

Artificial neural network applied for the forecast of critical PM_{2.5} events in the Aburra Valley

Daniela Baena-Salazar, José F. Jiménez, Carmen E. Zapata & Álvaro Ramírez-Cardona

Laboratorio de Calidad del Aire - CALAIRE, Facultad de Minas, Universidad Nacional de Colombia Sede Medellín, Colombia, dbaenas@unal.edu.co,
jffimene@unal.edu.co, cezapata@unal.edu.co, alramirezca@unal.edu.co

Received: December 04th, 2018. Received in revised form: March 08th, 2019. Accepted: March 25th, 2019.

Abstract

The great human health implications of exposure to atmospheric pollution events can have repercussions on the quality of life, economy, and the quality of city's ecosystems. With the possibility of predicting a critical event, the option of taking adequate measures for mitigation or even prevention of these impacts is enabled. In this paper, an Artificial Neural Networks (RNA) model was developed and tested to predict the daily concentration of particulate matter less than 2.5 microns (PM_{2.5}) in the Aburrá Valley (Colombia), with a day of anticipation, based on information from three stations of the Metropolitan Area Air Quality Monitoring Network.

Keywords: Keywords: air pollution; PM_{2.5} forecast; artificial neural network; meteorological data.

Red neuronal artificial aplicado para el pronóstico de eventos críticos de PM_{2.5} en el Valle de Aburrá

Resumen

Las grandes implicaciones que tiene en la salud humana la exposición a eventos de contaminación atmosférica, puede tener repercusiones en la calidad de vida, economía, y en la calidad de los ecosistemas de una ciudad. Con la posibilidad de prever un evento crítico, se habilita la opción de tomar medidas adecuadas para la mitigación o incluso la prevención dichos impactos. En este documento, se ha desarrollado y se ha probado un modelo de Redes Neuronales Artificiales (RNA) para pronosticar la concentración diaria del material particulado menor a 2.5 micras (PM_{2.5}) en el Valle de Aburrá (Colombia), con un día de anticipación, a partir de información de tres estaciones de la Red de Monitoreo de Calidad del Aire del Área Metropolitana.

Palabras clave: contaminación atmosférica; pronóstico de PM_{2.5}; red neuronal artificial; datos meteorológicos.

1. Introducción

La contaminación del aire es uno de los problemas más graves de la sociedad moderna, enmarcados de manera significativa por el desarrollo urbanístico e industrial de la sociedad. Los efectos que puede tener la contaminación en la salud humana ha sido tema de un intenso estudio; dado, su asociación con el aumento en la mortalidad y en enfermedades respiratorias y cardiovasculares [1]. Estos efectos pueden ser vistos a corto y largo plazo, con distintas repercusiones y afectaciones en cualquier etapa de la vida, desde la concepción hasta la vejez [2].

Recientemente, varias organizaciones de protección ambiental en diferentes países han establecido normas nacionales de calidad del aire para proteger así, la salud pública y el medio ambiente; además de los avances científicos que diariamente se presentan en pro de mejorar la calidad atmosférica en muchas ciudades [3]. Sin embargo a pesar de los importantes avances en la comprensión de las emisiones y los diferentes tipos de contaminantes, así como en la reducción de sus niveles en algunas zonas urbanas durante las últimas décadas; se estima que la contaminación atmosférica mata alrededor de 3 millones de personas al año en todo el mundo; en especial, el material particulado (PM) que ha estado relacionado directamente con varias

repercusiones en la calidad de vida y el estado de algunos ecosistemas [4]. Por lo tanto, un modelo preciso y fiable para pronosticar las concentraciones de los contaminantes en el aire proporcionaría información avanzada a una etapa temprana de tal manera que guíe las obras de las autoridades ambientales en favor del control y la protección de la calidad del aire y de la salud pública [5]

Existen dos factores importantes en el control de transferencia y difusión de los contaminante del aire; las emisiones de contaminantes que son las fuentes de contaminación y que pueden ser de origen natural y antropogénico; y las condiciones meteorológicas [6], jugando ambas un papel esencial en las fluctuaciones diarias de las concentraciones de algunos contaminantes en la atmósfera [5]

Los episodios de alta contaminación atmosférica en el Valle de Aburrá por material particulado y algunos aerosoles, se producen con frecuencia por que los niveles de la altura de la capa de mezcla son inferiores a los de las montañas dentro del valle; causado generalmente por periodos de alta estabilidad en la capa limite, en la que se inhibe el mezclado vertical y la dispersión de los contaminantes; esto, relacionado directamente con la cantidad de energía entrante (radiación) y el aire circulante dentro del Valle. La persistencia de la estabilidad atmosférica, a menudo conduce a concentraciones de 24 horas en promedio de PM_{2.5} que excedan los 35 $\mu\text{g}/\text{m}^3$, valor que según el Estándar Ambiental Nacional de Calidad del Aire (NAAQS), son niveles de concentración que pueden tener repercusiones en la salud humana en especial en población sensible como los asmáticos, los niños y los ancianos [7]

Otro factor que hace más crítica la situación de contaminación dentro de la región, es su condición geográfica. Dado, que la población está asentada sobre un estrecho valle, donde las cadenas montañosas bloquean las corrientes de aire que pudieran ventilarla y solo permiten el recorrido de vientos de baja y moderada velocidad procedentes del norte y cuya eficiencia en la remoción de contaminantes resulta insuficiente [8].

2. Metodología

En esta sección, la descripción de la zona de estudio y la metodología sistemática utilizada para el desarrollo del modelo, basado en las redes neuronales artificiales (RNA), es explicado en detalle; incluyendo los pasos a seguir para el desarrollo del modelo.

