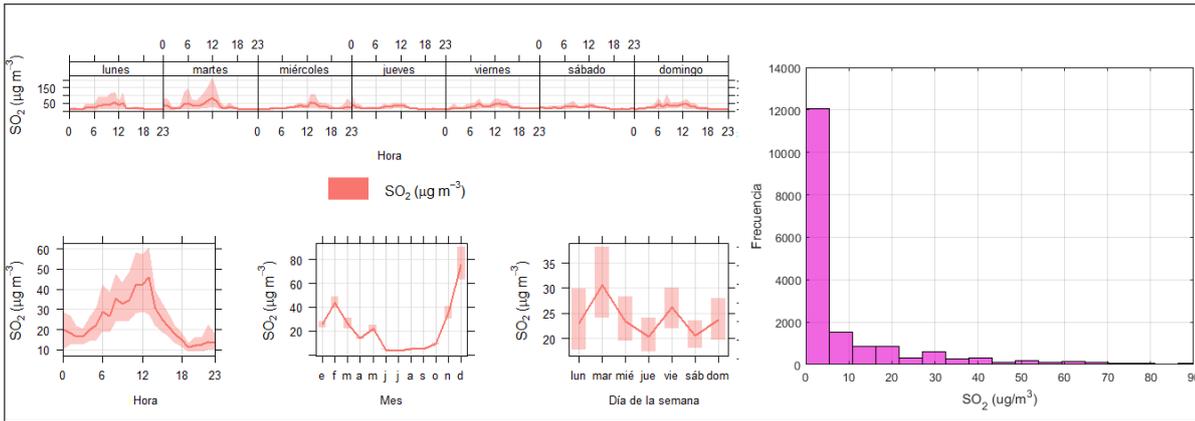


Figura 4.6: Histograma y tendencias mensual y horaria de SO_2

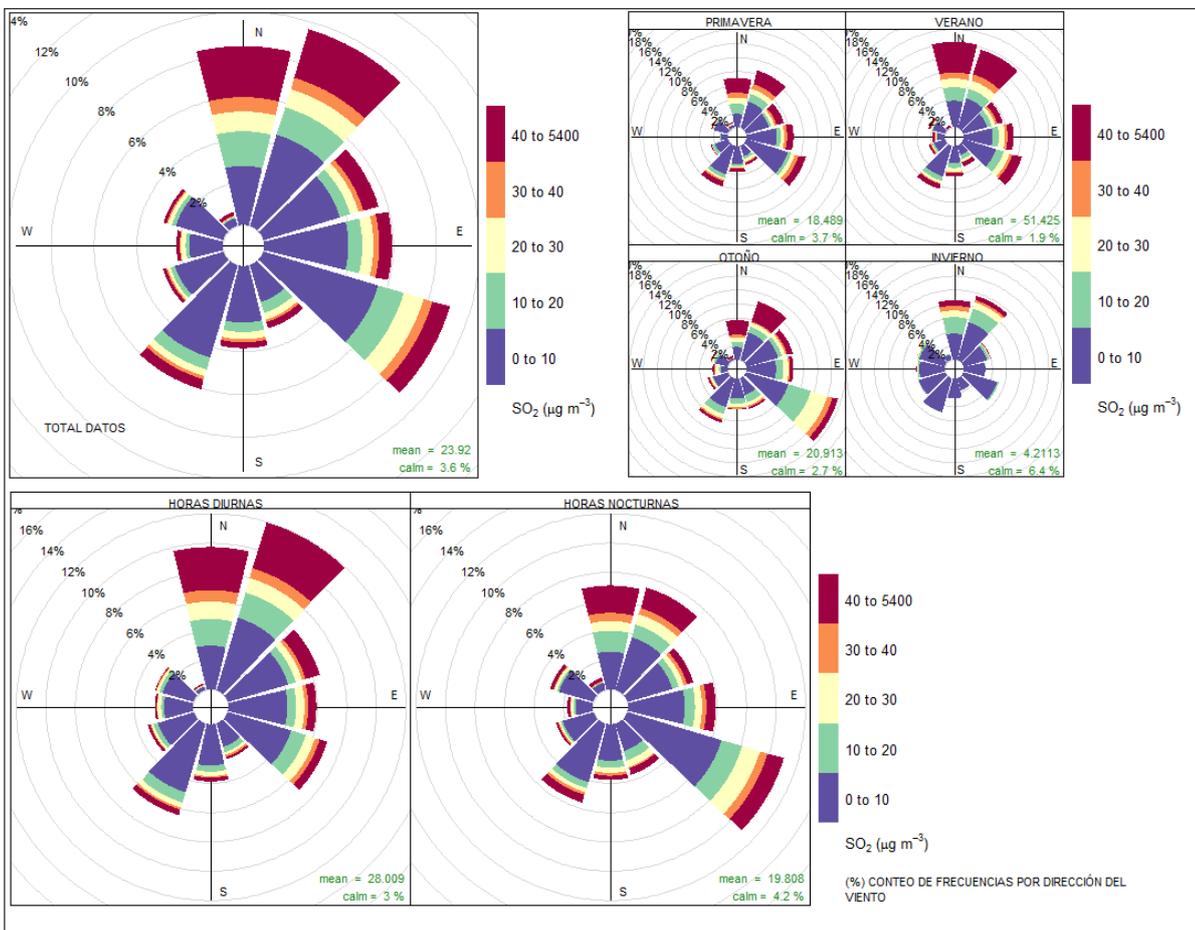


Figura 4.7: Rosas de distribución total, estacional y en horas diurnas y nocturnas de SO_2

El SO_2 registra un pico medio cercano a los $45 \mu g m^{-3}$ cerca del mediodía y presenta

menores concentraciones medias durante los meses de baja temperatura (**Figura 4.6**). Su principal fuente de emisión antropogénica es la quema de combustibles y es un gas propicio para unirse a núcleos de condensación, lo cual le otorga a las partículas formadas un nivel de acidez alto. Sus tendencias observadas en el período de estudio, posiblemente se encuentren más ligadas a la actividad industrial que al flujo vehicular. En un conteo de frecuencias con la totalidad de datos, se observa que las mayores concentraciones de SO_2 suelen coincidir con vientos del Norte y Noreste (**Figura 4.7**), fenómeno que se sostiene en las diferentes estaciones del año y en momentos del día (con mayores proporciones durante el verano y en horas diurnas). Esta dirección corresponde a la ubicación del Polo Petroquímico respecto a la estación de monitoreo, como se puede ver en la **Figura 3.3** del **Apartado 3.1.1**.

El NO_x posee su máximo medio (cercano a $70 \mu g m^{-3}$) a primeras horas de la mañana y su mínimo medio (cercano a $20 \mu g m^{-3}$) a media tarde, y sus concentraciones disminuyen drásticamente los fines de semana (**Figura 4.8**). Esto podría ligarse a los períodos de circulación de automotores, donde sus picos corresponden a emisiones de NO combinadas con producción fotoquímica de NO_2 . Estacionalmente, encuentra su pico durante el invierno.

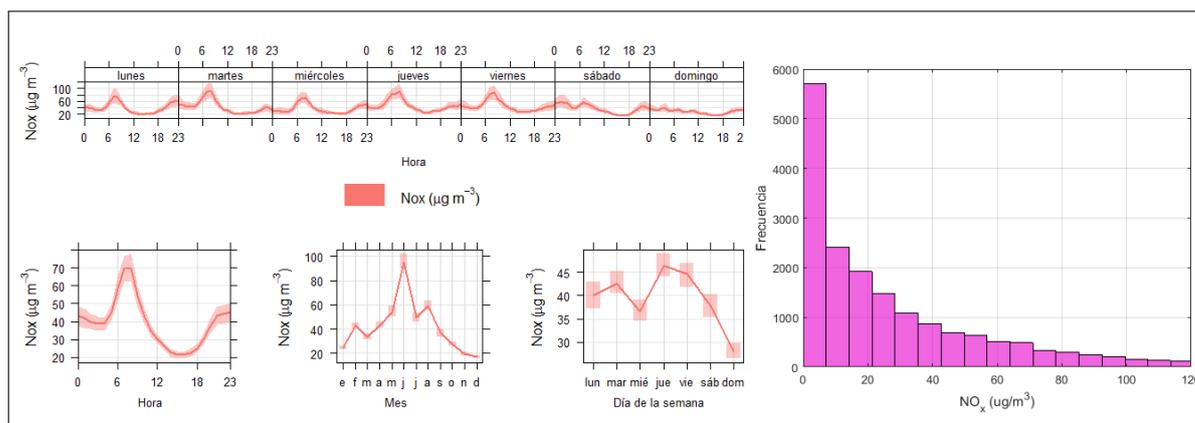


Figura 4.8: Histograma y tendencias mensual y horaria de NO_x

En la **Figura 4.9** se observa que también las mayores concentraciones se dan con vientos desde la dirección del sector industrial, pero también existe una componente apreciable desde el Sudoeste que corresponde a una zona residencial. Esto parece indicar que la estación se ve afectada por fuentes industriales y urbanas de NO_x .

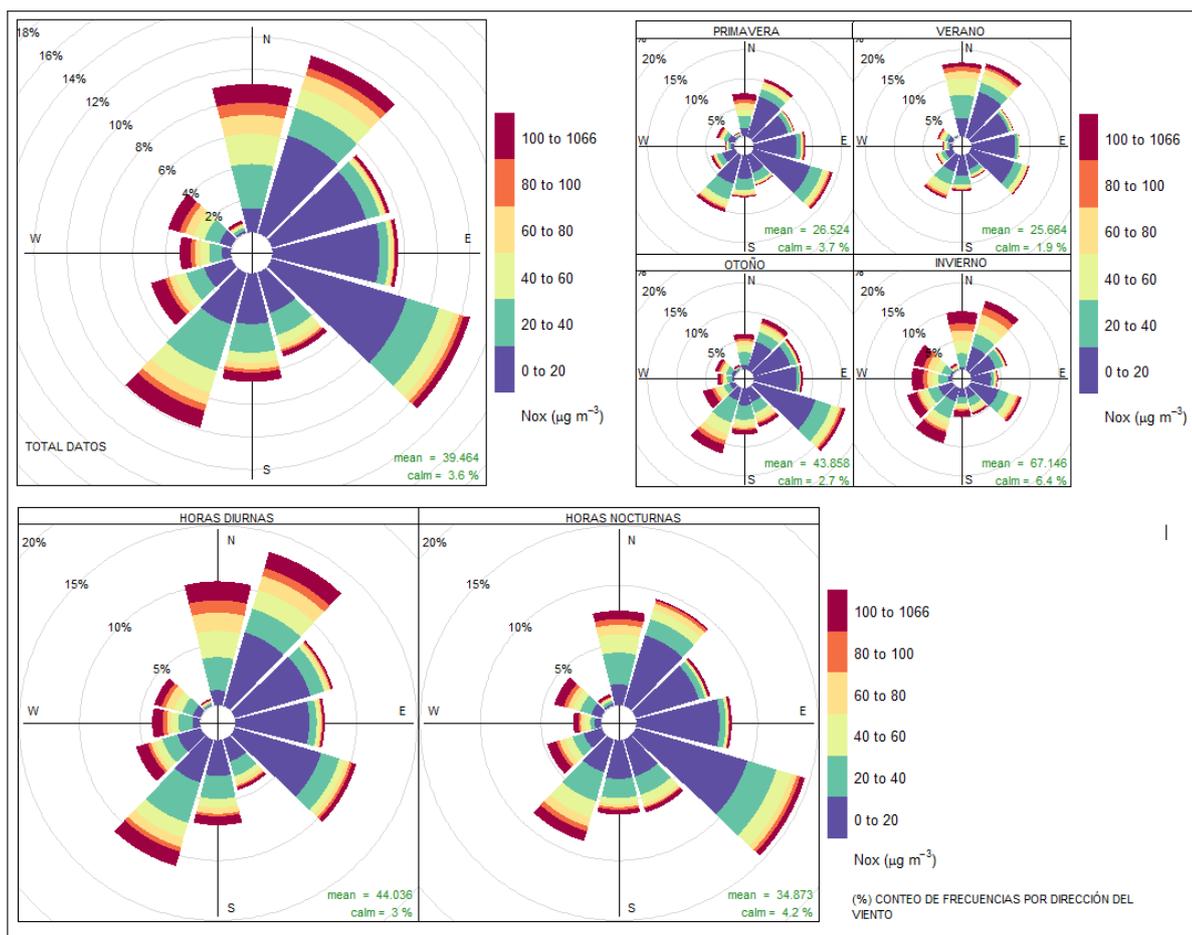


Figura 4.9: Rosas de distribución total, estacional y en horas diurnas y nocturnas de NO_x

La principal fuente antrópica de HCT es la quema de combustibles en vehículos. Por ello su tendencia diurna indica máximos (en promedio cercanos a $1,65 \mu\text{gm}^{-3}$) en primeras horas de la mañana y mínimos (cercanos a $1,35 \mu\text{gm}^{-3}$) después del mediodía y vuelve a crecer luego de las 18 h (**Figura 4.10**). Sus valores más frecuentes se encuentran entre $1 \mu\text{gm}^{-3}$ y $1,5 \mu\text{gm}^{-3}$. Estacionalmente, presenta las mayores concentraciones en otoño y las menores en primavera. Las rosas de la **Figura 4.11** indican que las mayores contribuciones provienen del sector industrial, seguida por una frecuencia algo menor desde el sector urbano durante todo el año y desde el Sudeste (principalmente en verano y otoño), donde se ubica la reserva natural “Eco Área Avellaneda” con espacios arbóreos y viñedos.

4 Potencialidad de las Redes Neuronales en calidad del aire

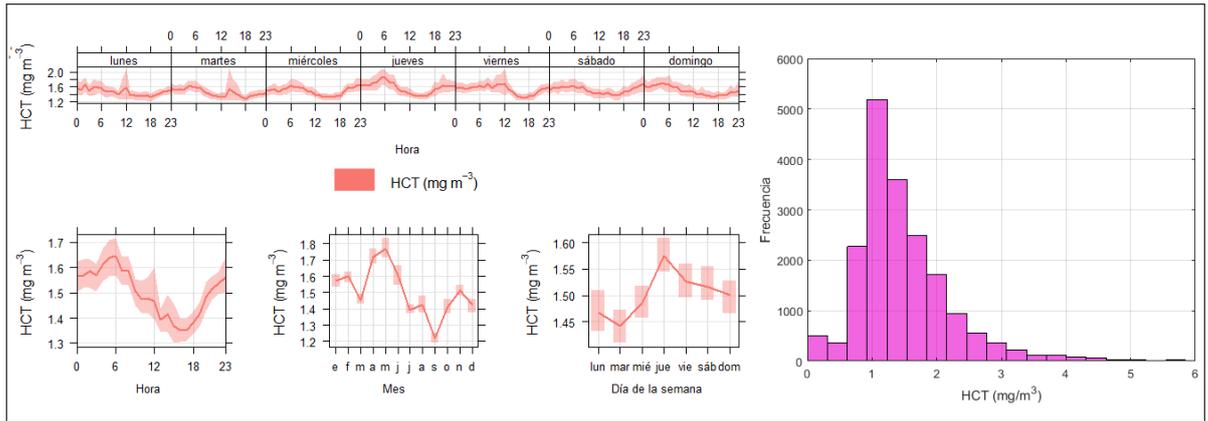


Figura 4.10: Histograma y tendencias mensual y horaria de HCT

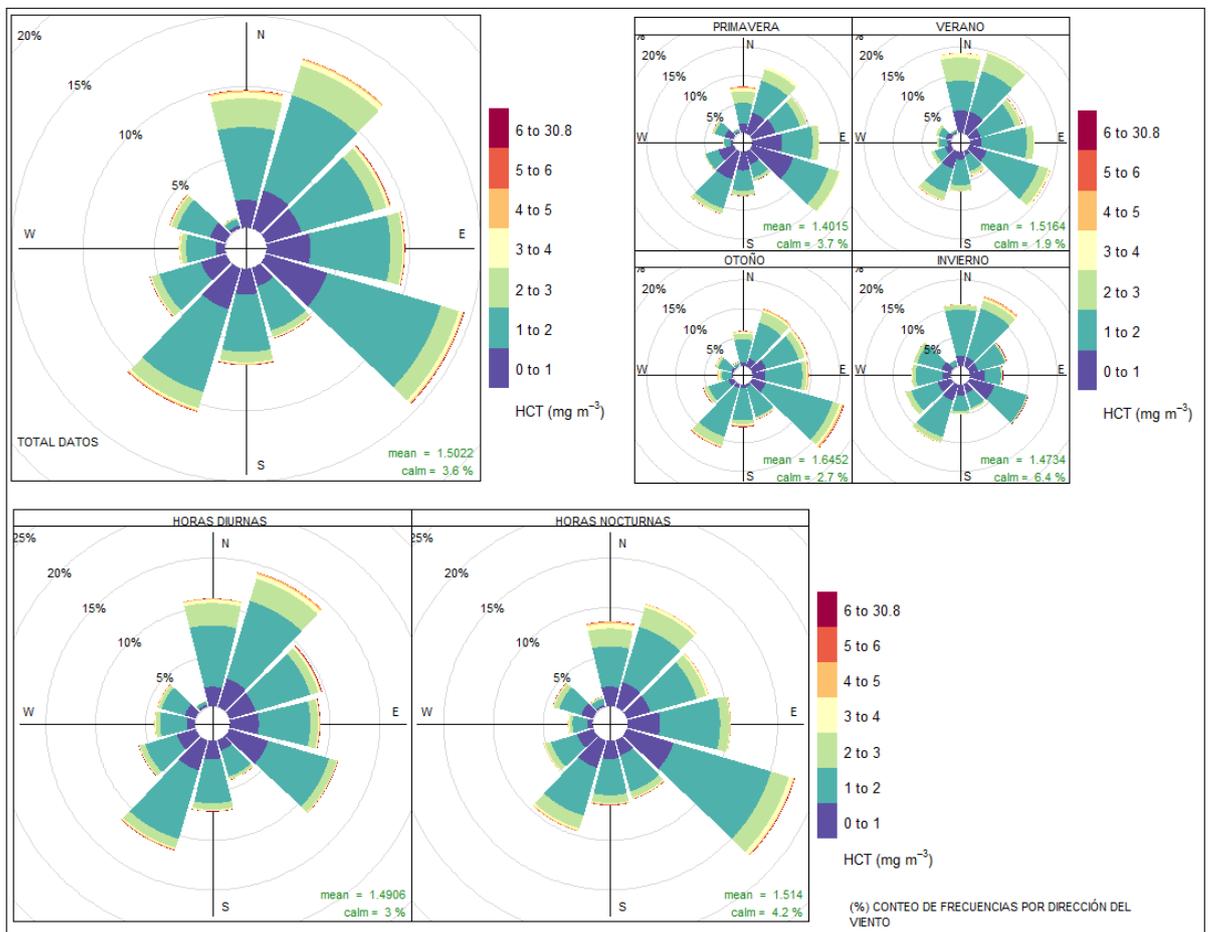


Figura 4.11: Rosas de distribución total, estacional y en horas diurnas y nocturnas de HCT

A continuación, se detallan las tendencias para el período de estudio de las series de

PM ($PM_{10}apra$, $PM_{25}acu$ y $PM_{10}acu$).

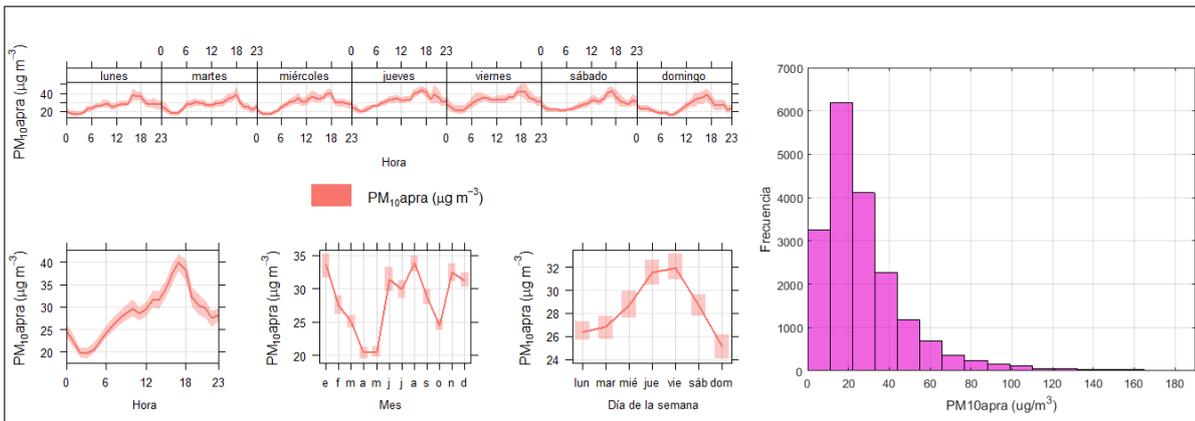


Figura 4.12: Histograma y tendencias mensual y horaria de $PM_{10}apra$

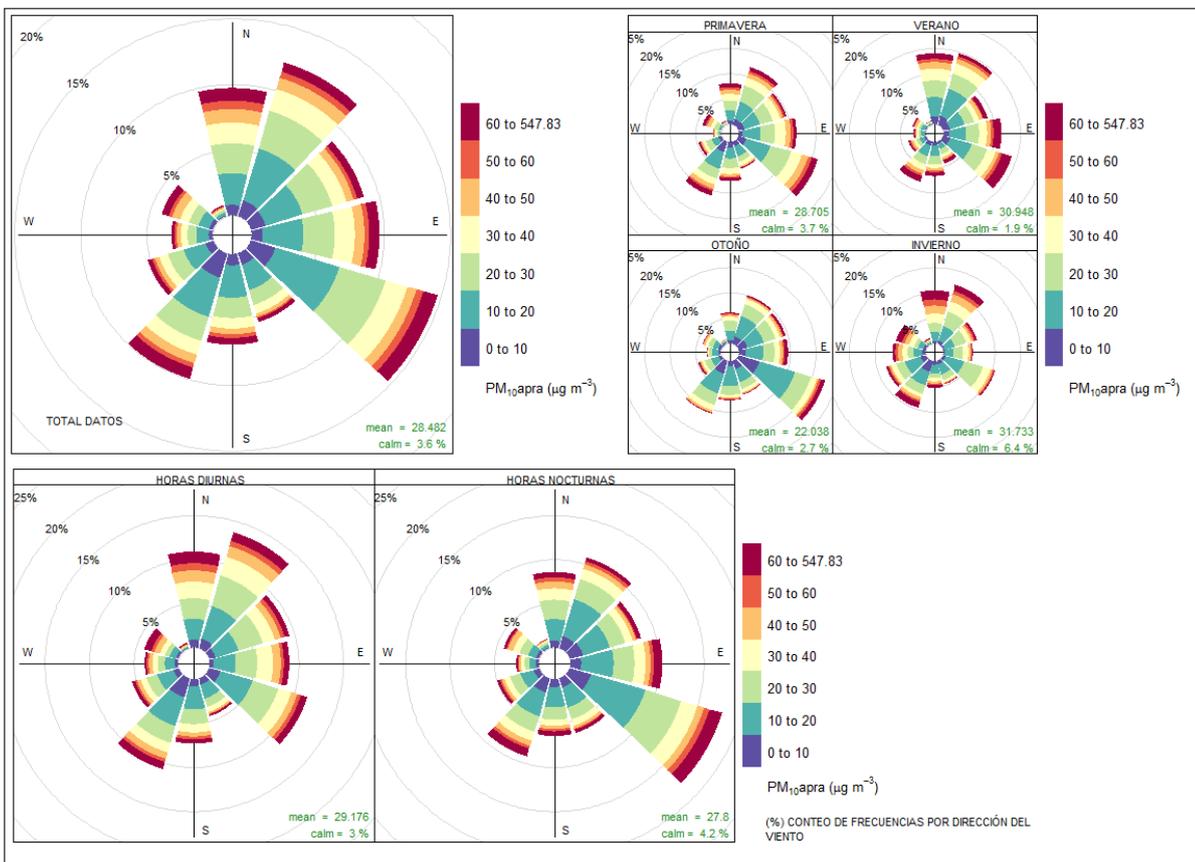


Figura 4.13: Rosas de distribución total, estacional y en horas diurnas y nocturnas de $PM_{10}apra$

$PM_{10}apra$ posee una tendencia diaria que comienza con valores mínimos a las 3h,

4 Potencialidad de las Redes Neuronales en calidad del aire

luego ascienden rápidamente hasta las 17h alcanzando su valor máximo (en promedio de $40 \mu g m^{-3}$), y finalmente decrecen los valores también a una tasa acelerada (**Figura 4.12**). Durante la semana, se observa un crecimiento sostenido desde los primeros días de la semana, para luego caer rápidamente en los fines de semana. Esto puede sugerir la permanencia del PM en la atmósfera que se acumula durante varios días y finalmente decae los días de menor actividad en la ciudad. Sus concentraciones más frecuentes se encuentran entre $15 \mu g m^{-3}$ y $30 \mu g m^{-3}$. La tendencia estacional no es clara o al menos no responde a causas fácilmente identificables, por lo que se precisa analizar un período de tiempo mayor para aseverar conclusiones al respecto. Las rosas de la **Figura 4.13** indican que las mayores concentraciones son comunes a todas las direcciones del viento.

El PM con tamaños menores a $2,5 \mu m$ monitoreado por la estación de ACUMAR durante el período de estudio (PM_{25acu}) presenta un mínimo a las 3h (en promedio de $6 \mu g m^{-3}$) y máximos (cerca de $16 \mu g m^{-3}$) a media mañana y cerca de las 23h (**Figura 4.14**). También se observa un mínimo local a media tarde. Semanalmente, se observa un nivel de base que aumenta los días jueves y viernes y cae rápidamente los fines de semana, lo cual parece estar ligado a las actividades en la ciudad. Sus concentraciones más frecuentes se encuentran cercanas a $10 \mu g m^{-3}$. En cuanto a las direcciones (**Figura 4.15**), se registran altas concentraciones desde todos los sectores, con mayor frecuencia desde el Sudoeste (sector urbano) durante todo el año y en especial en horas diurnas, y desde el Sudeste (sector de reserva natural) en primavera, verano y otoño y durante todo el día (con una frecuencia un poco mayor en horas nocturnas). La tendencia anual posee una forma similar a la de PM_{10apra} .

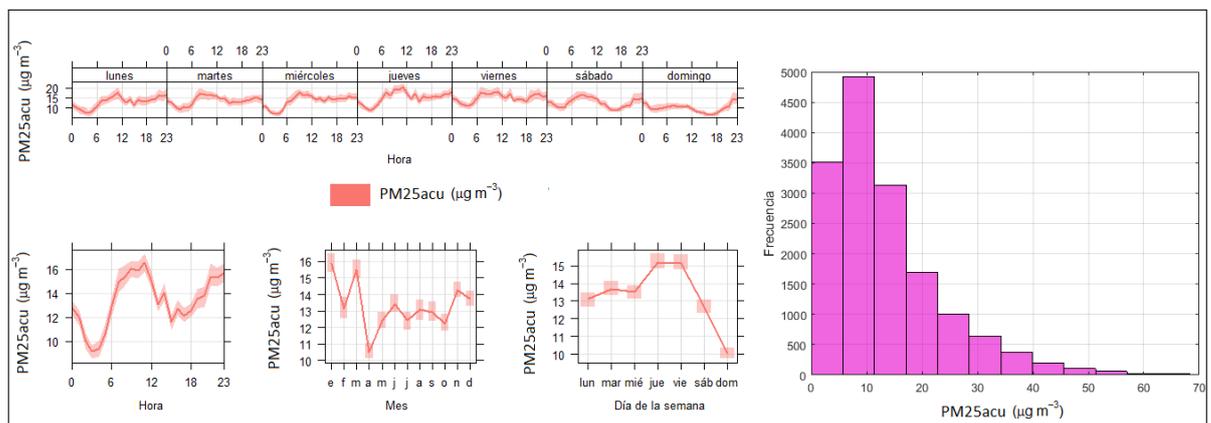


Figura 4.14: Histograma y tendencias mensual y horaria de PM_{25acu}

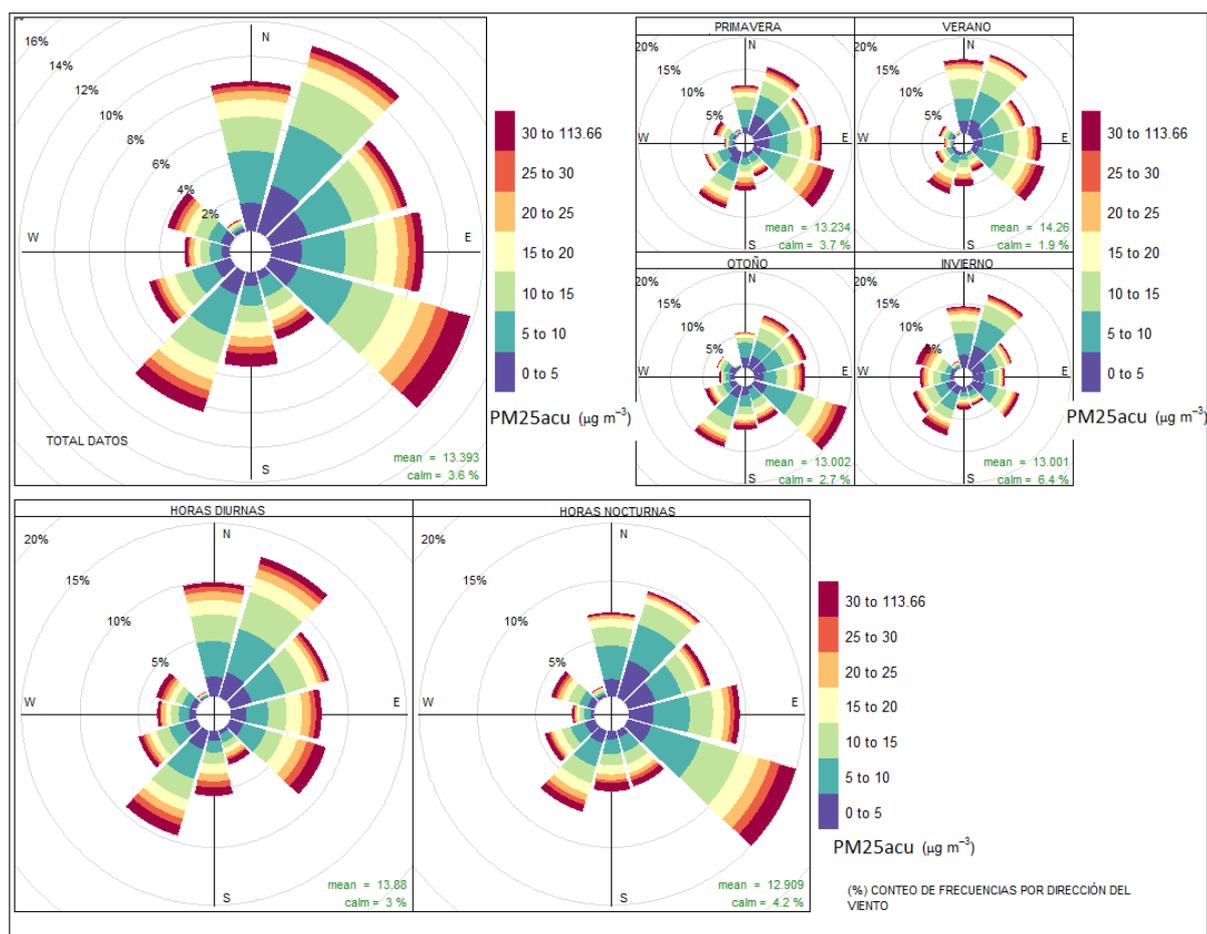


Figura 4.15: Rosas de distribución total, estacional y en horas diurnas y nocturnas de PM_{25acu}

Los ciclos diurno y semanal de PM_{10acu} parecen estar ligados a la actividad antrópica (**Figura 4.16**). Se registran mayores valores (medias cercanas a $25 \mu g m^{-3}$) a media mañana y durante las últimas horas de la tarde, lo que coincide con períodos de mayor circulación vehicular. Incluso se observa un decaimiento en las concentraciones medias durante los fines de semana, cuando la actividad industrial disminuye. Estacionalmente, muestra una tendencia más marcada que las otras series de PM y se encuentran mayores concentraciones en los meses cálidos. Sus valores de concentración más frecuentes se encuentran entre $5 \mu g m^{-3}$ y $20 \mu g m^{-3}$.

4 Potencialidad de las Redes Neuronales en calidad del aire

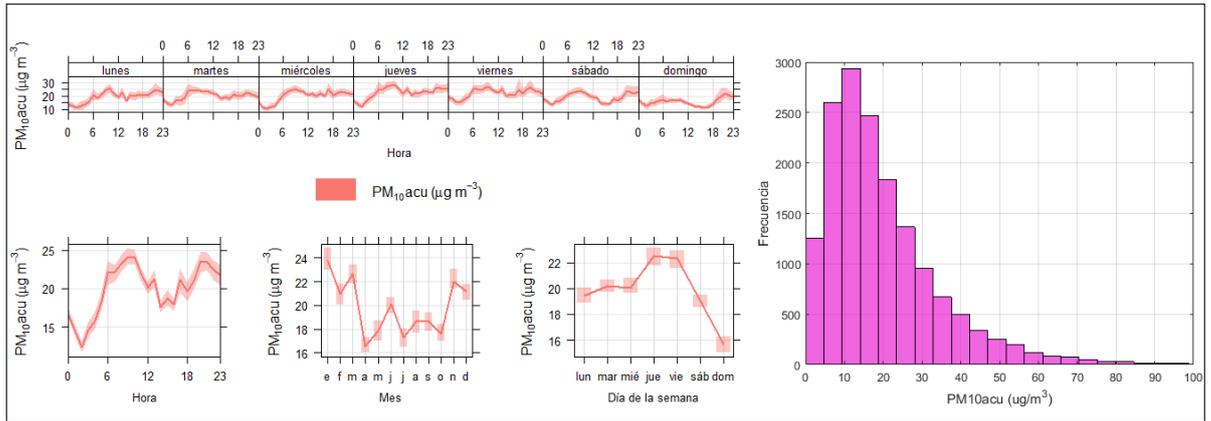


Figura 4.16: Histograma y tendencias mensual y horaria de $PM10acu$

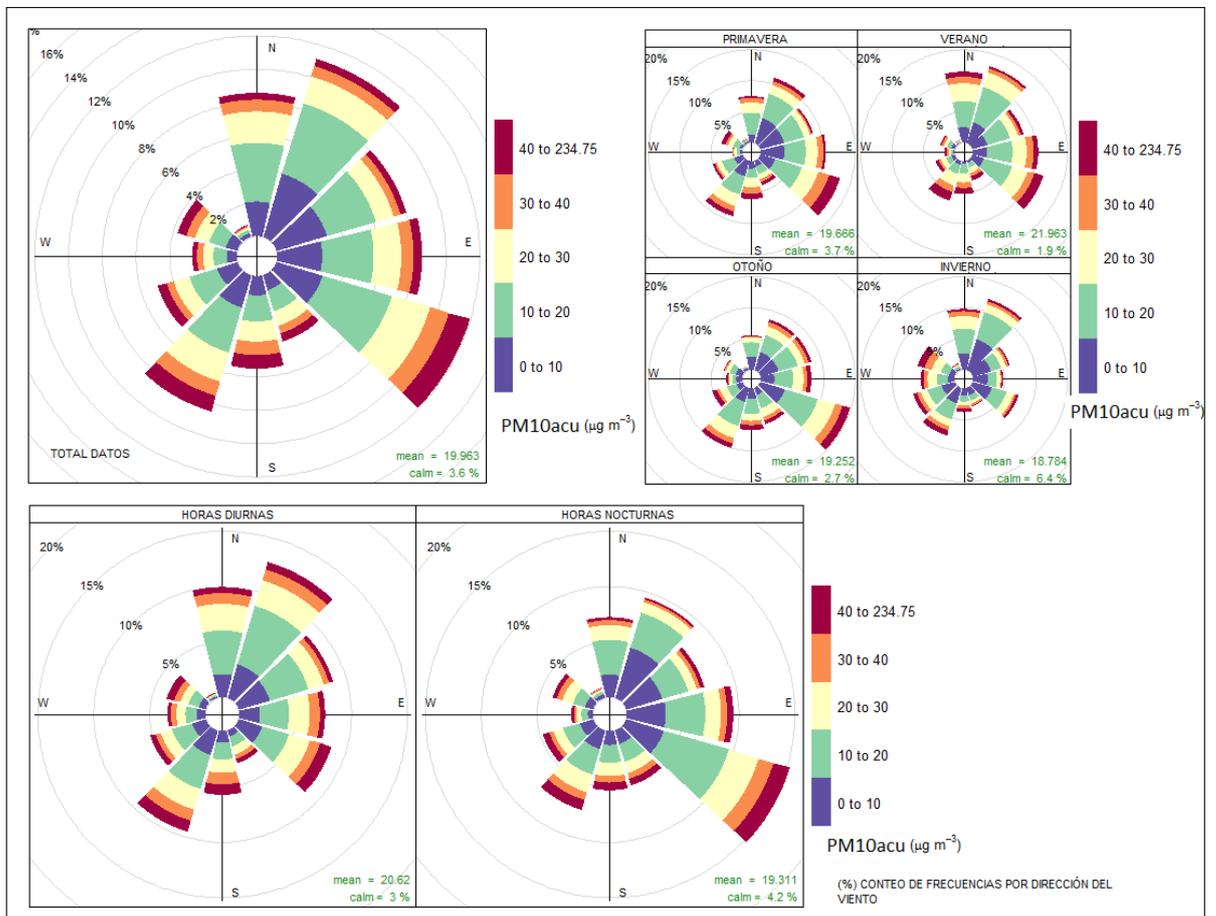


Figura 4.17: Rosas de distribución total, estacional y en horas diurnas y nocturnas de $PM10acu$

La **Figura 4.17** muestra que las principales contribuciones para $PM10acu$ se registran

desde el sector urbano (durante todo el año) y desde la reserva natural (todo el año, excepto en invierno).

A modo de completitud, se realizó un control de las series descritas en función de los límites establecidos en la normativa vigente para provincia de Buenos Aires (DECTO-2018-1074-GDEBA-GPBA)[202] y ningún contaminante excede dichos valores.

4.2.2. Modelado de PM

Las interpolaciones univariadas implementadas para completar la subserie de $PM_{2.5}$ con un 15% de datos faltantes generados artificialmente funcionan bien cuando estos faltantes están distribuidos al azar en la serie original, especialmente los métodos lineal y vecinos cercanos (**Figura 4.18**).

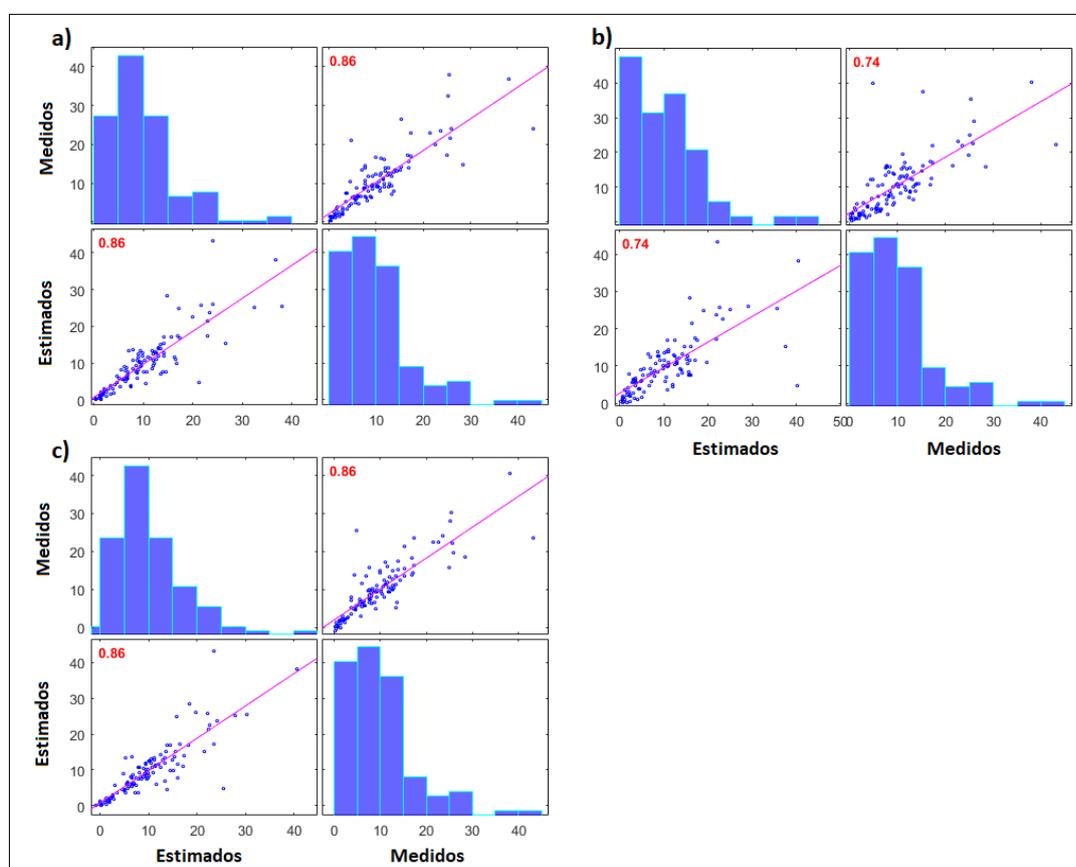


Figura 4.18: Desempeño de las imputaciones lineal (a), spline (b) y de vecinos cercanos (c) de la subserie de $PM_{2.5}$. Se muestran las distribuciones y las regresiones lineales dadas entre las series de valores estimados por los modelos y las mediciones reales

Los huecos (datos faltantes consecutivos) en una serie de un contaminante medido

generalmente no están distribuidos completamente al azar, sino que la distribución de tamaños varía de acuerdo a protocolos de mantenimiento, imponderables en el funcionamiento o la toma de datos, entre otros factores. Por ello, los resultados de la **Figura 4.19** son más realistas. En ella se puede observar que los errores relativos son bajos y las correlaciones aceptables cuando la longitud de huecos es pequeña, pero a medida que aparece una mayor cantidad de faltantes consecutivos los métodos univariados comienzan a fallar, alcanzándose errores de más de un 100 % para huecos longitud 8 o más. Es por ello que, dada la naturaleza de las distribuciones de faltantes en series de contaminantes, resulta necesario encarar la imputación a partir de métodos multivariados y no lineales que sean más robustos a la distribución de huecos.

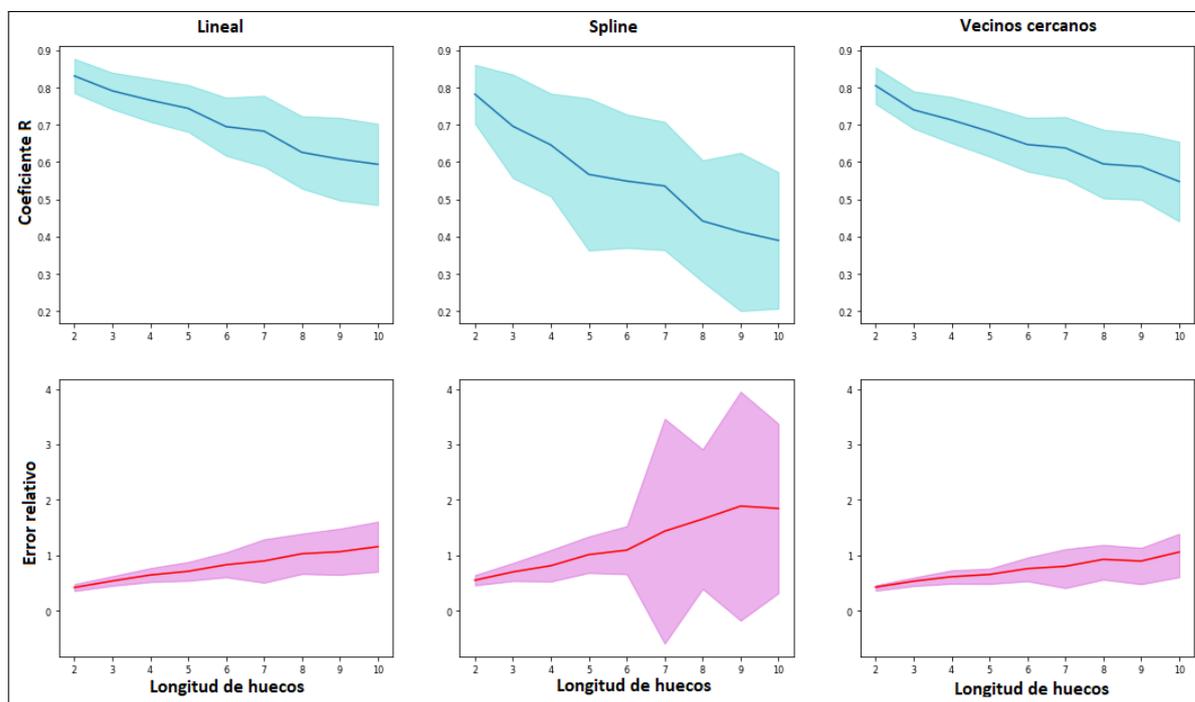


Figura 4.19: Coeficientes de correlación (R) con banda de incerteza de una desviación estándar en color cyan y errores relativos en rosa correspondientes a imputaciones con métodos lineal, spline y vecinos cercanos para diferentes longitudes de huecos distribuidos al azar en la subserie de prueba

Las series descritas en la **Tabla 3.1** del **Apartado 3.1.1**, además de *hora*, *diasem*, *mes* y *estab*, se utilizaron en el modelo RED1 (33 variables). La **Figura 4.20** muestra el coeficiente R alcanzado durante la validación para diferente cantidad de neuronas ocultas (en todos los casos el RMSE fue de $0,1 \mu\text{gm}^{-3}$). El mejor resultado en cuanto a desempeño y costo computacional se encontró con 24 neuronas ocultas.

