

# Calibracion de sensores de bajo costo

Lenna Pedroso

2025-11-27

## Índice

1.	Marco teórico . . . . .	1
	Objetivo general . . . . .	2
2.	Metodología . . . . .	2
	2.1. Tendencia temporal . . . . .	3
	2.2. Regresión lineal . . . . .	5
	2.3. Análisis de los errores del modelo . . . . .	7
3.	Conclusiones . . . . .	12
4.	Referencias bibliográficas . . . . .	12

## 1. Marco teórico

La exposición a material particulado fino, con un diámetro aerodinámico menor o igual a  $2,5 \mu\text{g}$  ( $\text{PM}_{2,5}$ ), puede provocar millones de muertes anuales, ya que estas partículas al ser inhaladas, pueden penetrar a través del sistema respiratorio y acumularse en diferentes partes del cuerpo (Li et al., 2020). Entre los riesgos asociados a su exposición se destacan: aumento en la probabilidad de desarrollar cáncer de pulmón, mortalidad prematura y enfermedades respiratorias (Borchers-Arriagada et al., 2024).

La Región Metropolitana de Santiago de Chile se encuentra entre las ciudades más contaminadas de Sudamérica, superando el límite máximo permisible referido a la concentración anual de ( $\text{PM}_{2,5}$ ) establecido para Chile ( $20 (\mu\text{g}/\text{m}^3)$ ) (Peralta et al., 2022). Esta situación se debe principalmente a su ubicación geográfica, cuyas características topográficas y condiciones meteorológicas restringen la dispersión de contaminantes favoreciendo su acumulación principalmente durante el invierno (Menares et al., 2021).

El control de la calidad del aire atmosférico desempeña un papel crucial en la prevención de la contaminación ambiental en la región. El Sistema de Información Nacional de Calidad del Aire (SINCA), en Santiago de Chile, presenta sensores remotos que estiman las concentraciones

de PM<sub>2,5</sub>. Estos sensores están distribuidos en un número limitado de sitios, (10 estaciones de monitoreo aproximadamente), por lo que ofrecen una resolución espacio-temporal limitada al aportar datos restringidos al área donde se encuentran estas estaciones.

Otra de las limitaciones es su elevado costo y su incompatibilidad, en países en vías de desarrollo, debido a su tamaño y gastos de mantenimiento. Además, requieren la incorporación de personal capacitado y calificado para su operación . Por otra parte, los modelos convencionales de distribución de estos sensores de monitoreo están sujetos a grandes incertidumbres debido a la falta de perfiles de emisión precisos de todas las fuentes contribuyentes y al número limitado de mediciones en los sitios receptores (Tarazona Alvarado et al., 2024). Como consecuencia, gran parte de la población permanece expuesta a fuentes locales no capturada por el SINCA, subestimándose la exposición real y, por ende, sus implicaciones en la salud humana.

La tecnología de detección de bajo costo es una alternativa viable de monitoreo de la calidad del aire debido a que proporcionan información sobre la variación temporal y espacial de la contaminación por PM<sub>2,5</sub>. Además, brindan información detallada sobre el grado de exposición de la población en la ciudad. Estos sensores se pueden utilizar para caracterizar variaciones de concentración de PM<sub>2,5</sub> e identificar puntos críticos de contaminación a escala local, incluso en condiciones extremas de contaminación del aire. Por otra parte, proporcionan información sobre la dinámica de los patrones de contaminación y permiten analizar el impacto de las fuentes locales, contribuyendo a la conceptualización de una gestión ambiental más encaminada a reducir la exposición humana a estos contaminantes (Kar et al., 2024; Tagle et al., 2020; Traviss et al., 2025) .

Una de las principales limitaciones de estos sensores es su baja presión en la determinación de la concentración del PM<sub>2,5</sub> ya que son muy sensibles a variaciones de temperatura y humedad. Además, tienden a sobreestimar los valores de concentración por lo que la calibración de estos prototipos constituye un paso esencial para garantizar un monitoreo complementario preciso en la ciudad.