2.1. Zona de Estudio

El Valle de Aburra se encuentra ubicado en la Cordillera central en el departamento de Antioquia (costado noroccidental de Suramérica); posee una extensión de 1152 Km² que hacen parte de la cuenca del Río Medellín; congrega 10 municipios del departamento (Barbosa, Girardota, Copacabana y Bello al norte; Medellín municipio núcleo; Envigado, Itagüí, Sabaneta, La Estrella y Caldas al sur). El Valle tiene una longitud aproximada de 60 kilómetros; está enmarcado por una topografía irregular y pendiente, que oscila entre los 1300 y 2800 metros sobre el nivel del mar.

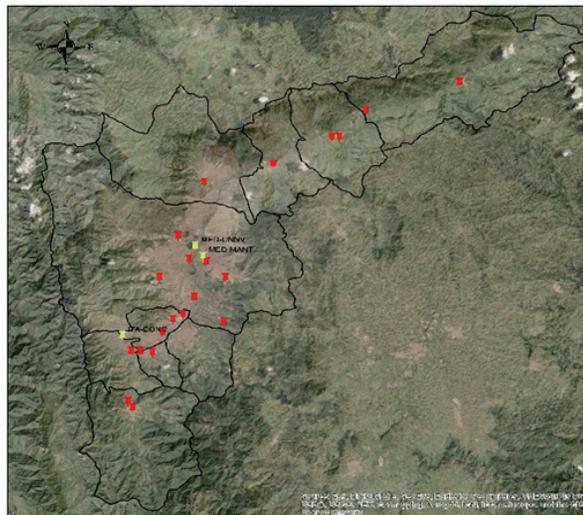


Figura 1. Mapa Valle de Aburrá y sus estaciones de monitoreo.
Fuente: Los Autores.

Las cordilleras que lo encierran, dan lugar a la formación de diversos microclimas, saltos de agua, bosques, sitios de gran valor paisajístico y ecológico [9].

La Red de Monitoreo de Calidad del Aire del Valle de Aburra, del Área Metropolitana, realimenta las bases de datos de manera continua, en material particulado PM₁₀ y PM_{2.5} en pro de la comunidad y la ciencia, haciendo los análisis necesarios para describir los procesos meteorológicos y de calidad atmosférica en la región. Se hace uso de los datos de tres estaciones de dicha red, Itagüí El concejo (ITA-CONC), Museo de Antioquia (MED-MANT) y Universidad Nacional de Colombia- Sede Medellín el Volador (MED-UNNV); con el fin de estudiar las variables meteorológicas y de calidad del aire para el desarrollo de un modelo estadístico para pronosticar eventos de contaminación atmosférica; modelo basado en la teoría de las redes neuronales explicado a continuación, junto con las variables predictores y la variable a pronosticar. Estas tres estaciones fueron seleccionadas en base al tiempo de monitoreo y variables disponibles, la cantidad de datos hace eficiente o ineficiente el modelo estadístico.

2.2. Redes Neuronales

Las Redes Neuronales son un modelo artificial y simplificado del cerebro humano, el cual se aplica perfectamente para sistemas que son capaz de adquirir conocimiento a través de la experiencia. Es un modelo matemático compuesto por un número de elementos procesales organizados en niveles, la capa de entrada, que recibe directamente la información proveniente de las fuentes externas de la red; la capa oculta, capa interna de la red que no tiene contacto con el entorno exterior y transforman los datos del nivel anterior; y la capa de salida, que transfiere la información de la red al exterior [10]. A su vez, cada nivel está compuesto de nodos de simple procesamiento o neurona, que según sus necesidades se organiza en diferentes capas.

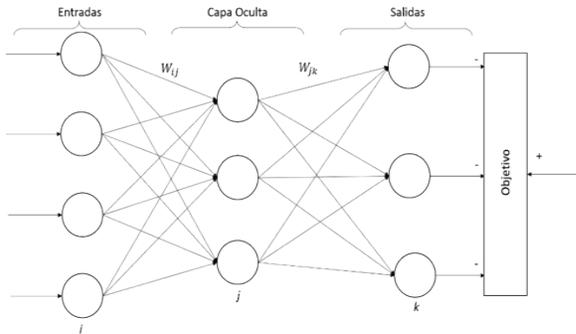


Figura 2.. Esquema de la estructura de una Red Neuronal Artificial.
Fuente [10]

Cada nodo calcula una combinación lineal de las entradas ponderadas de los enlaces que desembocan en él y las transforma mediante funciones, ya sean lineales o no lineales (Fig. 2).

La salida obtenida se hace pasar entonces como una entrada de otros nodos en la siguiente capa. Un importante requisito para este tipo de funciones no lineal es que se debe asignar a cualquier entrada un rango de salida finita, por lo general entre 0 y 1 o -1 y 1 [11].

Como se aprecia en la Fig. 2 en la arquitectura de las RNA, los flujos de información están estrictamente hacia adelante. Las expresiones básicas de las salidas de la capa oculta y la capa de salida son las siguientes ecuaciones (1) y (2):

$$h_j = f_{capa\ oculta} \left(\sum_{i=1}^m w_{ij} x_i \right) \quad (1)$$

$$Y_k = f_{capa\ de\ salida} \left(\sum_{j=1}^n w_{jk} h_j \right) \quad (2)$$

Donde $x = (x_1, x_2, \dots, x_i)$ ($i = 1, 2, \dots, m$) representan las entradas, $h = (h_1, h_2, \dots, h_j)$ ($j = 1, 2, \dots, n$) representan las salidas de la capa oculta, $Y = (y_1, y_2, \dots, y_k)$ ($k = 1, 2, \dots, n$) son las salidas de la red, w es la matriz de peso entre dos capas, $f_{(capa\ oculta)}$ y $f_{(capa\ de\ salida)}$, son las funciones de transferencia tanto para la capa oculta como para la capa de salida respectivamente [11].

El proceso para el desarrollo de la red neuronal artificial, puede resumirse como:

Paso 1. Tener una base de datos, con información histórica de la concentración del contaminante y de los datos la meteorología de la zona.

Paso 2. Descomponer la serie de tiempo en niveles detallados según su necesidad, por variable o parámetro de interés.