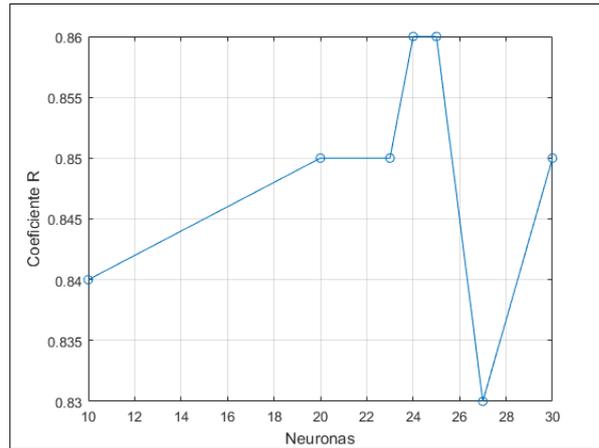


Figura 4.20: Coeficiente R de la RED1 en función de la cantidad de neuronas ocultas

Al aplicar un filtro de medianas a la serie PM_{25acu} se observó una mejoría en el desempeño de RED1, dado que de esa manera se logra reducir los efectos del ruido y valores extremos. Los mejores resultados se encontraron con 15 neuronas ocultas (RMSE de $0,1 \mu gm^{-3}$ y $R=0,9$). Si bien aumentó levemente el error general, el coeficiente de correlación aumentó significativamente con una menor cantidad de neuronas y tiempo de procesamiento.

La **Tabla 4.1** muestra el desempeño de la red de prueba REDP a la hora de estimar PM_{25acu} con cada predictor, ordenada en función del coeficiente de correlación obtenido. Es de esperar que los mejores predictores sean las series de PM dado que comparten fuentes y precursores con la variable objetivo. También se espera que los óxidos de nitrógeno estén en las primeras posiciones de la tabla, dado que son precursores de PM. La altura de capa de mezcla determina un límite superior donde quedan confinados los contaminantes y la velocidad del viento influye en la dispersión de los mismos, por ello son los predictores que siguen en importancia. Llamamos la atención los pobres resultados con $PPTacu$ y SO_2 , dado que el primero se relaciona a procesos de “lavado” del PM en el aire y el segundo es un precursor de la variable objetivo, pero al menos en el período y zona de estudio parecen tener poca influencia. También resultan notables los coeficientes de correlación altos de los predictores CO y SH_2 , dado que no se trata de precursores directos de PM, aunque pueden deberse a que compartan fuentes y tendencias con el objetivo.

La selección de predictores basada exclusivamente en conocimiento experto puede pasar por alto algunas relaciones debidas a mecanismos y fenómenos desconocidos. Por ello, se decidió utilizar los resultados de la **Tabla 4.1** como base para escoger a los me-

jores predictores. Cabe recordar que, como se mencionó en la **Sección 3.1**, los datos de vientos medidos en la estación de ACUMAR se utilizaron para calcular las componentes zonal ($Uacu$) y meridional ($Vacu$). Estas variables reemplazaron a $WSacu$ y $WDacu$, pues aportan la misma información combinada para el modelo.

R de validación	Predictor	R de validación	Predictor
0,85	$PM10acu$	0,17	$etilben$
0,39	$PM10apra$	0,16	$O3$
0,37	NOx	0,15	ben
0,36	NO	0,15	$diasem$
0,32	$NO2$	0,14	HCT
0,24	BLH	0,14	$Troc$
0,23	$SH2$	0,13	pla
0,23	$Uacu$	0,12	mes
0,22	$Vacu$	0,12	rad
0,22	$hora$	0,11	$Tacu$
0,22	oxi	0,07	nub
0,22	CO	0,06	$Pacu$
0,21	$estab$	0,05	vis
0,20	tol	0,04	$SO2$
0,19	$HRacu$	0,04	$HCNM$
0,19	HCM	0,01	$PPTacu$
0,18	$MPxi$	–	–

Tabla 4.1: R de validación en estimación de $PM25acu$ para el modelo REDP probado con cada predictor

La correlación cruzada entre predictores mostró que existe una alta equivalencia entre NO , $NO2$ y NOx , entre $Tacu$ y $Troc$, entre HCT y HCM , entre Oxi y tol , y entre $etilben$ y $MPxil$. De estos conjuntos, se descartaron los predictores que poseen un menor valor de coeficiente R en la **Tabla 4.1** (NO , $NO2$, $Tacu$, HCT , tol y $etilben$). La matriz de correlación puede consultarse en el **Anexo 8.1.4**.

El método directo gradual, ingresando uno a uno los predictores en el orden dado en la **Tabla 4.1**, determinó que el mejor desempeño del nuevo modelo llamado RED2 se obtiene con 15 variables ($PM10acu$, $PM10apra$, NOx , BLH , $SH2$, $Uacu$, $Vacu$, $hora$, oxi , CO , $estab$, $HRacu$, HCM , $MPxi$, $O3$) y con 15 neuronas ocultas (RMSE de $0,14 \mu g m^{-3}$ y $R=0,9$). Si bien el desempeño logrado no es superior al obtenido con las 33 variables iniciales, se consigue reducir el costo computacional y minimizar la linealización de la red al contar con una dimensión del espacio de entrada más pequeño y evitar información

redundante (lo que mejora la generalización). Las pruebas de sensibilidad de la RED2 demostraron que los mejores desempeños en cuanto a RMSE, R y costo computacional se alcanzan con los siguientes parámetros por defecto:

I. **Normalización con “mapminmax” (re-escalado entre -1 y 1)**. Los desempeños fueron inferiores implementando normalización con “mapstd” (valor medio cero y varianza uno), con método de imputación de datos faltantes a partir de interpolación lineal, y removiendo filas con información idéntica en los patrones de entrada.

II. **Función de entrenamiento Levenberg-Marquardt**. No se obtuvieron resultados superadores con funciones tales como Cuasi-Newton, Bayesiana, gradiente conjugado o gradiente descendente.

III. **Función de transferencia en neuronas ocultas “tansig” (Tabla 4.2)**.

IV. **Función de inicialización de pesos y sesgos Nguyen-Widrow**. Las funciones al azar y de punto medio mostraron desempeños muy inferiores.

Función	Descripción	RMSE (μgm^{-3})	R
tansig	Tangente hiperbólica sigmoide	0,14	0,90
radbasn	Base radial normalizada	0,14	0,90
satlins	Lineal saturada simétrica	0,14	0,90
tribas	Base triangular	0,14	0,90
poslin	Lineal positiva	0,14	0,89
radbas	Base radial	0,14	0,89
satlin	Lineal saturada	0,14	0,89
softmax	Exponencial normalizada	0,14	0,89
elliotsig	Sigmoide Elliot simétrica	0,14	0,88
logsig	Sigmoide logarítmica	0,17	0,87
purelin	Lineal	0,14	0,83
hardlims	Escalón simétrico	0,26	0,35
hardlim	Escalón	0,28	0,27
netinv	Inversa	0,32	0,16

Tabla 4.2: Coeficiente R de validación y RMSE en estimación de PM_{25acu} para el modelo RED2 con diferentes funciones de transferencia en neuronas ocultas. Las funciones “radbasn”, “satlins” y “tribas” igualaron en desempeño a la función por defecto “tansig”

El modelo dinámico recursivo con realimentación RED3 (**Figura 4.21**) igualó el desempeño de la RED2 con un menor costo computacional y menor cantidad de neuronas ocultas (9). La **Figura 4.22** muestra el desempeño de RED3 (RMSE y R) en función de la cantidad de neuronas ocultas y la **Figura 4.23** muestra las gráficas de desempeño,

entrenamiento y regresión.

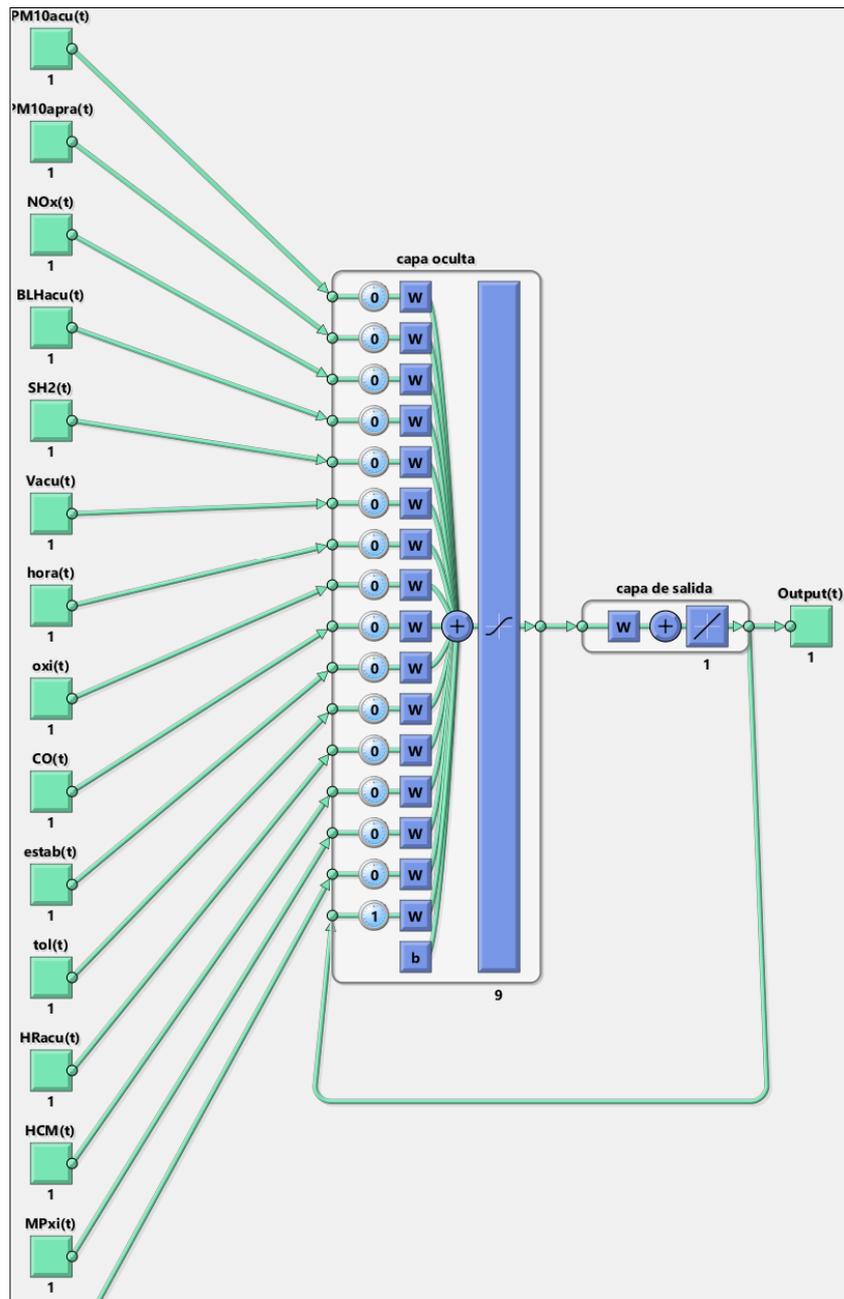


Figura 4.21: Esquema del modelo RED3

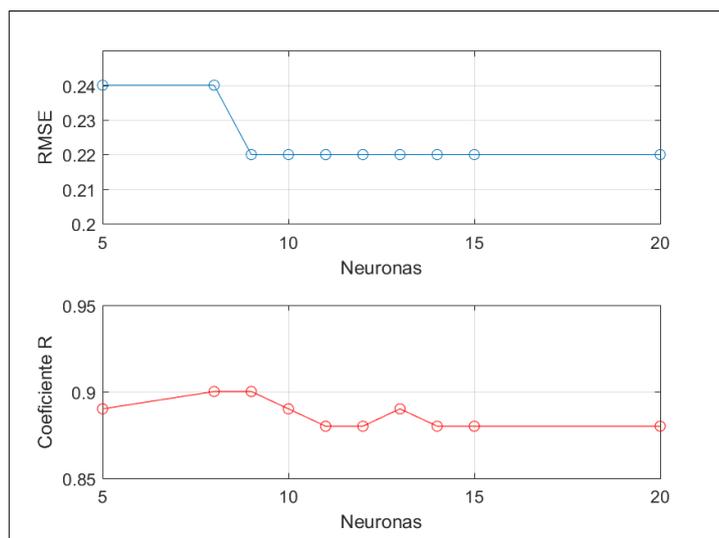


Figura 4.22: Desempeño de RED3 en función de cantidad de neuronas ocultas

En la **Figura 4.23** se puede observar que se activó la detención temprana del entrenamiento al alcanzar la cantidad máxima de fallas en la validación en 6 épocas consecutivas. También se muestra la raíz del error cuadrático medio (RMSE) de entrenamiento y validación alcanzan valores mínimos antes de la detención, lo cual implica un acierto en la convergencia, y que los coeficientes R para cada etapa de entrenamiento son igualmente altos (se logró una buena validación del modelo y una prueba final satisfactoria con nuevos patrones de entrada), lo cual indica una buena generalización. El modelo se entrenó en un total de 62 épocas y alcanzó su mejor desempeño a las 56 épocas.

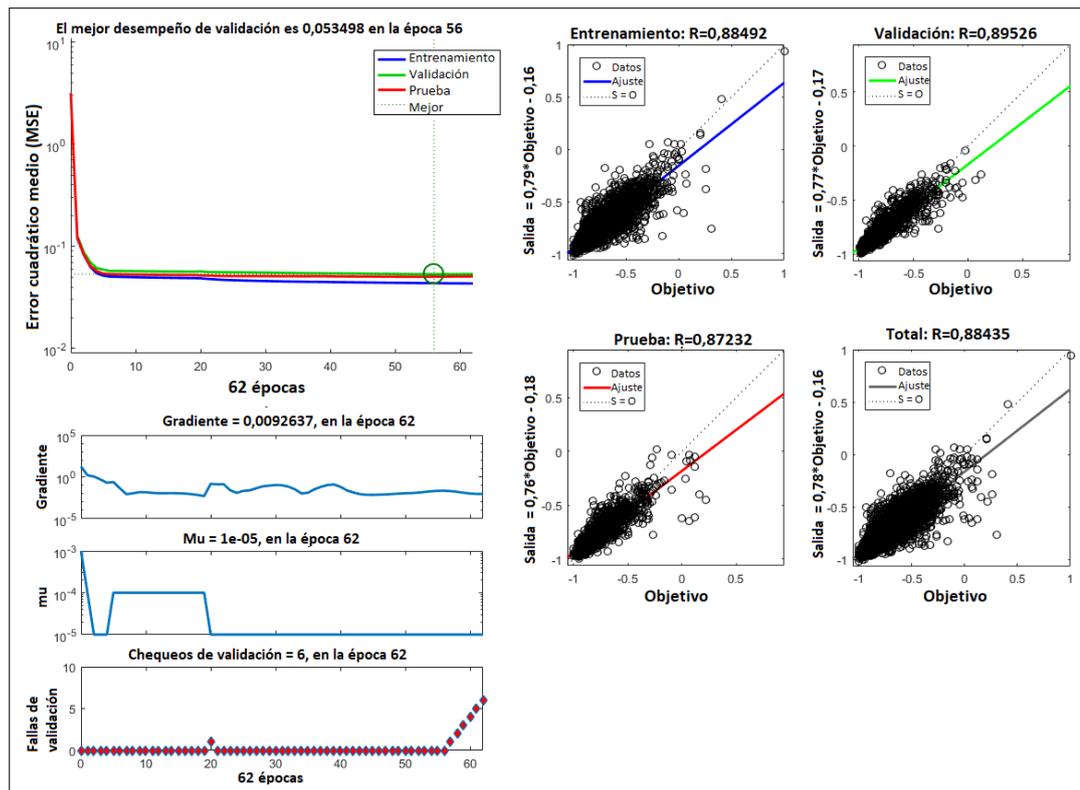


Figura 4.23: Gráficas de regresión, entrenamiento y desempeño de RED3

Se practicó la etapa de prueba del modelo RED3 con 1200 patrones elegidos al azar y separados antes de entrenar la red. En la **Figura 4.24** se muestran 4 subseries de 30 datos cada una elegidos al azar dentro de la serie de prueba con el fin de observar con mayor detalle los resultados. Se muestran los valores normalizados entre -1 y 1 para facilitar la comparación.

Se ha determinado que el modelo responde satisfactoriamente a cambios en las concentraciones y presenta un buen desempeño al computar con nuevos datos (con valores promedios de R entre 0,8 y 0,9, y de $RMSE$ cercanos a $0,1 \mu gm^{-3}$). Se ha notado una dificultad a la hora de estimar picos altos de contaminantes, que es una característica que suele acompañar a estos tipos de modelos. Esto se debe principalmente a que los eventos de altas concentraciones son los menos frecuentes y generalmente pueden deberse a situaciones poco comunes como incendios y fenómenos de transporte de material atípico. En este sentido, se cumple el condicionante de las ANN que les permite interpolar correctamente, pero mostrar ciertas falencias a la hora de extrapolar.

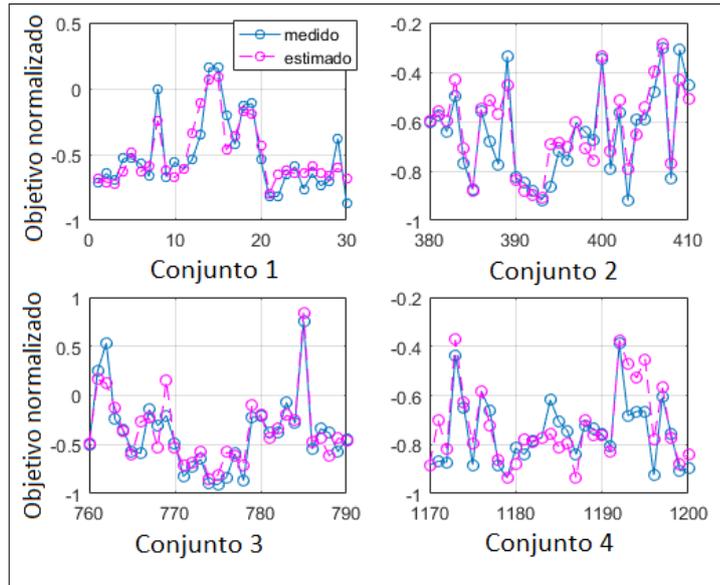


Figura 4.24: Comparación entre valores medidos (línea continua en azul) y estimados (línea discontinua en magenta) por el modelo RED3 en 4 conjuntos de datos elegidos al azar. Las escalas de los ejes de las ordenadas se autoajustaron para una mejor visualización de cada subserie, pero los valores normalizados se encuentran entre -1 y 1

Finalmente, se obtuvo un modelo RED3 de estimación de PM10acu con las siguientes características:

I. Modelo ANN-MLP-FF-BP dinámico recursivo.

II. Normalización con mapminmax en datos de entrada y objetivo. Esto implica que antes de ejecutar el modelo se debe normalizar de esta manera los datos de entrada y se debe denormalizar la salida para contar con las concentraciones reales.

III. 15 predictores de entrada. El modelo se ejecuta ingresando los valores de las 15 variables seleccionadas.

IV. Vector de 135 pesos implementados en 9 neuronas ocultas. Esta información se debe considerar para estimar el costo computacional, el cual resulta adecuado para cualquier computador convencional.

El código desarrollado y los datos para el diseño y entrenamiento del modelo RED3 se pueden consultar en el repositorio digital creado con productos de esta Tesis[36].

A continuación, se muestran los resultados de otro modelo ANN dinámico no recursivo RED4, con 25 neuronas ocultas. Este modelo pasó por la misma batería de pruebas y metodología de diseño que los modelos anteriores, con la diferencia de que fue entrenado sólo con variables meteorológicas (*BLHacu, Uacu, Vacu, estab, RHacu, rad, Troc, Pa-*

cu, visib, PPTacu) y temporales (valores medios por día de la semana, hora del día y estación del año). Esto se hizo con el fin de explorar la sensibilidad de estos modelos a la disponibilidad de datos y evaluar el desempeño de un modelo ANN de “bajo costo”.

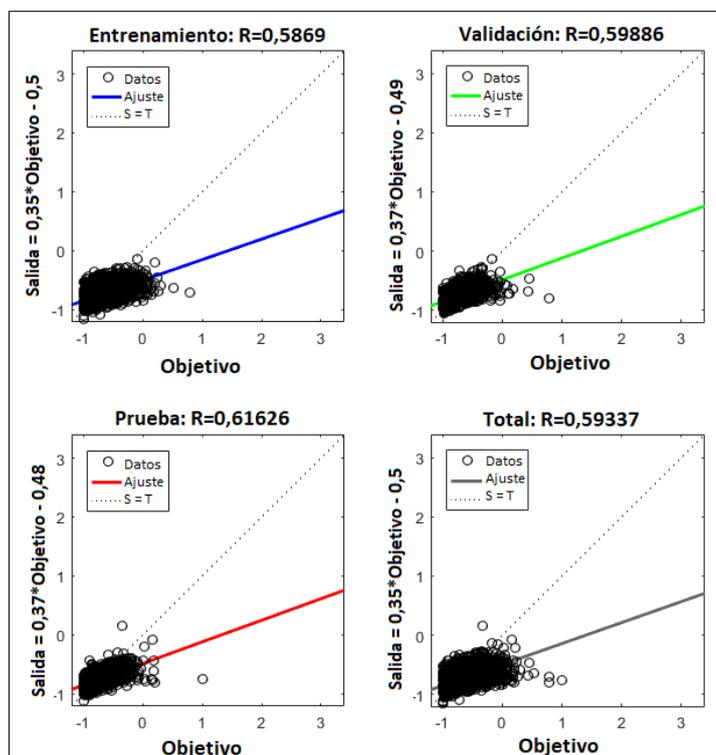


Figura 4.25: Desempeño del modelo RED4 en las diferentes etapas de entrenamiento

La **Figura 4.25** muestra los coeficientes de correlación obtenidos en las etapas de entrenamiento, validación y prueba, junto con las regresiones entre los valores medidos y estimados de la RED4 que se entrenó en un total de 63 épocas. Se pueden observar correlaciones más pobres que con RED3 pero aún así significativas, con una baja dispersión de los puntos ($RMSE = 0,14 \mu g m^{-3}$). La razón del mejor ajuste del modelo RED3 es su entrenamiento utilizando otros contaminantes precursores de PM y mediciones de PM de otros tamaños in situ y en cercanías. Resulta claro que una limitada disponibilidad de datos disminuirá el desempeño del modelo significativamente, pero el modelo RED4 ha demostrado que es posible desarrollar un enfoque de “bajo costo” que entregue buenas prestaciones y nueva información a sitios con limitada disponibilidad de mediciones.

4.3. Conclusiones preliminares

El análisis de tendencias de las variables indicaron que el mínimo local de PM_{25acu} durante la tarde coincide con vientos más intensos con dirección de Noreste respecto a la estación (sector industrial), la altura de la capa límite en ascenso acercándose a su máximo diario, condiciones neutrales a inestables de la atmósfera, valores bajos de humedad relativa y con concentraciones bajas de NO_x y HCT . En el verano se registran mayores concentraciones de PM_{25acu} , lo que coincide con mayores intensidades de velocidad del viento en dirección Sudeste (sector urbano), mayores alturas alcanzadas por la capa límite, condiciones atmosféricas cercanas a la inestabilidad, valores bajos de humedad y concentraciones bajas de NO_x . Se requiere de un estudio más profundo acerca de las vinculaciones entre PM_{25acu} y las variables descritas, con el fin de ajustar el modelo y ampliar sus capacidades de simulación bajo diferentes escenarios (tales como aquellos con picos máximos de concentración de PM_{25}). Estos resultados indican que será importante para la zona de la UNGS identificar las posibles fuentes urbanas e industriales respecto a la dirección del viento y que la altura de la capa límite es también un indicador a seguir, así como otras variables meteorológicas disponibles como la humedad relativa.

El diseño de modelos basados en ANN suele ser una tarea compleja debido a la cantidad de hiper-parámetros que poseen. No existe a la actualidad un protocolo procedimental o de toma de decisiones al respecto, por lo que un investigador debe utilizar conocimiento experto o reproducir otros procedimientos que son problema-dependientes. En este capítulo se muestran los resultados de una serie de pasos para el diseño que no pretende ocupar un lugar universal, pero ha demostrado ser efectivo para el problema aquí trabajado. Al respecto, los resultados muestran que muchos de los parámetros tomados por defecto aportan a un mejor desempeño del modelo y que se reduce el costo computacional agregando una realimentación de la salida a la capa de entrada con un retraso temporal. El filtrado de la señal objetivo mejoró el desempeño durante el entrenamiento y no influyó negativamente la predicción de las concentraciones medidas en las pruebas con nuevos datos sin filtrar.

La selección de predictores es un aspecto muy importante también, debido a que una alta dimensionalidad del espacio de entrada e información redundante puede provocar problemas de generalización y linealizar el modelo (perdiendo en parte sus propiedades de interpretar relaciones no lineales). En general, se puede elegir predictores con conocimiento experto o utilizar algún método intra o extra modelo. Para este desarrollo, se

decidió que abusar del conocimiento experto puede obviar cierta información implícita entre las variables (por ejemplo, la concentración de SO_2 y el régimen de precipitaciones suelen influenciar las concentraciones de PM, pero en este caso no afectaron la sensibilidad del modelo). Además, se propuso un método no lineal, gradual y fuera del modelo para seleccionar los mejores predictores, el cual arrojó mejores resultados que aquellos basados en métodos lineales.

El desarrollo de la ANN RED3 ha resultado satisfactorio, con un coeficiente R de 0,87 durante la etapa de prueba con nuevos datos y un RMSE de $0,22 \mu g m^{-3}$ (una proporción muy pequeña de los valores medios de las concentraciones de PM_{25acu}). Estos indicadores se encuentran entre los mejores desempeños indicados en la literatura para la predicción de valores de $PM_{2,5}$, como se mostró en la **Sección 2.1**. Por esto, se considera que el modelo también es adecuado para completar bases de datos de PM_{25} en la estación de monitoreo EMC I de ACUMAR para períodos sin medición. Asimismo, su desarrollo permite comprender con mayor profundidad las relaciones entre PM de pequeño tamaño y otros contaminantes y variables meteorológicas en el lugar de estudio. Algunas de estas características pueden ser extrapoladas al Centro-Norte de AMBA, pero otras responden puramente a cuestiones locales.

Tomando como objetivo el PM, este tipo de modelos se puede implementar en el Centro-Norte del conurbano con las mediciones que se vayan registrando. Mas allá de otros contaminantes, se ha determinado que predictores como la altura de la capa límite, la velocidad y dirección del viento, la humedad relativa y condiciones de estabilidad (datos con los que se cuenta en el AMBA, ya sea con estaciones de monitoreo o datos de asimilación) son importantes para estimar valores de PM. En efecto, el modelo RED4 demostró que se pueden implementar estos modelos utilizando exclusivamente variables meteorológicas y temporales (lo que se llamó aquí ANN de bajo costo), con un desempeño global menor pero aún así aceptable. La comparativa entre los resultados de los modelos RED3 y RED4 destacó lo valioso que es contar con mediciones de otros contaminantes, pues los precursores o aquellos que están sometidos a procesos similares que el PM especializan enormemente el modelo.

5 Caracterización de aerosoles

5.1. Objetivos e interés

Hasta la fecha, no existen registros de mediciones vinculadas a cantidad y tipos de aerosoles en el Centro-Norte del conurbano bonaerense, al igual que en la mayor parte del territorio argentino. Por ello, este capítulo presenta un primer estudio en la región al determinar sus principales propiedades ópticas con el fin de caracterizar la CA, en relación a los aerosoles. Esto se realiza en base a mediciones con equipamiento de bajo costo.

Por otro lado, las condiciones adecuadas para tomar mediciones de AOD no siempre están presentes, así como la disponibilidad de personal para realizar la medición, por lo que la cantidad de datos registrados con el fotómetro solar manual es limitada. Esto restringe el análisis que se puede realizar con este parámetro, y resulta necesario contar con un modelo explicativo que posea otras variables vinculantes. Es por ello que este capítulo también explora con ANN tipo SOM diferentes escenarios de contaminación atmosférica por aerosoles integrando diversas variables y las mediciones de bajo costo. Principalmente, se buscan vínculos entre la carga y clases de aerosoles con condiciones meteorológicas. Finalmente, se complementa el análisis con un estudio de trayectorias de masas de aire con el fin de determinar el origen regional de los principales tipos de aerosoles.

5.2. Resultados y discusiones

5.2.1. Exploración de datos meteorológicos

Las velocidades del viento registradas en Campo de Mayo entre marzo de 2018 a marzo de 2020 (**Figura 5.1**) resultan en promedio ser menores que los descritos para la estación EMCI de ACUMAR (**Figura 4.1**) en el **Capítulo 4**. Se debe considerar que ambas estaciones se encuentran distanciadas por más de 30 km y la de ACUMAR se

encuentra cerca del Río de la Plata y al borde de CABA, mientras que la de Campo de Mayo está más adentrada en el continente en una zona con mayores espacios verdes, por lo que se esperan diferencias meteorológicas.

En Campo de Mayo se observa un patron diurno y estacional similar al de ACUMAR, con mayores velocidades durante las primeras horas de la tarde y en los meses cálidos, pero con valores más frecuentes cercanos a 2 ms^{-1} y una distribución sesgada a valores bajos. También puede notarse que la banda de incerteza se ensancha para horas posteriores a las 21h y previas a las 6h, lo que se debe a que existen menos datos disponibles para todas las variables en esa banda horaria. Esto no representa un problema para la aplicación que se les dió en esta tesis a esos datos, dado que se emparejaron con mediciones fotométricas que se realizan fuera de esa banda horaria.

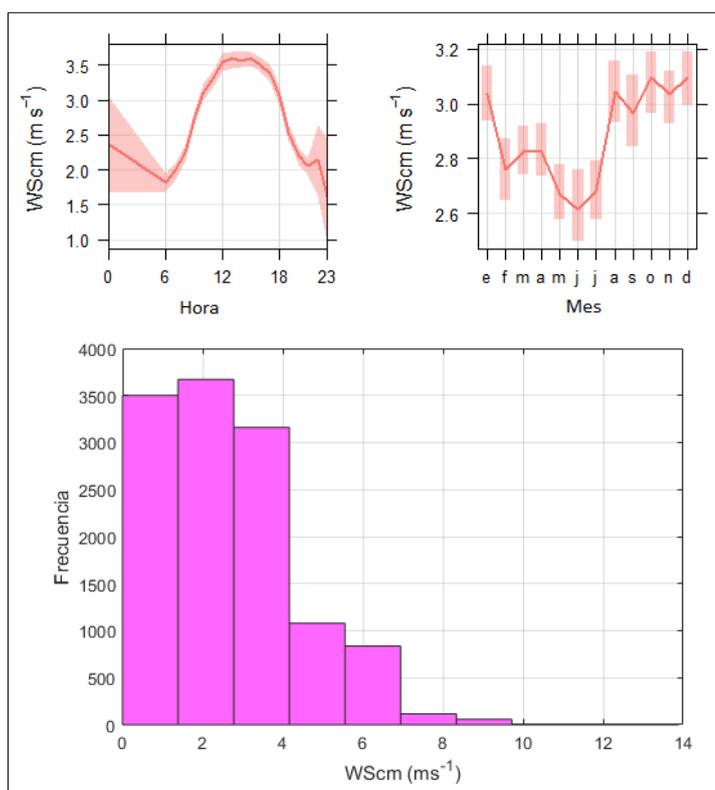


Figura 5.1: Histograma y tendencias mensual y horaria de datos de velocidad del viento del SMN en Campo de Mayo

En las horas diurnas parecen dominar los vientos del Norte y Noreste, mientras que durante la noche rotan más hacia el Este. Los sectores Este y Noreste tienen las mayores frecuencias durante primavera, otoño e invierno, mientras que en verano prepondera la dirección Sudoeste (**Figura 5.2**). Esta última observación es la que más se diferencia

respecto a la rosa de los vientos con datos de ACUMAR (**Figura 4.2**), donde los vientos del Oeste en ningún caso tienen mayor importancia.

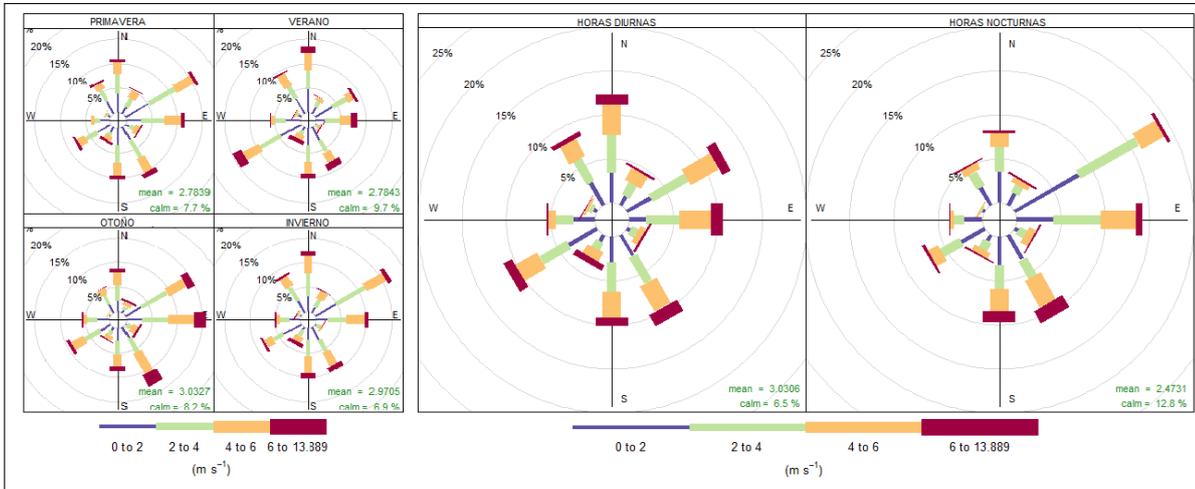


Figura 5.2: Rosas de los vientos estacionales y diferenciadas en horas diurnas y nocturnas con datos del SMN en Campo de Mayo

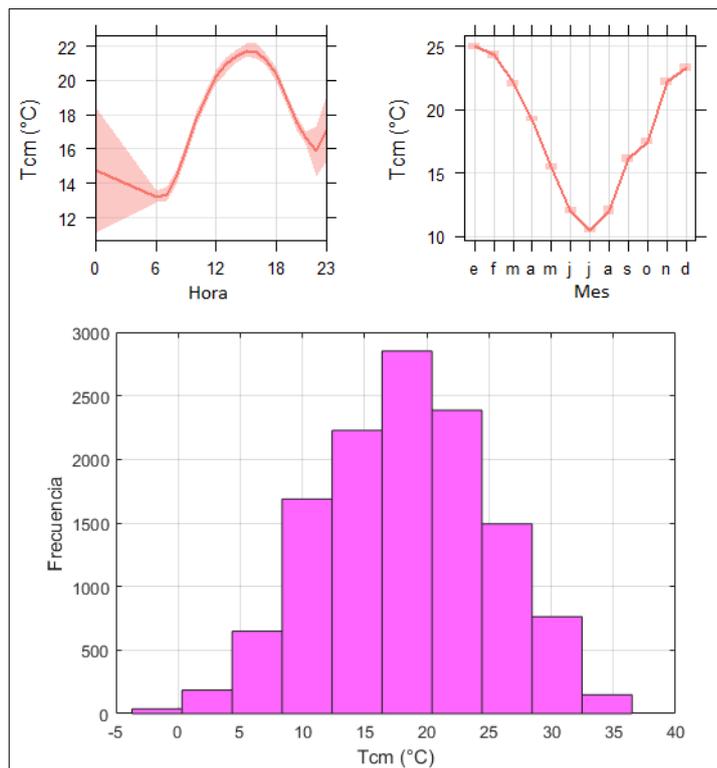


Figura 5.3: Histograma y tendencias mensual y horaria de datos de temperatura del SMN en Campo de Mayo

En cuanto a la temperatura, se observa en la **Figura 5.3** que posee una distribución normal centrada entre 15 °C y 20 °C con una marcha diaria y anual dentro de lo esperable.

La humedad relativa medida en Campo de Mayo y la altura de capa de mezcla del NOAA centrada en el sitio UNGS (consultar **Anexo 8.2.1**), tienen un comportamiento muy similar al mostrado en las **Figuras 4.3 y 4.4** con datos para CABA en el **Capítulo 4**.

Las precipitaciones acumuladas obtenidas del NOAA para los días de mediciones con el fotómetro son de 0 mm para la mayoría de los casos (lo que es lógico pensando en que las mediciones se realizaron siempre en días con escasa cobertura nubosa), con algunos días con lluvias registradas de entre 1 mm y 3 mm y muy pocos con lluvias mayores a 5 mm.

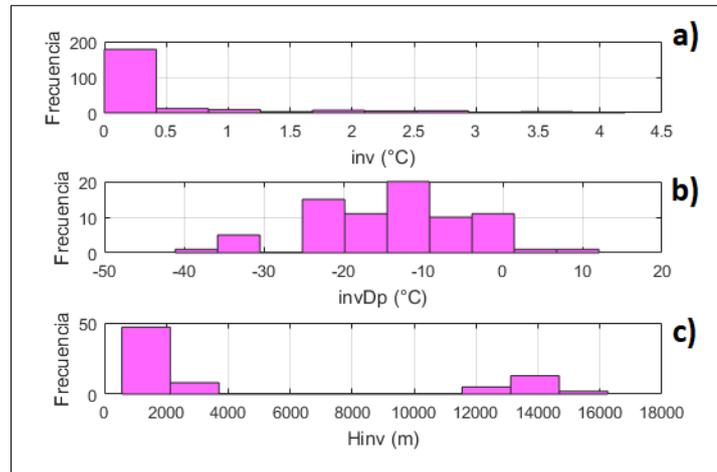


Figura 5.4: Histogramas de datos de inversión con sondeos del NOAA. (a) Diferencia de temperatura ambiente en la capa de inversión T_{inv} , (b) Diferencia de temperatura de rocío en la capa de inversión $invDp$, (c) Altura de la capa de inversión H_{inv}

La **Figura 5.4** muestra las distribuciones de valores de inversión calculados a partir de los sondeos del NOAA. T_{inv} , que es la diferencia de temperatura en la capa de inversión, tiene mayoría de valores igual a cero, pues son días en los cuales no se registraron capas de inversión, y luego presenta algunos valores de hasta 4 °C (**Figura 5.4(a)**). La disminución de la temperatura de rocío $invDp$ con la altura en las capas de inversión detectadas se encuentra mayoritariamente entre 10°C y 25°C (**Figura 5.4(b)**), y las alturas de base para las inversiones H_{inv} son mayoritariamente menores a 2000 m con algunos casos a gran altura entre 12000 m y 16000 m (**Figura 5.4(c)**). Las inversiones

tienen influencia en las capacidades dispersivas de la atmósfera, pero su contribución a la carga aerosólica es poco estudiada y depende de la locación. Su rol en la región se estudia más adelante, con los modelos SOM.

5.2.2. Mediciones fotométricas de bajo costo

Dentro del espectro de luz analizado, se espera que el valor medio de AOD disminuya con la longitud de onda, dado que los aerosoles típicos dispersan con mayor efectividad la luz en rangos cercanos al azul que al rojo. Esto no se cumple para el caso de la longitud de 619 nm en la **Figura 5.5**, medida con el fotómetro CALITOO. Esto puede deberse a una falla de calibración para ese canal en el fotómetro utilizado. Por este motivo, el coeficiente de Ångstrom fue calculado a partir de los datos de AOD de los canales azul (465 nm) y verde (540 nm).

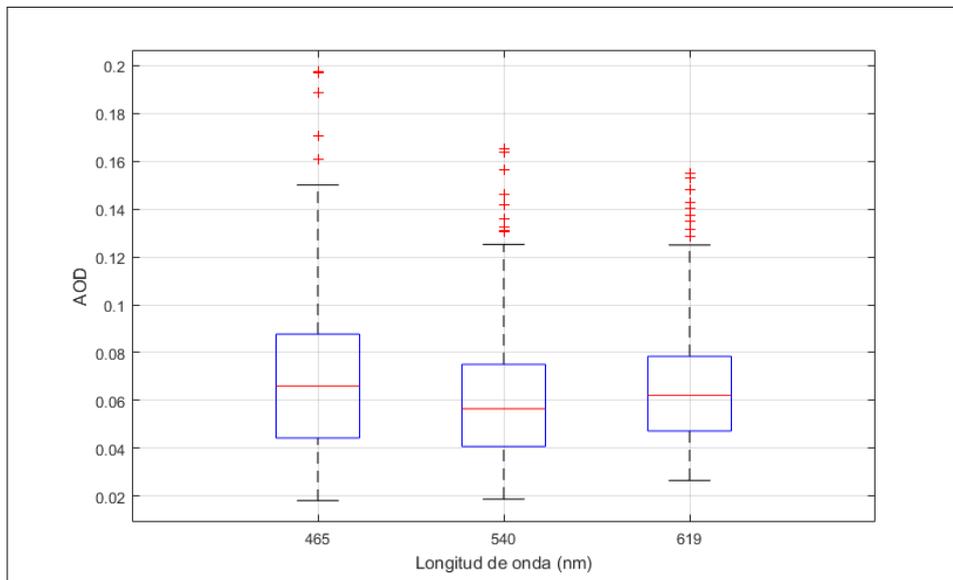


Figura 5.5: Gráfico de cajas con valores de AOD registrados para cada longitud de onda. El valor medio disminuye con la longitud de onda, con excepción del canal rojo (619 nm)

Las distribuciones de diferentes propiedades ópticas de los aerosoles pueden darnos información básica pero importante sobre la CA en la zona. La **Figura 5.6** muestra la gráfica del coeficiente de Ångstrom (α_{UNGS}) vs AOD medido en UNGS (AOD_{UNGS}), para determinar las clases de aerosoles según los datos registrados y la metodología descrita en la **Sección 3.2**. Se puede ver que las partículas de mayor tamaño corresponden principalmente a aerosoles de tipo “Marítimo” (M) y “Antártico” (A) y las de tamaño

medio a las especies continentales. Mientras que claramente los valores de AOD mayores a 0,1 caracterizan mayoritariamente a clases de mayor carga aerosólica, como “Continental Promedio” (CP). La distribución de clases muestra una mayor ocurrencia del tipo “Continental limpio” (CL) (alrededor de 50 %), seguido de “Continental promedio” (CP) (20 %) y “Antártico” (15 %). Hay algunos casos de “Continental contaminado” (CC) y “Marítimo”, y solo uno correspondiente a “Urbano / Industrial” (U/I) que se descartó por no ser estadísticamente significativo. Hay algunas estaciones del año en las que se producen quemas de biomasa en la región, pero parece no ser suficiente para que ese tipo de aerosoles sea la clase dominante (al menos no se detectó ningún caso de la clase “Quema de biomasa” (QB) con las mediciones). La clase “Continental Limpio” suele darse principalmente con partículas solubles en agua y minerales propias de zonas rurales. Esto se debe a que la UNGS se encuentra emplazada en una zona con amplios espacios verdes incluido el predio Campo de Mayo, de unas 8000 hectáreas de superficie. El siguiente grupo en importancia, “Continental Promedio”, se caracteriza con aerosoles de mezclas de hollín, polvo y compuestos solubles en agua. La intrusión de estos aerosoles en la zona se puede explicar a partir del flujo vehicular y la re-suspensión de partículas de polvo propias de calles no asfaltadas.

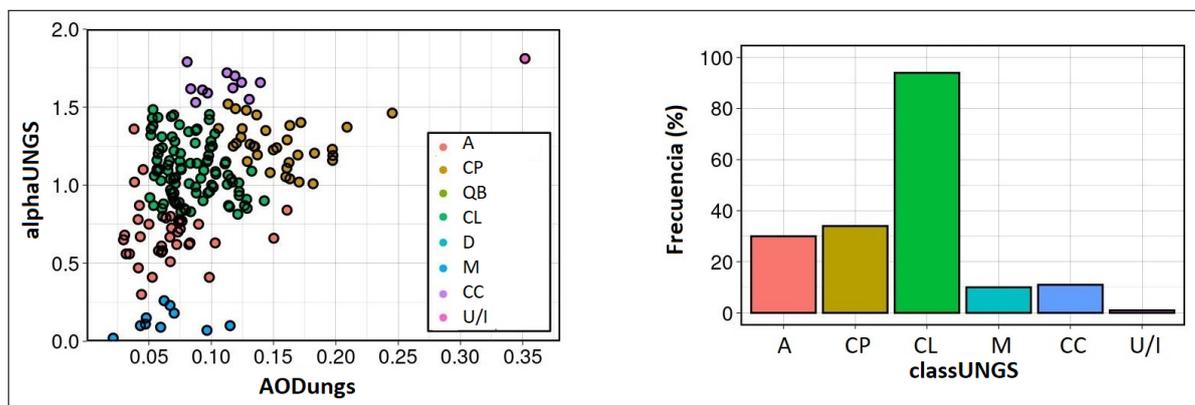


Figura 5.6: Clasificación de aerosoles con datos medidos en UNGS. A la izquierda, gráfico de α_{UNGS} vs AOD_{ungs} y los tipos de aerosoles diferenciados por colores. A la derecha, distribución de frecuencias por tipos de aerosoles. A: Antártico; CP: Continental Promedio; QB: Quema de Biomasa; CL: Continental Limpio; D: Desértico, M: Marítimo; CC: Continental Contaminado, U/I: Urbano/Industrial

Estos resultados indican que en la región se esperan valores bajos de AOD y condiciones de atmósfera limpia con respecto a los aerosoles, según criterios adoptados de diferentes antecedentes [93, 128, 203, 94]. Estudios previos realizados en la estación AE-

RONET más cercana (CEILAP BA) descartaron las clases “Antártica” y “Marítima”, marcándolas como clase “mixta” [93, 113]. Uno de los objetivos de este capítulo es explicar la presencia de esas clases a través de vínculos meteorológicos y retro-trayectorias de masas de aire.