## **Objetivo general**

El objetivo de esta investigación es realizar la calibración de sensores de bajo costo (Plantower) en la comuna Las Condes obteniendo un error de medición por debajo de 20  $\mu\text{g}/\text{m}^3$ .

## **2. Metodología**

Se realizó un monitoreo durante el periodo del 4 de junio al 23 de agosto de 2025. Los sensores de bajo costo se colocaron en la estación de referencia (SINCA) ubicada en la comuna Las Condes perteneciente a la ciudad de Santiago de Chile. Se colocaron 5 sensores de bajo costo para evaluar su capacidad de ,edir con presición las concentraciones de PM<sub>2,5</sub>.

## 2.1. Tendencia temporal

Primeramente, se analizará la tendencia temporal de las concentraciones de cinco sensores de la tecnología Plantower (P1-P5) y se compararán con los datos de referencia de la estación regulatoria (SINCA) para determinar si existe similitud en el comportamiento de ambos tipos de tecnología.

```
# Cargar librerias -----
library(dplyr)
library(tidyr)
library(rlang)
#Cargar datos -----
Condes_Plantower_SINCA_long3 <- read.csv2(
  "C:/Semestre 2-2025/Visualizacion/Página web/Pagina web/Datos/Condes_Plantower_SINCA_long3",
  header = TRUE
)

# Funcion para calcular las concentraciones semanales
Condes_summary_pm25 <- function(datos, tiempo, pm_col = "value") {
  datos %>%
    group_by(across(all_of(tiempo)), sensor) %>%
    summarise(
      n      = sum(!is.na(.data[[pm_col]])),
      mean   = mean(.data[[pm_col]], na.rm = TRUE),
      sd     = sd(.data[[pm_col]], na.rm = TRUE),
      se     = ifelse(n > 1, sd / sqrt(n), NA_real_),
      ymin   = mean - 1.96 * se,
      ymax   = mean + 1.96 * se,
      .groups = "drop"
    )
}
# Resumenes por dia de la hora y por dia de la semana
hour_wday <- Condes_summary_pm25(Condes_Plantower_SINCA_long3, c("wday", "hour"), "value_PM25")
hour <- Condes_summary_pm25(Condes_Plantower_SINCA_long3, "hour", "value_PM25")
wday <- Condes_summary_pm25(Condes_Plantower_SINCA_long3, "wday", "value_PM25")
# Especificar orden
# Especificar orden
hour_wday <- hour_wday %>%
  mutate(
    wday = factor(
      wday,
      levels = c(
        "lun\\\",",
        "mar\\\",",
        "mié\\\",",
        "jue\\\",",
        "vie\\\",",
        "sab\\\",",
        "dom\\\"
```

```

    "mar\\",
    "mié\\",
    "jue\\",
    "vie\\",
    "sáb\\",
    "dom\\"
)
)
)
)

# Cargar librerias
library(ggplot2)

# Obtener grafico
ggplot(hour_wday, aes(x = hour, y = mean, color = sensor, fill = sensor, group = sensor)) +
  geom_ribbon(aes(ymin = ymin, ymax = ymax), alpha = 0.15, colour = NA) +
  geom_line(linewidth = 1.4) +
  facet_wrap(vars(wday), nrow = 1) +
  labs(
    x = "Hora del día",
    y = expression(PM[2.5]~(mu*g/m^3))
  ) +
  scale_color_manual(
    values = c(
      "P1" = "gold",
      "P2" = "blue",
      "P3" = "#4EEE94",
      "P4" = "#87CEEB",
      "P5" = "coral",
      "SINCA" = "#EE7AE9"
    )
  ) +
  scale_fill_manual(
    values = c(
      "P1" = "gold",
      "P2" = "blue",
      "P3" = "#4EEE94",
      "P4" = "#87CEEB",
      "P5" = "coral",
      "SINCA" = "#EE7AE9"
    )
  )
)