Paso 3. Normalizar los parámetros [0,1] según la ecuación (3) respectivamente.

$$Z_i = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Paso 4. Aplicar a cada nodo la ecuación que transformará cada variable dentro de la RNA, algunas de ellas se pueden ver a continuación en la Tabla 1.

Tabla 1 Ecuaciones de transformación.

Función	Fórmula
Identidad	$y = x$
Escalón	$y = \begin{cases} +1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$ $y = \begin{cases} +1 & \text{si } x \geq 0 \\ -1 & \text{si } x < 0 \end{cases}$
Lineal por tramos	$y = \begin{cases} x & \text{si } -1 \leq x \leq 1 \\ +1 & \text{si } x > 1 \\ -1 & \text{si } x < -1 \end{cases}$
Sigmoidea	$y = \frac{1}{1 + e^{-x}}$ $y = \tanh(x)$
Sinusoidal	$y = \text{sen}(\varphi x + \varphi)$

Fuente [11]

Paso 5. Calcular los coeficientes que acompañan las entradas de cada nodo en las distintas capas de la RNA, a través de un análisis de sensibilidad.

Paso 6. Ejecución y salida de los resultados del pronóstico.

Al igual que cualquier otro modelo estadístico, las RNA deben en general, ser entrenados, calibrados y probados (validado) con dos conjuntos de datos independientes, proporcionando una mayor estabilidad en los resultados finales [12].

Para este trabajo la RNA tuvo como nodos de entrada las variables meteorológicas y las variables de calidad que con un peso dentro del modelo, definido por métodos tipo Montecarlo, aportan en mayor o menor medida información al pronóstico de la variable dependiente.

El interés primordial dentro de este modelo, es definir claramente los picos que tiene los altos niveles de PM2.5 promedio diarios dentro de la región, por tal, la capa oculta tuvo como función de transformación las funciones lineales, dado que no tiende a suavizar la serie, como las demás funciones de transformación; arrojando mejor precisión en la predicción de los altos niveles de material particulado.

2.3. Bases de datos y criterios de acción:

Como se mencionó en un inicio, La Red de Monitoreo de Calidad del Aire del Valle de Aburra a cargo del Área Metropolitana, está en un constante monitoreo de la calidad del aire dentro de la región, alimentando sus bases de datos; aprovechando eso, se tomaron 3 estaciones automáticas de PM2.5 y meteorología para hacer este modelo. Como método de elección de estas estaciones, en un principio, se basó en el límite mínimo de datos que deberá tener la red para almacenar como posibles experiencias; esto, para asegurar que la red tenga la suficiente información con la cual relacione los episodios de contaminación y las variables de entrada. Y como segundo filtro en la selección de las estaciones, es necesario que los puntos de muestreo cuenten con quipos que monitoreen todas las variables (PM2,5, precipitación, radiación, velocidad del viento, dirección del viento, presión, temperatura y humedad) , es decir no se

aceptaron aquellos puntos de monitoreo que no tenían mediciones de alguno de los parámetros de entrada al modelo.

Una medida adicional dentro de las entradas del modelo y que es de alta importancia, es la altura de la capa de mezcla, variable que se calculó con los datos obtenidos del radiómetro a cargo del Sistema de Alerta Temprana del Valle de Aburrá (SIATA). Dada la poca variabilidad en la capa de mezcla en una región, este calor de la altura de capa de mezcla es igual para todas las estaciones.

La base de datos de todos los parámetros que participan en este trabajo transcurre en un periodo de enero de 2013 hasta marzo de 2016 (excepto radiómetro que fue hasta el mes de febrero del mismo año).

Ahora, la selección de las variables que participan en la RNA como nodos de entrada; parte del análisis determinístico de cada una de ellas y sus correlaciones conjuntas; éste, permitiendo descartar alguna de ellas, si su correlación con otra es lo suficientemente dependiente que el modelo omita la participación de esta dentro del pronóstico. Las variables seleccionadas como nodos de entrada en la RNA son, la altura de la capa de mezcla, la temperatura, la humedad relativa, la radiación, la precipitación, la precipitación acumulada de los últimos 3 días, el viento y el material particulado de uno, dos y tres días anteriores.

2.3.1 Altura de capa de mezcla

La capa de mezcla es uno de esos parámetros claves para el análisis de la calidad del aire y de los contaminantes [13]. Es la parte de la atmósfera que está directamente afectada por los sucesos en la superficie y rige la mayoría de las variables meteorológicas. Su espesor es muy variado; puede estar entre decenas de metros a unidades de kilómetros (corresponde al 20 % de la troposfera en promedio); y generalmente responde a los forzamientos de esa superficies en periodos del orden horario o menos [14]. En general, la concentración de cualquier trazador dentro de la capa limite varia debido a los cambios en el volumen de mezcla [15]

La capa de mezcla determina el aire disponible para la dispersión de todos los componentes en la atmosfera, incluida la contaminación y el vapor en agua emitido en superficie [16]. Pero a pesar de su importancia crítica, la altura de la capa de mezcla no puede ser directamente medida dada su complejidad, pero es estimada con diferentes métodos matemáticos y de observación [17].

En este caso, se utiliza el método de la parcela, como herramienta para la determinación de la altura de la capa de mezcla; método que parte con la estabilidad atmosférica como referente para definirla, la cual, se determina provocando el ascenso o descenso de una parcela hipotética y luego compararla con las condiciones resultantes de la parcela con las condiciones del ambiente circundante (Fig. 3). Se supone que las parcelas o burbujas de aire que ascienden o descienden experimentan cambios de temperatura y humedad asociados con dos procesos primarios: cambios de presión y la liberación o absorción de calor latente debido a los procesos de evaporación o condensación. Además, se supone que la parcela no interactúa con el ambiente [19].