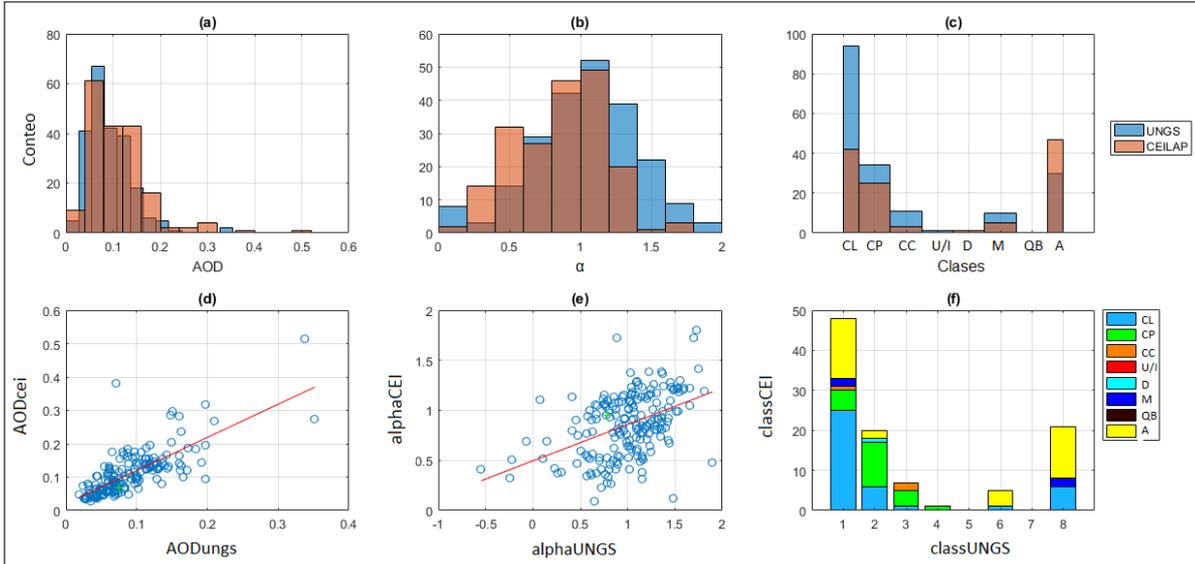


Figura 5.7: Comparación entre valores de los sitios UNGS y CEILAP: (a) histogramas superpuestos con valores de AOD, (b) histogramas superpuestos con valores de α (coeficiente de Ångstrom), (c) histogramas superpuestos con clases de aerosoles, (d) regresión lineal entre valores de AOD de ambos sitios, (e) regresión lineal entre los valores de α de ambos sitios, (f) diagrama de barras apiladas de ocurrencia de las clases en CEILAP para cada clase en UNGS

Los valores más frecuentes de coeficientes de Ångstrom (α) calculados en UNGS ($alphaUNGS$) se encuentran entre 0,7 y 1,3, correspondientes a una atmósfera limpia sin intrusión de aerosoles propios de eventos contaminantes de origen natural o antrópico [93, 128, 203, 94, 96, 129, 130]. La **Figura 5.7** muestra seis sub-gráficas que comparan los valores de AOD medidos en UNGS ($AODung$), los coeficientes de Ångstrom calculados ($alphaUNGS$) y la clasificación de aerosoles ($classUNGS$) con los datos equivalentes de la estación de AERONET CEILAP BA [34] ($AODcei$, $alphaCEI$ y $classCEI$). Las distribuciones son similares entre los datos de ambos sitios, con diferencias como un ligero corrimiento hacia valores de AOD más altos y valores más bajos de α (partículas más grandes) para CEILAP (**Figura 5.7 (a) y (b)**) que UNGS. La regresión lineal entre el AOD medido en CEILAP y en UNGS es significativa con un coeficiente de correlación alto ($R = 0,74$, dentro de un intervalo de confianza del 95 %), mientras que entre $alphaUNGS$ y $alphaCEI$ es menor ($R = 0,44$), como también puede verse en las

Figuras 5.7 (d) y (e). La **Figura 5.7 (f)** muestra la ocurrencia de las clases en CEILAP en función de las clases en UNGS, donde hay una similitud muy alta excepto para aquellas clases poco comunes durante el período de estudio como “Continental Contaminado”, “Urbano/Industrial” y “Marítimo”. Cabe señalar también que hay una mayor proporción de clase tipo “Antártico” en CEILAP, y una mayor proporción de clase tipo “Continental Limpio” en UNGS.

Estos resultados están dentro de las expectativas. La mayoría de datos en UNGS apuntan a una atmósfera limpia con baja carga de aerosoles, mientras que una menor cantidad, fuera del percentil 80, se corresponden a valores propios de una zona rural con intrusión de aerosoles típicos de actividades industriales. Las características de la zona de medición se condicen con estas observaciones, dado que los centros urbanos más importantes se encuentran aproximadamente a 2,5 km del campus de la UNGS y existen varias industrias aisladas a distancias menores. Los valores de AOD más altos y el mayor tamaño de partículas en CEILAP están vinculados a la mayor urbanización y flujo vehicular en ese sitio, aunque cabe aclarar que la fracción de tamaños de partículas que dominan ambos sitios parece ser la gruesa. La clase de aerosoles responde principalmente a procesos regionales y, por lo tanto, las similitudes entre los dos sitios son consistentes. La correlación significativa entre los valores de AOD registrados en ambos lugares posiblemente se deba principalmente a propiedades de la atmósfera superior que influyen en la concentración de aerosoles a escala regional.

Estacionalmente, en ambos sitios, los valores más bajos de AOD se registran durante el invierno y los valores más altos en verano (**Figura 5.8 (a) y (b)**). Se encontraron resultados similares en otros estudios para datos de varias estaciones pertenecientes a la red AERONET en Argentina [93, 203, 91]. Esto se debe a que, durante los meses cálidos, la radiación genera un mayor crecimiento de la capa mezclada convectiva en la que quedan atrapados los contaminantes [201]. Por otro lado, durante los meses cálidos es posible que ocurra una mayor quema de biomasa luego de las cosechas y, específicamente en la zona de la UNGS, quemadas en basurales, aunque no hay información de este tipo disponible en la zona como para aseverarlo. Por otro lado, parecería haber una tendencia estacional para α_{UNGS} , pero que no se observa con α_{CEI} (**Figura 5.8 (c) y (d)**).

Un ciclo diario con valores de AOD más altos durante las primeras horas de la mañana y las últimas horas de la tarde se puede ver para ambos sitios en la **figura 5.8 (e) y (f)**. El perfil horario para α_{UNGS} (**Figura 5.8 (g)**) muestra los valores más altos (tamaños de partícula más pequeño) registrados cerca del mediodía, coincidiendo con

valores bajos de AOD. Llamativamente, se observa una tendencia opuesta de $alphaCEI$ en la **figura 5.8(h)**. Estos resultados pueden estar relacionados con los efectos de la radiación solar, pero se necesitan más datos y análisis para confirmarlo. También pueden deberse a las horas pico de mayor actividad, pero no se encontraron patrones en relación a los días de la semana lo cual parece contradecir esa hipótesis.

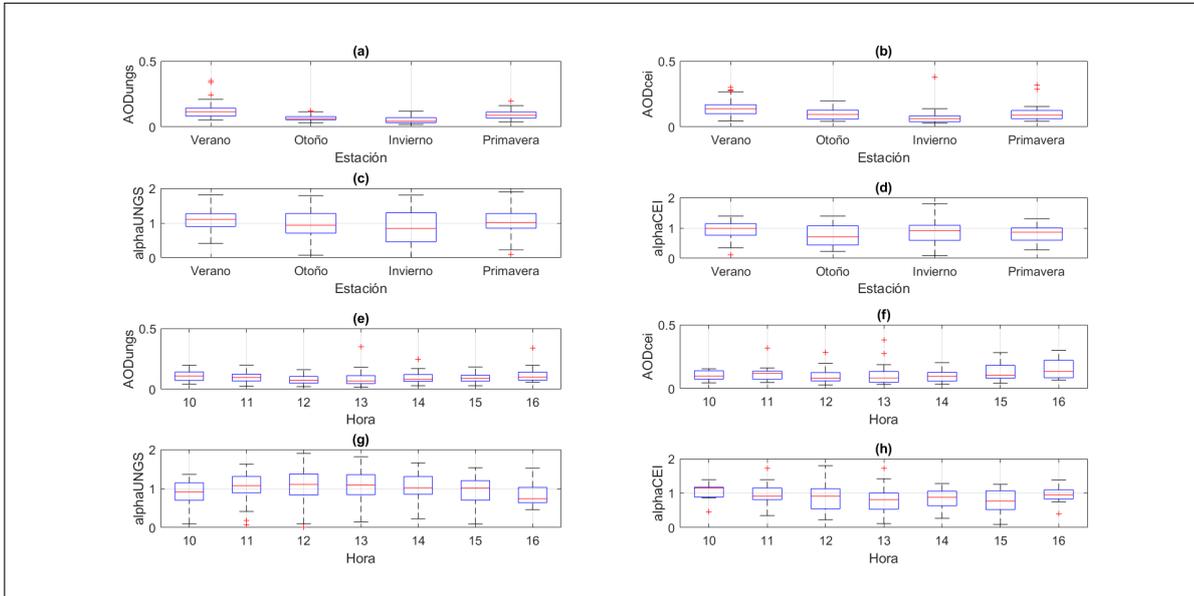


Figura 5.8: Tendencias medias para datos de UNGS y CEILAP: (a) estacional para AOD_{ungs} , (b) estacional para AOD_{cei} , (c) estacional para $alpha_{UNGS}$, (d) estacional para $alpha_{CEI}$, (e) horario para AOD_{ungs} , (f) horario para AOD_{cei} , (g) horario para $alpha_{UNGS}$, (h) horario para $alpha_{CEI}$

En otro orden de análisis, se muestran en mapas los resultados de las medias de AOD a 470 nm obtenidas del algoritmo MAIAC, que procesa imágenes del instrumento MODIS capturadas desde los satélites Terra y Aqua. La **Figura 5.9** muestra una visualización más regional con la estación de CEILAP incluida. Especialmente en verano, y un poco en otoño, pueden notarse las diferencias entre ambos sitios, con valores de AOD más altos cerca de CABA, en concordancia con lo visto en las distribuciones comparativas de la **Figura 5.7**. Así como se concluyó a partir de las mediciones con el fotómetro solar CALITOO en UNGS y datos de AERONET para CEILAP (**Figura 5.8**), los mayores valores de AOD se observan en verano con medias cercanas a 0,2 en CABA y 0,15 en la zona de la UNGS.

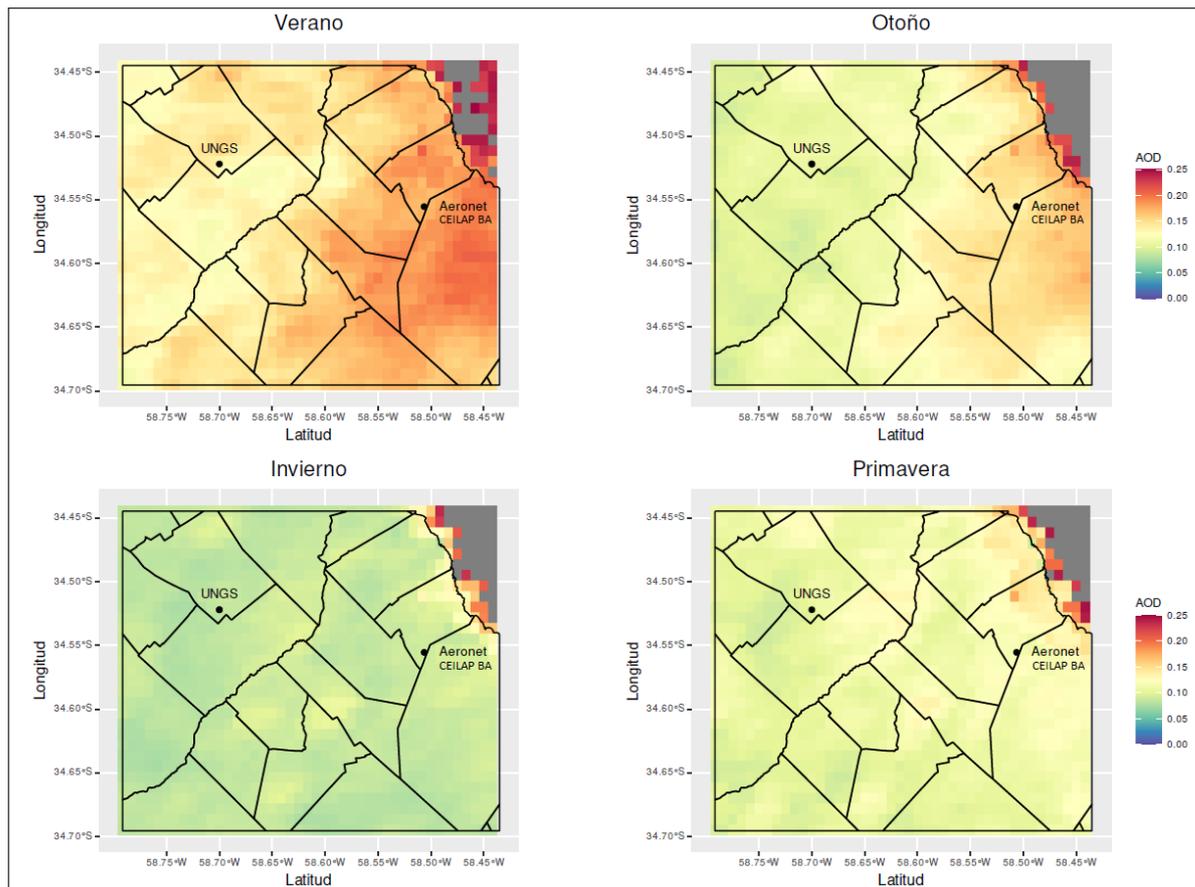


Figura 5.9: Mapa con píxeles coloreados en función del AOD medio a 470 nm para las estaciones del año adquirido del MAIAC. La visualización incluye a la estación de AERONET más cercana, CEILAP BA

El mapa de la **Figura 5.10** muestra la misma información que el anterior, pero con un acercamiento centrado en UNGS. Nuevamente, la estación con mayores valores medios permite observar mayores contrastes. Hay focos de mayor concentración en las ubicaciones de las ciudades más cercanas (San Miguel, J.C.Paz y Los Polvorines). Además puede observarse, también con mayor claridad durante el verano, el contraste con el gran espacio verde de Campo de Mayo al Sudeste de UNGS.

Un gran limitante que tienen los datos satelitales usados es que necesitan ser calibrados y validados con mediciones de AERONET para ser considerados como absolutos. Y en el conurbano, así como en gran parte del país, no hay estaciones de AERONET próximas que puedan ser representativas de los píxeles a validar. De todas formas, estos mapas permiten tener una visión más general y sacar conclusiones a nivel relativo.

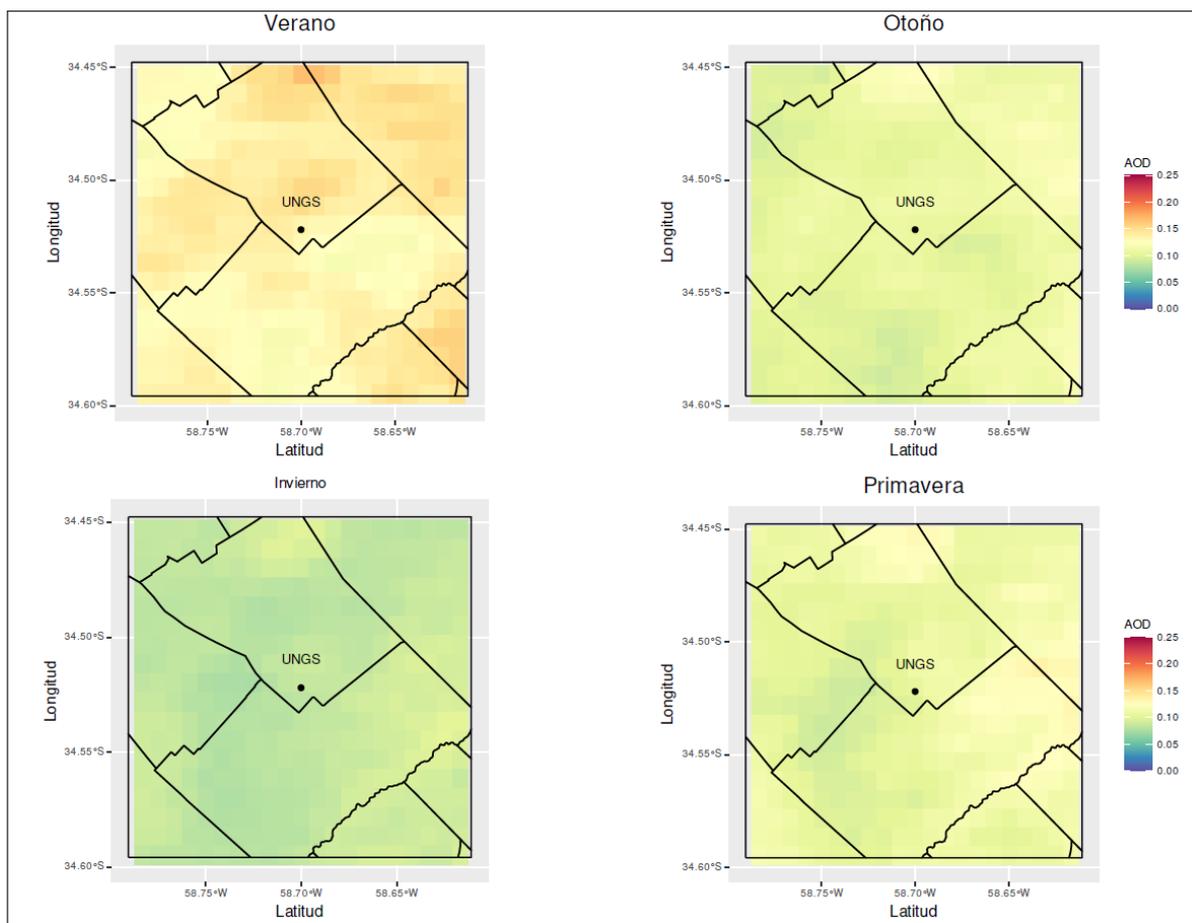


Figura 5.10: Mapa con píxeles coloreados en función del AOD medio a 470 nm para las estaciones del año adquirido del MAIAC. La visualización se centra en la UNGS mostrando las divisiones de partidos y los focos en las ciudades más cercanas

La **Figura 5.11** muestra proporciones similares en porcentajes de clases de aerosoles para UNGS (a) y CEILAP (b) por estación del año. La mayor proporción de “Continental Limpio” se encuentra en primavera, mientras que la clase “Continental Promedio” es más frecuente en verano, la clase “Antártico” es proporcionalmente mayor en otoño e invierno así como la clase “Marítimo”. La mayor ocurrencia de aerosoles de tipo “Continental Promedio” en verano tiene que ver con una atmósfera más contaminada debido al crecimiento de la capa mezclada, y la ocurrencia de clase “Antártico” está relacionada con la llegada de masas de aire frío originadas en la Antártica, que es más frecuente en los meses fríos.

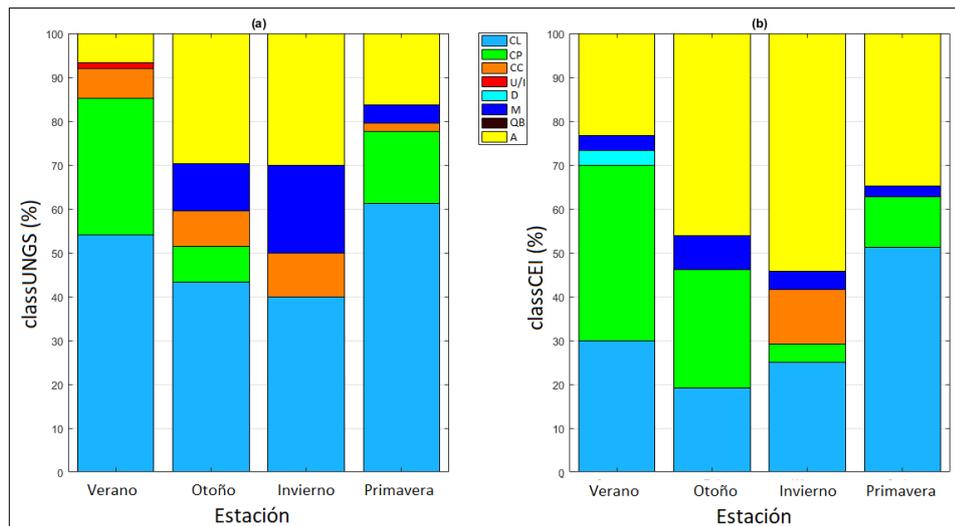


Figura 5.11: Gráficos de barras acumuladas de clases de aerosoles por estación del año: (a) *classUNGS*, (b) *classCEI*

5.2.3. Potenciales escenarios

Las comparaciones con los valores de CEILAP se implementaron durante la exploración de los datos medidos en UNGS, con el fin de contar con una referencia cercana. No obstante, los modelos SOM fueron implementados con las mediciones de bajo costo solamente, y la meteorología de la zona de estudio.

La serie de datos de los componentes del viento zonal y meridional calculadas con datos de viento del SMN en Campo de Mayo y de NOAA a 850 hPa están altamente correlacionadas ($R = 0,74$ con intervalo de confianza del 95 %), lo que indica que ambas alturas se ven afectadas por procesos similares en la atmósfera (**Figura 5.12**). Entonces, los datos del viento a 850 hPa se utilizaron como entrada para los modelos SOM.

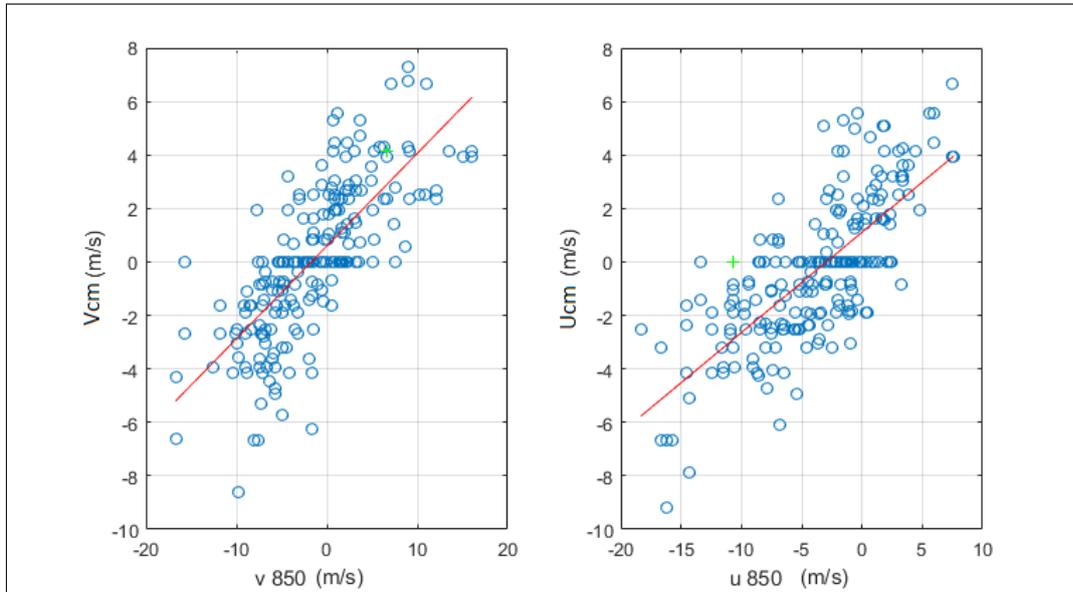


Figura 5.12: Regresiones lineales entre las componentes meridional y zonal calculadas con datos de viento del SMN y de NOAA a 850hPa

Las correlaciones lineales entre todas las variables de entrada sirven como un primer vistazo a la influencia de la meteorología en los valores de AOD y α . La **Figura 5.13** muestra que hay correlaciones débiles pero significativas entre la meteorología y las propiedades ópticas de los aerosoles en UNGS, aunque también pueden existir correlaciones no lineales más fuertes. Se puede ver que el AOD tiene una relación directa con $BLHungs$, $invDp$ y Tcm , y una inversa con $f(HR)$, $Hinv$, inv , Pcm y $ws850$. Mientras tanto, $alphaUNGS$ muestra una relación directa general con Tcm e inversa con Pcm y $ws850$. Se retoman estas observaciones y se abordan las posibles interpretaciones más adelante con los resultados de los SOM.

Los coeficientes R de las correlaciones entre las variables meteorológicas muestran que algunas de ellas pueden proveer información similar a los modelos (**Figura 5.13**). Con el fin de reducir la dimensionalidad del espacio de entrada para los SOM, de las variables que tienen una correlación significativa entre ellas, se descartaron aquellas con menor correlación con $AODung$ s. Ese fue el caso entre Tcm y Pcm (Pcm descartado), inv e $invDp$ (inv descartado).

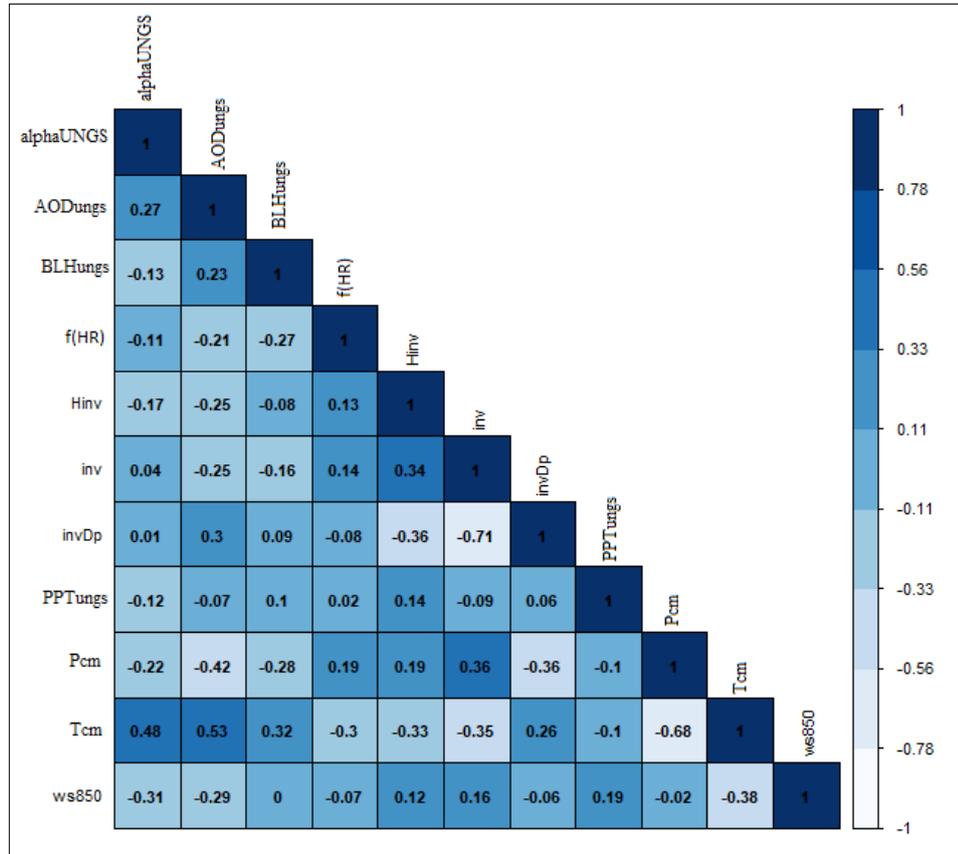


Figura 5.13: Matriz de correlaciones lineales entre las variables meteorológicas y los productos de aerosoles en UNGS. La intensidad del color indica el valor del coeficiente de correlación R (entre -1 y 1), el cual también se indica numéricamente en cada celda

Como base de datos de entrada, se elaboró finalmente una matriz con 226 filas (cada una coincidente en el tiempo con las mediciones del fotómetro en UNGS) y 11 columnas (variables elegidas). Entre todos los modelos probados con menos neuronas que patrones de entrada, el denominado SOM3 (mapa de 10x15, 150 neuronas en total) mostró el menor error topográfico (t_e), por lo que se utilizó esta configuración (**Figura 5.14 (a)**). El modelo se entrenó con un valor de radio de vecindad que va de 3 a 1 para la fase de lote e igual a 1 para la fase de ajuste fino. La distorsión calculada para cada neurona en el mapa SOM3 es alta en neuronas donde la influencia de *PPTungS* es preponderante, probablemente debido a que esa variable tiene valores de pesos elevados en los vectores correspondientes a esas neuronas y valores muy bajos en el resto (**Figura 5.14 (b)**).

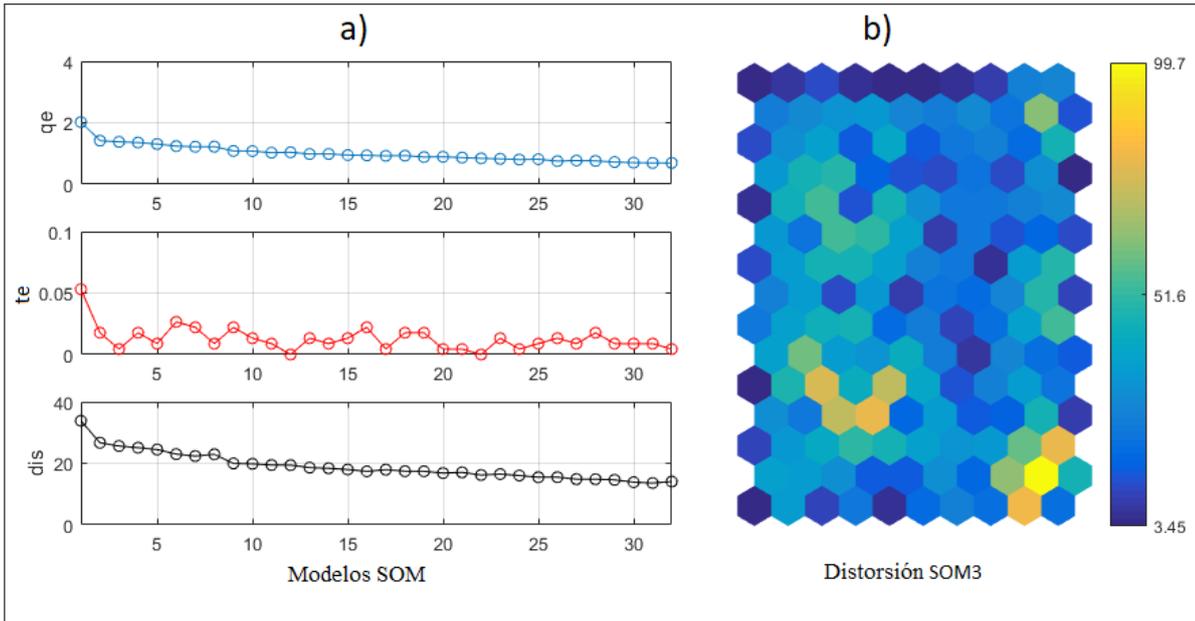


Figura 5.14: Indicadores de calidad del mapa final en los SOM diseñados con distintas cantidades de neuronas. (a) q_e , t_e y dis (distorsión) para cada SOM; (b) Distorsión del mapa final en el modelo seleccionado SOM3

El SOM3 entrenado devuelve una matriz de referencia, con los vectores de salida de cada neurona en el mapa. Cada uno de estos vectores de salida w_k tiene los valores de los pesos proyectados de cada variable en la capa de Kohonen del espacio de entrada. En la **Figura 5.15** se pueden observar sub-tramas con las neuronas del mapa coloreadas para cada variable de acuerdo con su peso en la matriz de salida. También se muestra la matriz de distancias medias entre neuronas vecinas (U-matrix) y se pueden observar conglomerados, como el que tiene las mayores distancias en la parte inferior derecha del mapa que es donde se colocaron los pesos más altos de *PPTungs*. Esto significa que el modelo asignó esas neuronas a la mayoría de los valores altos de precipitación.

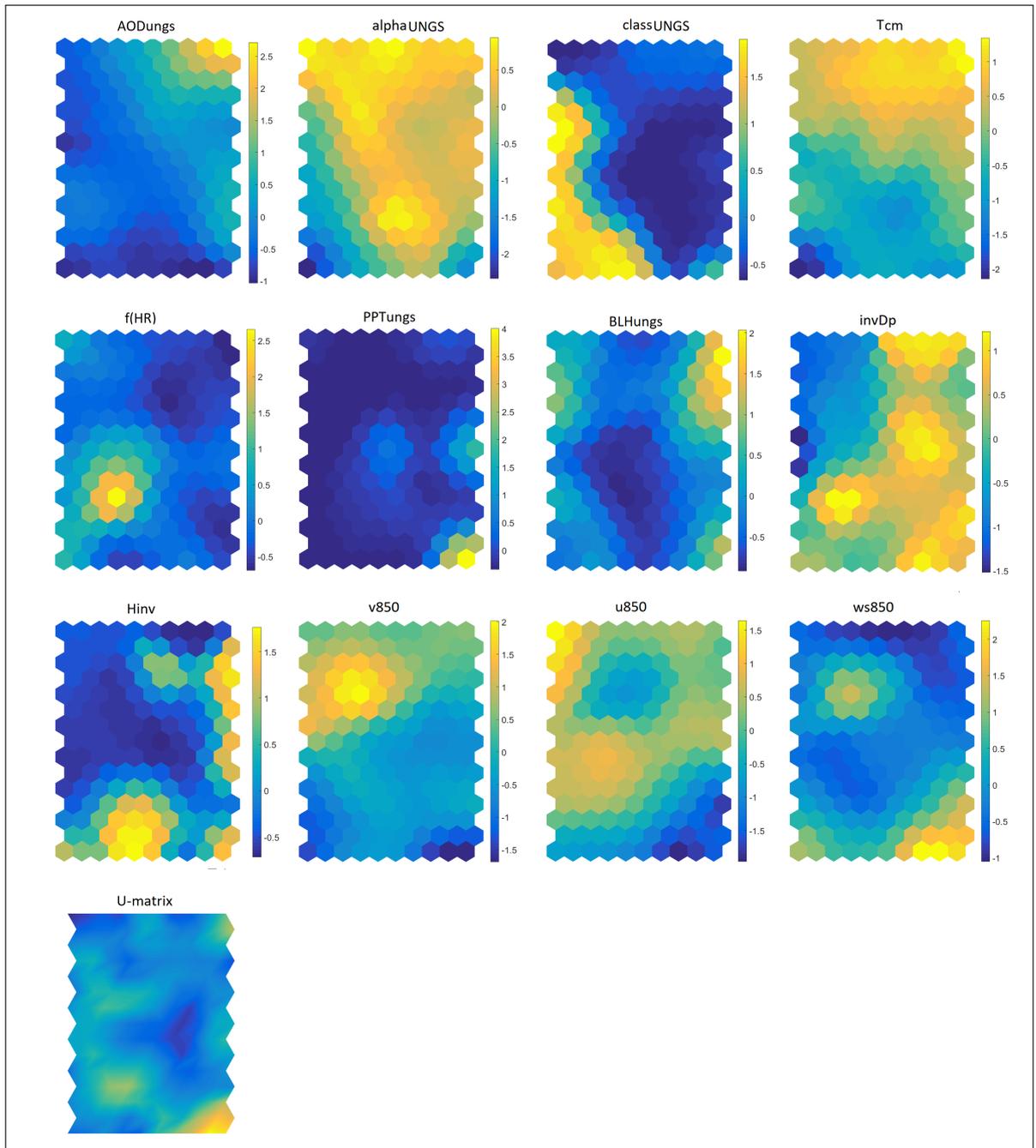


Figura 5.15: Representación de mapa y U-matrix de SOM3 para cada variable con sus pesos proyectados en las neuronas

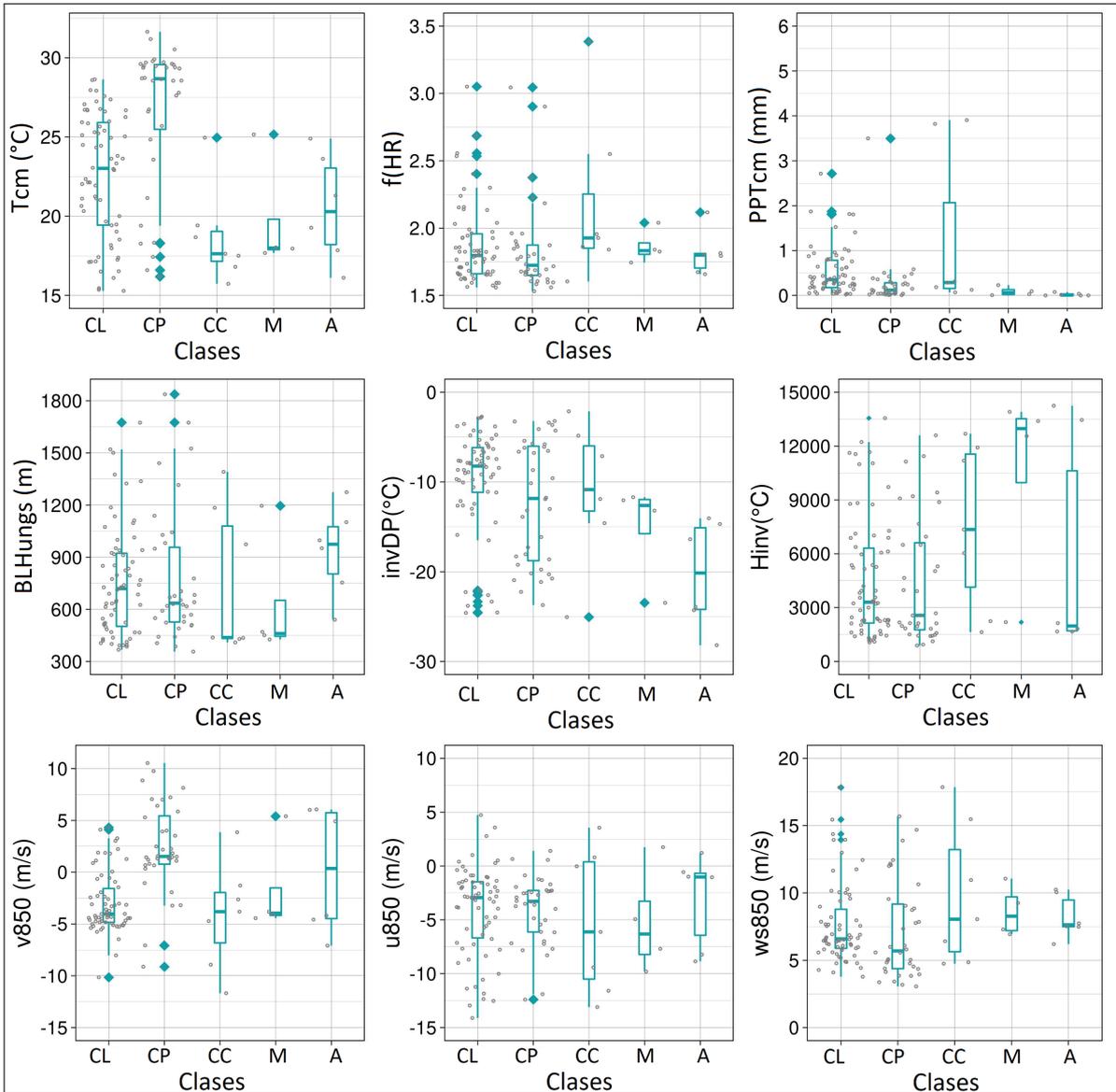


Figura 5.16: Diagramas de cajas con los pesos de las variables meteorológicas en la salida de SOM3 en función de la clase de aerosoles

La **Figura 5.16** muestra diagramas de caja con los valores de pesos en SOM3 de las variables meteorológicas para cada clase de aerosoles, con lo que se pueden extraer algunas características generales. Respecto a las clases continentales, el “Continental Limpio” se presenta generalmente con valores altos de PPT_{ung} , lo que es razonable debido a que una atmósfera limpia está relacionada con las lluvias debido al efecto de “barrido” que tienen sobre los aerosoles, e $invDp$ que implica presencia de fuertes capas de inversión. La clase “Continental Promedio” es consistente con una atmósfera ligeramente contaminada, y los resultados indican que coincide con valores altos de T_{cm}

(relacionados generalmente con una mayor carga de aerosoles [107, 109]) y valores bajos de ws_{850} , lo que puede estar asociado a débiles condiciones de dispersión. Los casos de “Continental Contaminado” se corresponden principalmente con altas velocidades de viento proveniente de los sectores Norte y Oeste (donde se ubican las industrias y el conglomerado urbano más poblado en cercanías a la UNGS, que es San Miguel), bajas temperaturas y una capa límite poco profunda. Los mapas mostrados en la **Figura 5.10** también mostraban mayores AOD en esas direcciones respecto a la UNGS.

Además, la ocurrencia de esta clase coincide con trayectorias de masas de aire con largas distancias dadas sobre el continente, como muestra la **Figura 5.17 (a)**.

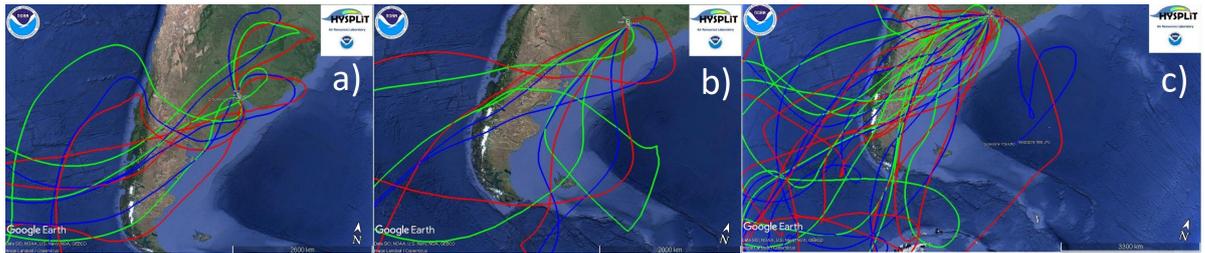


Figura 5.17: Vistas del Google Earth con una capa geo-referencial de trayectorias de masas de aire inversas de 7 días desde UNGS computadas con el modelo HYSPLIT de NOAA. El tiempo final de las simulaciones se eligió en base a la ocurrencia de las clases de aerosoles que mostraron mayor influencia regional: (a) Continental Promedio, (b) Marítimo, (C) Antártico. Las trayectorias en verde comienzan a los 2500 m sobre el nivel del mar, las azules a 1500 m y las rojas a 500 m

Valores bajos de T_{cm} , niveles medios de PPT_{ung} y valores altos de H_{inv} acompañan a la clase de aerosoles “Marítimo”. Las partículas de tipo marítimo parecen ser transportadas por masas de aire frío que atraviesan el Océano Atlántico con el viento proveniente del sector Suroeste (según las trayectorias HYSPLIT en la **Figura 5.17 (b)**), provocando algunas lluvias y la formación de una capa de inversión a gran altura. La meteorología a lo largo de las trayectorias de masas de aire muestra bajas temperaturas del aire (masas de aire frío), un aumento de la humedad para las trayectorias de 500 m cuando las masas viajan por encima del océano, y valles en la altura de la capa mezclada que marcan la transición océano-continente (un ejemplo puede verse en la **Figura 5.18**). Estas características son comunes durante el invierno, cuando se puede encontrar una mayor proporción de aerosoles de clase “Marítimo” (**Figura 5.11**).