```

```

theme_bw(base_size = 14) +
theme(
  legend.title = element_blank()
)

```

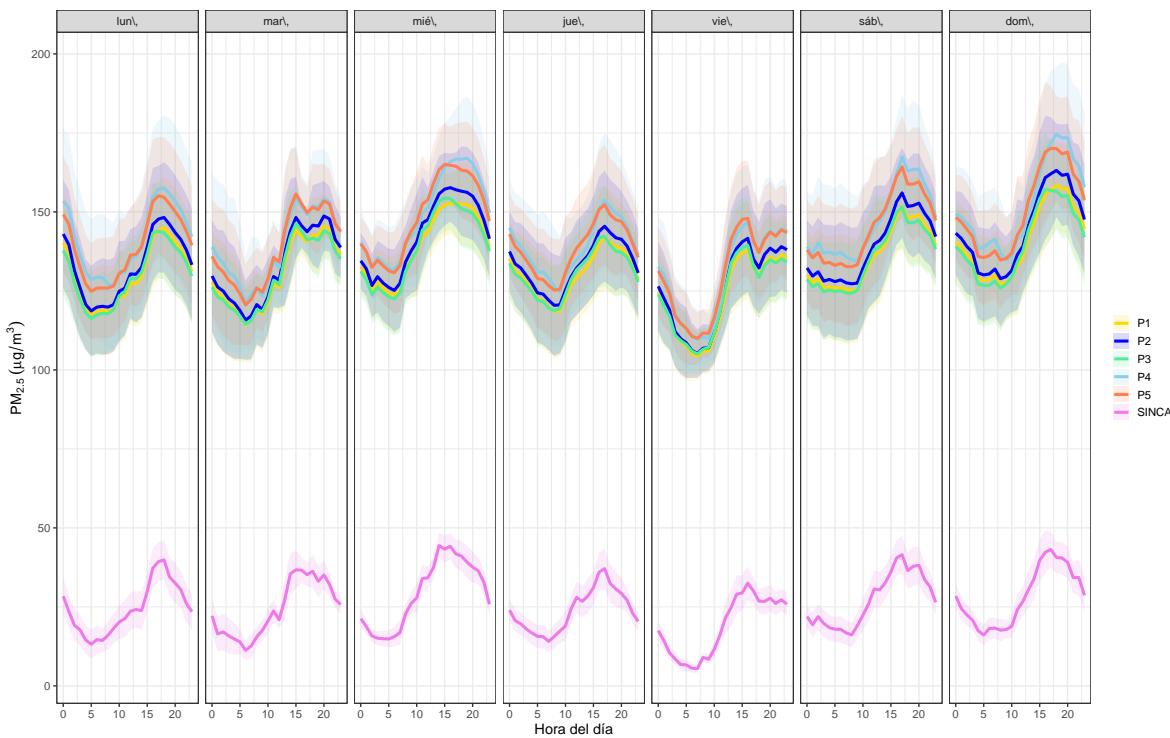


Figure 1: Variación horaria de por día de la semana y sensor

Al analizar el gráfico se observa que existe una marcada diferencia en el comportamiento de los sensores de bajo costo en comparación con los valores de la estación de referencia (SINCA). En este caso los sensores de bajo costo tienden a sobreestimar las concentraciones de material particulado demostrándose la importancia de su calibración para solucionar estas limitaciones.