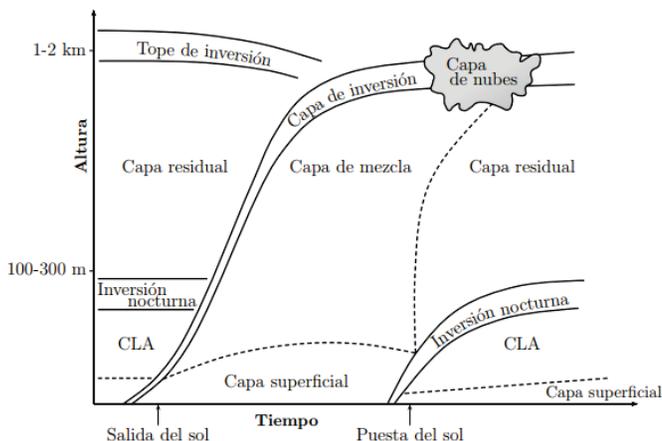


Figura 3. Capa de mezcla.

Fuente: [18]

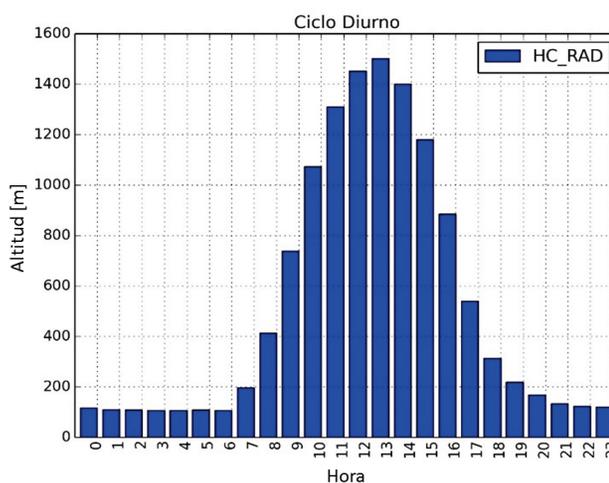


Figura 4. Ciclo diario típico de la altura capa de mezcla en el Valle de Aburrá, periodo febrero de 2013- febrero de 2016.

Fuente: Los Autores.

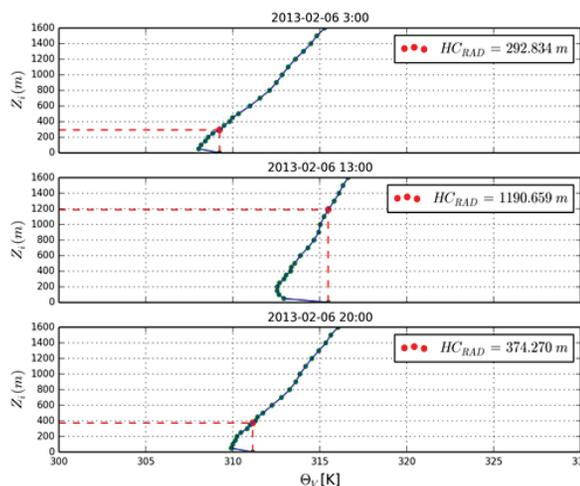


Figura 5. Resultados Método de la Parcela para el 2013-02-06.

Fuente:

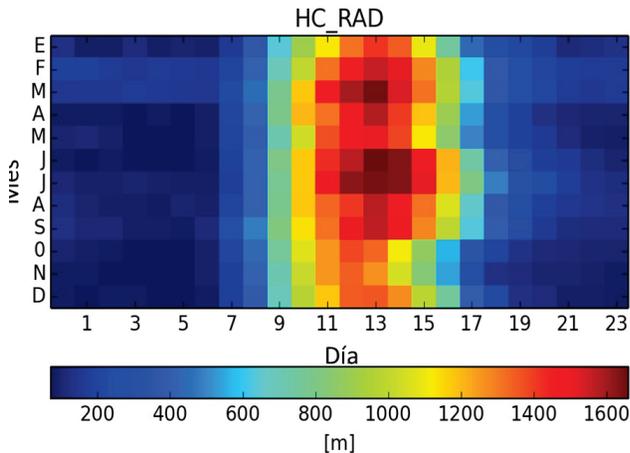


Figura 6. Ciclo diario anual altura capa de mezcla, periodo 2013-2016.
Fuente: Los Autores.

El método de la parcela, no es efectivo para el análisis de la altura de la capa de mezcla en horas nocturnas, debido a la complejidad que supone el análisis bajo condiciones de estabilidad atmosférica e inversión térmica. Para el análisis propuesto, se calculó la altura de la capa de mezcla para el periodo comprendido entre las 7 - 19 horas (local); posteriormente, se calculó el valor promedio diario de la altura de la capa de mezcla (Fig. 4).

La Fig. 5 representa un ejemplo de la aproximación en la altura de la capa de mezcla, a partir del método de la parcela, para el día 6 de febrero de 2013, a tres horas distintas (3, 13 y 20). Se evidencia que la altura máxima coincide con las horas donde se registra mayor radiación solar incidente.

El análisis del comportamiento de la altura de la capa de mezcla para el conjunto de datos disponibles (2013-2016), evidencia que los meses de junio y julio son los meses en los que se registra una altura promedio más alta; y los meses de octubre, noviembre y diciembre son los meses en registrar la altura promedio más baja (Ver Fig. 6). También se evidencia que los meses de junio, julio, agosto y septiembre tienden a registrar alturas superiores a los 1000 metros durante más tiempo en el ciclo diurno.

2.3.2. Temperatura en superficie

Algunos estudios han demostrado que los efectos de algunos contaminantes son más adversos durante determinadas estaciones climáticas, evidenciando dinámicas de los contaminantes en relación a la temperatura superficial y la radiación solar incidente [20]. Se evidencia además que, durante periodos cálidos en algunas regiones, se pueden presentar los niveles más altos de PM_{2.5} [21].