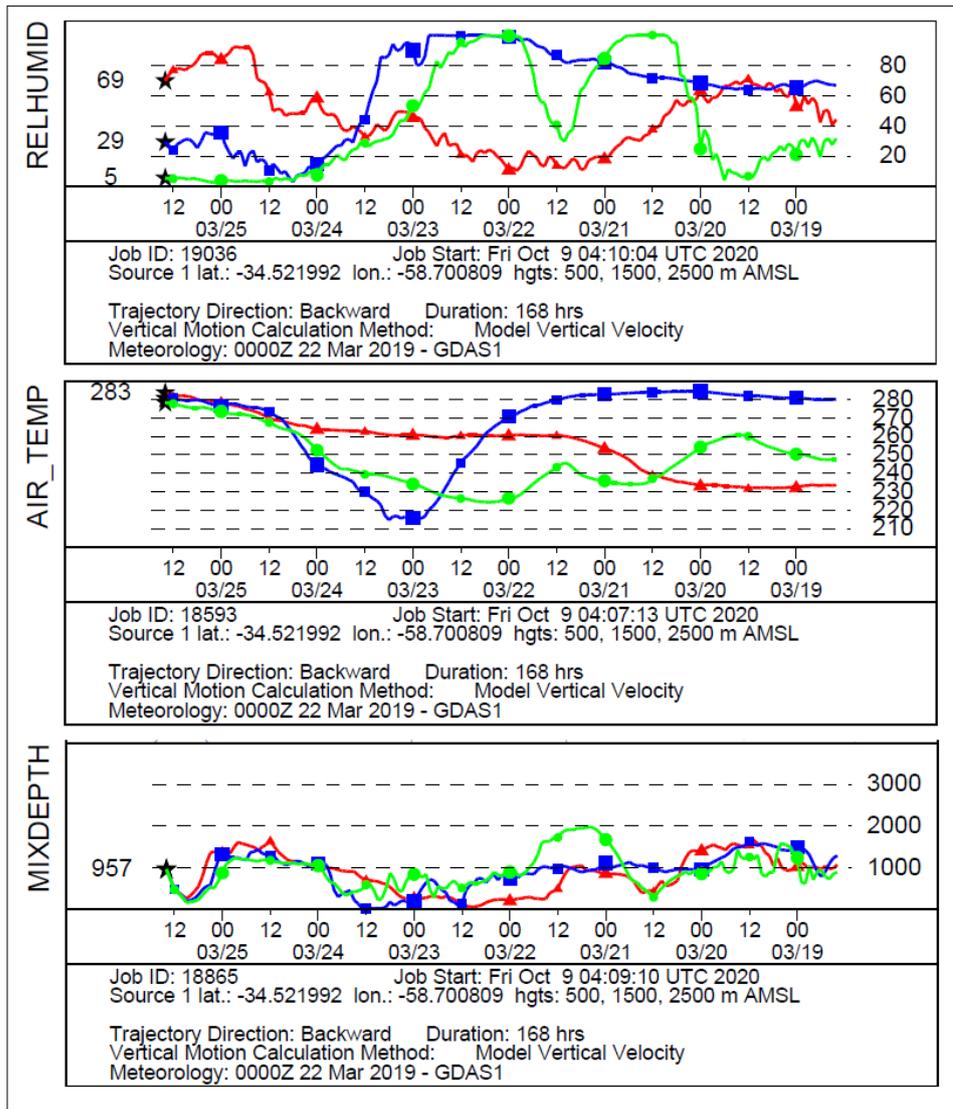


Figura 5.18: Evolución de la temperatura, humedad relativa y altura de capa mezclada para las trayectorias coincidentes con la clasificación de aerosoles tipo “Marítimo” para el día 22 de marzo de 2019, comenzando a las 0 h UTC. Imágenes descargadas de NOAA para las simulaciones de trayectorias realizadas. Las trazas verdes son datos a 2500 m de altura, las azules a 1500 m y las rojas a 500 m

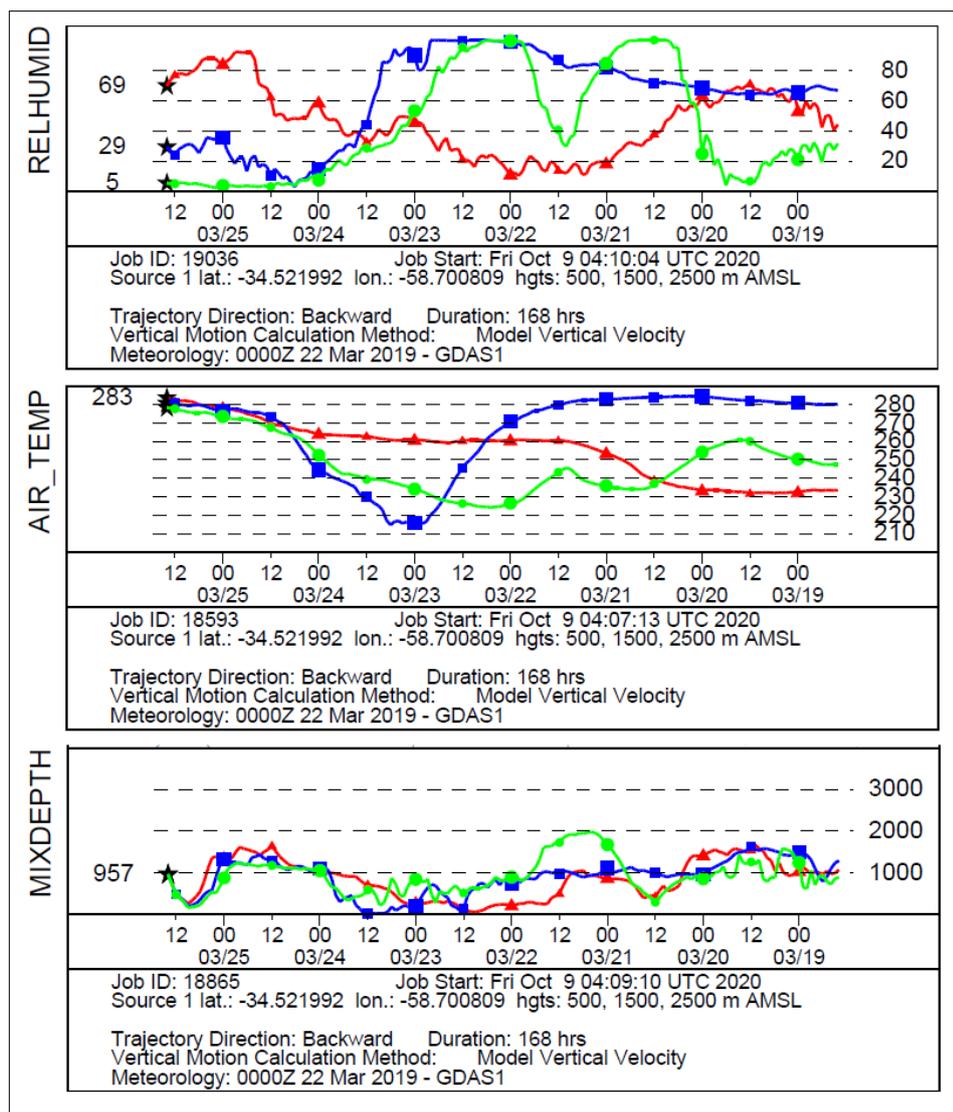


Figura 5.19: Evolución de la temperatura, humedad relativa y altura de capa mezclada para las trayectorias coincidentes con la clasificación de aerosoles tipo “Antártico” para el día 15 de abril de 2019, comenzando a las 0 h UTC. Imágenes descargadas de NOAA para las simulaciones de trayectorias realizadas. Las trazas verdes son datos a 2500 m de altura, las azules a 1500 m y las rojas a 500 m

Los aerosoles de tipo “Antártico” pueden ser transportados al sitio de medición por masas de aire polares. Los resultados indican que esta clase se encuentra principalmente con valores altos de BLH_{ungs} , valores bajos de $f(HR)$ y H_{inv} y valores absolutos altos de $invDp$, sin precipitaciones registradas (Figura 5.16). Esto parece coincidir con masas de aire de origen antártico que ingresan al continente atravesando la cordillera de los Andes (Figura 5.17 (c)), donde pierden su contenido de humedad (evidenciado parti-

cularmente en la trayectoria de mayor altura). La meteorología de la trayectoria muestra un aumento en la temperatura y una caída en la humedad relativa para las masas de aire después de cruzar los Andes (un ejemplo puede verse en **Figura 5.19**). Cuando estas masas llegan al sitio de medición, principalmente durante el otoño y la primavera según los registros, con una capa límite atmosférica profunda, generan una capa de inversión a bajas alturas aumentando la diferencia del punto de rocío sobre ella. Cabe aclarar que aunque el invierno presenta una proporción mayor de clase “Antártico” en comparación con otras estaciones (**Figura 5.11**), su ocurrencia es alta en número de casos durante el otoño y la primavera.

En el mapa final del modelo SOM3 se identificaron algunas regiones de interés, donde se observa una influencia de algunas variables en AOD_{Dungs} y $alphaUNGS$. Es decir, neuronas en el mapa que tienen una distancia apreciable de sus vecinas para las mismas variables. La **Figura 5.20** muestra los pesos desnormalizados de AOD dentro de cada neurona en el mapa de SOM3, coloreados de acuerdo con los pesos de las variables meteorológicas seleccionadas. El AOD se muestra como un gráfico de dos barras, la barra izquierda es el peso máximo desnormalizado en el mapa (para propósitos de contraste) y la derecha es el peso desnormalizado real correspondiente a cada neurona. La selección de agrupamientos se realizó en base a aquellas variables con pesos contrastantes en las mismas neuronas del mapa que AOD_{Dungs} . Se marcaron cuatro regiones de neuronas con elipses de diferentes colores, que fueron llamados “escenarios”.

El “ $AOD_{escenario1}$ ”, de cuatro neuronas marcadas con una elipse negra en la **Figura 5.20**, tiene valores altos de AOD (0,18 - 0,19) junto con los valores más altos de T_{cm} (29,5° C - 33,7° C) y BLH_{Dungs} (1040 m - 2120 m) y los valores más bajos de $f(HR)$ (correspondientes a humedades relativas entre 30 % y 40 %). Este escenario puede describirse como condiciones de clima cálido y seco, con una fuerte generación de turbulencia y empuje térmico en una atmósfera inestable, y un intenso crecimiento de la capa límite atmosférica. Estas características meteorológicas tienden a ocurrir principalmente en verano y son consistentes con los hallazgos detallados en las correlaciones de la **Figura 5.13**. Los efectos de la temperatura sobre los valores de AOD no son simples de interpretar, pero se puede afirmar que una temperatura más alta favorece la generación y el crecimiento de aerosoles secundarios, además de que un aumento de la temperatura implica un crecimiento en los procesos convectivos y en la profundidad de la capa límite atmosférica, lo que aumenta la carga de aerosoles y facilita la inclusión de más aerosoles naturales y antropogénicos en la columna atmosférica [107, 204, 205]. La **Figura 5.13** mostró una correlación lineal negativa entre AOD_{Dungs} y $f(HR)$, el cual

es un resultado difícil de explicar por sí mismo, debido a que se puede esperar que a mayor efecto higroscópico mayor sea el tamaño de las partículas y el correspondiente valor de AOD. Pero en este escenario, podemos ver que los valores altos de AOD_{Dungs} están relacionados con valores bajos de $f(HR)$, ya que las condiciones meteorológicas en verano con respecto a T_{cm} y BLH_{Dungs} juegan un papel más importante para AOD que $f(HR)$. Las condiciones de “ $AOD_{escenarioI}$ ” son consistentes con la presencia de aerosoles de tipo “Continental Promedio” (Figuras 5.11 y 5.16).

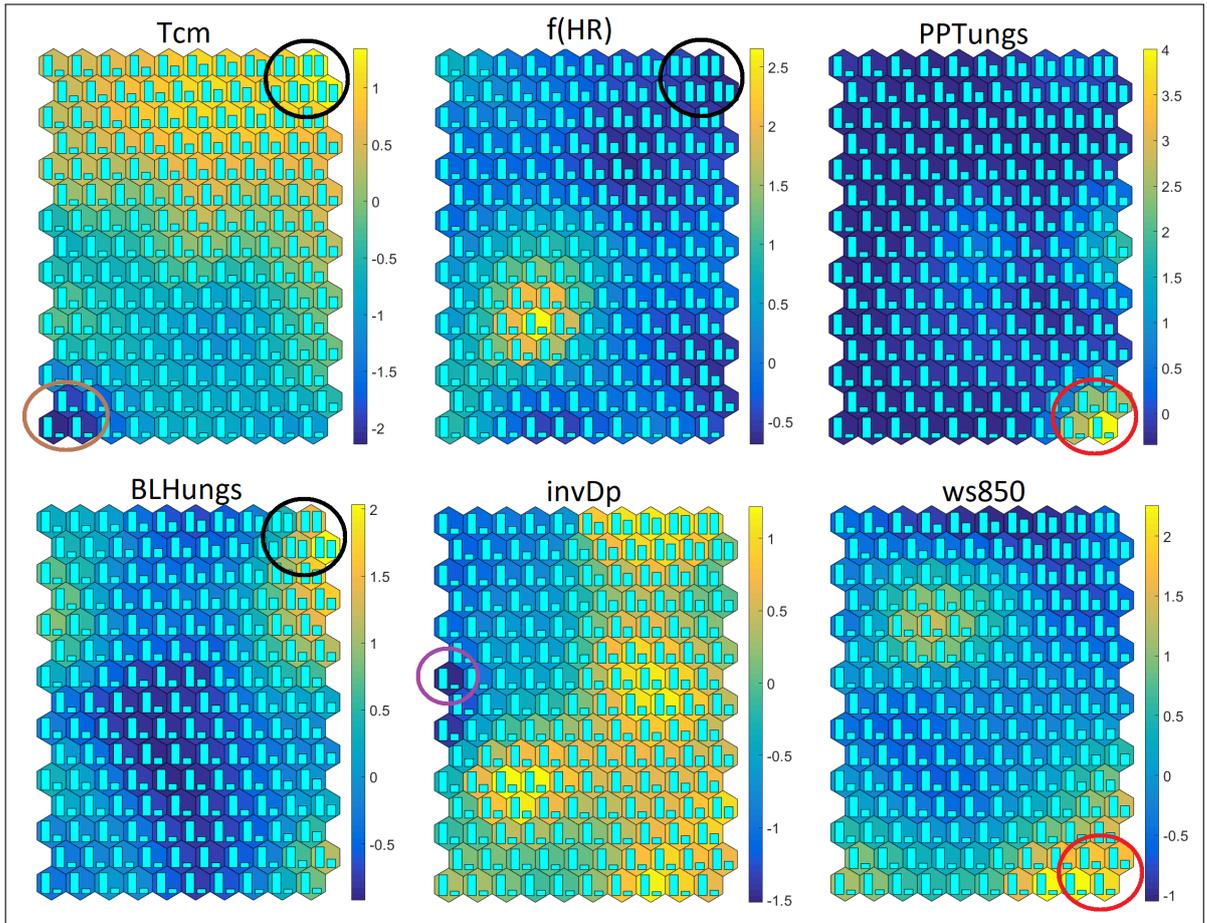


Figura 5.20: Representación de variables meteorológicas seleccionadas con sus pesos proyectados sobre las neuronas coloreadas del mapa de SOM3. Dentro de cada neurona hay una gráfica de dos barras con información de AOD, la barra izquierda corresponde al peso máximo desnormalizado de AOD_{Dungs} y la derecha al peso desnormalizado de AOD_{Dungs} en esa neurona. Hay cuatro regiones marcadas con elipses de diferentes colores que representan diferentes escenarios de AOD

Los valores de AOD más bajos (cerca de 0,03) están representados por “ $AOD_{escenarioI}$ ”, con cuatro neuronas en el mapa marcadas con una elipse roja. Este escenario corresponde a los valores más altos de $PPTungs$ (un promedio entre 3 mm y 6 mm de precipitación acumulada) y altas velocidades del viento (entre $15,4 \text{ ms}^{-1}$ y $17,9 \text{ ms}^{-1}$). Se puede encontrar evidencia contrastante en la literatura sobre la relación entre la abundancia de aerosoles y la cantidad de lluvia [107], pero este “escenario” muestra que $PPTungs$ tiene un efecto de remoción de aerosoles atmosféricos en el Centro-Norte del AMBA según los datos recopilados. Una disminución de los valores de AOD con la velocidad del viento también se ha observado en los resultados de las correlaciones (**Figura 5.13**), y ha sido indicada por varios autores en otros lugares del mundo [107, 109], dado que el viento tiene una influencia en los fenómenos de transporte y dispersión de aerosoles. En este escenario se esperan aerosoles de tipo “Continental Limpio” (**Figura 5.16**).

El “ $AOD_{escenarioIII}$ ”, de tres neuronas marcadas con una elipse marrón, tiene valores bajos de AOD y coinciden con los valores más bajos de Tcm . Esto es consistente con los hallazgos mostrados en la **Figura 5.8**, donde se observan valores bajos de AOD en invierno. Esto podría explicarse a través de la débil turbulencia térmica y la capa límite poco profunda que se dan durante los meses fríos.

Una última observación sobre la **Figura 5.20**, el “ $AOD_{escenarioIV}$ ” está marcado con una elipse púrpura y muestra valores de AOD muy bajos junto con los valores relativos más altos para $invDp$. Este escenario está relacionado con el efecto observado en los resultados de las correlaciones lineales (**Figura 5.13**), y generalmente se atribuye a una fuerte capa de inversión que se opone a la mezcla vertical y confina los aerosoles cerca de la superficie terrestre [206, 207]. Este escenario está relacionado con las características encontradas para el tipo de aerosoles “Continental Limpio” (**Figura 5.16**).

La **Figura 5.21** se desarrolló con la misma metodología que la **Figura 5.20**, pero con pesos desnormalizados de $alphaUNGS$ en el mapa. Se detectaron dos escenarios principales, donde Tcm y $ws850$ mostraron un contraste entre neuronas equivalentes con $alphaUNGS$.

El “ $alpha_{escenarioI}$ ”, identificado con una elipse negra en tres neuronas del mapa, muestra los valores más bajos de $alphaUNGS$ (cerca de 0) en coincidencia con los valores más bajos de Tcm (10° C a 11° C). En otras palabras, parece que se pueden encontrar partículas más grandes en días fríos. Además, una temperatura superficial más baja puede alcanzar con mayor frecuencia el nivel del punto de rocío aumentando el contenido de humedad en el aire, lo que lleva a un crecimiento higroscópico en los aerosoles.

Estos hallazgos son consistentes con la correlación directa entre α_{UNGS} y T_{cm} mostrada por la **Figura 5.13**, y parece ser más probable que se de en invierno junto con “ $AOD_{escenarioIII}$ ”.

El “ $\alpha_{escenarioII}$ ”, identificado con una elipse roja que encierra tres neuronas, tiene los valores más altos de α_{UNGS} y $ws850$. Los valores altos de velocidad del viento pueden transportar y resuspender pequeñas partículas en el aire. Además, este escenario parece coincidir con “ $AOD_{escenarioII}$ ”, en el cual los valores más bajos de AOD se relacionan con las precipitaciones.

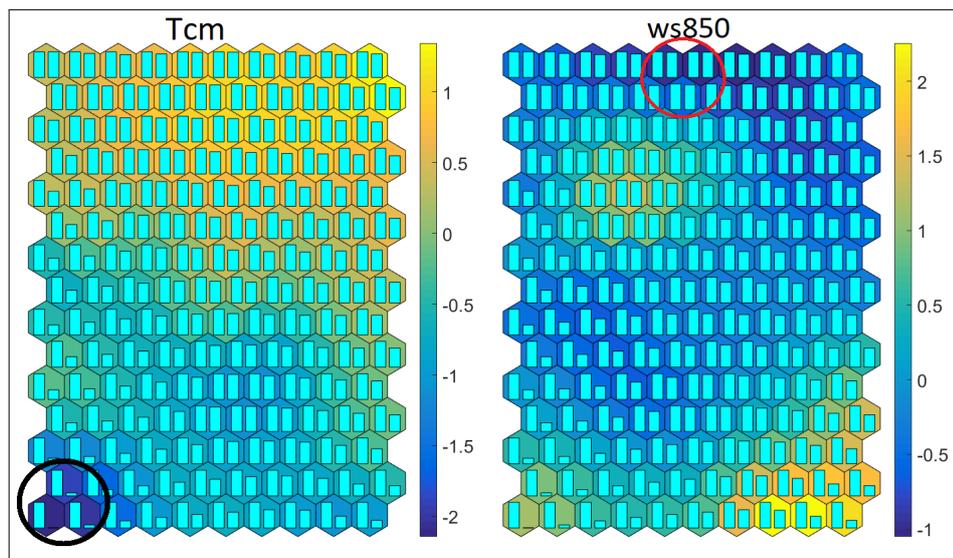


Figura 5.21: Representación de variables meteorológicas seleccionadas con sus pesos proyectados sobre neuronas coloreadas en el mapa de SOM3. Dentro de cada neurona hay una gráfica de dos barras con información del coeficiente de Ångstrom α , la barra izquierda corresponde al peso máximo desnormalizado de α_{UNGS} y la derecha al peso desnormalizado de α_{UNGS} en esa neurona. Hay dos regiones marcadas con elipses de diferentes colores que representan diferentes escenarios de α_{UNGS}

5.3. Conclusiones preliminares

Las mediciones de propiedades ópticas de aerosoles con el fotómetro solar de bajo costo en el sitio de la UNGS, se realizaron entre noviembre de 2018 y febrero de 2020 y son un primer antecedente en la región Centro-Norte del conurbano bonaerense. Este trabajo propone una integración a partir de ANN tipo SOM, para analizar y buscar

vínculos entre los datos medidos y las variables meteorológicas obtenidas del SMN y de NOAA READY.

La primera exploración de datos muestra valores bajos de AOD y una clase de aerosoles preponderante en la región de tipo “Continental Limpio”, seguida por la clase “Continental Promedio”, con tamaños de partículas generalmente grandes (α cercano a 1). No se han encontrado casos de clases de elevada contaminación. Además, se encontraron los valores más bajos de AOD principalmente durante el invierno y valores más altos durante el verano y la primavera, como lo han reportado otros estudios para diferentes lugares de Argentina. Los datos satelitales muestran mayor carga de aerosoles en las zonas altamente urbanizadas en torno a la UNGS.

Las regresiones lineales entre AOD y α registrados en UNGS, y las variables meteorológicas, muestran algunas relaciones generales. Para los datos analizados, se encontró que el AOD generalmente aumenta con la temperatura, la altura de la capa límite y la fuerza de la capa de inversión, mientras que disminuye con la humedad relativa, la presión y la velocidad del viento. Por otro lado, el tamaño de las partículas disminuye con la temperatura y aumenta con la humedad relativa, la presión y la velocidad del viento. Pero las regresiones lineales nos dan información limitada y las relaciones no lineales permanecen ocultas. Es por ello que se implementó un enfoque de clasificación de aprendizaje automático no supervisado con los datos (SOM), lo que permitió encontrar diferentes condiciones bajo las cuales suelen ocurrir las clases de aerosoles más frecuentes, y algunos escenarios de aerosoles típicos de la región. Estos resultados se complementaron con las retro-trayectorias de las masas de aire de HYSPLIT, que ayudaron a explicar algunos vínculos entre las condiciones meteorológicas y los tipos de aerosoles.

Los principales resultados indican que las condiciones de radiación solar estacional tienen un efecto determinante sobre el clima y el comportamiento de los aerosoles. Durante el verano, con condiciones secas, alta temperatura y una capa límite profunda, encontramos los valores más altos de AOD y el tipo de aerosoles “Continental Promedio” (“ $AOD_{escenarioI}$ ”). En invierno, las bajas temperaturas y una capa de mezcla poco profunda van acompañadas de valores bajos de AOD, partículas de gran tamaño y la clase “Continental Limpio” (“ $AOD_{escenarioI}$ ” y “ $alpha_{escenarioI}$ ”).

Se encontraron y clasificaron otros escenarios típicos. Uno de ellos mostró que las condiciones lluviosas con alta velocidad del viento implican valores bajos de AOD y tamaños de partículas pequeños (“ $AOD_{escenarioII}$ ” y “ $alpha_{escenarioII}$ ”). Por otro lado, las capas de inversión intensa se vincularon con valores bajos de AOD y la clase “Continental

Limpio” ($AOD_{escenarioIV}$).

Los casos de mayor contaminación se han registrado con vientos de los sectores Norte y Oeste (desde los polos industriales y un conglomerado altamente urbanizado), y en presencia de una fuerte capa de inversión (que en este caso retiene a los contaminantes emitidos desde esos sectores). En particular, la clase “Continental Contaminado” corresponde a la mayoría de las trayectorias de masas de aire dadas por encima del continente con bajas temperaturas y una capa límite poco profunda.

A pesar de que las clases de aerosoles “Marítimo” y “Antártico” suelen tratarse como escenarios de aerosoles “mixtos” en los trabajos citados para Buenos Aires, resultan ser los tipos más frecuentes después de los continentales (limpio y promedio) según los datos y la clasificación implementada, además de encontrarse algunas características generales en las condiciones meteorológicas relacionadas con ellos. La clase “Marítima” se da principalmente durante el invierno con masas de aire que llegan del Océano Atlántico provocando algunas lluvias, ya que el aire gana humedad durante su trayectoria, y la formación de una capa de inversión a gran altura a partir de la línea litoral. Los aerosoles clasificados como “Antárticos” están conectados con masas de aire del continente antártico que aumentan su temperatura y pierden contenido de humedad después de cruzar la cordillera de los Andes, principalmente durante el otoño y la primavera, y suelen ir acompañadas de una fuerte capa de inversión en altitudes bajas y profundas. Estos resultados pueden ayudar a explorar con más detalle en el futuro la presencia de las clases “Marítima” y “Antártica” de aerosoles en Buenos Aires.

Los hallazgos mostrados en este capítulo muestran que las relaciones no lineales entre las variables meteorológicas y las propiedades ópticas de los aerosoles con mediciones de bajo costo se pueden interpretar e integrar a través de la implementación de SOM. Este enfoque proporciona más información de la que pueden proporcionar las regresiones lineales o los métodos estadísticos tradicionales. Además, el seguimiento de las trayectorias de las masas de aire con el modelo HYSPLIT proporcionó información útil sobre el origen de los aerosoles y el comportamiento de las variables meteorológicas durante esas trayectorias. Cabe mencionar que se aplicó un método de clasificación de aprendizaje automático que da valores promediados de variables en diferentes agrupamientos, y el modelo final no es una regresión que pueda dar valores absolutos para la predicción. No obstante, se dieron diferentes escenarios de propiedades ópticas de los aerosoles en función de las condiciones meteorológicas y las observaciones se corresponden con las características de la zona de estudio y representan una primera aproximación a la determinación de la CA en la región con un control y mediciones sistemáticas.

Las mediciones in situ son una herramienta valiosa para una región sin antecedentes en estudios de contaminación atmosférica, y esta Tesis ofrece un primer aporte de información sobre cantidades y tipos de aerosoles relacionados con las condiciones meteorológicas de la zona. Sin embargo, se necesitan más mediciones y análisis más detallados con otros parámetros meteorológicos para lograr una mejor comprensión del comportamiento de los aerosoles en el conurbano bonaerense, de manera estacional y diaria. Los datos limitados recopilados con el fotómetro solar CALITOO se pueden complementar con datos satelitales (como los mostrados con el algoritmo MAIAC) y estaciones cercanas de AERONET. El fotómetro de bajo costo está calibrado con la red AERONET, por lo que puede ser útil a la hora de validar datos satelitales en la zona. Esta línea de trabajo posee gran potencial para continuarse, y con mayores datos y análisis se podrán separar las contribuciones locales de las regionales en la carga de aerosoles del área y recopilar más información y evidencia sobre la presencia de diferentes clases de aerosoles.

6 Desarrollo e implementación de equipos de bajo costo

6.1. Objetivos e intereses

Se presentan los resultados de los diseños experimentales llevados a cabo para probar y evaluar el desempeño de diferentes sensores ambientales de bajo costo (LCS) en base a criterios específicos. Estos resultados sirvieron al objetivo de seleccionar los sensores más adecuados para el desarrollo del prototipo de mini-estación ECAUNGS. Se muestran también simulaciones y comparación con equipo de referencia calibrado para el LCS de PM seleccionado. Los resultados permitieron conocer las limitaciones del sensor y desarrollar un algoritmo de gestión de calidad de datos que incluye advertencias sobre los datos medidos.

Asimismo, se presentan los resultados de campañas de medición de PM con ECAUNGS y un equipo móvil en torno a la UNGS, con resolución intra-urbana en pos de la búsqueda de fuentes de aerosoles y representar las principales tendencias. Estos resultados se ponen en diálogo con las propiedades ópticas de aerosoles mostradas en el **Capítulo 5**.

Finalmente, se integran los métodos de bajo costo desarrollados e implementados en esta Tesis con un modelo ANN alimentado con los datos de ECAUNGS, las mediciones fotométricas realizadas, y variables meteorológicas de otras fuentes. De esta manera, se modeliza $PM_{2,5}$ en el Centro-Norte del conurbano y se completa un análisis con valores relativos de aerosoles en superficie aportando nueva información de la CA en la zona. El desempeño del modelo es comparado con los resultados de los modelos de alto y bajo costo desarrollados (**Capítulo 4**).

6.2. Resultados y discusiones

6.2.1. Evaluación de sensores meteorológicos

En esta sección se presentan los resultados de la caracterización de los LCS 18b20, DHT, BMP y BME. Las comparaciones de todos estos sensores con las referencias arrojaron buenos resultados, como se muestra en la **Tabla 6.1**, con coeficientes de determinación (R^2) iguales o mayores que 0,94. La **Figura 6.1** muestra esto para las salidas de temperatura de los sensores contra PS (sensor termistor de referencia) con todas las pendientes cercanas a la unidad. La **Figura 6.2** muestra la misma característica para salidas de temperatura de sensores contra UNGS WS (estación meteorológica de referencia), donde DHT y BMP tienen pendientes cercanas a la unidad, pero BME y 18b20 muestran pendientes más altas. Los errores (RMSE) son significativamente grandes para las pruebas de humedad relativa, lo que puede verse como una alta dispersión en los ajustes de BME y DHT contra UNGS WS en la **Figura 6.3**. Esta figura también muestra una pendiente más cercana a uno para DHT, pero con comportamiento no lineal en valores extremos. Las comparaciones de temperatura contra PS dan como resultado un RMSE más alto y un R^2 más bajo que contra UNGS WS, esto puede explicarse con las temperaturas más altas utilizadas en el primer caso. La **Figura 6.4** muestra la linealidad de las salidas de presión de BMP y BME contra UNGS WS.

Sensor	RMSE/ R^2 (UNGS WS)	RMSE/ R^2 (PS)	RED (UNGS WS)	DER (PS)
BME (T)	1,4/0,99	2,0/0,95	2 %	6 %
18b20 (T)	1,1/0,99	2,0/0,95	2 %	4 %
BMP (T)	0,3/0,99	2,2/0,94	2 %	7 %
DHT (T)	1,5/0,99	1,9/0,94	3 %	6 %
BME (HR)	6,8/0,99	—	2 %	—
DHT (HR)	9,0/0,95	—	3 %	—
BME (P)	0,3/0,99	—	3 %	—
BMP (P)	0,3/0,99	—	1 %	—

Tabla 6.1: RMSE, R^2 y DER (desviación estándar relativa) de las comparaciones de los sensores con PS y WS UNGS

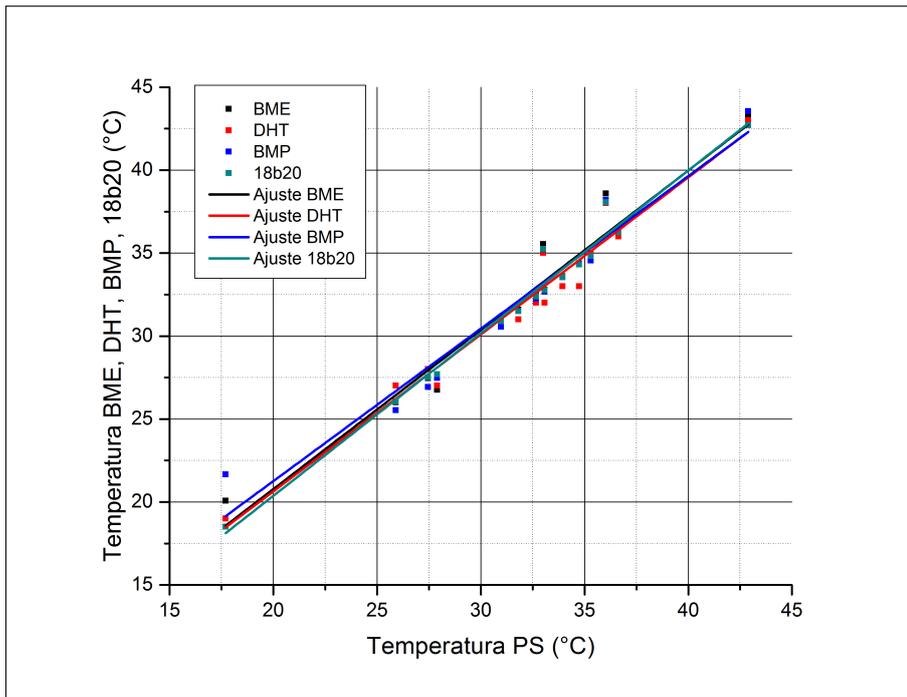


Figura 6.1: Salidas de temperatura de los sensores contra PS con ajuste lineal. Las pendientes resultan muy similares en todos los casos y cercanas a uno

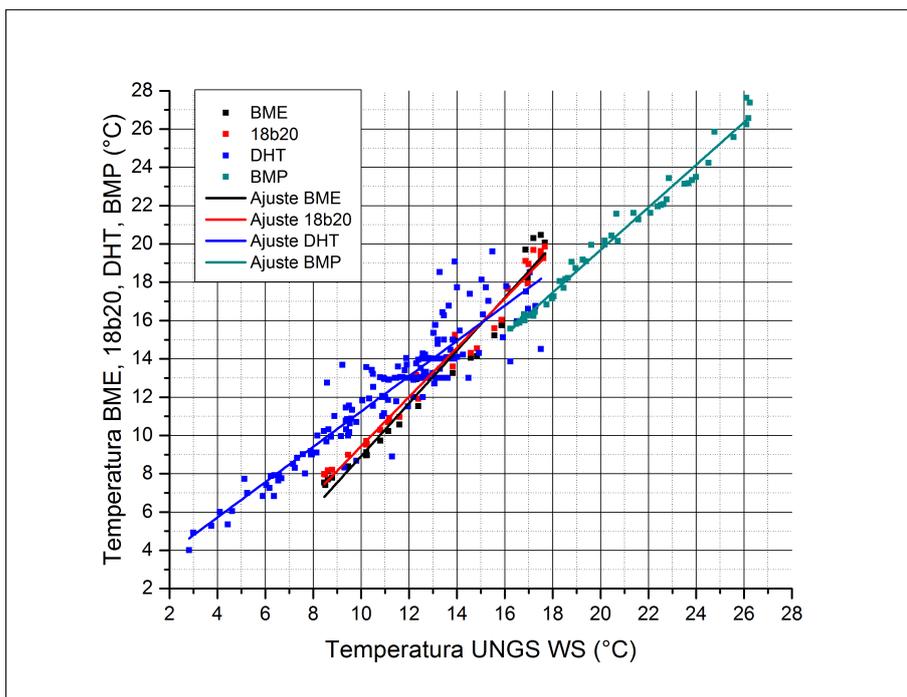


Figura 6.2: Salida de temperatura de los sensores contra UNGS WS con ajuste lineal. Se puede ver una mayor dispersión para las medidas de DHT

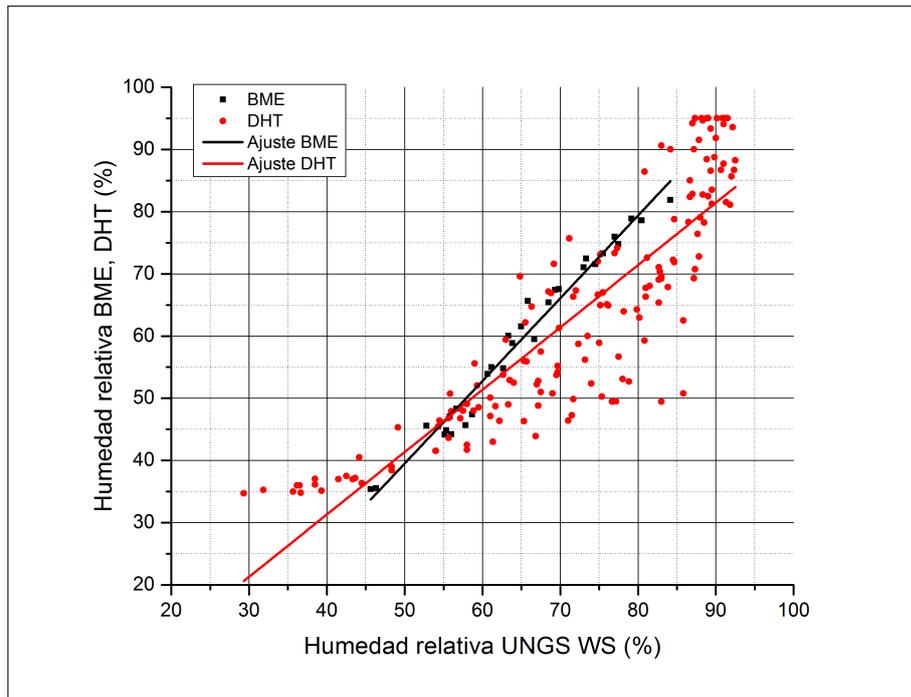


Figura 6.3: Salidas de humedad relativa de BME y DHT contra UNGS WS con ajuste lineal. DHT presenta una mayor dispersión y una pendiente diferente que BME

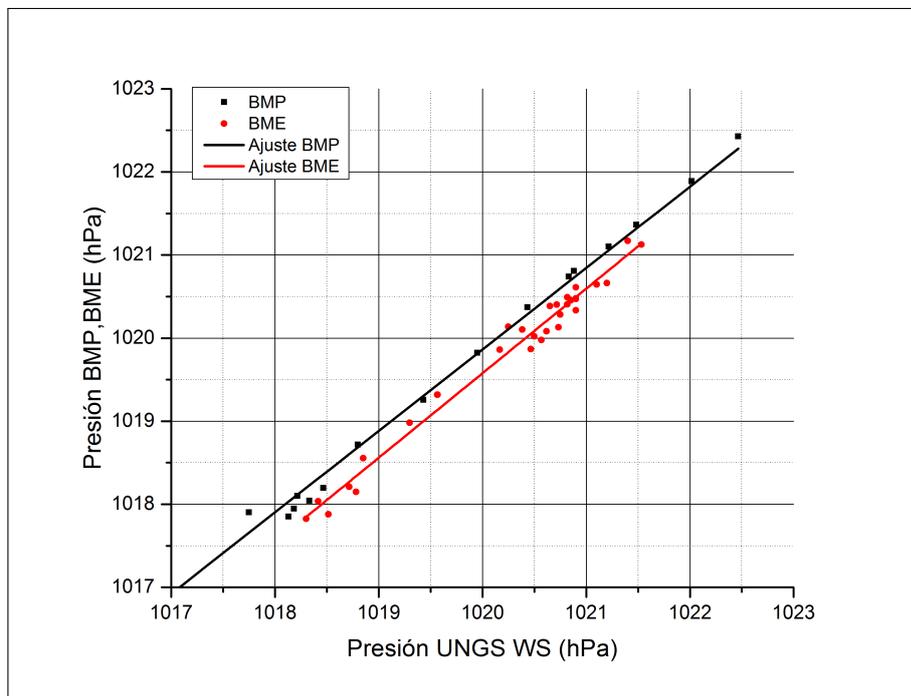


Figura 6.4: Salidas de presión de BMP y BME contra UNGS WS con ajustes lineales

Los resultados de exactitud se pueden ver también en la **Tabla 6.1** con los valores DER. Sobre la temperatura, con PS como referencia, el sensor 18b20 muestra la mayor exactitud. Pero con WS UNGS como referencia, no hubo diferencias significativas entre los sensores.

Como se puede ver en la **Tabla 6.2**, los resultados de precisión son muy altos en todas las mediciones de sensores. El sensor BME tiene la mejor resolución de temperatura, como el sensor BMP de presión. La baja resolución de DHT limitó los cálculos para este sensor, por lo que su precisión no es un resultado confiable y su resolución se informa como en la hoja de datos.

Sensor	Precisión (%)	Resolución	Tiempo de retraso (s)	Tiempo de subida (s)
BME (T)	98,7	0,01 °C	5	665,0
18b20 (T)	98,8	0,05 °C	1	253,5
BMP (T)	98,8	0,04 °C	1500	350,0
DHT (T)	98,8	1,00 °C	5	96,5
BME (HR)	98,7	0,09%	4	30,0
DHT (HR)	98,4	1,00%	58000	23,0
BME (P)	99,9	0,86 hPa	1	1,0
BMP (P)	99,9	0,05hPa	1	1,0

Tabla 6.2: Precisión, resolución, tiempo de retraso y tiempo de subida calculados para todos los sensores. La salida de datos de DHT está limitada por su resolución igual a la unidad.

En cuanto al tiempo de respuesta, hay diferencias notables entre los sensores debidas posiblemente a los diferentes tipos de encapsulados que presentan. El tiempo de retraso resulta muy alto para DHT, dado que su resolución sólo le permite trabajar con valores enteros se espera que la señal del sensor demorara en mostrar una reacción a los cambios. Es llamativo el tiempo de retraso de 1500 s para el sensor de temperatura BMP en comparación con el retraso de 1 s para 18b20 o de 5 s para BME en el mismo parámetro. Los sensores de presión también muestran un tiempo de retraso bajo, como la humedad relativa para BME. Los tiempos de retraso iguales a 1 ms no son resultados confiables porque esa es la frecuencia de muestreo. Pero se puede aseverar que esos tiempos de retraso son inferiores a 1 ms. El tiempo de subida da como resultado 1 s para los sensores de presión, 23 s para la humedad relativa con DHT y 30 s para la humedad relativa con BME. En el caso de las mediciones de temperatura, el tiempo de subida más corto fue

para DHT y el más alto para BME, como también se puede ver en la **Figura 6.5**.

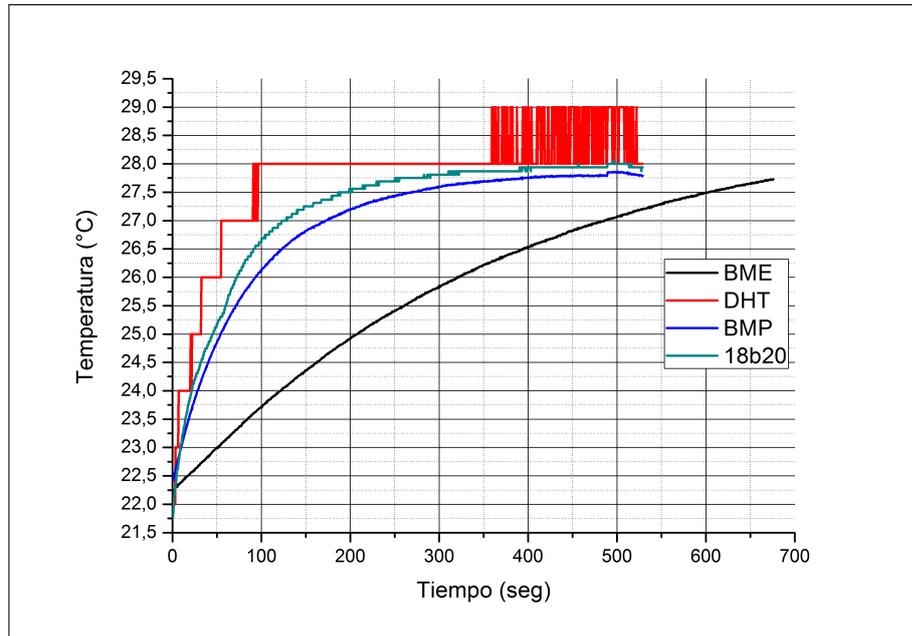


Figura 6.5: Salida de los sensores de temperatura en el tiempo sometidos a un gradiente positivo

Acerca de la respuesta a la pérdida de energía, no fue posible medir un tiempo de recuperación porque la variación de los datos para cada sensor después del corte de energía no fue significativa en comparación con la tolerancia de precisión correspondiente informada en la **Tabla 6.1**. En este sentido, puede afirmarse que el efecto de la pérdida de energía en los sensores es despreciable.

6.2.2. Caracterización de sensor de PM

Como se mencionó en la **Sección 3.3.2**, la precisión y resolución del sensor de PM PMS5003 fue evaluada mediante dos pruebas con valores cuasi-estacionarios de concentraciones de partículas. Ambas tuvieron una duración de 15 min, con una frecuencia muestral de aproximadamente 0,7 Hz, lo que resultó en 627 líneas de datos para la primera prueba, y 600 para la segunda.

La primera prueba, con valores medios de concentración de $3,1 \mu\text{g}/\text{m}^3$ para PM_{10} , $4,6 \mu\text{g}/\text{m}^3$ para $PM_{2,5}$ y $5,3 \mu\text{g}/\text{m}^3$ para PM_{1} , arrojó errores estándar despreciables (de $0,03$, $0,04$ y $0,06 \mu\text{g}/\text{m}^3$, respectivamente) y la precisión se encontró una o dos centésimas debajo del 100 %. La segunda prueba se ejecutó con valores medios de $6,1 \mu\text{g}/\text{m}^3$ para PM_{10} , $9,3 \mu\text{g}/\text{m}^3$ para $PM_{2,5}$ y $10,8 \mu\text{g}/\text{m}^3$ para PM_{1} , y arrojó errores estándar también

insignificantes (de 0,05, 0,09 y 0,11 $\mu\text{g}/\text{m}^3$, respectivamente) y misma precisión que la primer prueba (ver **Tabla 6.3**). Así como los errores levemente son mayores para la prueba con mayores valores medios de concentración de partículas, no fue posible realizar pruebas con concentraciones superiores porque el efecto de deposición hace más difícil sostener una concentración constante por un tiempo prolongado dentro de la cámara. Estos resultados no deben considerarse para evaluar la repetibilidad del sensor, ya que para ello se requiere conocer las condiciones en función de algún instrumento de referencia para repetir las pruebas con idénticos microambientes dentro de la cámara, pero permiten determinar que el sensor no presenta una variabilidad o ruido en sus señales que sean significativas durante un funcionamiento sostenido.

Respecto a las resoluciones para cada rango de tamaños, determinadas como las desviaciones estándar en ambas pruebas, los valores son mayores mientras mayores sean las concentraciones medias. En la **Tabla 6.3** se informan las resoluciones de la segunda prueba. La resolución que tienen los datos transmitidos por el puerto serial desde el sensor es de 1 $\mu\text{g}/\text{m}^3$ para todos los tamaños, pero este método demostró que la variabilidad de las mediciones del sensor en condiciones estacionarias invitan a considerar una resolución mínima más gruesa. Con el fin de asegurar una interpretación más confiable de los datos de este sensor en esta Tesis, se consideran como resolución a estos resultados de laboratorio.

Si bien los tiempos de respuesta del sensor PM5003 es informado como de 10 segundos según su hoja de datos[194], esto depende de la naturaleza de la perturbación y las condiciones en las que se encuentre midiendo el sensor. Las señales del sensor para la primer prueba con humo de sahumero dentro de la cámara se muestran en la **Figura 6.6**, donde se puede observar que los tiempos de retraso son insignificantes (el sensor inmediatamente reacciona a un cambio en las concentraciones, o al menos lo hace en un tiempo menor al período mínimo de muestreo que es 1 s). También se puede ver que el tiempo de subida hasta alcanzar el 95% del valor máximo de concentración es muy similar en las tres señales (entre 54 s y 57 s), para pasar de 4 a 113 $\mu\text{g}/\text{m}^3$ en PM_1 , de 6 a 158 $\mu\text{g}/\text{m}^3$ en $PM_{2,5}$ y de 7 a 169 $\mu\text{g}/\text{m}^3$ en PM_{10} . Los tiempos de bajada son iguales en los tres casos y notablemente mayor al de subida (273 s), lo cual implica que las partículas permanecen por un tiempo en la cámara de muestreo o en el circuito desde la toma de aire hasta la misma (ya que al tratarse de un sensor óptico sólo depende de la dispersión de la luz y no cuenta con una resiliencia que pueden aportar sistemas de transducción mecánicos o químicos). En este caso, la señal de PM_1 se toma ese tiempo para pasar de 110 a 13 $\mu\text{g}/\text{m}^3$, la de $PM_{2,5}$ de 155 a 14 $\mu\text{g}/\text{m}^3$ y la de PM_{10} de 164 a

$14 \mu\text{g}/\text{m}^3$.