## 2.2. Regresión lineal

Para obtener la ecuación de calibración para cada sensor de bajo costo se aplicó un modelo de regresión lineal mediante *scaterplot*. Este grafico es específico para analizar regresiones lineales correspondientes a contaminantes atmosféricos ya que permite observar la frecuencia de episodios de alta concentración en la ciudad. Los resultados se muestran en la siguiente figura.

```
# Cargar librerias -----
library(openair)
#Cargar datos -----
Condes_Plantower_SINCA_long_2 <- read.csv2(
  "C:/Semestre 2-2025/Visualizacion/Página web/Pagina web/Datos/Condes_P_SINCA_long.csv",
  header = TRUE
)
# Modelo de regresión lineal-----
scatterPlot(
  mydata = Condes_Plantower_SINCA_long_2,
  x      = "value",
  y      = "value_SINCA",
  method = "hexbin",
  col    = "jet",
  linear = TRUE,
  ci     = TRUE,
  fit    = TRUE,
  type   = "sensor",
  xlab   = "PM2.5 Sensor Plantower ( $\mu\text{g}/\text{m}^3$ )",
  ylab   = "PM2.5 SINCA ( $\mu\text{g}/\text{m}^3$ )",
  fontsize = 16
)
```

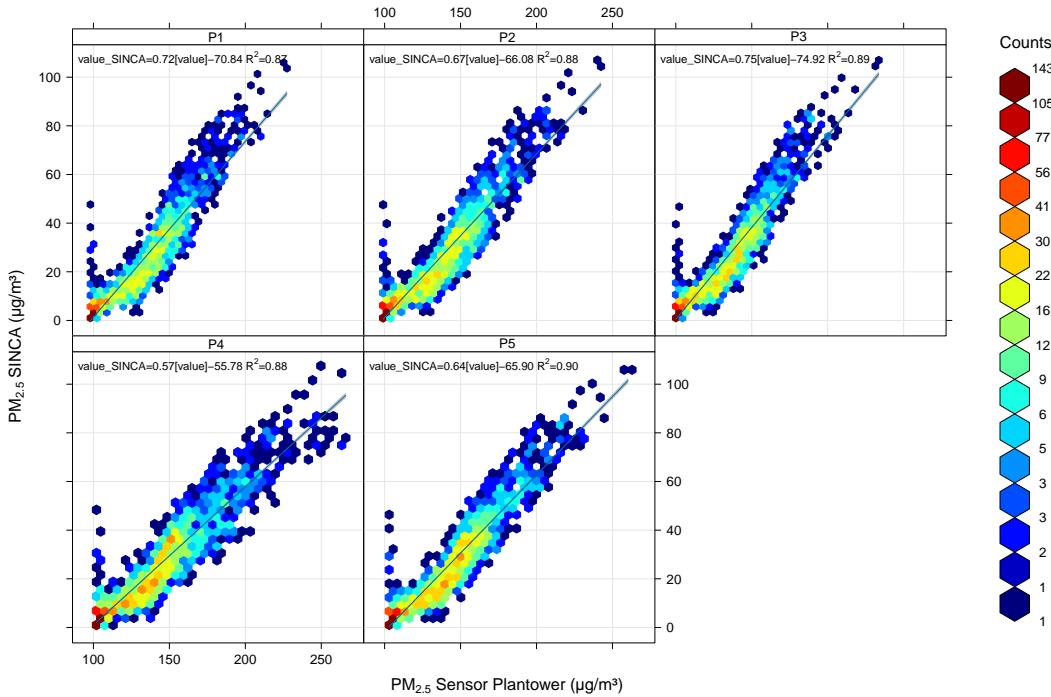


Figure 2: Análisis de regresión lineal SINCA vs Sensor de bajo costo(value)

En la figura se demuestra que a pesar de existir un defase en el comportamiento de estos sensores un modelo de regresión lineal puede explicar con un buen ajuste ( $R^2=0.87-0.9$ ) las concentraciones de material particulado. Además, los valores de alta concentración presentaban bajas frecuencias de aparición, desvirtuándose que estos episodios no fueron comunes en el periodo estudiado.