La inclusión de la temperatura al modelo, como la mayoría de las variables de entrada (excepto precipitación), se ingresa por su promedio diario. La Fig. 7 a) muestra el ciclo anual promedio que tiene la temperatura en superficie, en la cual se evidencia que los meses de junio, julio, agosto y diciembre se registran las mayores magnitudes promedio de la temperatura superficial sobre un tiempo continuo más

largo, mientras que los meses de abril, mayo, octubre, noviembre se registran las menores magnitudes de la temperatura superficial sobre un tiempo continuo menor.

2.3.3. Radiación en superficie

La variación de la radiación solar incidente en la superficie de la Tierra afecta profundamente los procesos dinámicos y termodinámicos a nivel global, debido a su relación con el balance de energía y los flujos de calor latente y sensible [24]. Por otra parte, éste parámetro cumple un papel importante en los procesos turbulentos y de dispersión de contaminantes en la atmosfera [25].

La Fig. 7 c) muestra el ciclo anual promedio de la radiación solar incidente. Los meses de mayor radiación solar incidente se registra durante los meses de julio, agosto y septiembre.

2.3.4. Humedad en superficie

La relación existente entre diferentes gases y material particulado con los porcentajes de humedad relativa en superficie es motivo de estudio en diferentes partes del mundo. Algunos estudios explican la relación positiva entre algunos de ellos (como son SO₂, PM_{2.5} y CO) con la humedad y otros una tendencia negativa, entre los que se mencionan al O₃ y NO₂ [13].

La humedad relativa, variable que presenta una alta correlación con la temperatura en superficie y tiene un comportamiento similar (bimodal), no fue descartada por su importante participación en la formación de PM_{2.5} secundario, al ser el medio en el que partículas ya existentes pueden ser conglomeradas y formar otras de igual tamaño [22][23]. La Fig. 8 b) muestra el ciclo anual de esta variable, indicando cómo los meses más secos son los meses comprendidos entre junio y septiembre.

2.3.5. Precipitación

En varias regiones de montaña se evidencia que, durante el invierno, éstas pueden registrar altas concentraciones de PM_{2.5}, fenómeno que está relacionado con la ausencia de la radiación solar y la nubosidad presente[26]; Sin embargo, la precipitación puede interferir con los procesos de expansión de la capa de mezcla y con ello el efectivo mezclado de los contaminantes. El Valle de Aburrá presenta dos épocas de precipitación centradas alrededor de los meses de abril y de octubre [27]. En la Fig. 8 d) se muestra el ciclo promedio diurno anual de éste parámetro. Se evidencia que los meses de mayor precipitación se localizan alrededor de mayo y noviembre, mientras que los meses de junio y julio se registran menores precipitaciones. También se evidencia que los niveles de concentración alta en contaminantes atmosféricos coinciden con las épocas previas a los meses de alta precipitación.

La inclusión de esta variable dentro del modelo, se toma desde dos puntos de vista; el primero, es la precipitación acumulada del día anterior, que implica una afectación directa en relación en la acumulación o lavado de los contaminantes en la atmosfera, lo cual se refleja en las

concentraciones de PM2.5; Y segundo, se toma como otra variable de entrada para la red, la precipitación acumulada durante los últimos tres días, dada la relación que tiene esos días de precipitación acumulada con otros parámetros

dependientes como son la presión atmosférica y la humedad relativa.