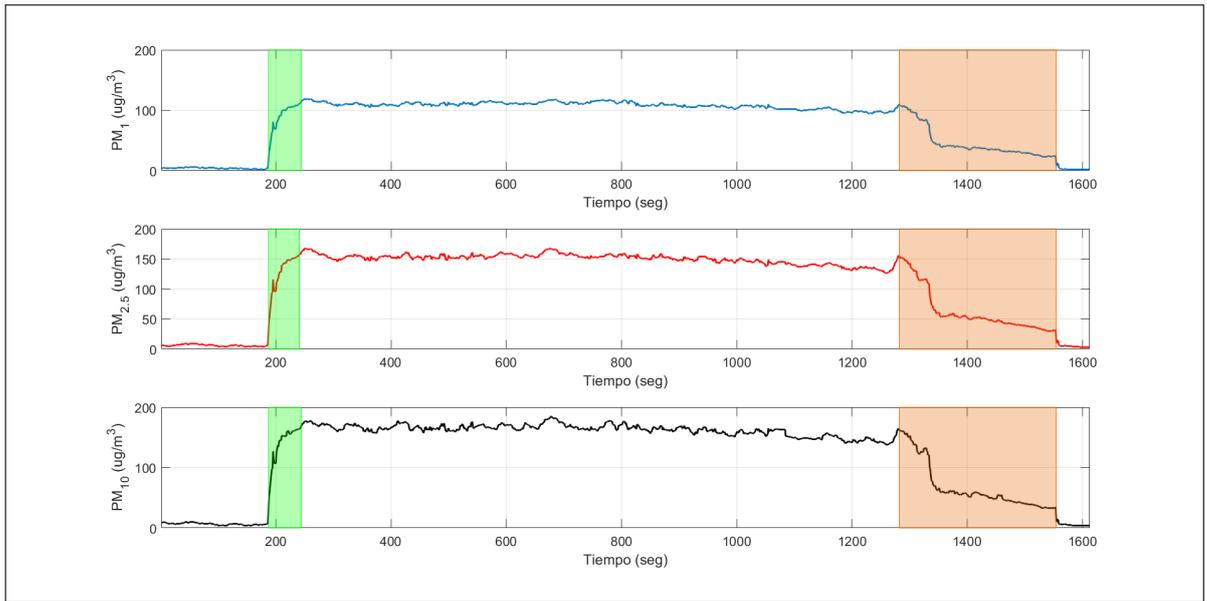


Figura 6.6: Señales de PM del sensor para la prueba 1 de respuesta con humo de sahuero en la cámara. El tiempo de subida está sombreado en verde, y el de bajada con color naranja

En la segunda prueba de tiempos de respuesta hecha con polvo de tiza, el tiempo de retraso nuevamente resultó menor a 1 s y las concentraciones máximas fueron más altas que en la prueba anterior y cercanas al límite superior de detección según la hoja de datos del sensor[194], especialmente las correspondientes a mayores tamaños de partículas. En este caso, los tiempos de bajada fueron similares para las tres señales (entre 326 s y 324 s), pero los tiempos de subida resultaron diferentes (**Figura 6.7**). La señal de PM_1 se tomó 20 s para llegar de 2 a $184 \mu\text{g}/\text{m}^3$, la de $PM_{2,5}$ 23 s para pasar de 3 a $546 \mu\text{g}/\text{m}^3$, y la de PM_{10} 30 s desde 4 a $987 \mu\text{g}/\text{m}^3$. Esta diferencia puede tener que ver con que la prueba se realizó con partículas más gruesas, por lo que hay mayor diferencia entre las concentraciones de las tres señales.

Se puede ver también en la **Figura 6.7** que a los 315 s se registró una perturbación en las señales, lo cual puede tener que ver con algún efecto de resuspensión de las partículas y fue algo que sucedió en mayor o menor medida en todas las pruebas hechas con polvo de tiza.

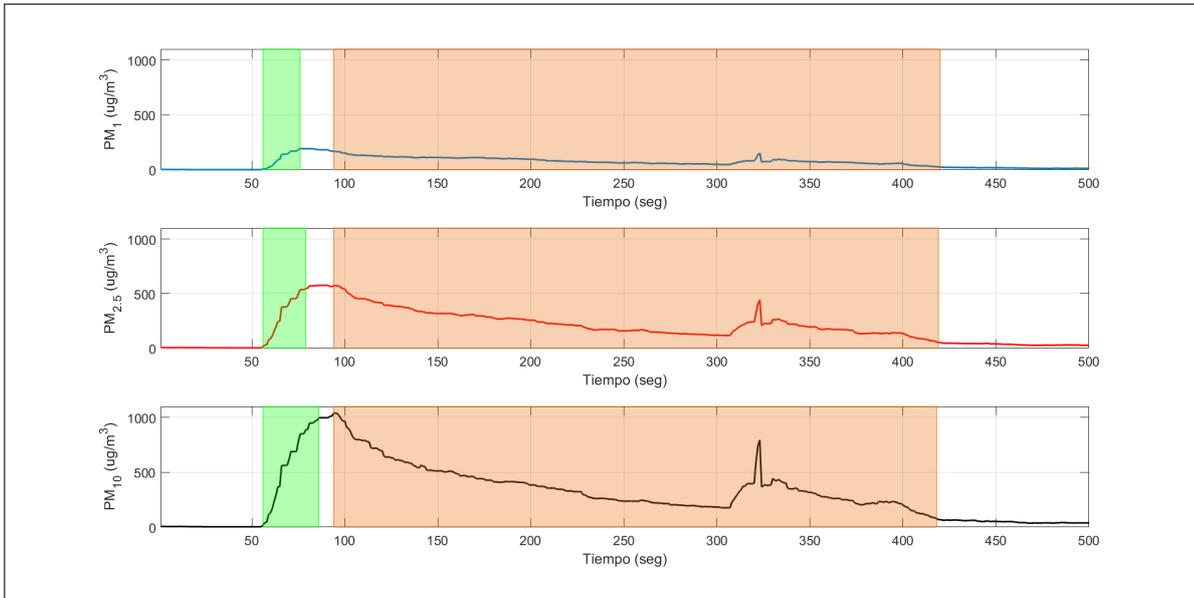


Figura 6.7: Señales de PM del sensor para la prueba 2 de respuesta con polvo de tiza en la cámara. El tiempo de subida está sombreado en verde, y el de bajada con color naranja

Estos resultados nos dan información valiosa a la hora de analizar a futuro las mediciones del sensor. En primer lugar, se concluye que el dispositivo responde inmediatamente a cualquier evento de corta duración, pero el nivel de concentración real de dicho evento puede recién verse reflejado en las señales medidas entre 20 s y 1 minuto después (dependiendo del tamaño de partículas del evento). Algo muy importante también es la duración final de ese evento a detectar, ya que el sensor puede tomarse para recuperar su señal a nivel de ambiente entre 4 y 5 minutos para un evento de partículas finas y entre 20 s y 1 minuto para un evento de partículas gruesas. Toda esta información debe ser tomada en cuenta al analizar señales históricas del sensor en busca de eventos de corta duración. Los resultados de estos experimentos también están indicados en la **Tabla 6.3**.

En todas las pruebas realizadas con cortes de energía, no se registraron cambios en las señales del sensor entre los valores previos y posteriores al corte. Según estos resultados, se puede concluir que el sensor no se ve afectado en sus mediciones luego de un corte del suministro eléctrico.

En las pruebas de laboratorio realizadas disminuyendo la presión de la cámara hasta 400 hPa con la bomba de vacío, se calcularon desviaciones estándar iguales a $1,1 \mu\text{g}/\text{m}^3$ para PM_1 y $1,5 \mu\text{g}/\text{m}^3$ para $PM_{2,5}$ y PM_{10} . Según la implementación de la **Ecuación 3.7** para estas pruebas, se determinaron límites de detección de $3,3 \mu\text{g}/\text{m}^3$ para PM_1 y

$4,4 \mu\text{g}/\text{m}^3$ tanto para $PM_{2,5}$ como para PM_{10} , como se informa en la **Tabla 6.3**.

Como se explica en la **Sección 3.3**, el sensor PMS5003 fue comparado con un instrumento calibrado del programa Cambike, el SPS30. La **Figura 6.8** muestra los resultados de las regresiones entre las señales de PM_1 ($PM1eca$ y $PM1cam$ para PMS5003 y SPS30, respectivamente), $PM_{2,5}$ ($PM25eca$ y $PM25cam$) y PM_{10} ($PM10eca$ y $PM10cam$) de ambos sensores en la campaña de 11 días con período muestral de 1 minuto (15131 líneas de datos).

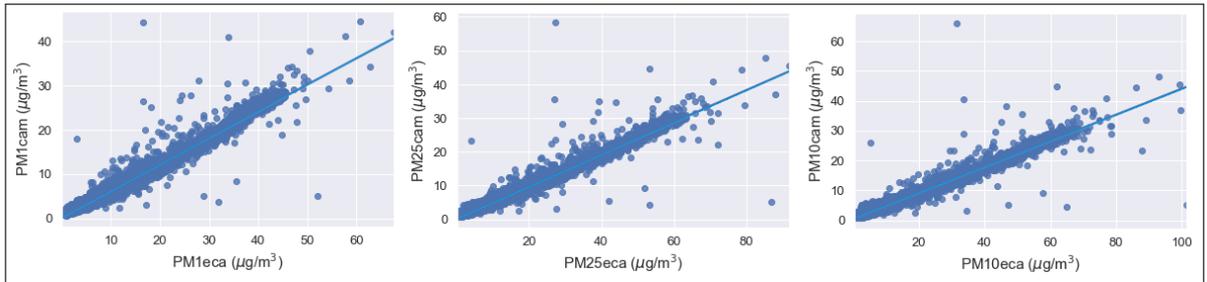


Figura 6.8: Gráficas de dispersión con recta de regresión lineal entre las señales de PMS5003 (PMeca) y SPS30 de Cambike (PMcam)

Los ajustes lineales pasan cerca del origen ($0,4 \mu\text{g}/\text{m}^3$ para PM_1 , $0,5 \mu\text{g}/\text{m}^3$ para $PM_{2,5}$ y $0,6 \mu\text{g}/\text{m}^3$ para PM_{10}), las pendientes ($0,6$ para PM_1 , $0,5$ para $PM_{2,5}$ y $0,4$ para PM_{10}) indican que las señales del PMS5003 están siempre por encima de las de Cambike. Los coeficiente de correlación (R) resultaron de $0,97$ en todos los casos con un valor p del orden de 10^{-16} . Además de estos indicadores de buen ajuste, la potencia de las regresiones indica que la certeza de las mismas es siempre cercana al 100% . El RMSE es un indicador de la dispersión que se puede observar en las gráficas, y resultó en $4,9 \mu\text{g}/\text{m}^3$ para PM_1 , $8,7 \mu\text{g}/\text{m}^3$ para $PM_{2,5}$ y $10,1 \mu\text{g}/\text{m}^3$ para PM_{10} . Esta información sobre la linealidad del sensor evaluado, se condensa en la **Tabla 6.4**.

La desviación estándar relativa de PMS5003 en torno a la referencia (DER), fue de $15,8\%$ para los valores de PM_1 , $19,8\%$ para $PM_{2,5}$ y $21,3\%$ para PM_{10} . La exactitud entonces en la **Tabla 6.4** es informada como la diferencia entre el 100% y esos valores.

Los sesgos del sensor en función del tiempo se muestran en la **Figura 6.9**, se observan pendientes de entre $0,12$ y $0,15 \%/h$ (sesgos de las señales en la **Tabla 6.4**). Esto implica que la señal del sensor duplicaría el valor de sus residuales respecto a la referencia en casi un mes de funcionamiento continuo, aunque se requeriría de una campaña de comparación más extensa para aseverar este hallazgo.

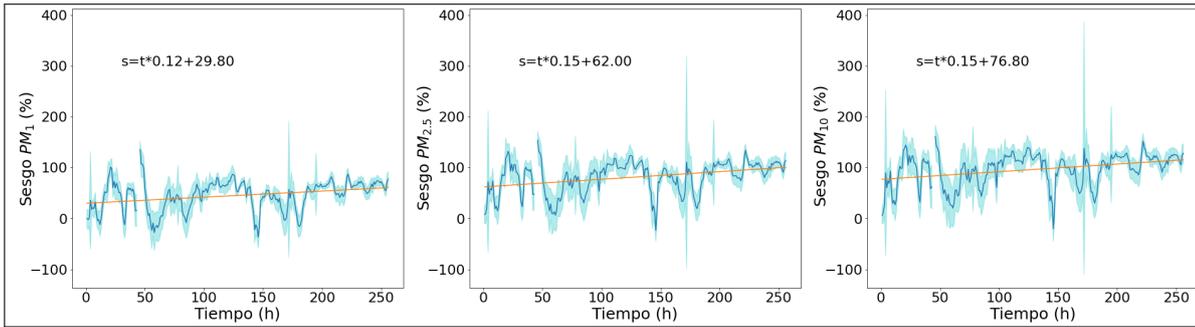


Figura 6.9: Gráficas de los sesgos de las tres señales del sensor PM5003 en base a los residuales con la referencia de Cambike. En sombreado celeste se muestra la desviación estándar del promedio horario realizado y en naranja la recta de ajuste. También se muestran las ecuaciones del ajuste

Los límites inferiores de detección (LLOD) fueron calculados con las pendientes de las regresiones y desviaciones estándar del PMS5003 cuando el sensor de Cambike midió menos de $1 \mu\text{g}/\text{m}^3$ y resultaron menores a los estimados con las pruebas de laboratorio, como se muestra en la **Tabla 6.4**.

Los cuantiles condicionales de los valores registrados por ambos sensores en la campaña de comparación arrojan nueva información (**Figura 6.10**). El hecho de que la mediana (trazada en rojo) se encuentre siempre por debajo del ajuste perfecto (línea azul), indica que el PMS5003 presenta sistemáticamente mayores valores de concentración que Cambike, para las tres señales. Los percentiles se encuentran en todo momento cercanos a la línea de la mediana, por lo que no hay demasiada dispersión al respecto y muestran la misma información que la mediana. Esa diferencia se observa también en los histogramas, donde se aprecian distribuciones similares para todos los casos, sólo que con un sesgo menor hacia la izquierda para el PMS5003, dado que los bins de rangos de mayor concentración tienen mayor conteo que para Cambike. Estos resultados muestran un error sistemático del sensor comparado con la referencia (con diferencias de concentración en promedio de $3,2 \mu\text{g}/\text{m}^3$ para PM_1 , $5,8 \mu\text{g}/\text{m}^3$ para $PM_{2,5}$ y $6,8 \mu\text{g}/\text{m}^3$ para PM_{10}). Si bien se necesita de campañas de comparación de mayor duración para abarcar diferentes épocas del año y determinar la influencia de otras variables, estos resultados se utilizaron para nivelar las señales del PMS5003 con las del equipo tomado de referencia y realizar una calibración preliminar eliminando esos errores sistemáticos.

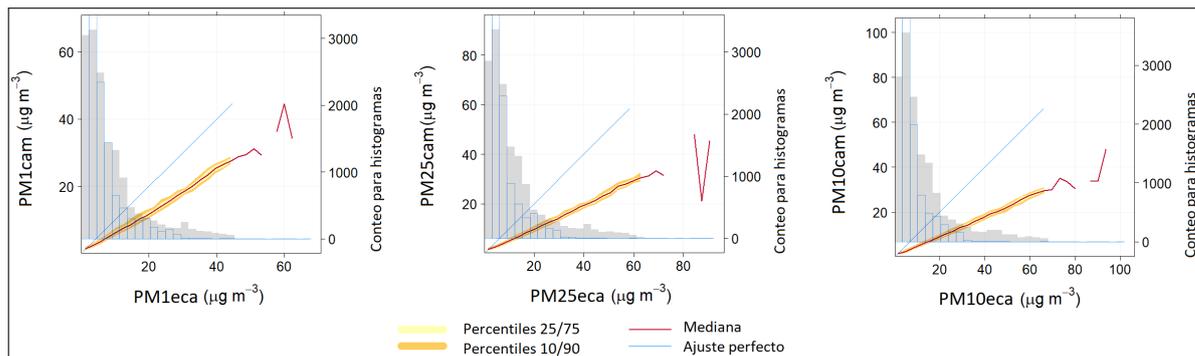


Figura 6.10: Cuantiles condicionales entre las mediciones del equipo de Cambike y el PMS5003 para las tres señales de PM. Se trazan el ajuste perfecto (1:1), las medianas y percentiles de Cambike en función del PMS5003 y los histogramas (sombreado en gris para el sensor en estudio y con línea azul para la referencia)

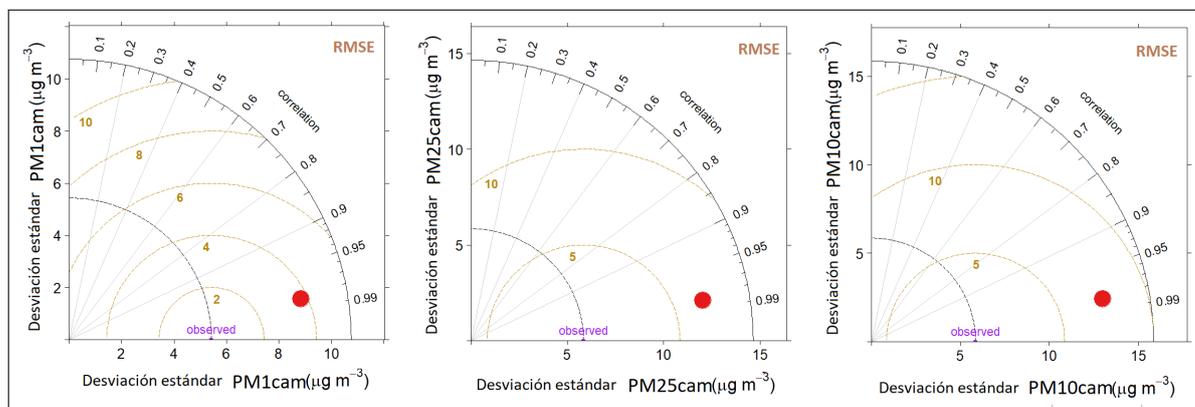


Figura 6.11: Diagramas de Taylor entre las señales de PMS5003 y de Cambike. Los puntos rojos en los diagramas indican las posiciones de las señales del PMS5003 respecto a los estadísticos. RMSE se marca con arcos en color marrón, el coeficiente R de las regresiones como las líneas radiales grises y la variabilidad se evalúa de acuerdo a la distancia radial del punto rojo en el espacio de desviaciones estándar de la referencia

La **Figura 6.11** muestra los diagramas de Taylor para las comparaciones entre las tres señales del sensor PMS5003 con la referencia. Estos gráficos permiten evaluar la variabilidad de las señales en torno a la referencia. Las posiciones del punto rojo en cada diagrama indica los valores del coeficiente de correlación R de las regresiones y el RMSE (datos ya informados previamente), pero más importante aún en este caso es la posición de ese punto en el espacio de desviaciones estándar de Cambike. Mientras más se aleje ese punto del arco de líneas punteadas (que indica la variabilidad del sensor de referencia

en ese espacio), mayor será su variabilidad y sus mediciones se alejarán de la referencia. Es esperable que esa variabilidad sea mayor para rangos de mayores partículas, dado que los valores absolutos de concentración son mayores, pero en ningún caso se separa más allá del doble de la desviación estándar de Cambike, por lo que se puede concluir que no existen componentes de variabilidad significativas para el PMS5003 en los datos adquiridos. Nuevamente, este análisis podría ser más rico en el caso de contar con campañas de comparación de mayor duración, pero la comparación realizada permite confiar en los valores arrojados por el sensor, al menos de forma relativa.

Indicador	PM_1	$PM_{2,5}$	PM_{10}
Precisión	$\sim 100\%$	$\sim 100\%$	$\sim 100\%$
Resolución ($\mu g/m^3$)	1,3	2,3	2,9
Tiempo de retraso (s)	<1	<1	<1
Tiempo de subida (s)	20-57	23-54	30-57
Tiempo de bajada (s)	273-326	273-325	273-326
LLOD ($\mu g/m^3$)	3,3	4,4	4,4

Tabla 6.3: Indicadores de desempeño del sensor PMS5003 determinados en las diferentes pruebas de laboratorio)

Indicador	PM_1	$PM_{2,5}$	PM_{10}
Linealidad (R^2 /RMSE ($\mu g/m^3$))	0,97/4,9	0,97/8,7	0,97/10,1
Exactitud (100% - DER)	84,2%	80,2%	78,7%
Sesgos (%/h)	0,12	0,15	0,15
LLOD ($\mu g/m^3$)	1,6	2,1	3,3

Tabla 6.4: Indicadores de desempeño del PMS5003 en comparación con el sensor del equipo de referencia Cambike

Las **Tablas 6.3** y **6.4** resumen los principales resultados de las pruebas realizadas en laboratorio y en comparación con la referencia para el sensor PMS5003. Estos indicadores de desempeño mejorarán con la calibración realizada (resta del error sistemático detectado en cada señal), pero se prefirió mostrar y tener en cuenta estos resultados sin la calibración con el fin de sobrestimar y trabajar en el marco de un intervalo de confianza mayor. Con el mismo criterio, los estadísticos fueron calculados con la frecuencia muestral utilizada (no con medias móviles que suavizan las señales y mejorarían los resultados de desempeño). Esto es cierto también en el caso de los sesgos, aunque las pendientes se calcularon luego en función de medias horarias.

Se toma como límites inferiores de detección del sensor a los informados en la **Tabla 6.3**. La resolución deberá considerarse a la hora de evaluar gradientes de concentración y la exactitud servirá para estimar las incertezas que acompañen a las mediciones de PM hechas con el equipo ECAUNGS. Asimismo, si bien la respuesta del sensor a una perturbación es menor a 1s, los tiempo de subida y bajada calculados serán indicadores importantes a la hora de estudiar eventos de corta duración. Finalmente, los valores de sesgos deberán corroborarse con campañas de comparación de mayor duración, pero si se ratifican los valores se debería recalibrar el sensor una vez por semana (con el sensor fuera de línea o de manera automática tomando en cuenta las pendientes calculadas).

Las siguientes pruebas del sensor se basan en la comparación con distribuciones de partículas programadas con la librería “opcsim”, como se indica en la **Sección 3.3**. La **Figura 6.12** muestra las distribuciones volumétricas en función del tamaño para diferentes tipos de aerosoles típicos. Todas están compuestas por 3 modos lognormales, pero a la vista las sumas finales tienen formas principalmente bimodales con un pico entre 0,1 y 1 μm y otro entre 2,5 y 10 μm , excepto para la distribución de aerosoles urbanos que muestra sólo uno distinguible entre 0,1 y 1 μm .

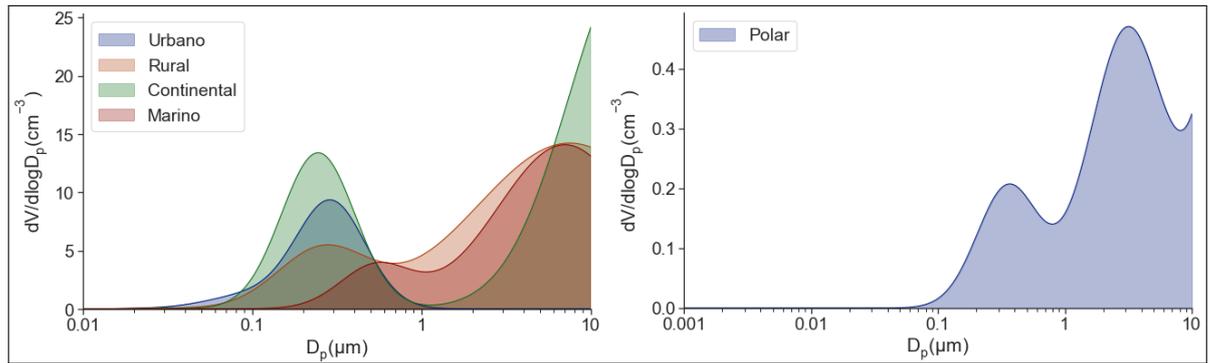


Figura 6.12: Distribuciones volumétricas en función del diámetro de partícula para tipos de aerosoles típicos. A la izquierda: urbano, rural, continental, marino. A la derecha en otra escala: polar (antártico). La escala de tamaños se limita a 10 μm por ser el mayor tamaño informado por el sensor de PM

Si se integran las distribuciones volumétricas se obtienen las cargas másicas en función del tamaño de partículas (normalizado por la densidad). La **Figura 6.13** muestra las funciones de distribución acumulada de masa para cada caso descrito anteriormente, normalizadas por la densidad que puede cambiar de manera temporal y espacial. Se pueden interpretar algunas características que influirán en las señales de un sensor de PM, especialmente las fracciones entre los valores de diferentes rangos de tamaños.

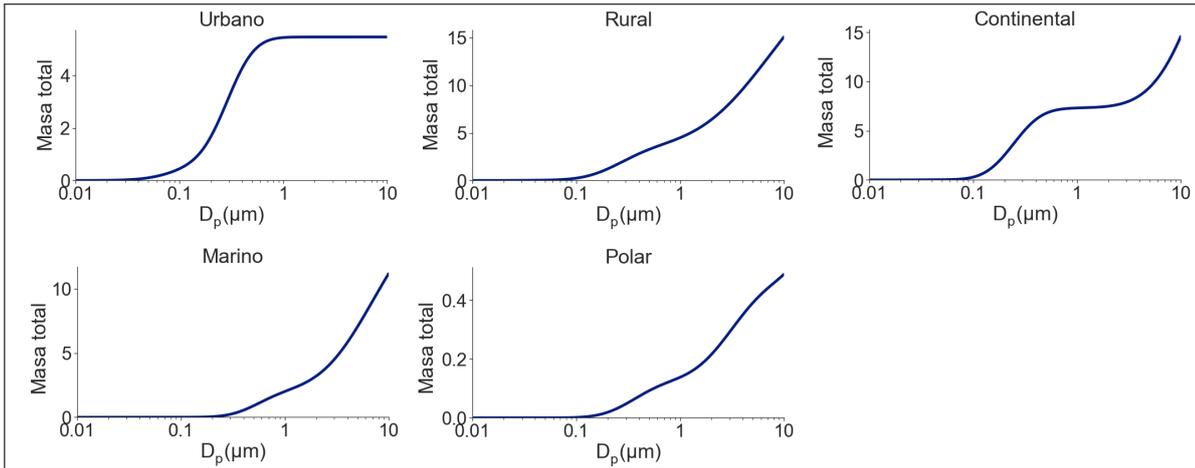


Figura 6.13: Funciones de distribución másica acumulada para los casos de aerosoles urbanos, rurales, continentales, marinos y polares (antárticos). Cada concentración másica se encuentra normalizada por la densidad, por lo que se informa la masa total de forma adimensional

Si se realiza el cociente entre los valores de masa total de partículas de tamaños menores a $10 \mu\text{m}$ (PM_{10}) y los de tamaños menores a $1 \mu\text{m}$ (PM_1), se obtiene una magnitud (que llamaremos $PM_{10}/1$) que aporta información sobre la fracción de tamaños de partículas presentes. En el caso urbano, se puede observar que la razón es muy cercana a 1, lo que significa que domina la fracción fina (como puede verse en la **Figura 6.13** ya que el valor máximo de masa se alcanza para diámetros cercanos a $1 \mu\text{m}$). Esto se corresponde con partículas provenientes de industrias, transporte, generación de energía y mecanismos de conversión de gases a aerosoles [164]. En el marco de las mediciones de propiedades ópticas de aerosoles descritas en el **Capítulo 5**, este tipo de aerosoles se corresponden con un coeficiente de Ångström elevado y el nivel de extinción de la luz dado por el AOD indica el grado de contaminación por este tipo de partículas. En el caso de las mediciones hechas en el Centro-Norte del conurbano, se encontraron pocos casos de aerosoles que se correspondan con esta categoría y, en general, con un nivel bajo de contaminación (bajo AOD). Las partículas de zonas rurales en general parecen tener una cantidad más considerable de partículas gruesas, por lo que la razón $PM_{10}/1$ es mayor y aproximadamente igual a 3. Estas son partículas de origen natural con una inclusión menor de aerosoles antropogénicos [164].

La distribución continental, referida a zonas profundamente adentradas en el continente, también tiene una fracción gruesa considerable como la rural, pero con un mayor pico en el modo fino (su $PM_{10}/1$ es cercano a 1,9). Estos aerosoles están compuestos por productos de oxidación secundaria y partículas primarias como polvo, polen y plantas

de tratamiento [164]. La región de estudio presenta aerosoles de tipo continental según los resultados del **Capítulo 5**, pero esta relación de tamaños se encuentra en zonas más dentro del continente.

Los aerosoles marinos poseen aun una mayor diferencia marcada entre los modos fino y grueso que el tipo rural ($PM_{10}/1$ cercano a 5,5), y provienen de la influencia del viento y de las burbujas de aire ascendente en la interfaz océano-aire [164]. En relación con propiedades ópticas, la presencia de estas partículas se indica con un valor bajo del coeficiente de Ångström y valores también relativamente bajos de AOD. Según los resultados del **Capítulo 5**, se encontraron una gran cantidad de casos de este tipo de partículas al estudiar las mediciones fotométricas, coincidentes con masas de aire provenientes del océano Atlántico. La distribución tipo polar está construida en base a mediciones en la región ártica, pero comparte muchas características en común con la antártica. Estos aerosoles poseen un $PM_{10}/1$ cercano a 3,8, y están compuestos por la neblina ártica/antártica y una leve inclusión de partículas de origen antropogénico [164]. En el **Capítulo 5** se encontró una alta ocurrencia de este tipo de partículas en la región coincidentes con masas de aire provenientes de la Antártica, coincidiendo con bajos coeficientes de Ångström (partículas grandes que se condice con el valor de $PM_{10}/1$ hallado) y valores también bajos de AOD.

Estos resultados de referencia se retomarán en el análisis de las campañas de medición de ECAUNGS en el **Apartado 6.2.3**.

La distribución de solución de sulfato de amonio programada también con “opcsim” está compuesta de un sólo modo correspondiente a la fracción ultrafina, como muestra la **Figura 6.14**. En la misma figura también se muestran los resultados de simulación del efecto de crecimiento higroscópico en la carga másica, donde la masa total crece un 800% entre los valores de humedad de 5% y 95%.

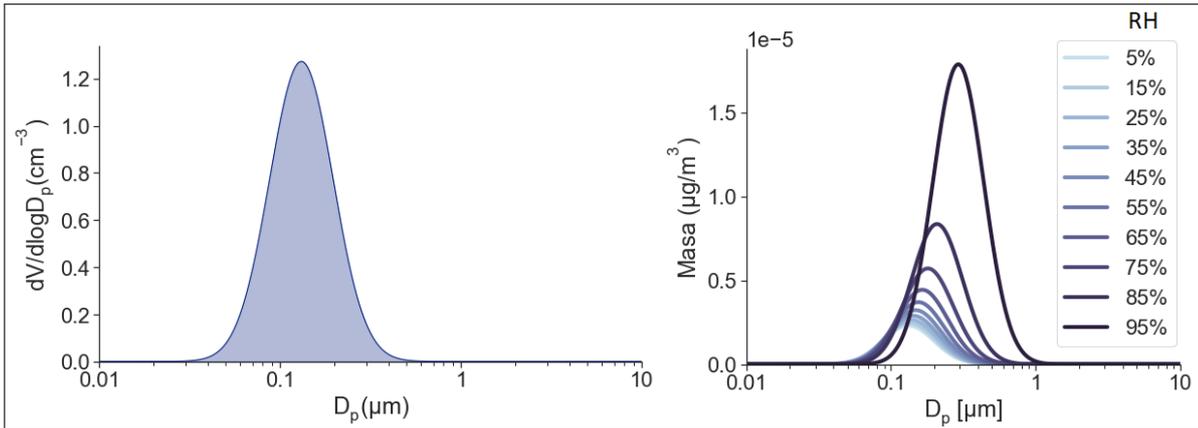


Figura 6.14: Solución programada de sulfato de amonio. A la izquierda, distribución volumétrica. A la derecha, efectos de la humedad relativa (HR) en la carga másica de la solución por unidad de volumen

Al compararse los efectos de la humedad relativa (HR) en las mediciones del sensor PMS5003 con los de la distribución programada de sulfato de amonio (la función de distribución acumulada másica) y de los nefelómetros “neph1” (calibrado con distribución urbana y 50% de HR) y “neph2” (calibrado con sulfato de amonio y 0% de HR), lo primero que llama la atención es que el sensor real mide diferentes valores para cada una de las tres fracciones de masa (**Figura 6.15**). Esto no ocurre con la distribución y los nefelómetros programados, ya que, como se vió en la **Figura 6.14**, la mayor carga másica se encuentra en partículas de tamaño menor a $1 \mu\text{m}$. Luego, se observan valores absolutos de las mediciones del sensor en un orden de magnitud mayor que para las simulaciones, lo cual es un indicativo de la baja confiabilidad de las mediciones del sensor para la fracción gruesa, ya mencionado en la literatura[20], pero también de las diferencias entre la solución simulada y las condiciones generadas dentro de la cámara real. De todas formas, la comparación permite determinar que la solución real presenta un crecimiento higroscópico con la HR y que eso se traduce en una mayor masa acumulada, lo que es medido también por el sensor.

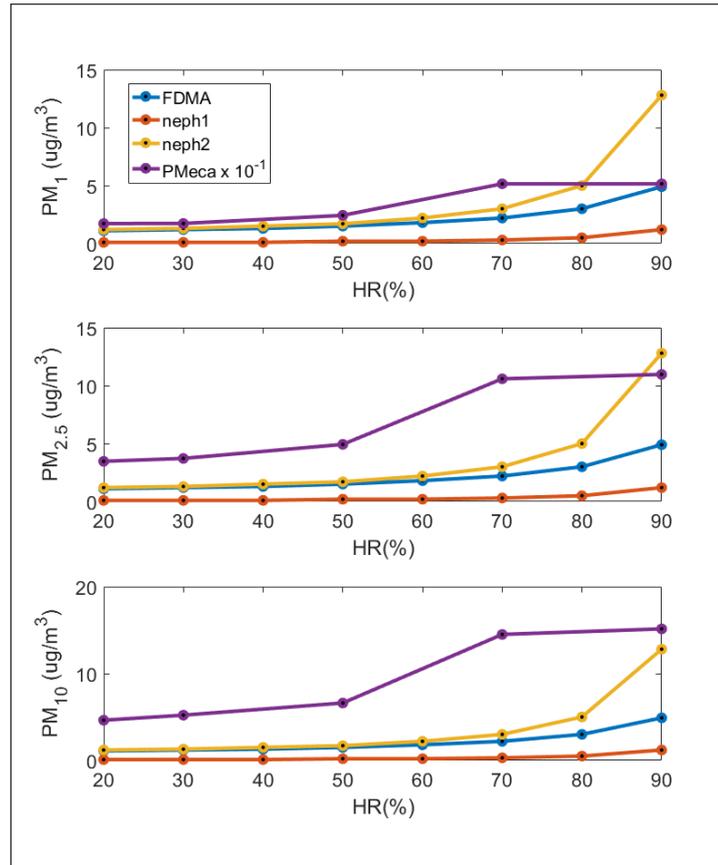


Figura 6.15: Efectos de la humedad relativa (HR) en simulaciones y en el sensor real PMS5003. Se muestran los valores de PM_1 , $PM_{2.5}$ y PM_{10} simulados para la solución de sulfato de amonio y para los nefelómetros programados (neph1 y neph2), junto con los medidos por el sensor en pruebas de laboratorio (PMeca) con misma solución bajo distintas condiciones de HR . Para la solución simulada se muestra la función de distribución másica acumulada evaluada en cada rango de tamaños. Los valores de PMeca fueron re-escalados para una mejor visualización

Mientras que el crecimiento higroscópico dado por un aumento de HR desde el 20 % al 90 % se calculó en un 400 % para la distribución programada, la dispersión de la luz para los nefelómetros modelados aumentó en un 1200 % (lo cual se tradujo en un similar porcentaje de aumento de masa en cada fracción de ambos). Es notable que el aumento en cada fracción de masa del sensor real fuese alrededor del 300 %, cercano a la de la distribución ideal, lo cual indica que la HR no parece afectar considerablemente su transducción sino que el sensor logra correctamente medir las concentraciones dadas por el crecimiento higroscópico. Hagan & Kroll (2020) encontraron resultados similares al estudiar el comportamiento de nefelómetros simulados frente a la HR [198]. Por otro

lado, “neph1” simuló concentraciones más similares a la distribución ideal que “neph2”, lo cual marca la importancia de calibrar estos sensores bajo las condiciones en las que serán utilizados. Esto también ha sido indicado en la literatura[208].

La EPA de EUA recomienda que las mediciones de PM se realicen con una *HR* entre 30 % y 40 %[209], lo cual deja un margen muy estrecho para trabajar. Según las simulaciones y pruebas realizadas, el crecimiento higroscópico comienza a afectar las mediciones del sensor con una *HR* a partir del 50 %, y más considerablemente a partir del 70 %. Otros estudios indican que LCS de PM ven afectado su rendimiento cuando la humedad relativa supera el 80-85 % [159].

Como se indicó en la **Sección 3.3**, con la información recogida en las pruebas, comparaciones y simulaciones, se diseñó un ciclo dentro del algoritmo original de control que controla el prototipo ECAUNGS. El mismo advierte al usuario sobre diferentes situaciones en las mediciones de PM a través del código que se detalla en la **Tabla 6.5**.

Notación	Advertencia
A	Debajo del límite de detección inferior para PM_{1eca} ($3,3 \mu g/m^3$)
B	Debajo del límite de detección inferior para PM_{25eca} ($4,4 \mu g/m^3$)
C	Debajo del límite de detección inferior para PM_{10eca} ($4,4 \mu g/m^3$)
D	Encima del límite de detección superior para PM_{25eca} ($500 \mu g/m^3$)
E	Curva ascendente ($PM_{25eca_t} - PM_{25eca_{t-1}} > 4,6 \mu g/m^3$)
F	Curva descendente ($PM_{25eca_t} - PM_{25eca_{t-1}} < -4,6 \mu g/m^3$)
G	$HR > 50 \%$
H	$HR > 70 \%$

Tabla 6.5: Códigos utilizados para la gestión de calidad de datos. Los marcadores se incluyen en la línea de datos cuando se verifica alguna de las condiciones y, en el caso de ser más de una, aparecerán consecutivas. Por ejemplo una línea de datos marcada como BFH significa que el valor de PM_{25eca} se encuentra por debajo del límite inferior de detección, que las lecturas están en una curva descendente y que la *HR* está por encima del 70 %

Los límites inferiores de detección considerados son los informados en la **Tabla 6.3**, el límite superior para PM_{25eca} es el indicado en la hoja de datos del sensor [194]. La detección de curvas ascendentes y descendentes son útiles para que el usuario pueda considerar los tiempos de subida y bajada encontrados, y los criterios se establecen como una diferencia entre el valor actual y el anterior de PM_{25eca} mayor a dos veces la resolución de la **Tabla 6.3** o menor al valor negativo del doble de la misma resolución, respectivamente. Finalmente, la *HR* tiene un efecto en las mediciones que debe ser tomada en cuenta, por ello se advierte cuando los valores superan el 50 % y cuando se encuentran

por encima del 70 % (donde ya el efecto es considerable).

6.2.3. Integración de metodologías de bajo costo

Como se explicó en la **Sección 3.3**, una vez realizadas las caracterizaciones de los sensores, se desarrolló el prototipo de mini-estación de CA ECAUNGS. Las mediciones realizadas en el exterior entre junio de 2019 y enero de 2021 con el equipo informan PM_1 ($PM1eca$) ($\mu g/m^3$), $PM_{2,5}$ ($PM25eca$) ($\mu g/m^3$), PM_{10} ($PM10eca$) ($\mu g/m^3$), temperatura ($Teca$) ($^{\circ}C$), humedad relativa ($HReca$) (%) y presión ($Peca$) (hPa), a un periodo muestral cercano a 1s. En el control de calidad se encontraron 2294149 líneas de datos con valores no numéricos y 958542 líneas de datos con valores fuera de los límites de detección (para el sensor de PM). Esto representa un 23,7% de la totalidad de datos que fueron etiquetados como no numéricos “NaN”. No se encontraron saltos abruptos entre los datos de PM. El análisis de calidad realizado en la base minutal y horaria no descartó ningún otro dato.

Brevemente, se analizaron las variables meteorológicas medidas por ECAUNGS. Se debe tener en cuenta que hubo períodos intermedios sin mediciones (por ejemplo, no hay mediciones durante meses de marzo, abril y mayo en todo el período), por lo que las tendencias pueden estar sesgadas. De todas formas, la totalidad de datos son valores que acompañan a las mediciones de PM y pueden resultar importantes a la hora de implementar el modelo de estimación con una red neuronal.

La humedad relativa ($HReca$) muestra un histograma centrado en valores cercanos al 50% (**Figura 6.16**), mientras que $HRacu$ (en CABA) y $HRcm$ (SMN de Campo de Mayo) se encontraban más sesgadas hacia valores altos (descritas en los **Capítulos 4** y **5**). Esto puede tener que ver con que las mediciones de ECAUNGS se realizaran principalmente en períodos secos o también con los niveles de incerteza del sensor BME280 a la hora de medir HR (como se muestra al comienzo de este capítulo). Se pueden observar también valores altos durante la noche y los más bajos en primeras horas de la tarde. La tendencia anual no es clara.

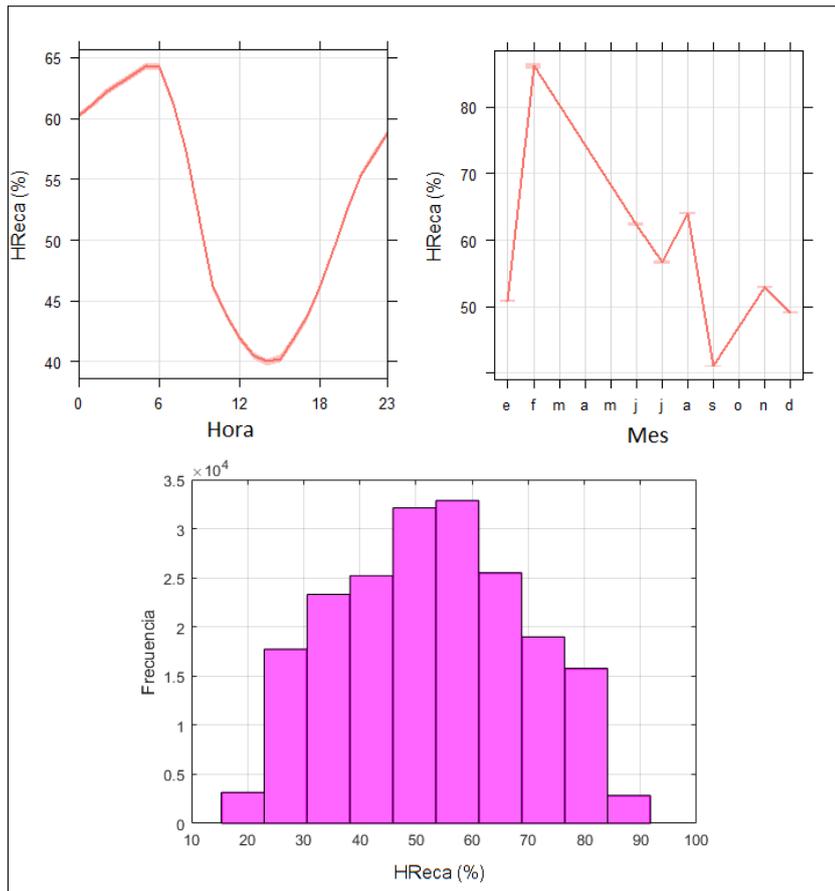


Figura 6.16: Histograma y tendencias mensual y horaria de $HReca$

La temperatura medida por ECAUNGS muestra tendencias muy similares a la medida en Campo de Mayo (T_{cm}), con una distribución normal y marchas diarias y anuales según lo esperado (**Figura 6.17**). La distribución se encuentra centrada en unos 20°C .

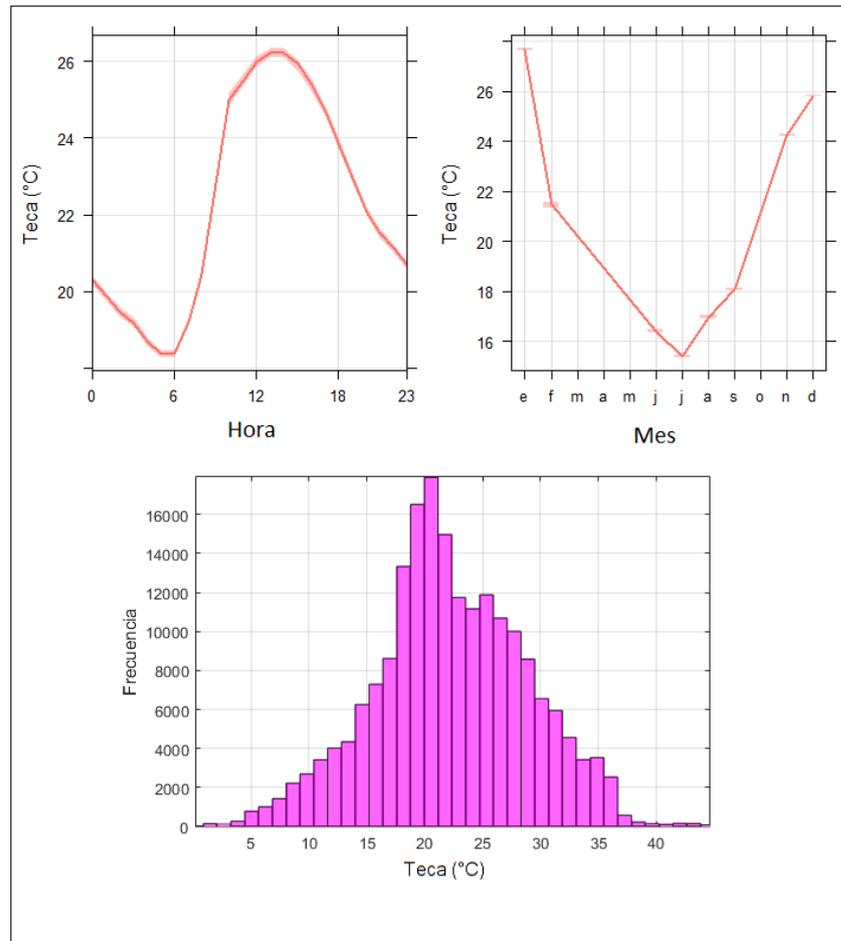


Figura 6.17: Histograma y tendencias mensual y horaria de *Teca*

La presión muestra también tendencias según lo esperado, con máximas en los meses fríos y una marcha diaria que disminuye en las horas de mayor radiación y luego vuelve a crecer posterior a la puesta del sol (**Figura 6.18**). Posee una distribución normal centrada en 1013 hPa.