### 2.3 Análisis de los errores del modelo

Uno de los aspectos a considerar para el uso de los sensores de bajo costo como complemento de los monitores regulatorios (SINCA) es el análisis de error. En este contexto, se determinó que los errores absolutos, así como las métricas específicas de error (MB, MGE) deberían estar por debajo de  $20 \mu\text{g}/\text{m}^3$ . Los resultados se muestran en la siguiente figura.

```
#Cargar datos -----
errores_completo <- read.csv2( "C:/Semestre 2-2025/Visualizacion/Página web/Página web/Datos",
  header = TRUE
)
```

```

# Cargar librerias -----
library(ggplot2)
library(dplyr)

# Seleccionar datos ----

errores_no_calibrado <- errores_completo %>% filter(Clasificacion == "No calibrados")

# Grafico de violín para errores no calibrados ----

violin_nocalibrado <- ggplot(errores_no_calibrado ,
                               aes(x = sensor, y = error_abs, fill = sensor)) +
  geom_violin(trim = FALSE, alpha = 0.35, linewidth = 1.1) +
  geom_boxplot(width = 0.10, alpha = 0.8, linewidth = 1.0,
                outlier.size = 1.5) +
  geom_hline(yintercept = 20, linetype = "dashed",
             color = "firebrick4", linewidth = 1) +
  scale_y_continuous(limits = c(0, NA)) +
  scale_fill_manual(values = c(
    "PM25_P1" = "firebrick1",
    "PM25_P2" = "firebrick2",
    "PM25_P3" = "firebrick3",
    "PM25_P4" = "firebrick4",
    "PM25_P5" = "#5A0F0F"
  )) +
  theme_bw(base_size = 16) +
  theme(
    plot.title = element_text(face = "bold", hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 1)
  ) +
  labs(title = "Error absoluto - No calibrados",
       x = "sensor", y = "Error absoluto ( $\mu\text{g}/\text{m}^3$ )",
       fill = "sensor")

# Grafico de violín para errores calibrados ----
# Seleccionar datos -----
errores_cal <- errores_completo %>% filter(Clasificacion == "Calibrados")

# Grafico de violín para errores no calibrados ----
violin_calibrado <- ggplot(errores_cal,
                           aes(x = sensor, y = error_abs, fill = sensor)) +
  geom_violin(trim = FALSE, alpha = 0.35, linewidth = 1.1) +

```

```

geom_boxplot(width = 0.10, alpha = 0.8, linewidth = 1.0,
             outlier.size = 1.5) +
geom_hline(yintercept = 20, linetype = "dashed",
            color = "firebrick4", linewidth = 1) +
scale_y_continuous(limits = c(0, NA)) +
scale_fill_manual(values = c(
  "PM25_P1" = "firebrick1",
  "PM25_P2" = "firebrick2",
  "PM25_P3" = "firebrick3",
  "PM25_P4" = "firebrick4",
  "PM25_P5" = "#5A0F0F"
)) +
theme_bw(base_size = 16) # Tema blanco y negro con tamaño de texto 16
theme_bw(base_size = 16) +
theme(
  plot.title = element_text(face = "bold", hjust = 0.5),
  axis.text.x = element_text(angle = 45, hjust = 1)
) +
labs(title = "Error absoluto - Calibrados",
     x = "sensor", y = "Error absoluto (pg/m³)",
     fill = "sensor")

```

```

# Cargar datos -----
Errores_anter_despues<- read.csv2( "C:/Semestre 2-2025/Visualizacion/Página web/Pagina web/Da
  header = TRUE
)
#Grafico de error antes calibración-----
barras_MGE <- ggplot(Errores_anter_despues,
                      aes(x = Sensor, y = MGE, fill = Clasificación)) +
geom_bar(stat = "identity", color = "black",
         position = position_dodge(width = 0.9)) +
geom_errorbar(aes(ymin = MGE - SD_MGE, ymax = MGE + SD_MGE),
              width = 0.2, linewidth = 0.8,
              position = position_dodge(width = 0.9)) +
geom_hline(yintercept = 20,
            linetype = "dashed",
            color = "firebrick4",
            linewidth = 1) +
scale_fill_manual(values = c(
  "No calibrados" = "#1E90FF",
  "Calibrados" = "firebrick1"
)) +