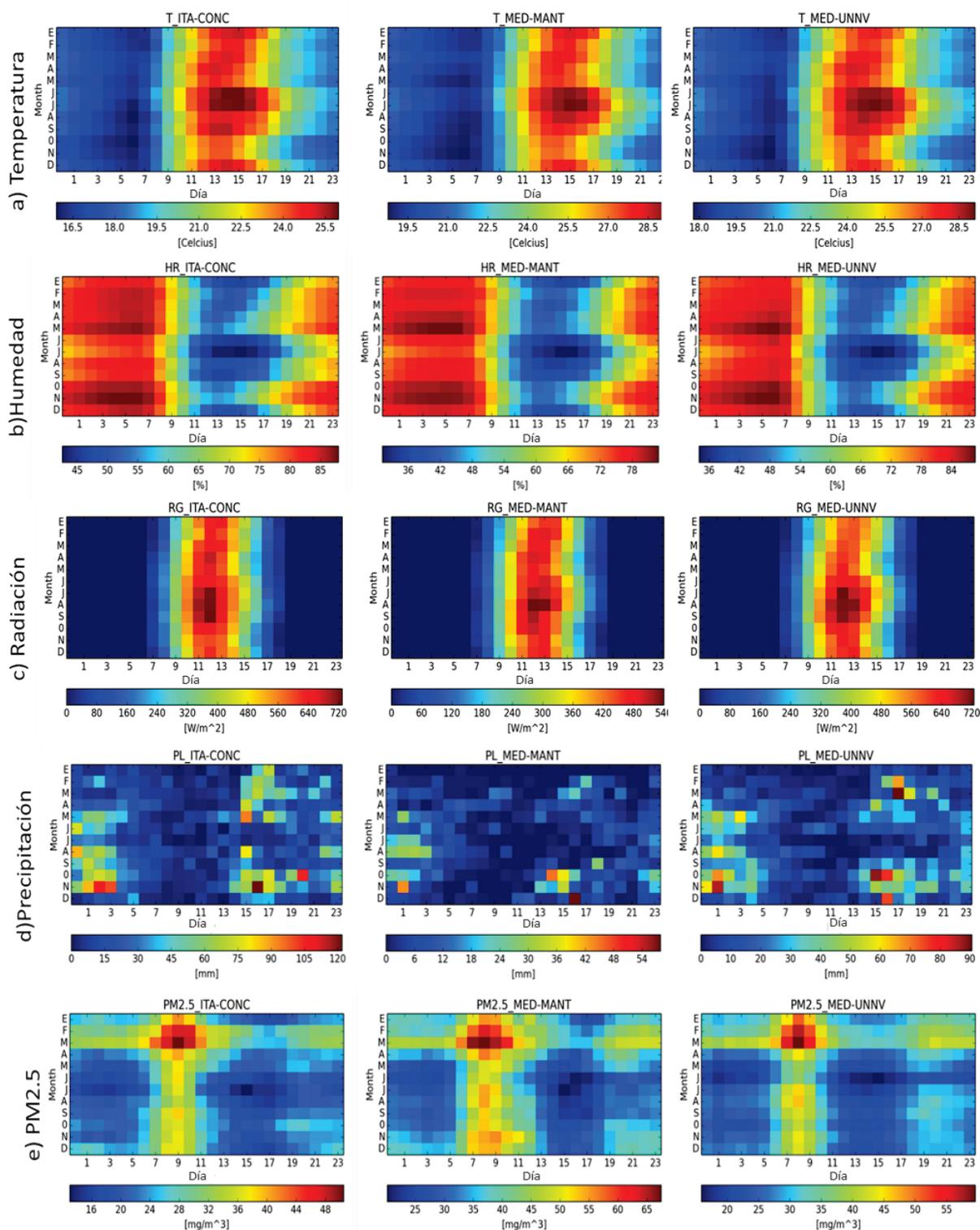


Figura 7 Ciclo anual de las variables predictoras, periodo 2013-2016.
 a) Temperatura, b) Humedad, c) Radiación, d) Precipitación y e) PM2.5.

Fuente: Los Autores.

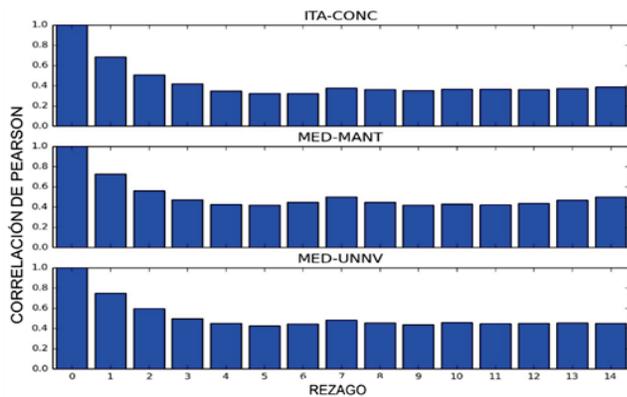


Figura 8. Autocorrelograma PM2.5, periodo 2013-2016.
Fuente: Los Autores.

2.3.6. Material particulado menor a 2.5 micras (PM2.5)

La Fig. 7 e) muestra el ciclo anual promedio que tiene el contaminante PM2.5, en la cual se evidencia que las mayores concentraciones se registraron durante los meses de febrero y marzo; por otra parte, el ciclo diurno manifiesta una tendencia a registrar las mayores concentraciones durante las horas de la mañana que, particularmente en el Valle de Aburrá, son horas pico de transporte vehicular, horas donde la gente va hacia su trabajo y actividades, y las horas donde las condiciones de la estabilidad atmosférica, y por ende una reducida altura de capa de mezcla, pueden estar presentes aún. Se evidencia también que las menores concentraciones se registran durante el periodo de meses comprendido entre mayo y octubre, cuya dinámica diurna tiende a registrar las menores concentraciones después de las horas del mediodía hasta las 6 del otro día. El mes de marzo es el que se registran más altas concentraciones horarias promedio; éste mes se caracteriza por ser un mes de transición de época seca a época húmeda y es un periodo crítico donde se pueden presentar eventos de contaminación debido a las condiciones meteorológicas y climáticas significantes

La Fig. 8 esquematiza el autocorrelograma rezagado 15 días para el PM2.5, en ésta se evidencia que la concentración tiene una correlación relativamente alta consigo mismo hasta con tres días de rezago, es decir, que los niveles de contaminación de los tres días anteriores influyen en la concentración del contaminante el día actual. Basado en esta relación, se ha tomado como variables de entrada al modelo las concentraciones de PM2.5 de los tres días anteriores.

2.3.7. Velocidad del Viento Superficial

Uno de los parámetros más importantes en la dispersión y transporte de los contaminantes, es la variable viento [28]. El transporte de los contaminantes en el interior de las urbes, está influenciada por numerosos factores como la configuración del dosel urbano, la fuente del contaminante y las condiciones del flujo de entrada [29]-[32]; en las urbes, las altas concentraciones de contaminantes se relacionan con

bajas velocidades del viento [31], [33]-[35].

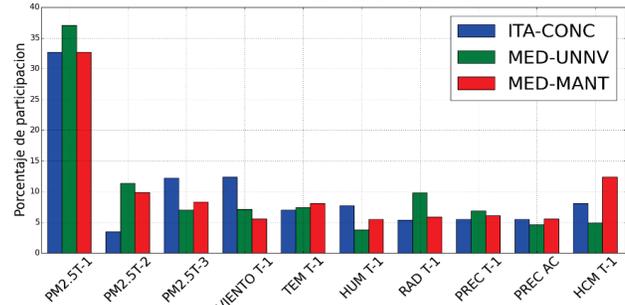


Figura 9. Porcentajes de participación, variables de entrada.
Fuente: Los Autores.

Dada la importancia de éste parámetro, en función del transporte y dispersión de los contaminantes, se tiene en cuenta como variables de entrada para la RNA.

3. Resultados y discusiones

La Fig. 9 esquematiza los pesos de los parámetros en términos de porcentaje de las variables de entrada (datos meteorológicos y PM2.5) a la RNA, en donde se le da mayor participación en las tres estaciones al PM2.5 registrado el día anterior, con un porcentaje promedio de influencia de 34.2%.

En el caso que nos ocupa, la validación del modelo se hizo con el pronóstico de los primeros 91 días el año 2016, periodo elegido por sus eventualidades durante el periodo marzo-abril en el cual ocurrió una contingencia ambiental dentro del Valle de Aburrá, en material particulado PM2.5 y otros gases.