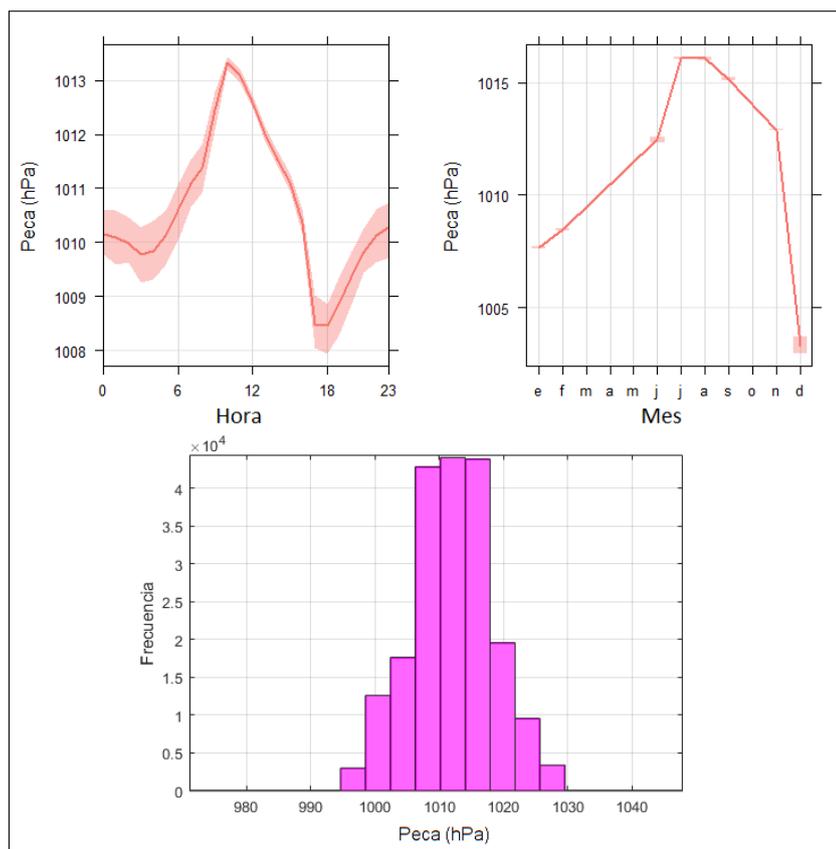


Figura 6.18: Histograma y tendencias mensual y horaria de *Peca*

Las **Figuras 6.19**, **6.20** y **6.21** muestran las tendencias horarias, mensuales y por día de la semana para las mediciones de ECAUNGS en las tres fracciones tamaños de PM. Se observan mínimos en los meses cálidos y en las horas de mayor radiación, además de valores altos los fines de semana. La serie de *PM_{25acu}* analizada en el **Capítulo 5** muestra, en contraste, un máximo a primeras horas de la mañana, una tendencia casi opuesta mensual y menores valores los fines de semana. Estas diferencias pueden deberse a las diferentes fuentes locales en ambos sitios, pero también debe tenerse en cuenta que las mediciones de ECAUNGS son preliminares y se requiere de un período más largo y continuo de mediciones para aseverar mayores conclusiones.

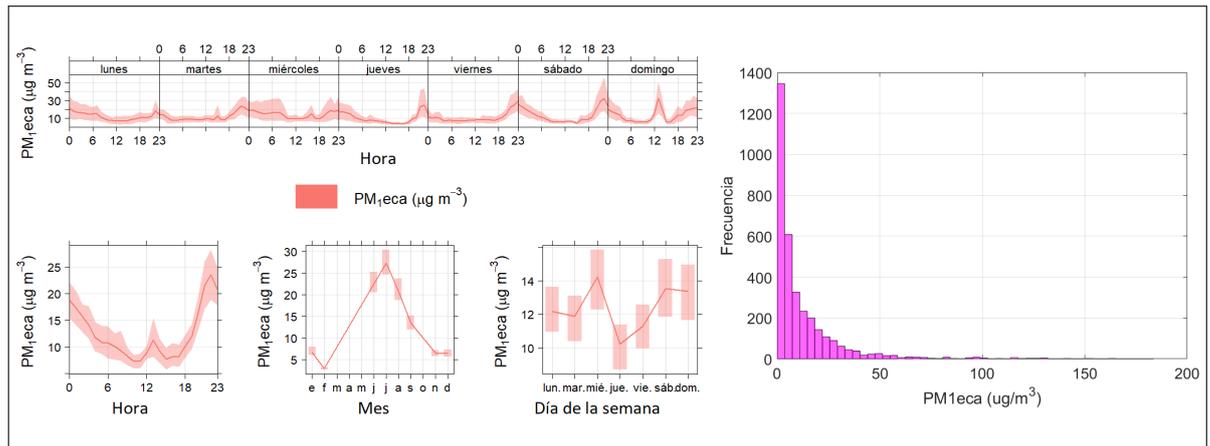


Figura 6.19: Histograma y tendencias mensual, horaria y por día de semana de PM_{10}

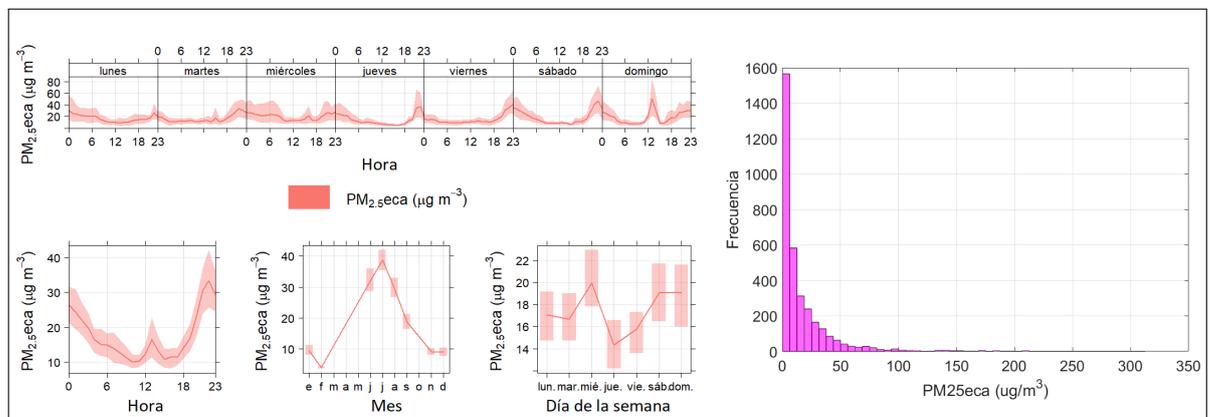


Figura 6.20: Histograma y tendencias mensual, horaria y por día de semana de $PM_{2.5}$

Los valores medios horarios se encuentran entre 5 y $24 \mu g/m^3$ para PM_{10} , entre 10 y $33 \mu g/m^3$ para $PM_{2.5}$ y entre 11 y $36 \mu g/m^3$ para PM_{10} . Si bien el prototipo debe ser validado con un equipo regulatorio y luego calibrado para tomar sus valores absolutos de PM como reales, cabe aclarar que las concentraciones mencionadas no superan los límites en la normativa vigente para provincia de Buenos Aires (DECTO-2018-1074-GDEBA-GPBA)[202]. Esto es razonable considerando que tampoco fueron superados en CABA (**Capítulo 4**) donde hay mayores fuentes urbanas de PM y donde las mediciones fotométricas muestran una mayor concentración de partículas que el Centro-Norte del conurbano (**Figura 5.9** en el **Capítulo 5**).

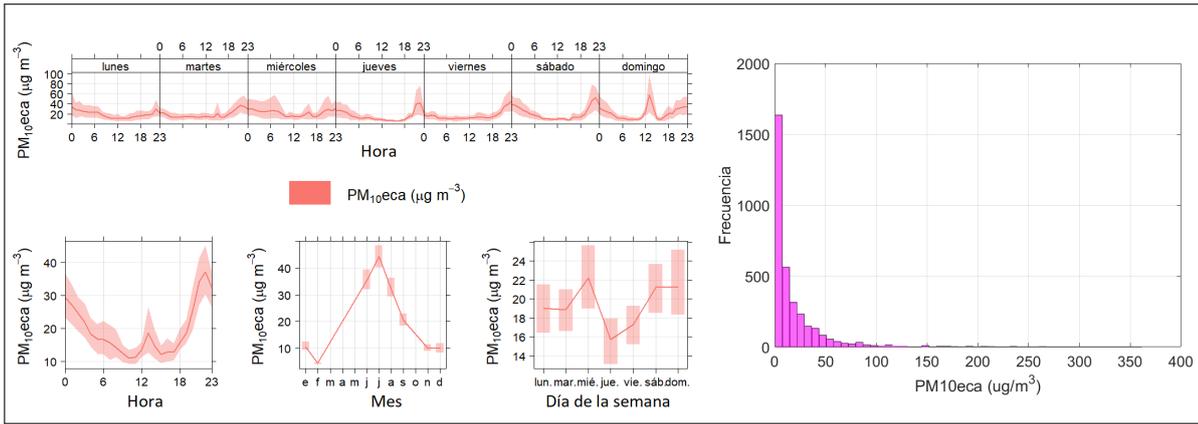


Figura 6.21: Histograma y tendencias mensual, horaria y por día de semana de PM_{10eca}

También cabe indicar que el percentil 75 no supera los $20 \mu g/m^3$ para PM_{1eca} y se encuentra apenas por encima de ese valor para PM_{25eca} y PM_{10eca} , como se muestra en la función de distribución acumulada de la **Figura 6.22**. Estos valores bajos de PM en superficie están acordes con las características de la zona y con los resultados del **Capítulo 5**, que indican una atmósfera limpia con bajas concentraciones de aerosoles en toda la columna vertical.

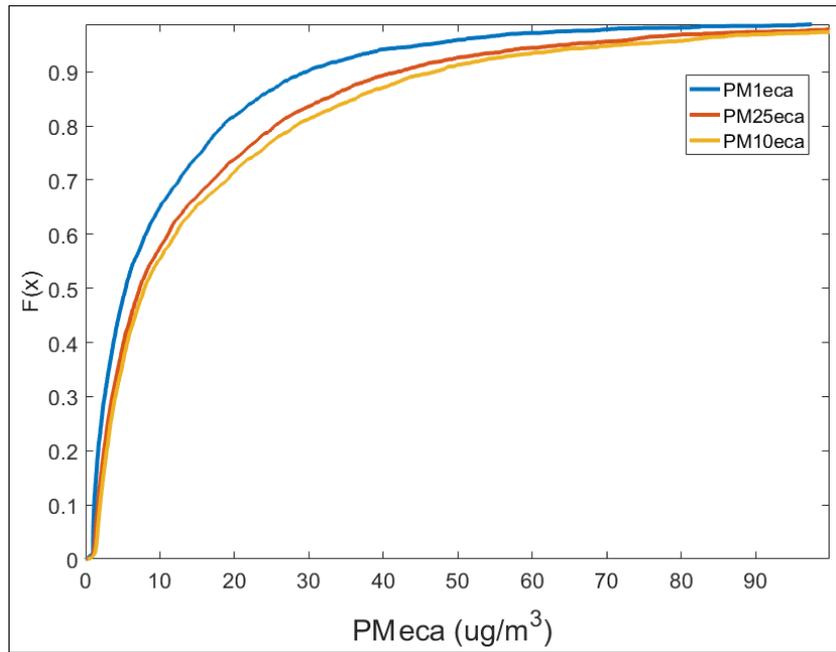


Figura 6.22: Función de distribución acumulada de PM_{eca} (las tres señales). El eje horizontal marca la concentración y el eje vertical la proporción (de 0 a 1)

Las frecuencias de las concentraciones de PM_{25eca} muestran gran variabilidad con

diversos picos (**Figura 6.23**). Si se toman los picos más importantes, se pueden identificar algunos modos principales para la distribución centrados aproximadamente en $16 \mu\text{g}/\text{m}^3$, $25 \mu\text{g}/\text{m}^3$ y $35 \mu\text{g}/\text{m}^3$. El primero siendo el principal, y los dos siguientes de similar importancia entre ellos.

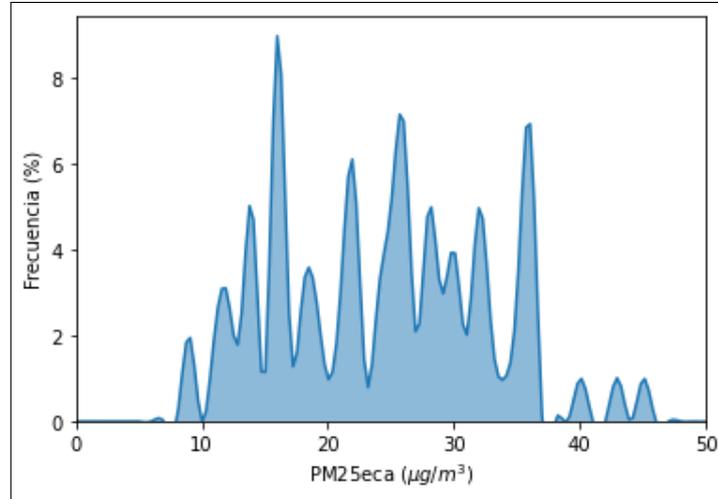


Figura 6.23: Frecuencia de concentraciones de PM_{25eca} para la totalidad de las mediciones

El mismo análisis para diferentes franjas horarias se muestra en la **Figura 6.24**, donde también se observan numerosos picos, aunque pueden identificarse algunos modos generales. La franja de 10 h a 14 h centrada al mediodía muestra una distribución que decrece casi continuamente (salvando los picos de las fluctuaciones) y tiene su máximo cerca de los $3 \mu\text{g}/\text{m}^3$. Las horas de la tarde (15 h a 18 h) muestran a bajas concentraciones un comportamiento similar a la franja anterior, pero con otro modo que aparece centrado en aproximadamente $30 \mu\text{g}/\text{m}^3$. Durante la tarde noche (19 h a 23 h) se ve una distribución similar a la total, con un modo principal cercano a $20 \mu\text{g}/\text{m}^3$ y otros dos secundarios de unos $3 \mu\text{g}/\text{m}^3$ y $32 \mu\text{g}/\text{m}^3$. Durante la noche, madrugada y primeras horas de la mañana (0 h a 10 h) se pueden identificar unos 3 modos, uno principal centrado en aproximadamente $34 \mu\text{g}/\text{m}^3$, y dos secundarios en $3 \mu\text{g}/\text{m}^3$ y $13 \mu\text{g}/\text{m}^3$. Esto complementa los resultados mostrados en la **Figura 6.20** e indica que existen mayores probabilidades de episodios de máximas concentraciones de PM durante las primeras horas del día.

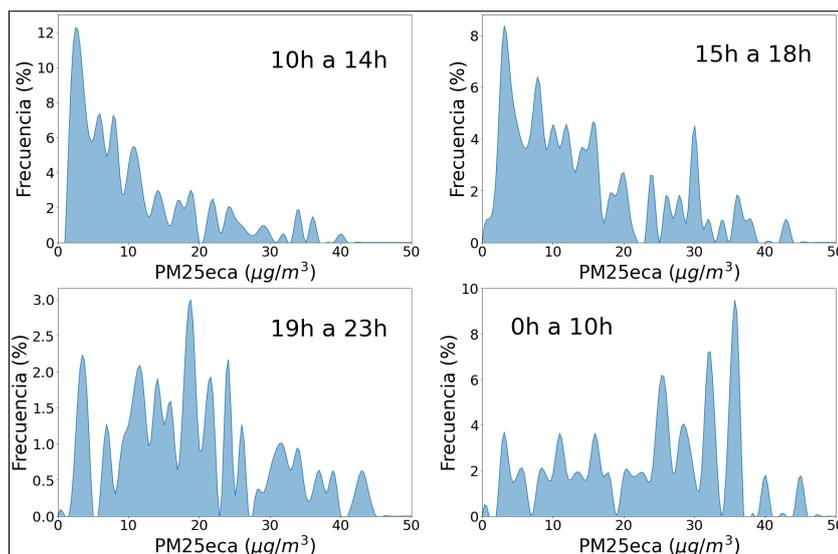


Figura 6.24: Frecuencia de concentraciones de $PM_{25}eca$ para 4 franjas horarias diferentes

El análisis de frecuencias de la razón $PM_{10}/1$ se muestra para la totalidad de los datos (Figura 6.25), ya que no hay diferencias entre las diferentes franjas horarias. En casi todo momento, parece que las concentraciones de $PM_{10}eca$ duplican las de $PM_{1}eca$, y otras pocas ocasiones las cuatriplican. Al parecer, la distribución másica por tamaño de partículas no responde a una zona urbanizada según las modelizaciones mostradas en la Figura 6.13, sino que el modo principal encontrado para $PM_{10}/1$ se corresponde con distribuciones de tipo rural, continental, marino o polar. Estos resultados concuerdan con las conclusiones hechas en el Capítulo 5 a partir de las mediciones fotométricas, donde se encontraron partículas gruesas y una atmósfera generalmente limpia (es decir pocos casos de contaminación urbana) con algunos casos típicos de una zona rural con intrusión de aerosoles industriales, además de varios casos de aerosoles de tipo continental, marino y antártico.

De todas formas, estos resultados deben corroborarse con mediciones validadas, ya que los LCS de PM no suelen medir directamente la fracción de PM_{10} y la informan con un mayor error[20]. Otros trabajos indican que el PMS5003 no distingue correctamente entre las diferentes fracciones de tamaños y que la señal de $PM_{2,5}$ es la más exacta a la hora de comparar con equipos de referencia[139, 161].

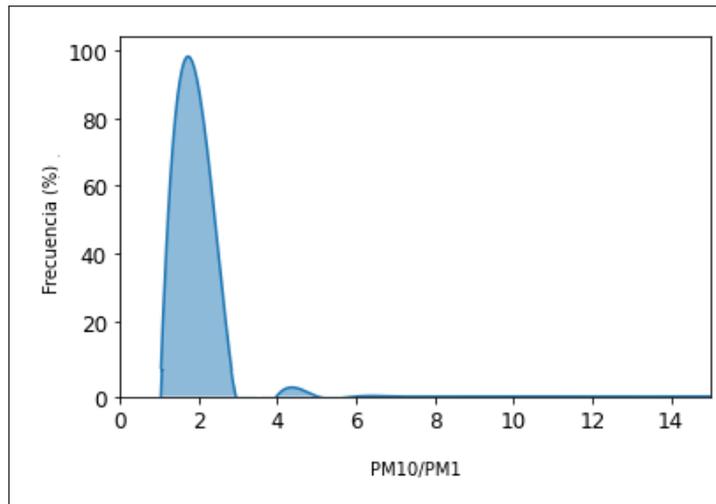


Figura 6.25: Frecuencia de concentraciones de la razón de PM10/1

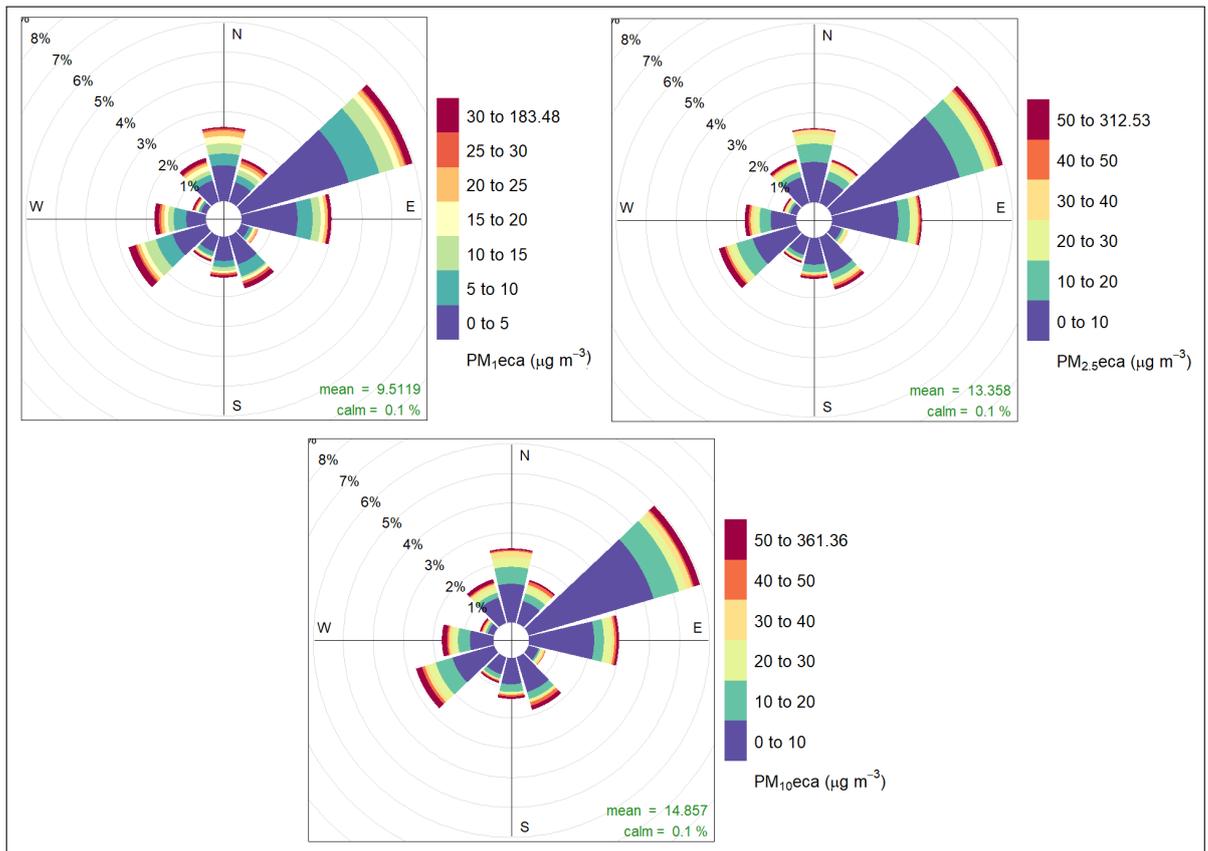


Figura 6.26: Rosas de contaminantes con los datos de dirección del viento de la estación de Campo de Mayo del SMN y los registros de ECAUNGS de PM en las tres fracciones de tamaños

La **Figura 6.26** muestra las direcciones de dispersión de PM para la región de estudio. Las direcciones prevalentes del viento durante las mediciones de ECAUNGS poseen una distribución similar a las mostradas en la **Figura 5.2** del **Capítulo 5**. En este último se encontraron casos de mayor contaminación (en cuanto a propiedades ópticas de aerosoles) con vientos de los sectores Oeste y Norte, pero las concentraciones medidas por ECAUNGS no parecen tener distribuciones diferentes de acuerdo a la dirección del viento. Mayores mediciones se requieren para evaluar esta diferencia entre las mediciones en superficie y las hechas con el fotómetro solar. Mientras que las mediciones fotométricas detalladas en el **Capítulo 5** permiten evaluar la calidad del aire en cuanto a cantidad y tipos de aerosoles en toda la columna atmosférica en la zona Centro-Norte del conurbano bonaerense y las mediciones de PM con ECAUNGS aportan información relativa a partículas en el entorno del punto de medición en el partido de San Miguel, mediciones con equipos móviles entregan una resolución espacial intra-urbana a la calidad del aire. La **Figura 6.27** muestra un mapa trazado con mediciones de $PM_{2,5}$ (las mediciones con otras fracciones de tamaños muestran los mismos patrones) realizadas con el equipo de Cambike en diferentes calles en la zona de la UNGS. La escala se corresponde a las concentraciones medias para cada punto de coordenadas, ya que se completó más de un recorrido en la mayoría de las calles.

Los registros de mayores concentraciones están vinculados al tráfico. Las calles de mayor circulación vehicular poseen mayores concentraciones medias que las calles internas, además los segmentos en rojo corresponden a aquellos donde generalmente hay una alta densidad de automóviles detenidos o con marcha lenta por semáforos y/o embotellamientos. Lo mismo se evidencia en las vías principales contiguas a Campo de Mayo, pese a ser un gran espacio verde con menores emisiones y condiciones de mejor dispersión. Esto se concluyó también en el único antecedente en la zona a partir de mediciones manuales pasivas de PM[31]. El centro de San Miguel es la zona más urbanizada en el mapa, sin embargo muestra valores medios relativamente bajos comparados con las zonas de atascos de tráfico en las rutas principales. Esto apoya la hipótesis de que el tráfico es una fuente importante de partículas en la zona, lo cual también pudo verse con los datos satelitales de AOD mostrados en el **Capítulo 5** (**Figuras 5.9** y **5.10**), donde los sitios más urbanizados poseen un promedio de carga aerosólica mayor que su entorno (en contraste con, por ejemplo, Campo de Mayo).

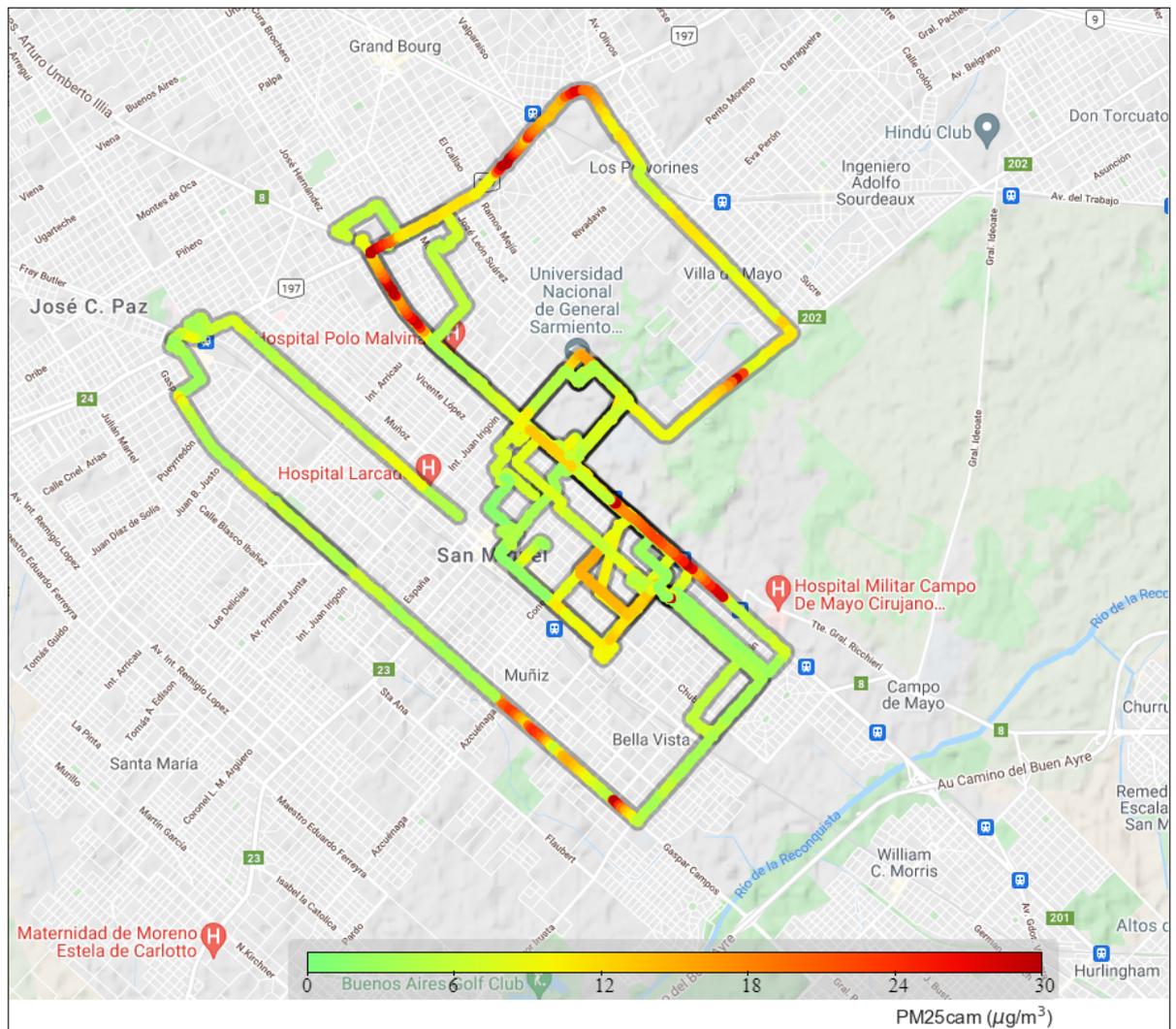


Figura 6.27: Mapa con concentraciones de $PM_{2.5}$ medidas con el equipo de Cambike en una campaña de 2 meses de duración realizada en bicicleta. Como se realizó más de un recorrido por la mayoría de las calles, se representan los valores medios en cada coordenada

El desarrollo del modelo integrador ANN con las mediciones de bajo costo de ECAUNGS y el fotómetro solar CALITOO, y con datos del SMN de Campo de Mayo, ACUMAR en Dock Sud y NOAA, tiene la doble finalidad de completar el estudio de calidad de aire en el Centro-Norte del conurbano, compararlo con lo mostrado en el **Capítulo 4** de CABA (además de utilizar lo aprendido en ese capítulo) y evaluar la calidad de los datos medidos con ECAUNGS. En primer lugar, se muestran en la **Tabla 6.6** los coeficientes de correlación (R) dados entre la salida de una red ANN univariada (probada con cada una de las variables de entrada de manera independiente) con parámetros por defecto y

los valores de $PM25eca$ deseados. Las variables consideradas para la entrada del modelo son las mismas que fueron presentadas en la **Tabla 3.6** del **Apartado 3.3.3**.

Variable	R
<i>BLHungs</i>	0,61
<i>estab</i>	0,54
<i>Teca</i>	0,50
<i>mes</i>	0,48
<i>alphaUNGS</i>	0,48
<i>AODungs</i>	0,48
<i>Uacu</i>	0,44
<i>PM10acu</i>	0,42
<i>Vacu</i>	0,38
<i>hora</i>	0,37
<i>RADacu</i>	0,34
<i>Peca</i>	0,32
<i>PM25acu</i>	0,31
<i>Ucm</i>	0,3
<i>HReca</i>	0,28
<i>Vcm</i>	0,22
<i>diasem</i>	0,08
<i>PPTungs</i>	0,04

Tabla 6.6: Coeficientes de correlación R para las regresiones entre las salidas de la RED ANN de prueba implementada con cada una de las variables consideradas de manera independiente y los valores medidos por $PM25eca$

En comparación con los resultados del **Capítulo 4** con datos de CABA mostrados en la **Tabla 4.1**, la altura de capa de mezcla y la estabilidad nuevamente tuvieron un peso importante para el modelo. En este caso el mes tuvo mayor preponderancia que la hora del día (a la inversa del modelo para CABA) y nuevamente el día de la semana no tuvo influencia. La temperatura tuvo un alto coeficiente, seguido por los datos fotométricos y luego por las componentes del viento y $PM10acu$. Si bien sólo hay 137 h de datos de $AODungs$ y $alphaUNGS$, se decidió dejarlos en el modelo dada la estrecha relación con el PM en superficie. Así como con los datos de CABA, la precipitación no tuvo correlación con el PM en superficie, al contrario de lo visto con las propiedades ópticas de aerosoles en la columna atmosférica en el **Capítulo 5 (Figuras 5.16 y 5.20)**.

Las correlaciones lineales realizadas entre todas las variables, permitió descartar las que aporten información similar a otra variable con una mejor correlación con $PM25eca$, y además extraer algunas conclusiones a nivel ambiental-meteorológico en la zona. $PM25acu$

y *estab* fueron descartados por su similitud con *PM10acu* y *BLHungs*, respectivamente, como también se vio en el **Capítulo 4**. También fueron descartados *Teca* y *RADacu* por su relación con *BLHungs*, lo cual es esperable (sucedió lo mismo con *HReca*, aunque se debe estudiar con mayor detenimiento este vínculo para lograr explicarlo). También ocurrió con las velocidades del viento entre ACUMAR y Campo de Mayo (fueron descartadas entonces *Ucm* y *Vcm* por tener menor cantidad de datos y menor impacto en el modelo). Finalmente, entre *AODungs* y *alphaUNGS* fue descartada esta última. Se puede consultar en el **Anexo 8.3.2** la matriz de correlaciones.

Las pruebas con el método directo gradual para incorporar variables de entrada determinaron cuáles son las de mayor peso. Entonces, aquellas con las que la red logra una buena convergencia son: *BLHungs*, *meses*, *AODungs*, *Uacu*, *PM10acu*, *Vacu* y *hora*. Adicionalmente, el mejor desempeño del modelo, que fue llamado REDungs, se consiguió practicando un filtro de medianas a *PM25eca*, una normalización a las variables de tipo “mapstd” (varianza 0 y desviación estándar igual a la unidad), función de entrenamiento de gradiente conjugado (a diferencia de los modelos con datos de CABA donde se usó Levenberg-Marquardt), función de transferencia de base radial para un total de 30 neuronas ocultas y realimentación con un retraso temporal de 1 h (recurrencia). La **Figura 6.28** muestra la forma básica de REDungs, con los 7 predictores elegidos y la realimentación de la salida a la capa oculta.

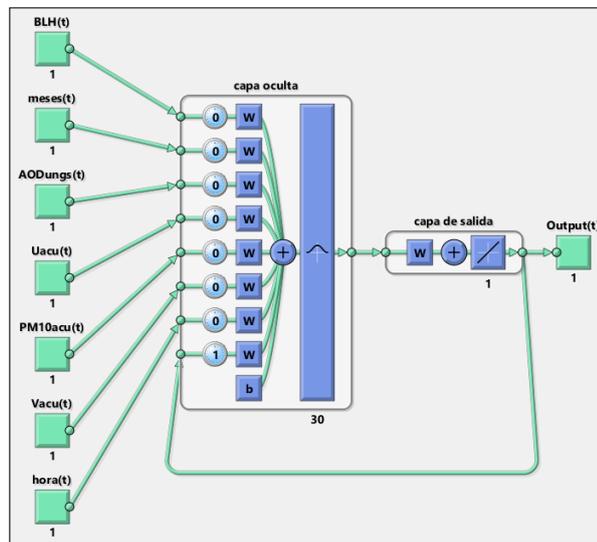


Figura 6.28: Esquema de REDungs. Se observa la realimentación con retraso temporal de 1 h

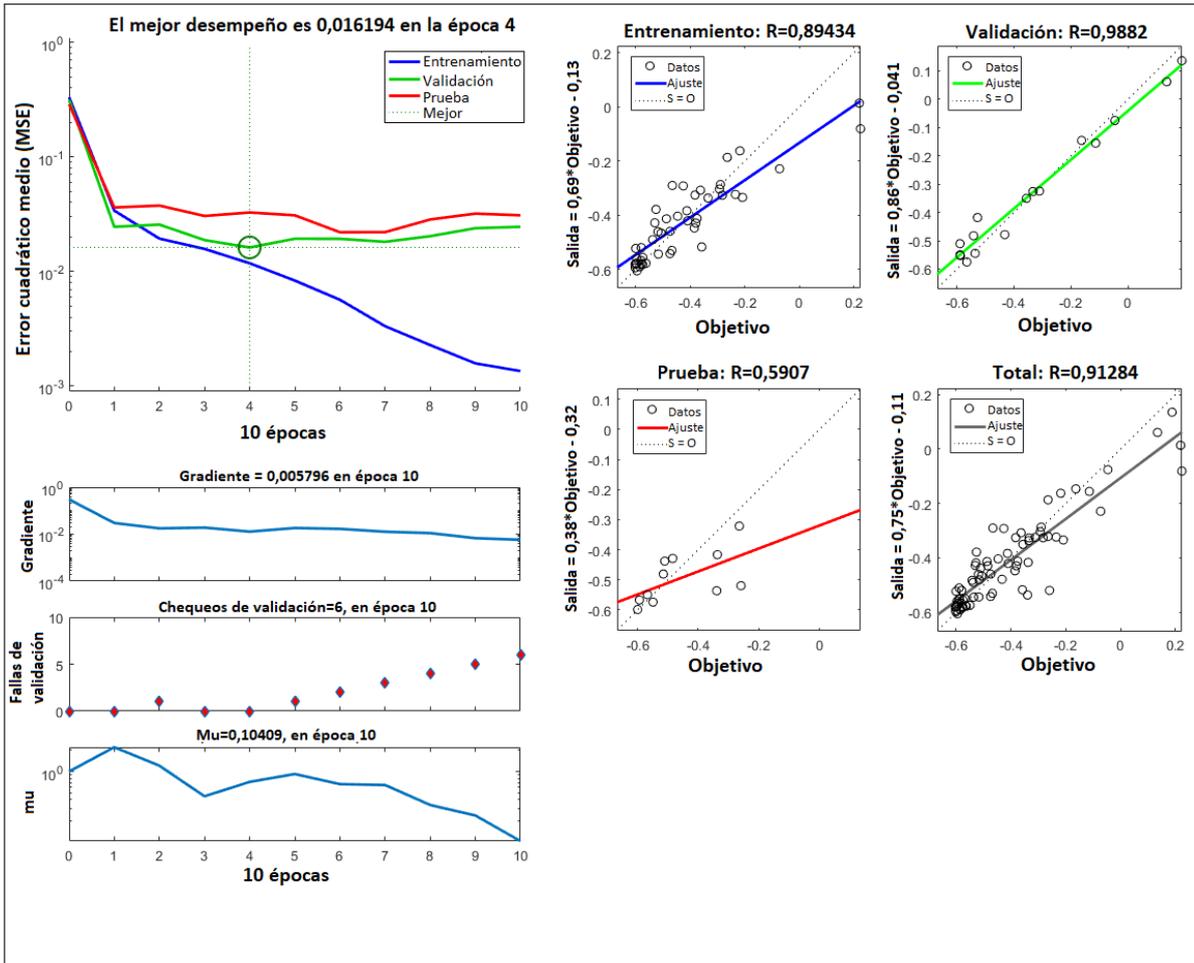


Figura 6.29: Gráficas de regresión, entrenamiento y desempeño de REDungs

Como se puede ver en la **Figura 6.29**, el entrenamiento de REDungs se detuvo en 10 épocas, alcanzando un error MSE mínimo cercano a 0,02 (eso implica un bajo error RMSE cercano a $0,14 \mu\text{g}/\text{m}^3$) y un R cercano a 0,99 en la validación y 0,60 en la etapa de prueba. La diferencia de rendimiento entre estas etapas está vinculado a la división de los datos (60 % entrenamiento, 20 % validación y 20 % prueba), ya que quedaron pocos datos para probar el modelo. En este sentido, se prefirió entrenar con buena cantidad de datos para que la red pueda converger con indicadores de buen desempeño a coste de dejar pocos datos para validar y probar el modelo. Aún así, tomando un $R = 0,6$ como el más bajo posible para REDungs, se alcanzó un desempeño muy bueno (semejante al detallado para el modelo de bajo costo RED4 en el **Capítulo 4**, alimentado sólo con datos meteorológicos) considerando los datos (en cantidad y calidad) disponibles. En el **Capítulo 2, Apartado 2.1.1**, se explicó que el método de aprendizaje por gradiente conjugado demanda mayor tiempo que el Levenberg-Marquardt, pero de todas formas

el modelo se completó en un tiempo corto de procesamiento con una cantidad aceptable de neuronas y con una baja cantidad de predictores.

Si bien este modelo no es suficiente para validar los datos registrados por ECAUNGS, sus indicadores alientan a considerar estas mediciones de PM como datos relativos que aportan información cualitativa valiosa sobre la CA en la zona. Se encontraron buenas correlaciones y vínculos esperados entre la meteorología, las mediciones de PM en ACUMAR y las propiedades ópticas de aerosoles en UNGS con las mediciones de PM de ECAUNGS. Con una mayor cantidad de mediciones se podrá avanzar en un modelo más robusto y completo.

Como se mencionó, hay pocas mediciones de ECAUNGS para trabajar con ANN, pero de todas formas se realizó una prueba posterior del modelo con datos nuevos de un período de 11 h (**Figura 6.30**). Si bien son pocos datos para sacar mayores conclusiones, se puede observar que los valores estimados siguieron a los medidos correctamente con excepción de los valores entre las horas 4 y 6. En este caso, así como se observó con las pruebas de la RED3 en el **Capítulo 4, Figura 4.24**, el modelo pierde exactitud cuando intenta estimar valores elevados de concentraciones de PM. Los indicadores estimados de rendimiento en esta prueba del modelo fueron R cercano a 0,5 y un RMSE aproximado de $0,1 \mu\text{g}/\text{m}^3$.

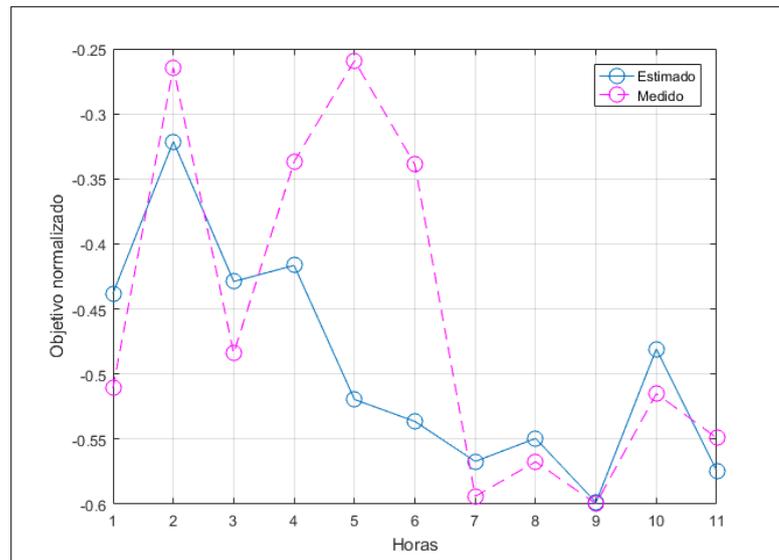


Figura 6.30: Comparación entre valores medidos (línea continua en azul) y estimados (línea discontinua en magenta) por el modelo en un conjuntos de 11 datos elegidos al azar fuera del entrenamiento del modelo. Las escalas de los ejes de las ordenadas se autoajustaron para una mejor visualización de cada subserie, pero los valores normalizados se encuentran entre -1 y 1

A modo de sumario para el modelo integrador de bajo costo desarrollado, se destacan las siguientes características:

- I. **Modelo ANN-MLP-FF-BP dinámico recurrente.**
- II. **Normalización con “mapstd” en datos de entrada y objetivo.** Esto implica que antes de ejecutar el modelo se debe normalizar de esta manera los datos de entrada y se debe denormalizar la salida para contar con las concentraciones reales.
- III. **7 predictores de entrada.** El modelo se ejecuta ingresando los valores de las 7 variables seleccionadas.
- IV. **Vector de 210 pesos implementados en 30 neuronas ocultas.** Esta información se debe considerar para estimar el costo computacional, el cual resulta adecuado para cualquier computador convencional.

El código desarrollado en lenguaje MATLAB y los datos para el diseño y entrenamiento del modelo REDungs se puede consultar en el repositorio digital de los productos de esta Tesis[36].

6.3. Conclusiones preliminares

Aunque las características de rendimiento y la facilidad de uso de los sensores meteorológicos evaluados aquí varían ampliamente, existe una fuerte evidencia de que tienen un uso inmediato en una amplia gama de aplicaciones. Estos dispositivos pueden proporcionar información útil sobre las características ambientales a nivel científico con un marco apropiado que evalúe sus rendimientos. Las pruebas realizadas en laboratorio y en exteriores permitieron contar con resultados bajo diferentes condiciones experimentales, lo que conduce a una mejor comprensión del comportamiento de los sensores. Los resultados obtenidos al respecto no pretenden reemplazar la información de las hojas de datos, pero reflejan comparativamente el funcionamiento de los diferentes sensores presentados en las mismas condiciones.

Los resultados de linealidad muestran que los sensores de humedad relativa presentan algunas dificultades en términos de la dispersión de sus datos reflejados en altos valores de RMSE, especialmente con DHT que tiene regiones no lineales cercanas a sus límites de detección. Los resultados en las mediciones de temperatura fueron satisfactorios para todos los sensores y los mejores ajustes lineales se dieron en las comparaciones de los valores de presión. Las diferencias en las pendientes indican que es necesaria una calibración previa para usar los datos de estos sensores en diferentes proyectos. La exactitud

demostró ser satisfactoria en todos los casos, con una pequeña ventaja del 18b20 en experimentos de laboratorio. La precisión también fue buena para todos los sensores. Los resultados de resolución fueron similares a los reportados por las hojas de datos con el BME como el de división más fina, aunque las diferencias entre 18b20 y BMP resultaron mucho menores en estos experimentos. En cuanto a la presión, el BMP tuvo una mejor resolución que el BME.

Estos resultados de la intercomparación de los sensores de humedad, presión y temperatura permitieron seleccionar de forma fundamentada al BME280 como parte de la estación de calidad de aire ECAUNGS, pero también establecieron diferentes criterios para la elección del sensor que mejor se ajuste a diferentes posibles proyectos.