```

```

theme_bw(base_size = 16) +
theme(
  plot.title = element_text(face = "bold", hjust = 0.5),
  axis.text.x = element_text(angle = 45, hjust = 1),
  legend.title = element_blank()
) +
labs(title = "MGE antes / después de calibrar",
  x = "Sensor", y = "MGE (µg/m³)")

# Gráfico de barras MB -----
barras_MB <- ggplot(Errores_anteriores_despues,
                     aes(x = Sensor, y = MB, fill = Clasificación)) +
  geom_bar(stat = "identity", color = "black",
           position = position_dodge(width = 0.9)) +
  geom_errorbar(aes(ymin = MB - SD_MB, ymax = MB + SD_MB),
                width = 0.2, linewidth = 0.8,
                position = position_dodge(width = 0.9)) +
  geom_hline(yintercept = 20,
             linetype = "dashed",
             color = "firebrick4",
             linewidth = 1) +
  scale_fill_manual(values = c(
    "No calibrados" = "#1E90FF",
    "Calibrados"     = "firebrick1"
  )) +
  theme_bw(base_size = 16) +
  theme(
    plot.title = element_text(face = "bold", hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.title = element_blank()
) +
  labs(title = "MB antes / después de calibrar",
    x = "Sensor", y = "MB (µg/m³)")

# Cargar librerías -----
library(patchwork)

# Especificar una leyenda para los violines -----
figura_violines <- (violin_nocalibrado | violin_calibrado) +
  plot_layout(guides = "collect") # Una leyenda
  theme(legend.position = "right") # leyendas a la derecha

```

```

# Especificar una leyenda para los graficos de barras -----
figura_barras <- (barras_MGE | barras_MB) +
  plot_layout(guides = "collect") +
  theme(legend.position = "right")

# Unir figuras y especificar orden ----

figura_final <- figura_violines / figura_barras

# Ver figura ----

figura_final