En La Fig. 10, se muestra los resultados del pronóstico para las 3 estaciones, el periodo usado para calentar y calibrar la RNA está acotado por la línea amarilla; por lo tanto, el periodo de validación de los resultados es posterior a la línea amarilla. Se evidencia que, para las tres estaciones, los resultados del valor pronosticado se caracterizan por seguir la tendencia del valor real modelo. También se evidencia que los picos de altas concentraciones son pronosticados con una mejor precisión para las estaciones de MED-UNNV y MED-MANT.

Para estudiar la relación entre los valores reales y durante el proceso de calibración y validación, las Figs. 11, 12 y 13 muestran los gráficos de dispersión de los datos, en donde la correlación entre los dos conjuntos de valores, los puntos tienden a concentrarse en la línea de identidad. Se evidencia que la más alta correlación corresponde a la estación MED-UNNV con un $r^2=0.864$, mientras que con un $r^2=0.806$ ITA-CONC es la estación con menor correlación entre los valores reales y los predichos.

En la Fig. 14 Se muestra la validación de los primeros 91 días del año, en la que el modelo tiende a aumentar su error para el último mes, esto debido a que en ese mes, una de las variables de entrada no tenía datos (altura de capa de mezcla, esto asociados a la ausencia de datos del radiómetro, con el cuál se calcula estas alturas), entonces el modelo independientemente de la ausencia de esta variable da un reporte para la predicción del PM2.5, partiendo de los demás

parámetros, pero aumenta su error a medida que se deja de



alimentar la RNA en cada una de sus entradas.

Figura 10. Pronóstico estaciones.

Fuente: Los Autores.

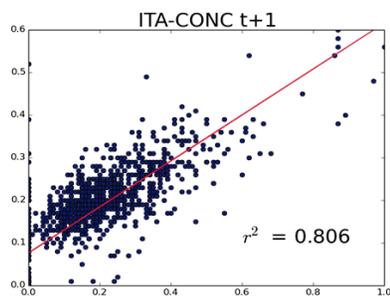


Figura 11. Correlación valores previsto y previstos ITA-CONC.

Fuente: Los Autores.

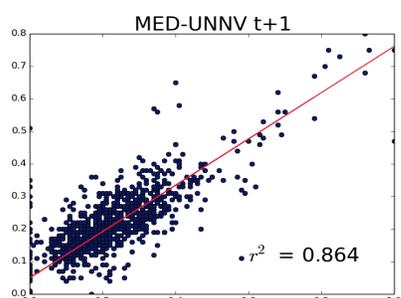


Figura 12 Correlación valores previsto y previstos MED- UNNV.

Fuente: Los Autores.

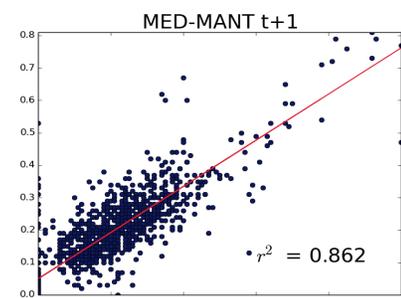


Figura 13. Correlación valores reales y previstos MED-MANT.

Fuente: Los Autores.

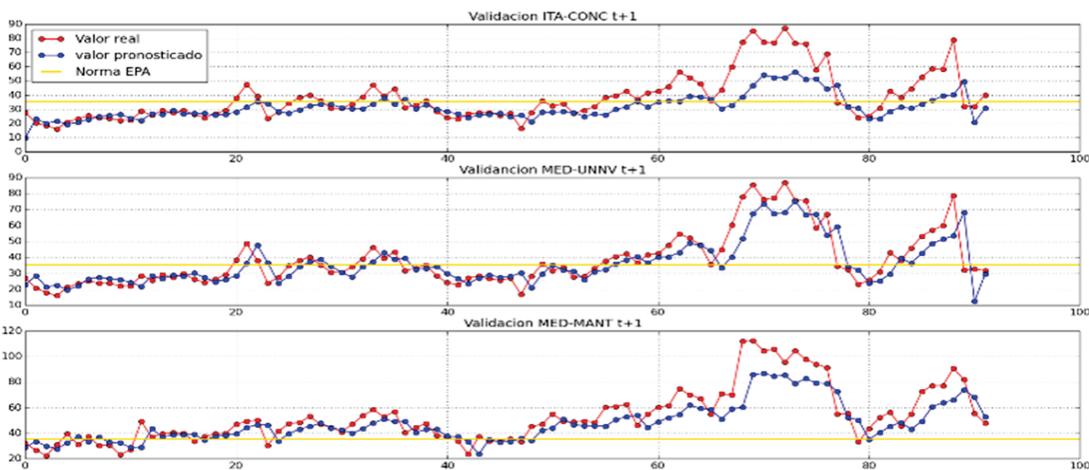


Figura 14. Validación RNA.

Fuente: Los Autores

La Fig. 14 muestra en la línea amarilla, el umbral en el que la EPA considera que las concentraciones PM2.5 superior a $35 \mu\text{g}/\text{m}^3$, comienza a tener repercusiones a la salud pública; es decir, es el límite en el que se define un evento de contaminación crítica para este trabajo. El modelo, identifica dentro del periodo de validación los posibles eventos de contaminación crítica, pero difiere con la concentración del PM2.5. Con un porcentaje superior al 70% de la eficiencia del modelo en los pronósticos de posibles eventos críticos, el error del modelo para la estación ITA-CON durante esta validación fue de 18.52%; y para la estación MED-UNNV y MED-MANT fueron de 12.35% y 12.05% respectivamente

Se recomienda seguir trabajando en este modelo de pronóstico, pues sin duda alguna una, queda encontrar el factor que correlacione con menor error el valor pronosticado con el valor observado.