Las pruebas de laboratorio, campo, comparaciones con el equipo de Cambike y simulaciones con la librería “opcsim” arrojaron resultados que sirvieron para caracterizar al sensor de PM seleccionado para ECAUNGS, el PMS5003. Éste mostró una buena precisión y linealidad, así como un tiempo de respuesta casi inmediato y aceptables niveles de resolución y límites inferiores de detección (diferentes a los informados en la hoja de datos). Los tiempos de subida y bajada son un aspecto importante a considerar cuando se detectan episodios de corta duración, ya que el sensor puede demorarse hasta 1 minuto en alcanzar el valor real de medición y más de 5 minutos en alcanzar nuevamente los valores estables una vez finalizado el evento. Una preocupación puede ser el sesgo del sensor, una deriva que con el paso del tiempo en un funcionamiento continuo puede afectar las mediciones, aunque se necesita comparar con equipo regulatorio para verificarlo. Por otro lado, las pruebas con las soluciones de sulfato de amonio demostraron que el sensor se ve afectado por la humedad, aunque al mismo nivel que las simulaciones siguiendo el crecimiento higroscópico de las partículas. Resta para futuros trabajos realizar pruebas con controles más precisos de humedad y también verificar su comportamiento en torno a la temperatura.

Cabe dedicar unas líneas a la importancia que tuvo el desarrollo de la cámara de calibración para pruebas de laboratorio. Si bien no se le dio lugar en este capítulo de resultados, es otro producto tecnológico de bajo costo en el marco de esta Tesis que ha permitido evaluar sensores meteorológicos y de PM con una buena versatilidad y simple funcionamiento. La pandemia de COVID-19 interrumpió los proyectos experimentales, pero se encuentra pendiente y próximo a completarse la modificación de esta cámara para evaluar sensores de CO_2 .

El desarrollo del equipo ECAUNGS también se vio interrumpido por la pandemia, pero el prototipo final demostró ser compacto y liviano, de fácil colocación, estable en

sus mediciones y autosuficiente en su alimentación eléctrica. Las características determinadas para el sensor de PM permitieron desarrollar un código de gestión de calidad de datos que permite advertir al usuario sobre diferentes situaciones a considerar (como valores fuera de los límites de detección, detección de eventos y valores de humedad que pueden vulnerar las mediciones). Esto completa el control del equipo ECAUNGS, al cual se le incorporará en el futuro un sistema de comunicación de radiofrecuencia con equipo Lora. Este aspecto del desarrollo comenzó a trabajarse, pero también debió postergarse por la pandemia, al igual que las gestiones con agencias regulatorias para realizar comparaciones con equipos oficiales. Hasta que este último aspecto se concrete, las mediciones de ECAUNGS de PM en superficie aportarán información más a un nivel relativo, pero que aun así permite elaborar estudios sin precedentes en la zona.

Con las mediciones de ECAUNGS se detectaron valores bajos de PM en relación con la radiación solar (entre el mediodía y la tarde y en los meses cálidos para la marcha anual), lo cual está vinculado a la dinámica de la capa límite que también se detectó en el **Capítulo 4**. Los percentiles 75 de las concentraciones de PM se encuentran cercanos a los valores más frecuentes medidos en CABA y el modo principal de *PM_{25eca}* iguala al valor medio máximo horario en CABA, aunque nuevamente no se puede determinar con total seguridad que estos valores sean los reales. De todas formas, ninguna de estas concentraciones superan los límites establecidos por ley. Las mediciones de ECAUNGS indican una mayor proporción de la fracción gruesa de partículas en comparación con CABA y esto, según las simulaciones, implican distribuciones de partículas típicas de ambientes rurales, continental, marítimo o antártico. Esto último verifica los hallazgos descritos en el **Capítulo 5**, donde se encontró una atmósfera limpia (sin demasiada intrusión de aerosoles de tipo urbano) con alta frecuencia de ocurrencia de partículas de tipo antártico y marítimo

No se detectaron direcciones del viento con principales contribuciones de PM en la zona, al parecer esto es más evidente para las propiedades ópticas de aerosoles que poseen una componente local y otra más regional (**Capítulo 5**). Posiblemente las fuentes de PM a nivel de superficie en el sitio de medición interfieren más que las contribuciones de las fuentes previamente detectadas al Oeste (centro urbano de San Miguel) y al Norte (sector industrial). En relación con las fuentes, en este capítulo también se presentaron los resultados de la campaña de medición con el equipo de Cambike. Se logró determinar que el tráfico posee una influencia muy importante en las concentraciones de partículas, dado que los mayores valores se corresponden con las rutas principales y sectores donde suele haber vehículos detenidos o en marcha lenta por la presencia de semáforos o congestiones

usuales.

Si bien las mediciones de ECAUNGS aun deben ser consideradas como relativas, el buen desempeño del modelo integrador REDungs indica que los valores de PM medidos se encuentran vinculados de forma lógica con la meteorología de capa límite y otras variables a nivel regional y local. Si bien las correlaciones lineales entre el AOD medido en UNGS y el PM de ECAUNGS son ruidosas y poco claras, el modelo ANN implementado logra interpretar y sacar provecho de vínculos no lineales entre estas variables. Este modelo permite actualmente conocer los valores relativos de PM de forma horaria en la zona, y su aceptable desempeño (comparable con el del modelo RED4 presentado en el **Capítulo 4**) demuestra que la integración de diferentes métodos de bajo costo es posible y superadora a lo que pueda conseguirse con cada parte individual.

El modelo desarrollado para la zona podrá ser mejorado cuando se cuente con una mayor cantidad de mediciones, tanto en superficie como fotométricas. Esto es sólo cuestión de tiempo, ya que el equipo ECAUNGS continua en funcionamiento y las mediciones con el fotómetro solar CALITOO aún se realizan regularmente.

Otro aspecto que enriquecerá enormemente la base de datos del equipo y la mejora del modelo REDungs será el agregado de sensores de gases a la estación ECAUNGS. Los sensores probados no fueron adecuados, pero el MHZ14 de CO_2 posee el potencial para ser calibrado e incorporado. En relación a esto, recientemente se cuenta con un equipo “Allin-Wayra” desarrollado por un grupo del Centro de Investigación en Informática para la Ingeniería de la UTN Regional Córdoba[210], como parte de una colaboración. Este equipo permite realizar mediciones de diferentes gases que complementarán los mediciones de ECAUNGS y alimentarán posteriormente el modelo.

7 Conclusiones generales

La contaminación del aire en entornos urbanos es uno de los mayores problemas en la actualidad a nivel mundial. Multitud de estudios han demostrado el riesgo de la exposición humana a ciertos niveles de concentración y tamaños de partículas en el aire, sin embargo aún existe un gran desconocimiento al respecto en muchas partes, como sucede en la megaciudad que conforman CABA y su conurbano.

La situación se agrava en poblaciones con bajos niveles socioeconómicos, dado que la frágil infraestructura, el acceso restringido a servicios de salud y una calidad de vida desfavorable las hacen especialmente vulnerables a la contaminación del aire. Este es el caso de muchos sectores del AMBA, como lo es para gran parte del Centro-Norte del conurbano bonaerense. Esta región de estudio puede ver deteriorada su calidad del aire, más allá del nivel socioeconómico, por factores tales como un crecimiento urbano poco planificado, instalación de industrias con escasos controles y presencia de basurales improvisados con eventuales quemas.

Los desafíos a enfrentar para determinar la calidad del aire en la región de estudio no son pocos. No hay antecedentes de trabajos con este fin (con excepción de una breve campaña de mediciones pasivas llevada a cabo hace 20 años), dado que no existen mediciones sistemáticas de contaminantes en el aire. La estación meteorológica de Campo de Mayo del SMN es el único sistema de medición de variables relevantes, pero aun así no posee registros en el rango horario de las 21 h a las 6 h. Los equipos regulatorios de medición de calidad del aire tienen costos prohibitivos para gobiernos de recursos económicos limitados como los municipales en torno a la UNGS, y mucho más inalcanzables para los financiamientos destinados a la investigación científica en economías como la de Argentina.

Considerando estas observaciones, la presente Tesis se concibió con el fin de dar respuesta a esta problemática en base a conocimiento científico y tecnológico que logre soslayar estas limitaciones. Se propuso un enfoque de bajo costo, con un pilar puesto en el modelado optimizado de Redes Neuronales que logren extraer la mayor cantidad de información posible de una base de datos acotada, y otro en el diseño y desarrollo

7 Conclusiones generales

tecnológico cimentado en los recursos de los que se pudo disponer para contar con las primeras mediciones sistemáticas en la región.

En primer lugar, se llevó a cabo un modelado con Redes Neuronales de tipo Perceptrón Multicapa para estimar valores de concentraciones de material particulado $PM_{2,5}$ en CABA, con el fin de estudiar el potencial de estas herramientas. La elección de esta zona de estudio se basó en ser la más cercana al Centro-Norte del conurbano con datos de contaminantes disponibles, por lo que fue un primer paso para generar conocimiento acerca de las relaciones entre el $PM_{2,5}$ y la meteorología y para evaluar los alcances de estos modelos cuando se cuenta con una base de datos histórica y multivariada. El diseño de Redes Neuronales es una tarea compleja debido a la cantidad de hiperparámetros que poseen y no hay un protocolo establecido ya que su comportamiento y desempeño es problema-dependiente. Se propuso entonces un procedimiento y una red de prueba fuera del modelo que permita evaluar los vínculos (incluidos los no lineales) entre los potenciales predictores y la variable objetivo. De esta manera, se logró desarrollar un modelo con indicadores de desempeño similares a los mejores casos mostrados en la literatura que es capaz de completar datos de la estación EMC I de ACUMAR en CABA y que permitió extraer algunas conclusiones útiles para su posterior implementación en el Centro-Norte del conurbano. Por un lado, mostró la importancia de identificar fuentes urbanas e industriales respecto a la dirección del viento y, por otro, se identificaron relaciones significativas entre el material particulado y meteorología dada principalmente por la altura de capa de mezcla, la estabilidad, y la humedad relativa. También se desarrolló un modelo de menor costo computacional y de requerimientos de datos como otro punto de referencia.

En segundo lugar, se mostraron los resultados de mediciones fotométricas en la zona de la UNGS con un equipo de bajo costo calibrado. Se encontraron valores bajos de espesor óptico (baja carga aerosólica) con tamaños generalmente grandes de partículas, correspondiendo con aerosoles de tipo continentales, marítimos y antárticos, en ese orden de importancia. Estacionalmente, se midieron los mayores valores de espesor óptico en verano (al igual que lo reportado en la literatura para otros sitios de Argentina), lo cual fue corroborado también con datos satelitales. Otro resultado destacable de los datos satelitales son las “manchas de calor” de mayor concentración de aerosoles en la columna atmosférica sobre los centros urbanos más importantes en la región de estudio y, en contraste, las menores en Campo de Mayo que es el mayor espacio verde de la región. Las regresiones lineales entre las propiedades ópticas de aerosoles y la meteorología dan información limitada y sesgada, por ello se integraron las mediciones de bajo costo con

Mapas Auto-organizados que interpretaron además relaciones no lineales entre las variables. Este enfoque complementado con estudio de trayectorias de masas de aire permitió identificar escenarios típicos de aerosoles en función de las condiciones meteorológicas, además de comprender la presencia de los aerosoles de tipo “Marítimo” y “Antártico”. Uno de esos escenarios es en el verano, con condiciones secas, altas temperaturas y una capa límite profunda que condicionan una carga alta de aerosoles distribuidos verticalmente. Todo lo contrario ocurre en el invierno, además de la presencia de partículas de mayor tamaño. También se verificaron los efectos de “lavado” con las precipitaciones y de dispersión con la velocidad del viento. Respecto a la dirección del viento, los casos de mayor contaminación se vincularon con los sectores al Norte de UNGS (industrias) y al Oeste (conglomerado de San Miguel).

En tercer lugar, se presenta el desarrollo y pruebas de sensores ambientales de bajo costo. Se diseñó una cámara de calibración para realizar pruebas de laboratorio y se realizaron pruebas con otros equipos calibrados, para finalmente seleccionar los sensores que formaron parte del prototipo ECAUNGS. A su vez, se agregó un ciclo de gestión de calidad al algoritmo de control de la estación advirtiéndole sobre posibles situaciones en las que los datos de PM requieran una posterior revisión por parte del usuario. Las mediciones con ECAUNGS mostraron que las partículas a nivel de superficie tienen una relación inversa con la radiación solar diaria y anual (lo cual se explica con una capa límite estrecha que concentra los aerosoles cerca de su base), además de corroborar las observaciones con las propiedades ópticas de aerosoles de la prevalencia de una fracción gruesa de partículas propias de ambientes continentales, marítimos o antárticos (pero poco urbanizados). También se realizó una campaña de mediciones móviles de PM en la zona de la UNGS con un equipo calibrado, mostrando que el tráfico posee una influencia muy determinante en las concentraciones de partículas en superficie. Esto también tiene su correlación con el AOD de acuerdo a los datos satelitales.

Finalmente, se integraron las diferentes metodologías y desarrollos de bajo costo con un modelo de aprendizaje automático. Para ello se utilizaron las mediciones de ECAUNGS, del fotómetro solar de bajo costo y meteorología de diversas fuentes, logrando una buena convergencia y aceptable rendimiento, además de demostrar nuevamente la importancia de la capa límite atmosférica, las condiciones de estabilidad, el espesor óptico de aerosoles y la velocidad del viento. Este modelo permite predecir valores de $PM_{2,5}$ en el Centro-Norte del conurbano bonaerense con una exactitud media que mejorará con mayor cantidad de mediciones.

Los resultados y la experiencia general para el desarrollo de esta Tesis permiten con-

7 Conclusiones generales

cluir que los enfoques de bajo costo (tanto computacional como tecnológicos) son de gran utilidad a la hora de determinar la CA en una región sin antecedentes ni historia de mediciones sistemáticas. Los modelos han demostrado ser muy potentes, entregando información nueva en base a los datos disponibles y estimando valores de PM con una buena exactitud. La disponibilidad actual de plataformas “open source” permiten desarrollos con excelentes prestaciones para mediciones ambientales y posterior manejo de los datos, pero los sensores poseen aun sus limitaciones que deben ser estudiadas en el contexto del proyecto que se pretenda llevar a cabo. Estos dispositivos requieren ser calibrados y validados frente a equipos regulatorios para considerar sus mediciones como absolutas, pero sin ello aun así pueden entregar información valiosa en términos de valores relativos para regiones sin cobertura, como se comprobó para la región de estudio. La integración de los diferentes métodos de bajo costo empleados y adaptados (mediciones de bajo costo con Redes Neuronales Artificiales optimizadas) demostró potenciar los beneficios que posee cada parte por separado. Esto es, la clasificación de escenarios de aerosoles con Mapas Auto-Organizados y mediciones del fotómetro solar de bajo costo, y la predicción de PM a partir de Redes Perceptrón Multicapa con datos de ECAUNGS y del fotómetro CALITOO. La propuesta puede ser adaptada para cualquier otro sitio, y con eso en mente se creó un reservorio digital de libre acceso.

En base a los resultados y observaciones indicadas, el estado de calidad del aire de la región, en cuanto a aerosoles, puede ser considerado como bueno ya que ningún valor supera los límites impuestos por la ley (referida a niveles de PM) y la atmósfera presenta características de una zona poco urbanizada sin mayores episodios de contaminación y con buenas condiciones de dispersión. Sin embargo, se debe considerar que los centros urbanos y el tráfico son las principales fuentes locales de partículas, por lo que se requieren de campañas focalizadas en las zonas de mayor circulación para aseverar los riesgos sobre la salud. Además, si no se planifica el crecimiento de la mancha urbana en el futuro podrán encontrarse episodios cada vez más frecuentes de altas concentraciones de partículas en el aire.

La situación dada por la pandemia de COVID-19 condicionó parte del desarrollo de esta Tesis, ya que debieron postergarse las comparaciones con equipos regulatorios, la incorporación de equipo de transmisión de datos por radiofrecuencia y de sensores de gases para ECAUNGS, además de que las campañas de mediciones se vieron restringidas. De todas formas, las observaciones en base a los resultados obtenidos permitieron dar respuesta a las preguntas iniciales y cumplir con el objetivo principal. Por supuesto que se encuentra en proceso un estudio sobre los efectos de la reducción de actividades

durante las fases más restrictivas dadas por la pandemia en los parámetros trabajados, que no fue incorporado aquí por exceder los alcances propuestos.

Las mediciones fotométricas continúan realizándose de manera sistemática y el equipo ECAUNGS se encuentra en funcionamiento. Por otro lado, una colaboración con la UTN Regional Córdoba permitirá contar con mediciones de gases en la zona. Todos estos datos enriquecerán los modelos desarrollados y mejorarán su desempeño, además de permitir confirmar algunas observaciones que quedaron pendientes de explicación.

El objetivo a largo plazo es contar con una red de monitoreo validada y con múltiples indicadores. Pero los estudios presentados aquí representan un primer esfuerzo y acercamiento a determinar la CA en la zona, y se espera que los resultados expuestos puedan tener un efecto en la toma de decisiones a nivel de planificación urbana.

8 ANEXOS

8.1. Potencialidad de las redes neuronales en calidad del aire

8.1.1. Conversión de datos de plafond

Código Synop en Plafond	
Cifra	Rangos en metros
0	[0;50]
1	[50;100]
2	[100;200]
3	[200;300]
4	[300;600]
5	[600;1000]
6	[1000;1500]
7	[1500;2000]
8	[2000;2500]
9	Más de 2500m, o sin nubes

Tabla 8.1: Altura de la base de capa de nubes de acuerdo al código Synop

8.1.2. Método para el cálculo de estabilidad atmosférica

Se describe a continuación la metodología para el cálculo de la estabilidad atmosférica de acuerdo a Turner[168]. El método determina las clases de estabilidad en función de la velocidad del viento y del índice de radiación neta (RN), como se muestra en la **Tabla 8.2**.

Luego, se define un índice de insolación “ii” en función del ángulo de elevación solar, de acuerdo a la **Tabla 8.3**.

RN							ws (m/s)
4	3	2	1	0	-1	-2	Rangos
1	1	2	3	4	6	7	$0 < ws < 0,8$
1	2	2	3	4	6	7	$0,8 \leq ws < 1,9$
1	2	3	4	4	6	7	$1,9 \leq ws < 2,9$
2	2	3	4	4	5	6	$2,9 \leq ws < 3,4$
2	2	3	4	4	4	5	$3,4 \leq ws < 3,9$
2	3	3	4	4	4	5	$3,9 \leq ws < 4,9$
3	3	4	4	4	4	5	$4,9 \leq ws < 5,5$
3	3	4	4	4	4	4	$5,5 \leq ws < 6$
3	4	4	4	4	4	4	$ws \geq 6,0$

Tabla 8.2: Clasificación de estabilidad atmosférica en función de la velocidad del viento (ws) y el índice de radiación neta (RN)

h (°)	ii
$60 < h$	4
$35 < h < 60$	3
$15 < h < 35$	2
$h < 15$	1

Tabla 8.3: Índices de insolación (ii) en función de la altura del sol sobre el horizonte (h)

El índice “ii” se corrige por cantidad de cielo cubierto (nub) y plafond (pla) para obtener los índices de radiación neta RN, de acuerdo a los criterios establecidos por Santa Cruz (2000) [211]:

I. Si $nub=8/8$ y $pla < 2200m$, RN=0

II. En horas nocturnas (determinadas desde una hora antes de la puesta del sol hasta una hora después de la salida):

a. Si $nub \leq 3/8$, RN= -2

b. Si $nub > 3/8$, RN= -1

III. En horas diurnas:

a. Si $nub \leq 4/8$, RN= ii

b. Si $nub > 4/8$, RN se determina según los siguientes criterios:

b1. Si $h < 2200m$, RN= ii-2

b2. Si $2200m < h < 4900m$, RN= ii-1

b3. Si $nub=8/8$ y $h \leq 2200m$, RN= ii-1

c. Si resultase $RN \leq 0$, RN= 1

Posteriormente, se determina el ángulo de declinación solar δ_s en radianes, ángulo entre la línea Sol-Tierra y el plano ecuatorial celeste, como se indica en la **Ecuación 8.1** [201]:

$$\delta_s = \phi_r \cos \left[\frac{2\pi (d - d_r)}{d_y} \right] \quad (8.1)$$

Siendo ϕ_r la latitud del Trópico de Cancer ($23,45^\circ = 0,409$ radianes), d es el número del día del año, d_r es el día del solsticio de invierno en el hemisferio sur (igual a 173) y d_y es el número promedio de días por año (365,25).

Luego, el ángulo de elevación local h se determinó de acuerdo a la **Ecuación 8.2**:

$$\text{sen}(h) = \text{sen}(\phi) \text{sen}(\delta_s) - \text{cos}(\phi) \text{cos}(\delta_s) \text{cos} \left[\left(\frac{\pi t_{UTC}}{12} - \lambda_e \right) \right] \quad (8.2)$$

Donde ϕ y λ_e son la latitud y longitud geográficas de la estación de medición de ACUMAR (fuente de los datos utilizados en estos cálculos) en radianes, y t_{UTC} es el tiempo coordinado universal en horas.

8.1.3. Tendencias de las variables de entrada del modelo RED3 no incluidas en el Capítulo 4

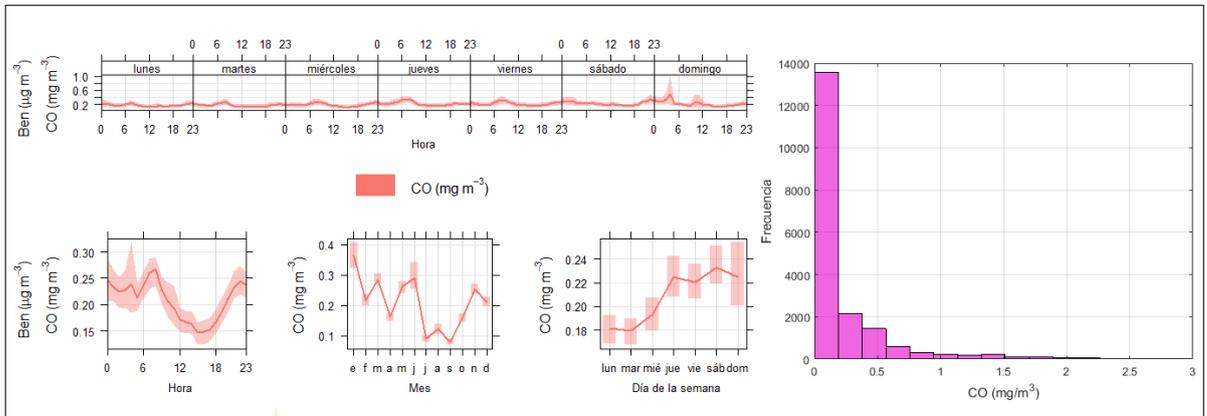


Figura 8.1: Histograma y tendencias diaria y anual de CO

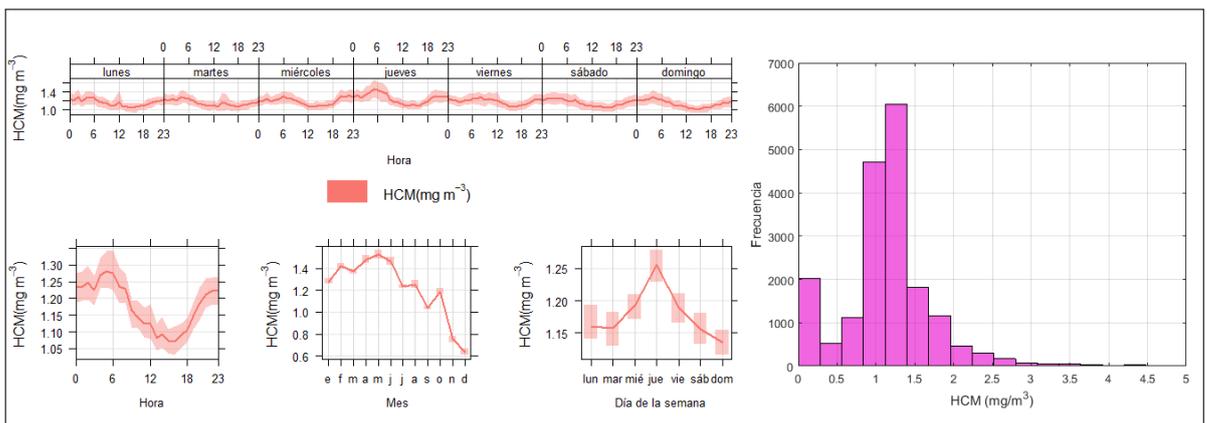


Figura 8.2: Histograma y tendencias diaria y anual de HCM

8.1 Potencialidad de las redes neuronales en calidad del aire

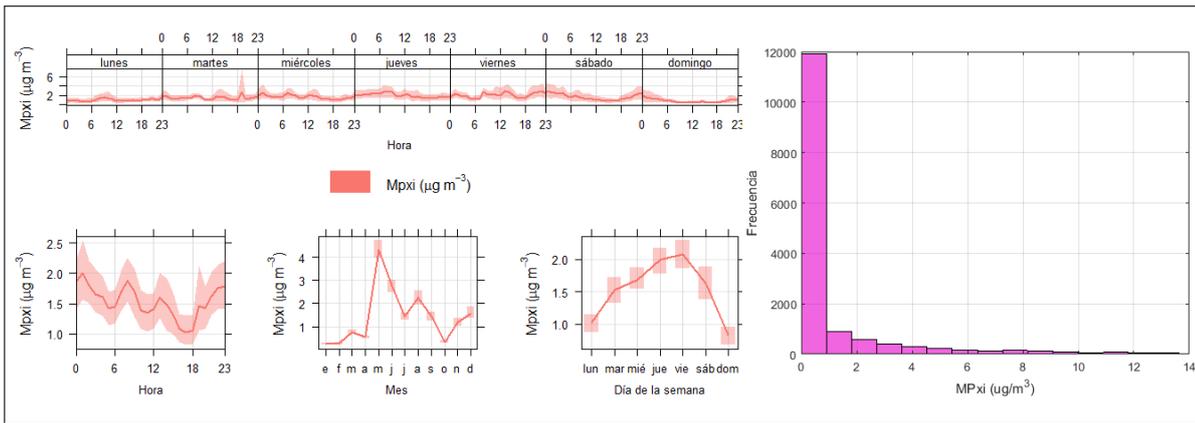


Figura 8.3: Histograma y tendencias diaria y anual de MPxi

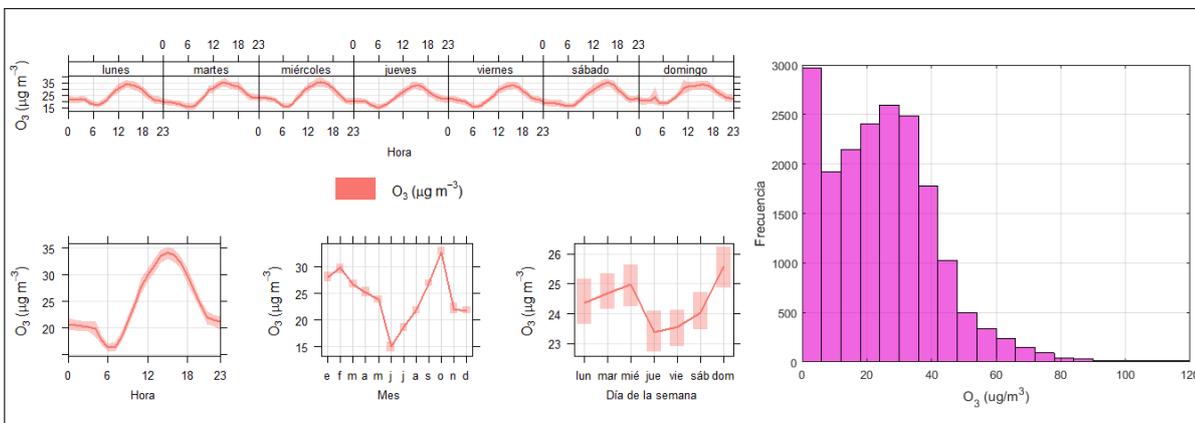


Figura 8.4: Histograma y tendencias diaria y anual de O₃

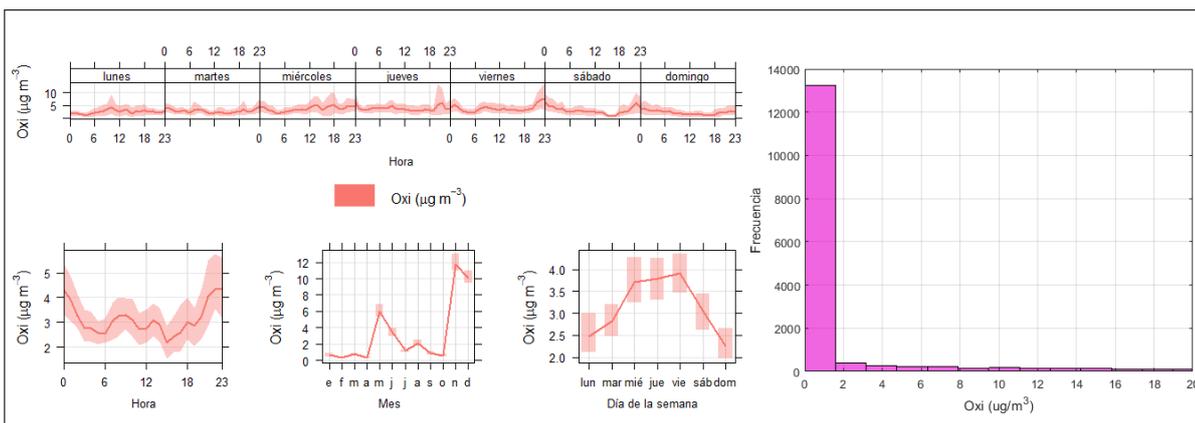


Figura 8.5: Histograma y tendencias diaria y anual de Oxi

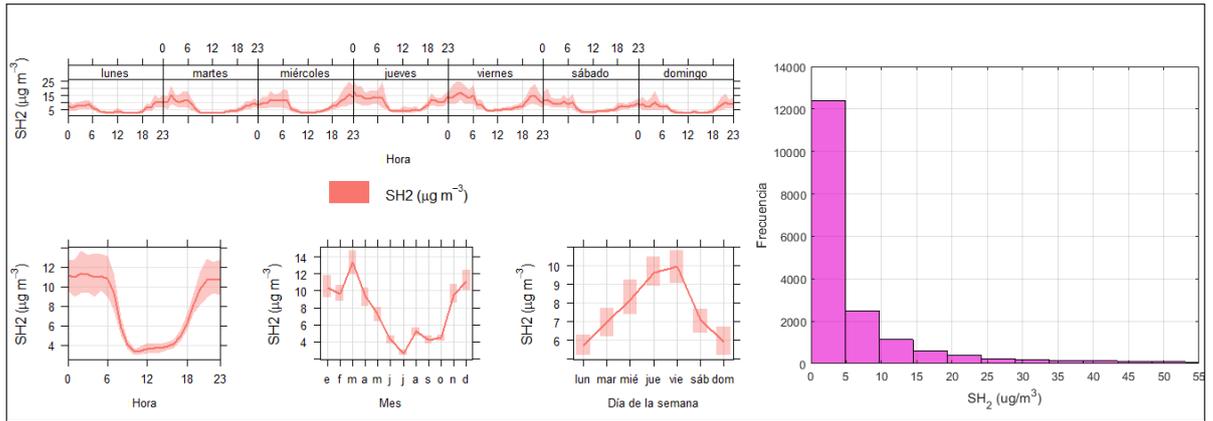


Figura 8.6: Histograma y tendencias diaria y anual de SH_2

8.1.4. Matriz de correlación entre todas las variables consideradas como potenciales entradas del modelo RED3

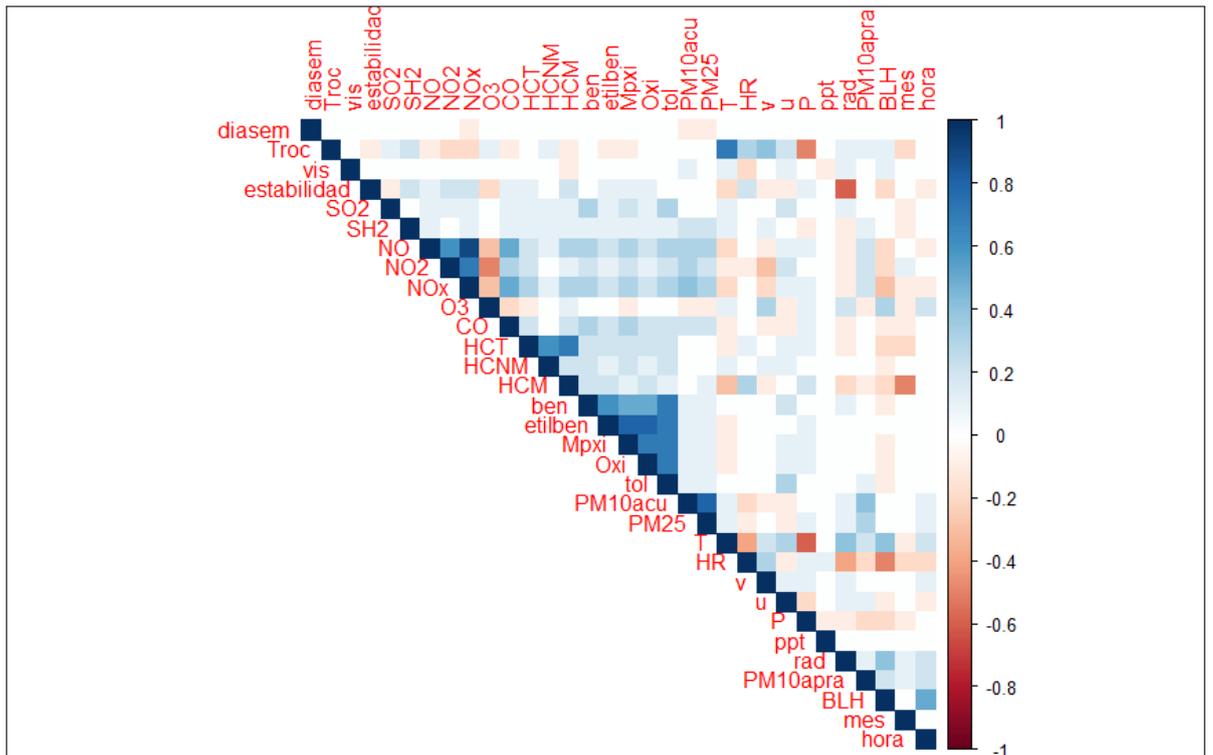


Figura 8.7: Matriz de correlación lineal entre todas las variables. La escala de colores indica el valor del coeficiente de correlación R

8.2. Caracterización de aerosoles

8.2.1. Tendencias de variables no incluidas en el análisis del Capítulo 5

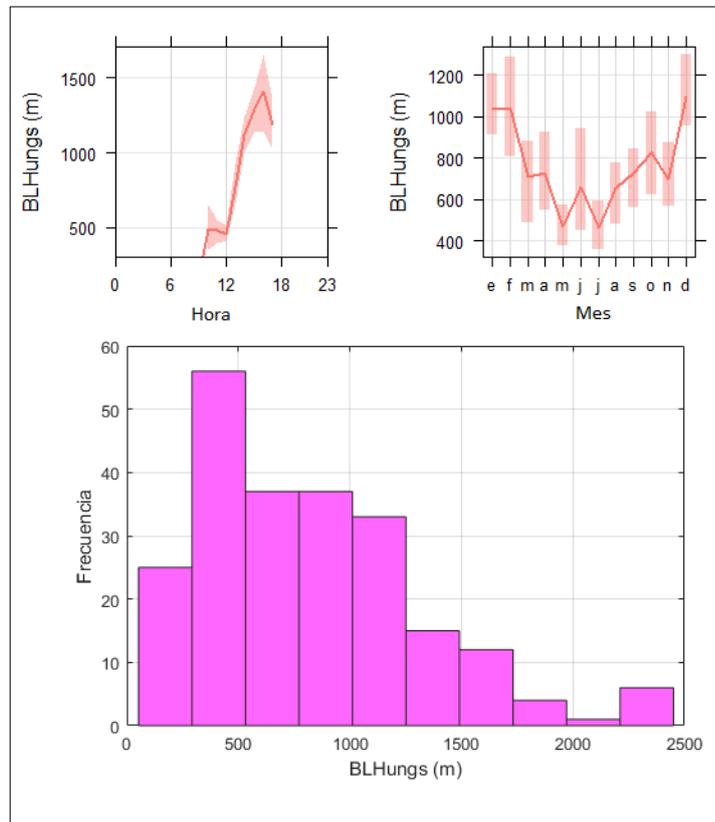


Figura 8.8: Histograma y ciclos diarios y estacionales de la altura de capa mezclada en UNGS (*BLHungs*)

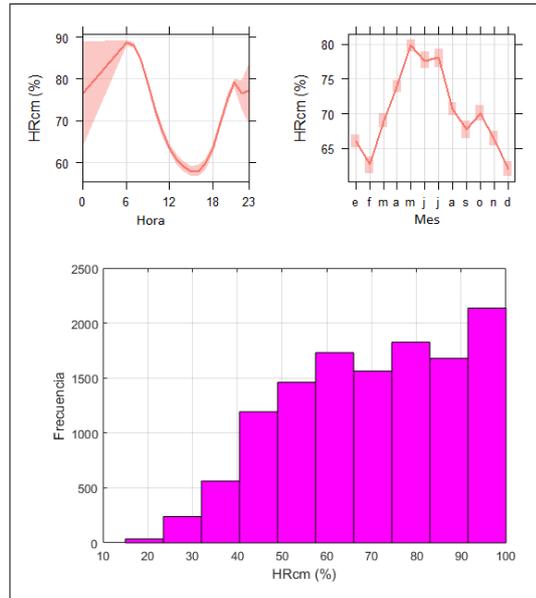


Figura 8.9: Histograma y ciclos diarios y estacionales de la humedad relativa en Campo de Mayo (*HRungs*)

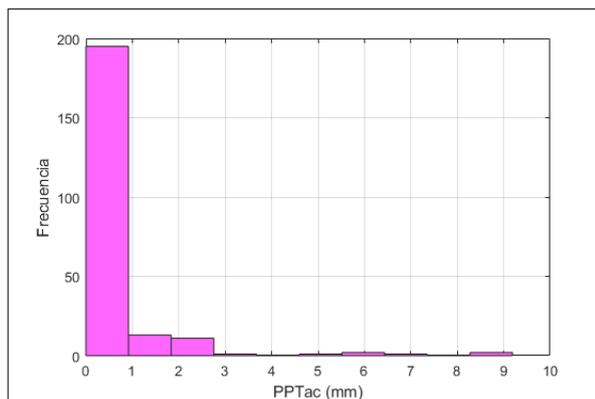


Figura 8.10: Histograma de la precipitación acumulada en la zona de la UNGS (*PP-Tungs*)

8.2.2. Calibración de fotómetro solar CALITOO

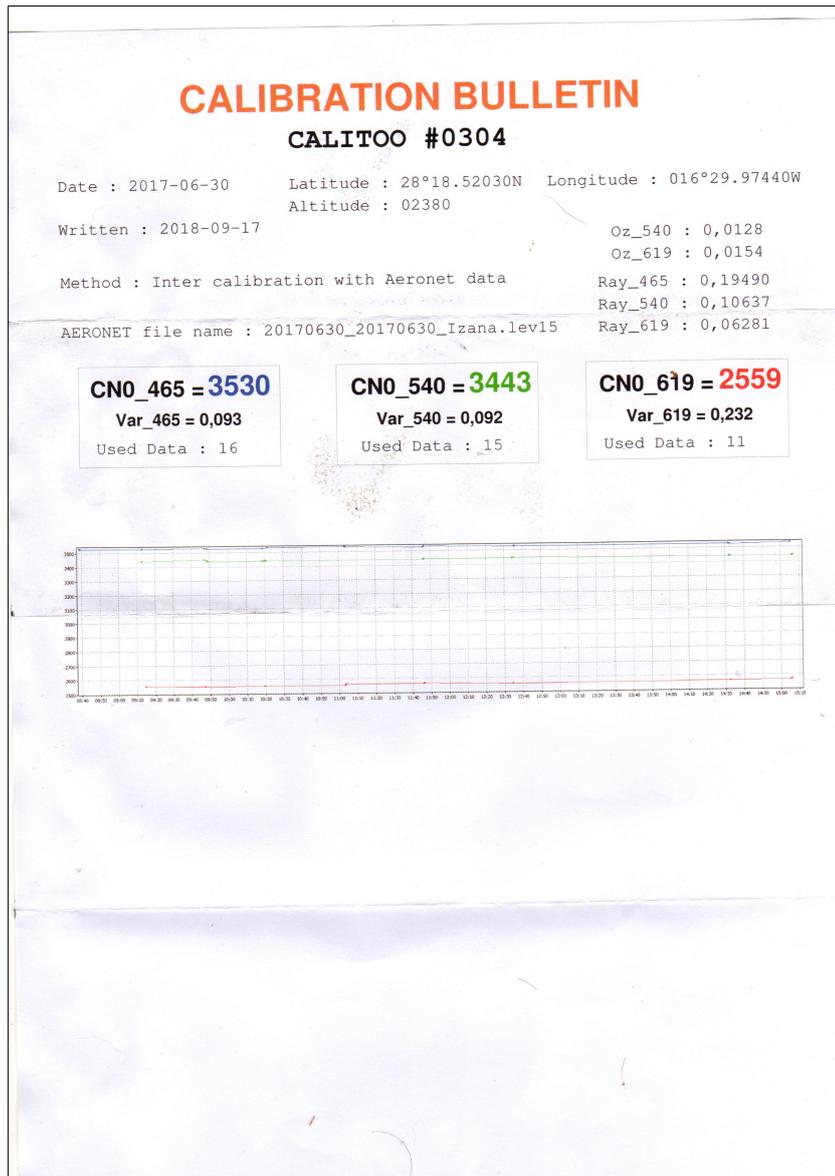


Figura 8.11: Boletín de calibración de fotómetro CALITOO con parámetros informados

8.3. Desarrollo e implementación de equipos de bajo costo

8.3.1. Informe de calibración de Open-Seneca para equipo de Cambike



air quality sensing powered by citizen science

Quality standards of open-seneca sensors

The European Air Quality Directive 2008/50/EC¹ defines reference measurement methods for air quality monitoring and data quality objectives that supplementary sensor platforms should meet to be used for regulatory purposes. The data quality objective for indicative measurements of PM_{2.5} is set at an uncertainty of $\pm 50\%$.

Open-seneca monitoring devices use the Sensirion SPS30 Particulate Matter Sensor. This sensor has received the MCERTS Performance Standards for Indicative Ambient Particulate Monitors for PM_{2.5} in the range of 0 - 75 $\mu\text{g}/\text{m}^3$. The uncertainty of this sensor is $\pm 8.9\%$, complying with the data quality objective for indicative measurements.

The EPA Air sensors guideline² classifies the performance of air quality sensors in five types (Tiers I-V). The suggested performance goals are summarised in the following table:

Tier	Application area	Uncertainty
I	Education and information	$\pm 50\%$
II	Hotspot identification and characterisation	$\pm 30\%$
III	Supplemental monitoring	$\pm 20\%$
IV	Personal exposure	$\pm 30\%$
V	Regulatory monitoring	$\pm 10\%$

According to this classification, the SPS30 sensor used in the open-seneca monitor would be suitable for any of the applications of Tier I-V. The aim of the open-seneca initiative is to empower citizens with data about their personal exposure to particulate pollution (Tier IV), to raise awareness and drive behaviour change (Tier I), and to provide high spatial and temporal resolution pollution maps that highlight hotspots of particulate pollution (Tier II and III) to inform policy. However, in a citizen science setting where the sensors are mobile, the uncertainty required for Tier V might be compromised, and the initiative does not aim to ensure regulatory compliance of cities to current air quality regulations.

In order to ensure the quality of the data collected with open-seneca monitors, a two-stage sensor calibration is performed. The first stage is done by the manufacturer (laboratory calibration under controlled conditions) and the second one is performed by open-seneca (co-localisation with reference station in the pilot city).

A previous pilot run in Buenos Aires, Argentina, shows the feasibility of using these sensors in a mobile network (citizens carrying the sensor on their bikes). The sensors were co-located with a reference station at the US embassy in Buenos Aires (Figure 1). The sampling rate of the sensors was 5s and the data was binned in hourly intervals for comparison with reference data (Figure 2). All open-seneca sensors were extremely coherent with each other (as expected from the manufacturer's laboratory calibration), meaning that their data is comparable even if placed in different locations. The sensors under-reported PM_{2.5} concentrations, but overall followed a similar trend to the reference.

¹ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32008L0050&from=en>

² <https://www.epa.gov/air-sensor-toolbox/how-use-air-sensors-air-sensor-guidebook#pane-1>

1/3

Quality standards of open-seneca sensors

The European Air Quality Directive 2008/50/EC¹ defines reference measurement methods for air quality monitoring and data quality objectives that supplementary sensor platforms should meet to be used for regulatory purposes. The data quality objective for indicative measurements of PM_{2.5} is set at an uncertainty of $\pm 50\%$.

Open-seneca monitoring devices use the Sensirion SPS30 Particulate Matter Sensor. This sensor has received the MCERTS Performance Standards for Indicative Ambient Particulate Monitors for PM_{2.5} in the range of 0 - 75 $\mu\text{g}/\text{m}^3$. The uncertainty of this sensor is $\pm 8.9\%$, complying with the data quality objective for indicative measurements.

The EPA Air sensors guideline² classifies the performance of air quality sensors in five types (Tiers I-V). The suggested performance goals are summarised in the following table:

Tier	Application area	Uncertainty
I	Education and information	$\pm 50\%$
II	Hotspot identification and characterisation	$\pm 30\%$
III	Supplemental monitoring	$\pm 20\%$
IV	Personal exposure	$\pm 30\%$
V	Regulatory monitoring	$\pm 10\%$

According to this classification, the SPS30 sensor used in the open-seneca monitor would be suitable for any of the applications of Tier I-V. The aim of the open-seneca initiative is to empower citizens with data about their personal exposure to particulate pollution (Tier IV), to raise awareness and drive behaviour change (Tier I), and to provide high spatial and temporal resolution pollution maps that highlight hotspots of particulate pollution (Tier II and III) to inform policy. However, in a citizen science setting where the sensors are mobile, the uncertainty required for Tier V might be compromised, and the initiative does not aim to ensure regulatory compliance of cities to current air quality regulations.