```

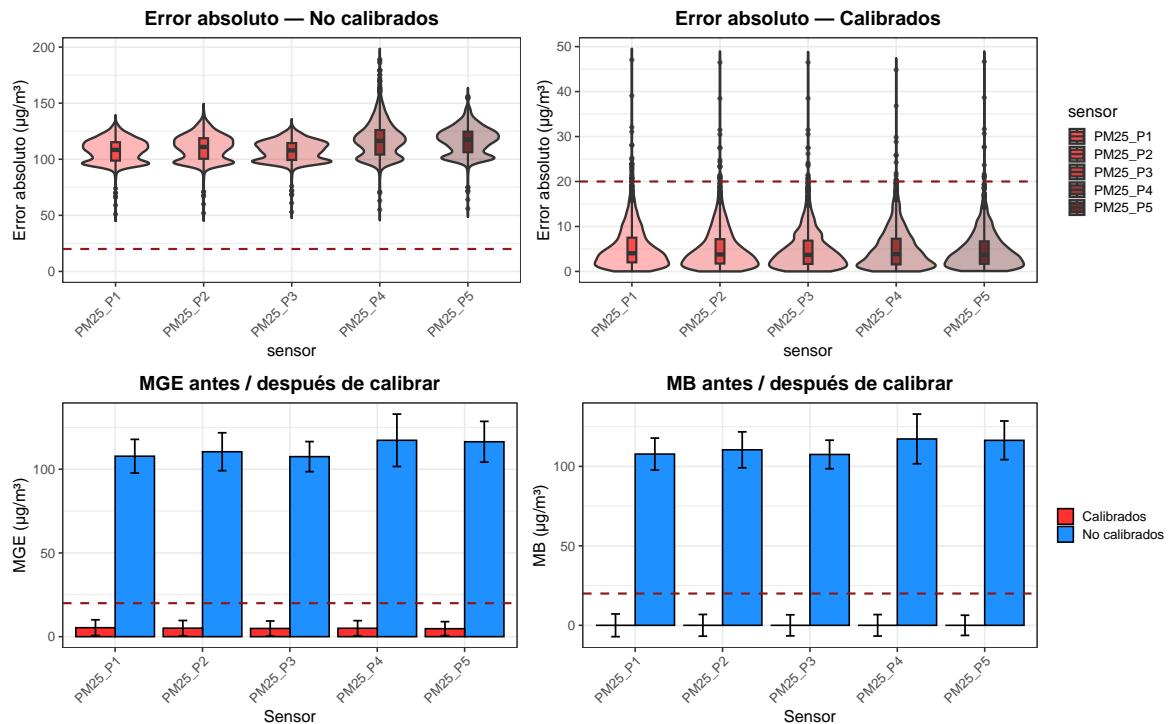


Figure 3: Análisis de errores antes y después de la calibración

La figura se evidencia que al aplicarle un modelo de regresión lineal a los sendores de bajo costo todos los valores de errores medios estuvieron por debajo de  $20 \mu\text{g}/\text{m}^3$ .

Los resultados pueden consultarse en la siguiente página web:

### 3. Conclusiones

1. Los sensores de bajo costo tienden a sobreestimar la concentración de PM<sub>2,5</sub>.
2. Un modelo de regresión lineal puede representar el comportamiento de las concentraciones de PM<sub>2,5</sub> con un ajuste de ( $R^2=0,87-0,9$ ), permitiendo obtener valores de error de estimación del modelo por debajo de 20  $\mu\text{g}/\text{m}^3$ .

### 4. Referencias bibliográficas

- Borchers-Arriagada, N., Vander Hoorn, S., Cope, M., Morgan, G., Hanigan, I., Williamson, G., & Johnston, F. H. (2024). The mortality burden attributable to wood heater smoke particulate matter (PM<sub>2,5</sub>) in Australia. *Science of The Total Environment*, 921, 171069.
- Kar, A., Ahmed, M., May, A. A., & Le, H. T. (2024). High spatio-temporal resolution predictions of PM<sub>2,5</sub> using low-cost sensor data. *Atmospheric Environment*, 326, 120486.
- Li, J., Zhang, H., Chao, C.-Y., Chien, C.-H., Wu, C.-Y., Luo, C. H., Chen, L.-J., & Biswas, P. (2020). Integrating low-cost air quality sensor networks with fixed and satellite monitoring systems to study ground-level PM<sub>2,5</sub>. *Atmospheric Environment*, 223, 117293.
- Menares, C., Perez, P., Parraguez, S., & Fleming, Z. L. (2021). Forecasting PM<sub>2,5</sub> levels in Santiago de Chile using deep learning neural networks. *Urban Climate*, 38, 100906.
- Peralta, B., Sepúlveda, T., Nicolis, O., & Caro, L. (2022). Space-time prediction of PM<sub>2,5</sub> concentrations in Santiago de Chile using LSTM networks. *Applied Sciences*, 12(22), 11317.
- Tagle, M., Rojas, F., Reyes, F., Vásquez, Y., Hallgren, F., Lindén, J., Kolev, D., Watne, Å. K., & Oyola, P. (2020). Field performance of a low-cost sensor in the monitoring of particulate matter in Santiago, Chile. *Environmental Monitoring and Assessment*, 192(3), 171.
- Tarazona Alvarado, M., Salamanca-Coy, J., Forero-Gutiérrez, K., Núñez, L., Pisco-Guabave, J., Escobar-Díaz, F., & Sierra-Porta, D. (2024). Assessing and monitoring air quality in cities and urban areas with a portable, modular and low-cost sensor station: Calibration challenges. *International Journal of Remote Sensing*, 45(17), 5713–5736.
- Traviss, N., Stanway, J., Woodward, J., Webler, T., Allen, G., & Ahmadi, M. (2025). Four-year community-wide PM<sub>2,5</sub> exposure characterization using a low-cost sensor network in a rural valley influenced by residential wood smoke. *Atmospheric Environment*, 121398.