In order to ensure the quality of the data collected with open-seneca monitors, a two-stage sensor calibration is performed. The first stage is done by the manufacturer (laboratory calibration under controlled conditions) and the second one is performed by open-seneca (co-localisation with reference station in the pilot city).

A previous pilot run in Buenos Aires, Argentina, shows the feasibility of using these sensors in a mobile network (citizens carrying the sensor on their bikes). The sensors were co-located with a reference station at the US embassy in Buenos Aires (Figure 1). The sampling rate of the sensors was 5s and the data was binned in hourly intervals for comparison with reference data (Figure 2). All open-seneca sensors were extremely coherent with each other (as expected from the manufacturer's laboratory calibration), meaning that their data is comparable even if placed in different locations. The sensors under-reported PM_{2.5} concentrations, but overall followed a similar trend to the reference.

¹ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32008L0050&from=en>

² <https://www.epa.gov/air-sensor-toolbox/how-use-air-sensors-air-sensor-guidebook#pane-1>

Quality standards of open-seneca sensors

The European Air Quality Directive 2008/50/EC¹ defines reference measurement methods for air quality monitoring and data quality objectives that supplementary sensor platforms should meet to be used for regulatory purposes. The data quality objective for indicative measurements of PM_{2.5} is set at an uncertainty of $\pm 50\%$.

Open-seneca monitoring devices use the Sensirion SPS30 Particulate Matter Sensor. This sensor has received the MCERTS Performance Standards for Indicative Ambient Particulate Monitors for PM_{2.5} in the range of 0 - 75 $\mu\text{g}/\text{m}^3$. The uncertainty of this sensor is $\pm 8.9\%$, complying with the data quality objective for indicative measurements.

The EPA Air sensors guideline² classifies the performance of air quality sensors in five types (Tiers I-V). The suggested performance goals are summarised in the following table:

Tier	Application area	Uncertainty
I	Education and information	$\pm 50\%$
II	Hotspot identification and characterisation	$\pm 30\%$
III	Supplemental monitoring	$\pm 20\%$
IV	Personal exposure	$\pm 30\%$
V	Regulatory monitoring	$\pm 10\%$

According to this classification, the SPS30 sensor used in the open-seneca monitor would be suitable for any of the applications of Tier I-V. The aim of the open-seneca initiative is to empower citizens with data about their personal exposure to particulate pollution (Tier IV), to raise awareness and drive behaviour change (Tier I), and to provide high spatial and temporal resolution pollution maps that highlight hotspots of particulate pollution (Tier II and III) to inform policy. However, in a citizen science setting where the sensors are mobile, the uncertainty required for Tier V might be compromised, and the initiative does not aim to ensure regulatory compliance of cities to current air quality regulations.

In order to ensure the quality of the data collected with open-seneca monitors, a two-stage sensor calibration is performed. The first stage is done by the manufacturer (laboratory calibration under controlled conditions) and the second one is performed by open-seneca (co-localisation with reference station in the pilot city).

A previous pilot run in Buenos Aires, Argentina, shows the feasibility of using these sensors in a mobile network (citizens carrying the sensor on their bikes). The sensors were co-located with a reference station at the US embassy in Buenos Aires (Figure 1). The sampling rate of the sensors was 5s and the data was binned in hourly intervals for comparison with reference data (Figure 2). All open-seneca sensors were extremely coherent with each other (as expected from the manufacturer's laboratory calibration), meaning that their data is comparable even if placed in different locations. The sensors under-reported PM_{2.5} concentrations, but overall followed a similar trend to the reference.

¹ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32008L0050&from=en>
² <https://www.epa.gov/air-sensor-toolbox/how-use-air-sensors-air-sensor-guidebook#pane-1>

8.3.2. Correlaciones lineales entre todas las potenciales variables de entrada para el modelo REDungs

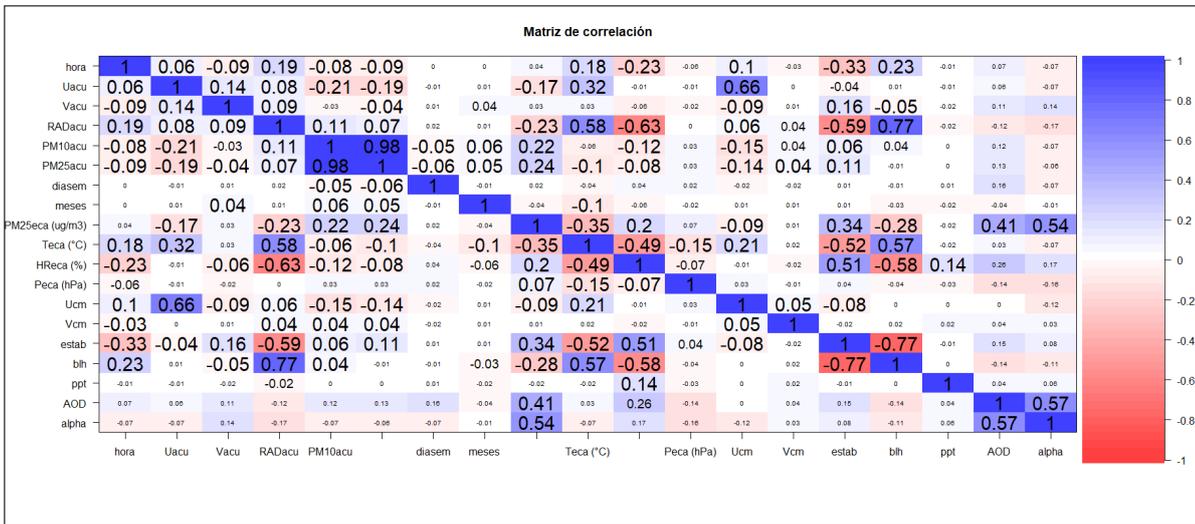


Figura 8.12: Matriz de correlación lineal entre todas las variables consideradas para entrada del modelo REDungs. Los números en cada celda indican el R de Spearman, lo cual escala del color y tamaño de letra

Bibliografía

- [1] Malek Bentayeb y col. “Association between long-term exposure to air pollution and mortality in France: a 25-year follow-up study”. En: *Environment international* 85 (2015), págs. 5-14.
- [2] Mathilde Pascal y col. “Assessing the public health impacts of urban air pollution in 25 European cities: results of the Aphekom project”. En: *Science of the Total Environment* 449 (2013), págs. 390-400.
- [3] Ole Raaschou-Nielsen y col. “Particulate matter air pollution components and risk for lung cancer”. En: *Environment international* 87 (2016), págs. 66-73.
- [4] Congbo Song y col. “Health burden attributable to ambient PM_{2.5} in China”. En: *Environmental pollution* 223 (2017), págs. 575-586.
- [5] WHO. *Health and air pollution*. 2020. URL: <https://www.who.int/health-topics/air-pollution> (visitado 23-05-2020).
- [6] Matthew Limb. *Half of wealthy and 98 % of poorer cities breach air quality guidelines*. 2016.
- [7] Yves Rybarczyk y Rasa Zalakeviciute. “Machine learning approaches for outdoor air quality modelling: A systematic review”. En: *Applied Sciences* 8.12 (2018), pág. 2570.
- [8] Raghavendra V Kulkarni y Ganesh Kumar Venayagamoorthy. “Particle swarm optimization in wireless-sensor networks: A brief survey”. En: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 41.2 (2010), págs. 262-267.
- [9] O Dubovik y col. “Accuracy assessments of aerosol optical properties retrieved from Aerosol Robotic Network (AERONET) Sun and sky radiance measurements”. En: *Journal of Geophysical Research: Atmospheres* 105.D8 (2000), págs. 9791-9806.

- [10] Brent N Holben y col. “AERONET—A federated instrument network and data archive for aerosol characterization”. En: *Remote sensing of environment* 66.1 (1998), págs. 1-16.
- [11] Yoram J Kaufman, Didier Tanré y Olivier Boucher. “A satellite view of aerosols in the climate system”. En: *Nature* 419.6903 (2002), págs. 215-223.
- [12] Houda Yahi. “Méthodologie de prévision de la pollution atmosphérique particulaire: vers une télédétection satellitaire”. En: Littoral. 2010.
- [13] LA Jimoda. “Effects of particulate matter on human health, the ecosystem, climate and materials: a review”. En: *Facta universitatis-series: Working and Living Environmental Protection* 9.1 (2012), págs. 27-44.
- [14] IARC. *IARC: Outdoor Air Pollution a Leading Environmental Cause of Cancer Deaths*. 2021. URL: <https://www.iarc.who.int/news-events/iarc-outdoor-air-pollution-a-leading-environmental-cause-of-cancer-deaths> (visitado 01-03-2021).
- [15] EEA. *EEA: Air quality in Europe - 2020 report*. 2021. URL: <https://www.eea.europa.eu/publications/air-quality-in-europe-2020-report> (visitado 01-03-2021).
- [16] Feng Niu y Zhanqing Li. “Systematic variations of cloud top temperature and precipitation rate with aerosols over the global tropics”. En: *Atmospheric Chemistry and Physics* 12.18 (2012), págs. 8491-8498.
- [17] K Raghavendra Kumar y col. “Aerosol climatology and discrimination of aerosol types retrieved from MODIS, MISR and OMI over Durban (29.88 S, 31.02 E), South Africa”. En: *Atmospheric Environment* 117 (2015), págs. 9-18.
- [18] M Calvello y col. “Physical and optical properties of atmospheric aerosols by in-situ and radiometric measurements”. En: *Atmospheric Chemistry and Physics* 10.5 (2010), págs. 2195-2208.
- [19] Yuanyuan Chu y col. “A review on predicting ground PM_{2.5} concentration using satellite aerosol optical depth”. En: *Atmosphere* 7.10 (2016), pág. 129.
- [20] Luigi y Del Giudice Antonio y De Vito Saverio y Di Francia Girolamo y Esposito Elena y Formisano Fabrizio y Massera Ettore y Miglietta Maria Lucia y Polichetti Tiziana Alfano Brigida y Barretta. “Una revisión de los sensores de material particulado de bajo costo desde la perspectiva de los desarrolladores”. En: *Sensores* ().

- [21] Fabiana Franceschi, Martha Cobo y Manuel Figueredo. “Discovering relationships and forecasting PM10 and PM2.5 concentrations in Bogotá, Colombia, using artificial neural networks, principal component analysis, and k-means clustering”. En: *Atmospheric Pollution Research* 9.5 (2018), págs. 912-922.
- [22] EPA. *US EPA models*. 2020. URL: <https://www.epa.gov/scram/air-quality-models> (visitado 23-05-2020).
- [23] UK MET OFFICE. *MET Office models*. 2020. URL: www.metoffice.gov.uk/research/approach/modelling-systems (visitado 23-05-2020).
- [24] CIAM. *EMEP model*. 2020. URL: www.emep.int/models (visitado 23-05-2020).
- [25] NOAA. *NOAA model WRF-CHEM*. 2020. URL: <https://ruc.noaa.gov/wrf/wrf-chem/> (visitado 23-05-2020).
- [26] Xiao Feng y col. “Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation”. En: *Atmospheric Environment* 107 (2015), págs. 118-128.
- [27] Imad A Basheer y Maha Hajmeer. “Artificial neural networks: fundamentals, computing, design, and application”. En: *Journal of microbiological methods* 43.1 (2000), págs. 3-31.
- [28] Hezhong Tian y col. “Current status and future trends of SO2 and NOx pollution during the 12th FYP period in Guiyang city of China”. En: *Atmospheric environment* 69 (2013), págs. 273-280.
- [29] UNGS. *Universidad Nacional de General Sarmiento*. 2021. URL: <https://www.ungs.edu.ar/> (visitado 10-04-2021).
- [30] G Miño ML & Alsina. *Diagnóstico Ambiental del Partido de San Miguel*. UNGS Ediciones, 2006.
- [31] LA Fagundez y col. “Preliminary air pollution monitoring in San Miguel, Buenos Aires”. En: *Environmental monitoring and assessment* 71.1 (2001), págs. 61-70.
- [32] APrA. *Agencia de Protección Ambiental de Buenos Aires*. 2021. URL: <https://www.buenosaires.gob.ar/agenciaambiental> (visitado 03-04-2021).
- [33] Autoridad de la Cuenca Matanza-Riachuelo. *Monitoreo de calidad de aire ACUMAR*. 2020. URL: <http://www.acumar.gob.ar/monitoreo-ambiental/calidad-de-aire/> (visitado 25-05-2020).

- [34] NASA. *Watch Programme*. 2020. URL: <https://aeronet.gsfc.nasa.gov/> (visitado 19-05-2020).
- [35] MAyDS. *Red FEMA*. 2020. URL: <https://redfema.ambiente.gob.ar/monitor/aire> (visitado 23-05-2020).
- [36] SCAGLIOTTI. *REPOSITORIO TESIS*. 2021. URL: <https://github.com/ascaglio/Estrategias-de-bajo-costo-para-la-exploracion-de-la-calidad-del-aire.git> (visitado 03-08-2021).
- [37] Neil D Bennett y col. “Characterising performance of environmental models”. En: *Environmental Modelling & Software* 40 (2013), págs. 1-20.
- [38] Jessica Spate y col. “Data mining as a tool for environmental scientists”. En: *Proceedings of the iEMSs Third Biennial Meeting: Summit on Environmental Modelling and Software*. International Environmental Modelling y Software Society. 2006.
- [39] Zhi-Hua Zhou. *Three perspectives of data mining*. 2003.
- [40] Sheen Mclean S Cabaneros, John Kaiser Calautit y Ben Richard Hughes. “A review of artificial neural network models for ambient air pollution prediction”. En: *Environmental Modelling & Software* (2019).
- [41] Ángel Arroyo y col. “Neural models for imputation of missing ozone data in air-quality datasets”. En: *Complexity* 2018 (2018).
- [42] Antje M Moffat y col. “Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes”. En: *Agricultural and Forest Meteorology* 147.3-4 (2007), págs. 209-232.
- [43] W Macculloch y W Pitts. “A logical Calculus of Ideas Immanent in Nervous Activity. Bull. Mathematical Biophysics. Vol. 5”. En: (1943).
- [44] Frank Rosenblatt. “Perceptron simulation experiments”. En: *PROCEEDINGS OF THE INSTITUTE OF RADIO ENGINEERS*. Vol. 47. 3. 1959, págs. 468-468.
- [45] Marvin Minsky y Seymour Papert. “An introduction to computational geometry”. En: *Cambridge tiass., HIT* (1969).
- [46] Rudolf Kruse y col. “Multilayer perceptrons”. En: *Computational Intelligence*. Springer, 2016, págs. 47-92.

- [47] James L McClelland, David E Rumelhart, PDP Research Group y col. “Parallel distributed processing”. En: *Explorations in the Microstructure of Cognition 2* (1986), págs. 216-271.
- [48] Jean-François Mas y col. “Modelling deforestation using GIS and artificial neural networks”. En: *Environmental Modelling & Software* 19.5 (2004), págs. 461-471.
- [49] Vladan Babovic. “Data mining in hydrology”. En: *Hydrological Processes: An International Journal* 19.7 (2005), págs. 1511-1515.
- [50] Andréa Oliveira Souza Da Costa y col. “Use of neural networks for monitoring surface water quality changes in a neotropical urban stream”. En: *Environmental monitoring and assessment* 155.1-4 (2009), págs. 527-538.
- [51] Friedrich Recknagel. “Applications of machine learning to ecological modelling”. En: *Ecological modelling* 146.1-3 (2001), págs. 303-310.
- [52] Maurice Dixon y col. “Experience with data mining for the anaerobic wastewater treatment process”. En: *Environmental Modelling & Software* 22.3 (2007), págs. 315-322.
- [53] Ashif Panakkat y Hojjat Adeli. “Neural network models for earthquake magnitude prediction using multiple seismicity indicators”. En: *International journal of neural systems* 17.01 (2007), págs. 13-33.
- [54] Caren Marzban y Gregory J Stumpf. “A neural network for tornado prediction based on Doppler radar-derived attributes”. En: *Journal of applied meteorology* 35.5 (1996), págs. 617-626.
- [55] Donald W McCann. “A neural network short-term forecast of significant thunderstorms”. En: *Weather and Forecasting* 7.3 (1992), págs. 525-534.
- [56] HD Navone y HA Ceccatto. “Predicting Indian monsoon rainfall: a neural network approach”. En: *Climate Dynamics* 10.6-7 (1994), págs. 305-312.
- [57] David Elizondo, Gerrit Hoogenboom y RW McClendon. “Development of a neural network model to predict daily solar radiation”. En: *Agricultural and Forest Meteorology* 71.1-2 (1994), págs. 115-132.
- [58] David W Aha y Richard L Bankert. “Feature selection for case-based classification of cloud types: An empirical comparison”. En: *Proceedings of the AAAI-94 workshop on Case-Based Reasoning*. Vol. 106. 1994, pág. 112.

- [59] Gavin C Cawley y Stephen R Dorling. “Reproducing a subjective classification scheme for atmospheric circulation patterns over the United Kingdom using a neural network”. En: *International Conference on Artificial Neural Networks*. Springer. 1996, págs. 281-286.
- [60] Dimitris Voukantsis y col. “Intercomparison of air quality data using principal component analysis, and forecasting of PM10 and PM2.5 concentrations using artificial neural networks, in Thessaloniki and Helsinki”. En: *Science of the Total Environment* 409.7 (2011), págs. 1266-1276.
- [61] Z Ceylan y SEROL Bulkan. “Forecasting PM10 levels using ANN and MLR: A case study for Sakarya City”. En: *Global Nest Journal* 20.2 (2018), págs. 281-90.
- [62] M Asghari y H Nematzadeh. “Predicting air pollution in Tehran: Genetic algorithm and back propagation neural network”. En: *Journal of AI and Data Mining* 4.1 (2016), págs. 49-54.
- [63] Aaron Daly y Paolo Zannetti. “Air pollution modeling—An overview”. En: *Ambient air pollution* (2007), págs. 15-28.
- [64] Teuvo Kohonen. “Self-organized formation of topologically correct feature maps”. En: *Biological cybernetics* 43.1 (1982), págs. 59-69.
- [65] Craig M Risien y col. “Variability in satellite winds over the Benguela upwelling system during 1999–2000”. En: *Journal of Geophysical Research: Oceans* 109.C3 (2004).
- [66] Patricia Melin y col. “Analysis of spatial spread relationships of coronavirus (COVID-19) pandemic in the world using self organizing maps”. En: *Chaos, Solitons & Fractals* 138 (2020), págs. 109917.
- [67] Judith E Dayhoff. *Neural network architectures: an introduction*. Van Nostrand Reinhold Co., 1990.
- [68] Tatiana Tambouratzis y George Tambouratzis. “Meteorological data analysis using self-organizing maps”. En: *International Journal of Intelligent Systems* 23.6 (2008), págs. 735-759.
- [69] Christian Jensen y Björn Trägårdh. “Temporära organisationer för permanenta problem. Om implementering av samverkansprojekt för unga som står långt från arbetslivet, Lund, Företagsekonomiska institutionen”. En: *Management & Organisation* (2012).

- [70] BC Hewitson y Robert George Crane. “Self-organizing maps: applications to synoptic climatology”. En: *Climate Research* 22.1 (2002), págs. 13-26.
- [71] Mikko Kolehmainen y col. “Forecasting air quality parameters using hybrid neural network modelling”. En: *Environmental Monitoring and Assessment* 65.1-2 (2000), págs. 277-286.
- [72] Antonio Neme y Leticia Hernández. “Visualizing patterns in the air quality in Mexico City with self-organizing maps”. En: *International Workshop on Self-Organizing Maps*. Springer. 2011, págs. 318-327.
- [73] K Srinivasa Raju y D Nagesh Kumar. “Classification of Indian meteorological stations using cluster and fuzzy cluster analysis, and Kohonen artificial neural networks”. En: *Hydrology Research* 38.3 (2007), págs. 303-314.
- [74] Björn A Malmgren y Amos Winter. “Climate zonation in Puerto Rico based on principal components analysis and an artificial neural network”. En: *Journal of climate* 12.4 (1999), págs. 977-985.
- [75] Dantong Liu y col. “Size distribution, mixing state and source apportionment of black carbon aerosol in London during wintertime”. En: *Atmospheric Chemistry and Physics* 14.18 (2014), págs. 10061-10084.
- [76] Jure Zupan y Johann Gasteiger. *Neural networks for chemists: an introduction*. John Wiley & Sons, Inc., 1993.
- [77] Matt W Gardner y SR Dorling. “Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences”. En: *Atmospheric environment* 32.14-15 (1998), págs. 2627-2636.
- [78] MT Hagan, HB Demuth y MH Beale. “Neural network design, PWS Pub”. En: *Co., Boston, Mass., USA* (1995).
- [79] Kurt Hornik. “Approximation capabilities of multilayer feedforward networks”. En: *Neural networks* 4.2 (1991), págs. 251-257.
- [80] Kurt Hornik. “Some new results on neural network approximation”. En: *Neural networks* 6.8 (1993), págs. 1069-1072.
- [81] Daniel Svozil, Vladimir Kvasnicka y Jiri Pospichal. “Introduction to multi-layer feed-forward neural networks”. En: *Chemometrics and intelligent laboratory systems* 39.1 (1997), págs. 43-62.

- [82] LE Scales. *Introduction to non-linear optimization*. Macmillan International Higher Education, 1985.
- [83] Hagan Demuth Beale, Howard B Demuth y MT Hagan. “Neural network design”. En: *Pws, Boston* (1996).
- [84] Gabriel Ibarra-Berastegi y col. “From diagnosis to prognosis for forecasting air pollution using neural networks: Air pollution monitoring in Bilbao”. En: *Environmental Modelling & Software* 23.5 (2008), págs. 622-637.
- [85] Sandhya Samarasinghe. *Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition*. Crc Press, 2016.
- [86] Patricio Perez. “Prediction of sulfur dioxide concentrations at a site near downtown Santiago, Chile”. En: *Atmospheric Environment* 35.29 (2001), págs. 4929-4935.
- [87] Teuvo Kohonen. “Exploration of very large databases by self-organizing maps”. En: *Proceedings of international conference on neural networks (icnn'97)*. Vol. 1. IEEE. 1997, PL1-PL6.
- [88] Alfred Ultsch. “Kohonen’s self organizing feature maps for exploratory data analysis”. En: *Proc. INNC90* (1990), págs. 305-308.
- [89] Teuvo Kohonen. *Self-organization maps*. 1995.
- [90] Aaron Van Donkelaar y col. “Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application”. En: *Environmental health perspectives* 118.6 (2010), págs. 847-855.
- [91] LA Otero y col. “Estadística de la evolución de los aerosoles medidos en las estaciones de la red AERONET en Argentina desde 1999”. En: *Anales AFA*. Vol. 15. 2003, págs. 23-328.
- [92] L Otero y col. “Caracterización de un proceso de intrusión de aerosoles en Córdoba mediante un fotómetro solar de la Red AERONET”. En: *Proceeding of AFA* 16 (2005), págs. 41-44.
- [93] Lidia Otero y col. “Espesor óptico de aerosoles durante el año 2002 para diez estaciones pertenecientes a la red AERONET–NASA”. En: *Opt. Pura Apl* 39.4 (2006), págs. 355-364.
- [94] Lidia Ana Otero y col. “ANÁLISIS DE 12 AÑOS DE MEDICION DE UN FOTOMETRO SOLAR DE LA RED AERONET INSTALADO EN FALDA DEL CARMEN, PROVINCIA DE CORDOBA”. En: *ANALES AFA*. Vol. 24. 1. 2013.

- [95] Julio Miguel Angeles Suazo. “Efecto del espesor óptico de aerosol en el forzamiento radiativo en la estación antártica peruana Machu Picchu”. En: (2013).
- [96] Thomas F Eck y col. “Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols”. En: *Journal of Geophysical Research: Atmospheres* 104.D24 (1999), págs. 31333-31349.
- [97] World Meteorological Organization. *Watch Programme*. 2020. URL: <https://community.wmo.int/activity-areas/gaw> (visitado 19-05-2020).
- [98] David R Brooks y Forrest M Mims III. “Development of an inexpensive handheld LED-based Sun photometer for the GLOBE program”. En: *Journal of Geophysical Research: Atmospheres* 106.D5 (2001), págs. 4733-4740.
- [99] Nicolas Sifakis y Pierre-Yves Deschamps. “Mapping of air pollution using SPOT satellite data”. En: *Photogrammetric Engineering and Remote Sensing* 58 (1992), págs. 1433-1433.
- [100] Diofantos G Hadjimitsis, CRI Clayton y Adrianos Retalis. “The use of selected pseudo-invariant targets for the application of atmospheric correction in multi-temporal studies using satellite remotely sensed imagery”. En: *International Journal of Applied Earth Observation and Geoinformation* 11.3 (2009), págs. 192-200.
- [101] Jerónimo José Escribano Alisio y col. “Satellite Retrievals of Aerosol Optical Depth over a Subtropical Urban Area: The Role of Stratification and Surface Reflectance”. En: (2014).
- [102] Lorraine A Remer y col. “The MODIS aerosol algorithm, products, and validation”. En: *Journal of the atmospheric sciences* 62.4 (2005), págs. 947-973.
- [103] Robert C Levy y col. “Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance”. En: *Journal of Geophysical Research: Atmospheres* 112.D13 (2007).
- [104] RC Levy y col. “The Collection 6 MODIS aerosol products over land and ocean”. En: *Atmospheric Measurement Techniques* 6.11 (2013), págs. 2989-3034.
- [105] Alexei Lyapustin y col. “MODIS collection 6 MAIAC algorithm”. En: *Atmospheric Measurement Techniques* 11.10 (2018), págs. 5741-5765.
- [106] F Toledo y col. “AOT Retrieval Procedure for Distributed Measurements With Low-Cost Sun Photometers”. En: *Journal of Geophysical Research: Atmospheres* 123.2 (2018), págs. 1113-1131.

- [107] Gohar Ali y col. “Spatiotemporal trends of aerosols over urban regions in Pakistan and their possible links to meteorological parameters”. En: *Atmosphere* 11.3 (2020), pág. 306.
- [108] Foued Chabane, Ali Arif y Said Benramache. “The Estimate of Aerosol Optical Depth for Diverse Meteorological Conditions The Estimate of Aerosol Optical Depth for Diverse Meteorological Conditions”. En: ().
- [109] Xingfa Gu y col. “The impacts of regional transport and meteorological factors on aerosol optical depth over Beijing, 1980–2014”. En: *Scientific reports* 8.1 (2018), págs. 1-12.
- [110] Bushra Khalid y col. “Estimation of aerosol optical depth in relation to meteorological parameters over eastern and western routes of China Pakistan economic corridor”. En: *Journal of Environmental Sciences* 99 (2021), págs. 28-39.
- [111] Qi-Xiang Chen y col. “Verification of aerosol classification methods through satellite and ground-based measurements over Harbin, Northeast China”. En: *Atmospheric Research* 216 (2019), págs. 167-175.
- [112] Julien Djossou y col. “Mass concentration, optical depth and carbon composition of particulate matter in the major southern West African cities of Cotonou (Benin) and Abidjan (Côte d’Ivoire)”. En: *Atmospheric Chemistry and Physics* 18.9 (2018), págs. 6275-6291.
- [113] Lidia Ana Otero y col. “Siete fotómetros de la red AERONET instalados en territorio argentino: análisis estadísticos de los datos y caracterización de los aerosoles”. En: *ANALES AFA*. Vol. 29. 3. 2018, págs. 78-82.
- [114] Mohammad Rezaei y col. “Discrimination of aerosol types over the Tehran city using 5 years (2011–2015) of MODIS collection 6 aerosol products”. En: *Journal of Environmental Health Science and Engineering* 17.1 (2019), págs. 1-12.
- [115] Lidia Ana Otero y col. “Clasificación de los aerosoles para la Estación AERONET de Comodoro Rivadavia durante 2015”. En: (2019).
- [116] Facundo Ismael Casasola y col. “Clasificación de los aerosoles en la estación AERONET de Tucuman, Argentina”. En: (2020).
- [117] AF Stein y col. “NOAA’s HYSPLIT atmospheric transport and dispersion modeling system”. En: *Bulletin of the American Meteorological Society* 96.12 (2015), págs. 2059-2077.

- [118] G Rolph, A Stein y B Stunder. “Real-time Environmental Applications and Display sYstem: READY. Environmental Modelling & Software”. En: (2017).
- [119] DG Calinoiu y col. “Sun photometer–remote sensing for aerosol assessment”. En: *Journal of Environmental Protection and Ecology* 19.2 (2018), págs. 462-470.
- [120] Ali M Al-Salihi. “Characterization of aerosol type based on aerosol optical properties over Baghdad, Iraq”. En: *Arabian Journal of Geosciences* 11.20 (2018), págs. 1-15.
- [121] Shokufeh Delfi y col. “Investigation of aerosols pollution across the eastern basin of Urmia lake using satellite remote sensing data and HYSPLIT model”. En: *Journal of Environmental Health Science and Engineering* 17.2 (2019), págs. 1107-1120.
- [122] S Papandrea y col. “Diseño y desarrollo de un fotómetro solar basado en tecnología LED”. En: *ANALES AFA*. Vol. 26. 2. 2015.
- [123] Kuo-Nan Liou. *An introduction to atmospheric radiation*. Elsevier, 2002.
- [124] Anders Ångström. “On the atmospheric transmission of sun radiation and on dust in the air”. En: *Geografiska Annaler* 11.2 (1929), págs. 156-166.
- [125] Oleg Dubovik y col. “Variability of absorption and optical properties of key aerosol types observed in worldwide locations”. En: *Journal of the atmospheric sciences* 59.3 (2002), págs. 590-608.
- [126] Kusiel S Shifrin. “Simple relationships for the Ångström parameter of disperse systems”. En: *Applied optics* 34.21 (1995), págs. 4480-4485.
- [127] P Ristori y col. “Characterization of the atmospheric aerosols in the Buenos Aires area using sun photometer measurements”. En: *4th Iberoamerican Meeting on Optics and 7th Latin American Meeting on Optics, Lasers, and Their Applications*. Vol. 4419. International Society for Optics y Photonics. 2001, págs. 506-509.
- [128] Brent N Holben y col. “An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET”. En: *Journal of Geophysical Research: Atmospheres* 106.D11 (2001), págs. 12067-12097.
- [129] Guillaume A D’Almeida, Peter Koepke y Eric P Shettle. *Atmospheric aerosols: global climatology and radiative characteristics*. A Deepak Pub, 1991.
- [130] Anna Raquel Esteve Martínez. “Estudio de los aerosoles atmosféricos en Valencia mediante medidas del nefelómetro de integración”. Tesis doct. Universitat de València, 2010.

- [131] Juan Francisco Rojas García. “Monitoreo de cuatro contaminantes atmosféricos a nivel individual”. Tesis doct. Universidad Veracruzana. Facultad de Ciencias Químicas. Región Xalapa., 2019.
- [132] DJ von Lehmden y C Nelson. *Quality assurance handbook for air pollution measurement systems. Volume II. Ambient air specific methods*. Inf. téc. Environmental Protection Agency, Research Triangle Park, NC (USA . . . , 1977.
- [133] ME NZ. *Good-practices guide for air quality monitoring and data managment*. Inf. téc. Ministry of the Environment, New Zealand, 2009.
- [134] Prashant Kumar y col. “The rise of low-cost sensing for managing air pollution in cities”. En: *Environment international* 75 (2015), págs. 199-205.
- [135] Eric S Hall y col. “Integrating sensor monitoring technology into the current air pollution regulatory support paradigm: Practical considerations”. En: *Am. J. Environ. Eng* 4.6 (2014), págs. 147-154.
- [136] Alastair Lewis, W Richard Peltier y Erika von Schneidemesser. “Low-cost sensors for the measurement of atmospheric composition: overview of topic and future applications”. En: (2018).
- [137] John L Campbell y col. “Quantity is nothing without quality: automated QA/QC for streaming environmental sensor data”. En: *BioScience* 63.7 (2013), págs. 574-585.
- [138] Marc Kamionka, Philippe Breuil y Christophe Pijolat. “Calibration of a multi-variate gas sensing device for atmospheric pollution measurement”. En: *Sensors and Actuators B: Chemical* 118.1-2 (2006), págs. 323-327.
- [139] Yang Wang y col. “Laboratory evaluation and calibration of three low-cost particle sensors for particulate matter measurement”. En: *Aerosol Science and Technology* 49.11 (2015), págs. 1063-1077.
- [140] Saverio De Vito y col. “On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario”. En: *Sensors and Actuators B: Chemical* 129.2 (2008), págs. 750-757.
- [141] Laurent Spinelle y col. “Calibration of a cluster of low-cost sensors for the measurement of air pollution in ambient air”. En: *SENSORS, 2014 IEEE*. IEEE. 2014, págs. 21-24.
- [142] Laurent Spinelle y col. “Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO₂”. En: *Sensors and Actuators B: Chemical* 238 (2017), págs. 706-715.

- [143] Jingjin Shi y col. “Validation of a light-scattering PM2.5 sensor monitor based on the long-term gravimetric measurements in field tests”. En: *PloS one* 12.11 (2017).
- [144] C Lin y col. “Evaluation and calibration of Aeroqual series 500 portable gas sensors for accurate measurement of ambient ozone and nitrogen dioxide”. En: *Atmospheric Environment* 100 (2015), págs. 111-116.
- [145] Ricardo Piedrahita y col. “The next generation of low-cost personal air quality sensors for quantitative exposure monitoring”. En: *Atmospheric Measurement Techniques* 7.10 (2014), pág. 3325.
- [146] Walter Fuertes y col. “Distributed system as internet of things for a new low-cost, air pollution wireless monitoring on real time”. En: *2015 IEEE/ACM 19th International Symposium on Distributed Simulation and Real Time Applications (DS-RT)*. IEEE. 2015, págs. 58-67.
- [147] NANO-TERA. *Air quality Programme*. 2020. URL: <http://opensense.epfl.ch/> (visitado 22-05-2020).
- [148] EU proyect. *Air quality Programme*. 2020. URL: <http://www.citi-sense.eu/> (visitado 22-05-2020).
- [149] EU proyect. *Air quality Programme*. 2020. URL: <http://www.everyaware.eu/> (visitado 22-05-2020).
- [150] EU Council Directive. “50/EC of the European Parliament and of the Council on ambient air quality and cleaner air for Europe”. En: *Official Journal of the European Communities (2008) L 152* (2008).
- [151] John McNeill y col. “Wearable wireless sensor patch for continuous monitoring of skin temperature, pressure, and relative humidity”. En: *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE. 2017, págs. 1-4.
- [152] Jenny Aitken-Christie, Toyoki Kozai y Mary Ann Lila Smith. *Automation and environmental control in plant tissue culture*. Springer Science & Business Media, 2013.
- [153] Matthias Bartholmai, Enrico Köppe y Patrick P Neumann. “Monitoring of hazardous scenarios using multi-sensor devices”. En: *The Fourth International Conference on Sensor Device Technologies and Applications*. 2013, págs. 9-13.

- [154] Fabián Rolando Jiménez López y Andrés Jiménez López. “Field variables monitoring in real time (GPS, soil moisture, temperature) with precision farming applications”. En: *Proceedings of the 6th Euro American Conference on Telematics and Information Systems*. 2012, págs. 89-93.
- [155] Andrea Polidori, Vasileios Papapostolou y Hang Zhang. “Laboratory Evaluation of Low-Cost Air Quality Sensors”. En: *South Coast Air Quality Management District: Diamondbar, CA, USA* (2016).
- [156] C Borrego y col. “Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise”. En: *Atmospheric Environment* 147 (2016), págs. 246-263.
- [157] Claudio Castellini y Vikram Ravindra. “A wearable low-cost device based upon force-sensing resistors to detect single-finger forces”. En: *5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*. IEEE. 2014, págs. 199-203.
- [158] Sinan Sousan y col. “Inter-comparison of low-cost sensors for measuring the mass concentration of occupational aerosols”. En: *Aerosol Science and Technology* 50.5 (2016), págs. 462-473.
- [159] Leigh R Crilley y col. “Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring”. En: *Atmospheric Measurement Techniques* (2018), págs. 709-720.
- [160] T Sayahi, A Butterfield y KE Kelly. “Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors”. En: *Environmental Pollution* 245 (2019), págs. 932-940.
- [161] Florentin MJ Bulot y col. “Long-term field comparison of multiple low-cost particulate matter sensors in an outdoor urban environment”. En: *Scientific reports* 9.1 (2019), págs. 1-13.
- [162] Joel Kuula y col. “Laboratory evaluation of particle-size selectivity of optical low-cost particulate matter sensors”. En: *Atmospheric Measurement Techniques* 13.5 (2020), págs. 2413-2423.
- [163] Gustav Mie. “Contributions to the optics of turbid media, particularly of colloidal metal solutions”. En: *Contributions to the optics of turbid media* 25.3 (1976), págs. 377-445.

- [164] John H Seinfeld y Spyros N Pandis. *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons, 2016.
- [165] Junwei Zhou y Andrew Mason. “Communication buses and protocols for sensor networks”. En: *Sensors* 2.7 (2002), págs. 244-257.
- [166] Servicio Meteorológico Nacional. *Datos meteorológicos SMN*. 2020. URL: <https://www.smn.gob.ar/> (visitado 25-05-2020).
- [167] Laboratorio de Recursos del Aire. *Datos meteorológicos NOAA READY*. 2020. URL: <https://www.ready.noaa.gov/READYcmet.php> (visitado 25-05-2020).
- [168] D Bruce Turner. “A diffusion model for an urban area”. En: *Journal of Applied Meteorology* 3.1 (1964), págs. 83-91.
- [169] Chengcai Li y col. “Application of MODIS satellite products to the air pollution research in Beijing”. En: *Science in China Series D(Earth Sciences)* 48 (2005), págs. 209-219.
- [170] Emanuele Emili y col. “PM10 remote sensing from geostationary SEVIRI and polar-orbiting MODIS sensors over the complex terrain of the European Alpine region”. En: *Remote sensing of environment* 114.11 (2010), págs. 2485-2499.
- [171] Muhammad Iqbal. *An introduction to solar radiation*. Elsevier, 2012.
- [172] GLOBE. *CALITOO photometer*. 2021. URL: <https://www.calitoo.fr/index.php?page=en> (visitado 13-04-2021).
- [173] NASA. *NASA EARTH*. 2021. URL: <https://search.earthdata.nasa.gov> (visitado 20-04-2021).
- [174] Juha Vesanto y col. *SOM toolbox for Matlab 5*. Inf. téc. Citeseer, 2000.
- [175] Teuvo Kohonen. *Self-organizing maps*. Vol. 30. Springer Science & Business Media, 2012.
- [176] Ariel Fabricio Scagliotti y Guillermo Antonio Jorge. “Inter-comparison of environmental low-cost sensors on Arduino platform”. En: *Journal of the IEST* 63.1 (2020), págs. 35-45.
- [177] R Williams y col. “Sensor Evaluation Report US Environmental Protection Agency”. En: *Washington, DC.[Google Scholar]* (2014).
- [178] Bhupesh Aneja y col. “Review of Temperature Measurement and Control”. En: *Intl. Journal of Electrical and Electronics Engineers* 3.2 (2011), págs. 29-37.

Bibliografía

- [179] Anindya Ananda Hapsari, Asif Iqbal Hajamydeen y Muhammad Irsyad Abdullah. “A review on indoor air quality monitoring using iot at campus environment”. En: *International Journal of Engineering & Technology* 7.4.22 (2018), págs. 55-60.
- [180] Maxim Integrated. *18b20 Datasheet*. 2020. URL: <https://datasheets.maximintegrated.com/en/ds/DS18B20.pdf> (visitado 22-05-2020).
- [181] Gediminas Gričius y col. “Advanced approach of multiagent based buoy communication”. En: *The Scientific World Journal* 2015 (2015).
- [182] DF ROBOT. *DHT11 Datasheet*. 2020. URL: <https://image.dfrobot.com/image/data/KIT0003/DHT11%20datasheet.pdf> (visitado 22-05-2020).
- [183] ADAFRUIT. *BMP180 Datasheet*. 2020. URL: <https://cdn-shop.adafruit.com/datasheets/BST-BMP180-DS000-09.pdf> (visitado 22-05-2020).
- [184] BOSCH. *BME280 Datasheet*. 2020. URL: https://ae-bst.resource.bosch.com/media/_tech/media/datasheets/BST-BME280-DS002.pdf (visitado 22-05-2020).
- [185] METEOSTAR. *WT1081 Datasheet*. 2020. URL: <http://www.meteostar.com.ar/descargas/estacinmeteorologica-wt1081-meteostar.pdf> (visitado 22-05-2020).
- [186] PASCO. *PS2135 Datasheet*. 2020. URL: https://www.pasco.com/file_downloads/Downloads_Manuals/PASPORT-Fast-Response-Temperature-Probe-Manual-PS-2135.pdf (visitado 22-05-2020).
- [187] RIGOL. *DM3068 Datasheet*. 2020. URL: <http://beyondmeasure.rigoltech.com/acton/attachment/1579/f-0052/0/-/-/-/-/file.pdf> (visitado 22-05-2020).
- [188] ARDUINO. *ARDUINO UNO R3 Datasheet*. 2020. URL: <http://static6.arrow.com/aropdfconversion/97b354c3cd11851042ebd1fd252fbd6b6def5adafruit2877arduinouno.pdf> (visitado 22-05-2020).
- [189] OKY-3001. *SD card module OKY-3001 Datasheet*. 2021. URL: <http://www.agspecinfo.com/pdfs/O/OKY3001.PDF> (visitado 04-06-2021).
- [190] DS3231. *RTC DS3231 Datasheet*. 2021. URL: <https://datasheets.maximintegrated.com/en/ds/DS3231.pdf> (visitado 04-06-2021).

- [191] GP2Y1014AU0F. *GP2Y1014AU0F dust sensor Datasheet*. 2021. URL: https://datasheet.lcsc.com/szlcsc/1905131608_Sharp-Microelectronics-GP2Y1014AU0F_C390729.pdf (visitado 04-06-2021).
- [192] HKA5. *HK-A5 PM sensor Datasheet*. 2021. URL: https://www.diyelectronics.co.za/store/index.php?controller=attachment&id_attachment=483 (visitado 04-06-2021).
- [193] PMS1003. *PMS1003 sensor Datasheet*. 2021. URL: http://www.aqmd.gov/docs/default-source/aq-spec/resources-page/plantower-pms1003-manual_v2-5.pdf (visitado 04-06-2021).
- [194] PMS5003. *PMS5003 sensor Datasheet*. 2021. URL: https://www.aqmd.gov/docs/default-source/aq-spec/resources-page/plantower-pms5003-manual_v2-3.pdf (visitado 04-06-2021).
- [195] Heinrich Kaiser y Hermann Specker. “Bewertung und vergleich von analysenverfahren”. En: *Fresenius’ Zeitschrift für analytische Chemie* 149.1 (1956), págs. 46-66.
- [196] Open-Seneca. *Open-Seneca organization*. 2021. URL: <https://open-seneca.org/> (visitado 07-06-2021).
- [197] Karl E Taylor. “Summarizing multiple aspects of model performance in a single diagram”. En: *Journal of Geophysical Research: Atmospheres* 106.D7 (2001), págs. 7183-7192.
- [198] David H Hagan y Jesse H Kroll. “Assessing the accuracy of low-cost optical particle sensors using a physics-based approach”. En: *Atmospheric Measurement Techniques* 13.11 (2020), págs. 6343-6355.
- [199] Meilu He, Nueraili Kuerbanjiang y Suresh Dhaniyala. “Performance characteristics of the low-cost Plantower PMS optical sensor”. En: *Aerosol Science and Technology* 54.2 (2020), págs. 232-241.
- [200] OPEN SENECA. *OPEN SENECA APP*. 2019. URL: <http://app.open-seneca.org/> (visitado 20-11-2019).
- [201] Roland B Stull. *An introduction to boundary layer meteorology*. Vol. 13. Springer Science & Business Media, 2012.
- [202] GBA. *LEY AMBIENTAL*. 2021. URL: <https://normas.gba.gob.ar/ar-b/decreto/2018/1074/17866> (visitado 19-03-2021).

- [203] Michael Hess, Peter Koepke e I Schult. “Optical properties of aerosols and clouds: The software package OPAC”. En: *Bulletin of the American meteorological society* 79.5 (1998), págs. 831-844.
- [204] Manu Mehta y col. “Recent global aerosol optical depth variations and trends—A comparative study using MODIS and MISR level 3 datasets”. En: *Remote sensing of environment* 181 (2016), págs. 137-150.
- [205] A Sen y col. “Variations in particulate matter over Indo-Gangetic Plains and Indo-Himalayan Range during four field campaigns in winter monsoon and summer monsoon: role of pollution pathways”. En: *Atmospheric environment* 154 (2017), págs. 200-224.
- [206] T Banerjee, SB Singh y RK Srivastava. “Development and performance evaluation of statistical models correlating air pollutants and meteorological variables at Pantnagar, India”. En: *Atmospheric Research* 99.3-4 (2011), págs. 505-517.
- [207] Tirthankar Banerjee y Rajeev Kumar Srivastava. “Evaluation of environmental impacts of Integrated Industrial Estate—Pantnagar through application of air and water quality indices”. En: *Environmental monitoring and assessment* 172.1 (2011), págs. 547-560.
- [208] Tongshu Zheng y col. “Field evaluation of low-cost particulate matter sensors in high-and low-concentration environments”. En: *Atmospheric Measurement Techniques* 11.8 (2018), págs. 4823-4846.
- [209] Judith C Chow, John G Watson y col. “Guideline on speciated particulate monitoring”. En: *Report prepared for US Environmental Protection Agency, Research Triangle Park, NC, by Desert Research Institute, Reno, NV* (1998).
- [210] Sebastián Diez y col. “Allin-Wayra: una estación experimental para el monitoreo continuo de la calidad del aire empleando sensores de bajo costo”. En: *Revista de la Facultad de Ciencias Exactas, Físicas y Naturales* 5 (2018), pág. 45.
- [211] A Santa Cruz. “Difusión de contaminantes gaseosos en la atmósfera”. En: *Argentina: Universidad Tecnológica Nacional* (2000).