4. Conclusiones

Este documento, se construye para la previsión de las concentraciones diarias de PM2.5, teniendo en cuenta información de variables meteorológicas y de calidad del aire para el Valle de Aburrá en escala diaria. Las variables de entrada se definen con base en análisis de correlación. Sus señales son transformadas con las funciones de activación que no suavizaran excesivamente las salidas del modelo.

La precisión de la predicción del número de días con eventos de contaminación crítica, es relativamente buena, pero sin embargo no se descarta la posibilidad de que el modelo pueda mejorar si se toman medidas estadísticas más estrictas de descarte. Ejemplo es. PM2.5 T-2 variable que el modelo omitió, probablemente por su alta relación con el PM2.5 T-1.

Los modelos de RNA funcionan mejor cuando la información disponible corresponde a las distintas variables relacionadas con el fenómeno en cuestión, son más larga y completa. En el caso de la predicción de eventos críticos de contaminación urbana, son importantes tanto las variables meteorológicas como los registros de los contaminantes.

Por su sencillez, los modelos estadísticos pueden proveer una primera aproximación al pronóstico de eventos críticos de contaminación atmosférica debida a diferentes contaminantes [material particulado, dióxido de azufre (SO₂), dióxido de nitrógeno (NO₂), monóxido de carbono (CO) y ozono (O₃)], y por esta vía mitigar su efecto.

El modelo, identifica dentro del periodo de validación los posibles eventos de contaminación crítica, pero tiene un margen de error entre los valores pronosticados y los valores observados. Aunque presentó una eficiencia del 70 % una falencia de este modelo es que basado en los días anteriores, no ha tenido mucha efectividad en pronosticar el primer día de eventos críticos

Bibliografía

- [1] Bert, B. and Stephen, H.T., Air pollution and health, *Air Pollut. Heal.*, 360, pp. 115-148, 1999. DOI: 10.1016/S0140-6736(02)11274-8
- [2] Ruckerl, R., Schneider, A., Breitner, S., Cyrus, J. and Peters, A., Health effects of particulate air pollution: a review of epidemiological evidence, *Inhal. Toxicol.*, 23(10), pp. 555-592, 2011. DOI: 10.3109/08958378.2011.593587
- [3] Cantelli, A., D'Orta, F., Cattini, A., Sebastianelli, F. and Cedola, L., Application of genetic algorithm for the simultaneous identification of atmospheric pollution sources, *Atmos. Environ.*, 115, pp. 36-46, 2015. DOI: 10.1016/j.atmosenv.2015.05.030
- [4] Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C. and Baklanov, A., Real-time air quality forecasting, part I: history, techniques, and current status, *Atmos. Environ.*, 60, pp. 632-655, 2012. DOI: 10.1016/j.atmosenv.2012.06.031
- [5] Bai, Y., Li, Y., Wang, X., Xie, J. and Li, C., Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions, *Atmos. Pollut. Res.*, 7(3), 2016. DOI: 10.1016/j.apr.2016.01.004
- [6] He, J., Yu, Y., Liu, N. and Zhao, S., Numerical model-based relationship between meteorological conditions and air quality and its implication for urban air quality management, *Int. J. Environ. Pollut.*, 53(3/4), pp. 265, 2013. DOI: 10.1504/IJEPP.2013.059921
- [7] Silcox, G.D., Kelly, K.E., Crosman, E.T., Whiteman, C.D. and Allen, B.L., Wintertime PM2.5 concentrations during persistent, multi-day cold-air pools in a mountain valley, *Atmos. Environ.*, 46, pp. 17-24, 2011. DOI: 10.1016/j.atmosenv.2011.10.041
- [8] Bedoya, J., Air quality in the Aburrá Valley, Antioquia-Colombia, pp. 7-15, 2008.
- [9] Área Metropolitana del Valle de Aburrá, Atlas Metropolitano. 2010.
- [10] Match, D.J., Redes neuronales: conceptos básicos y aplicaciones. Historia Santiago, 2001, 55 P.
- [11] Trigo, R.M. and Palutikof, J.P.P., Simulation of daily temperatures for climate change scenarios over Portugal: a neural network model approach, *Clim. Res.*, 13(1), pp. 45-59, 1999. DOI: 10.3354/cr013045
- [12] Tangang, F.T., Tang, B., Monahan, A.H. and Hsieh, W.W., Forecasting ENSO events: a neural network-extended EOF approach. *J. Clim.*, 11(1), pp. 29-41, 1998. DOI: 10.1175/1520-0442(1998)011<0029:FEEANN>2.0.CO;2
- [13] Elminir, H.K., Dependence of urban air pollutants on meteorology., *Sci. Total Environ.*, 350(1-3), pp. 225-237, 2005. DOI: 10.1016/j.scitotenv.2005.01.043
- [14] Stull, R., *Meteorology*, Second Ed. Garland, Gary, 2000.
- [15] Chambers, S.D. et al., Quantifying the influences of atmospheric stability on air pollution in Lanzhou, China, using a radon-based stability monitor, *Atmos. Environ.*, 107, pp. 233-243, 2015. DOI: 10.1016/j.atmosenv.2015.02.016
- [16] Seibert, P.P. et al., Review and intercomparison of operational methods for the determination of the mixing height, *Atmos. Environ.*, 34(7), pp. 1001-1027, 2000. DOI: 10.1016/S1352-2310(99)00349-0
- [17] Collaud-Coen, M., Praz, C., Haefele, A., Ruffieux, D., Kaufmann, P.P. and Calpini, B., Determination and climatology of the planetary boundary layer height above the Swiss plateau by in situ and remote sensing measurements as well as by the COSMO-2 model, *Atmos. Chem. Phys.*, 14(23), pp. 13205-13221, 2014. DOI: 10.5194/acp-14-13205-2014
- [18] Jiménez, J.F., Altura de la capa de mezcla en un área urbana, montañosa y tropical. Caso de estudio: Valle de Aburrá (Colombia), Tesis Doctoral, Universidad de Antioquia, Medellín, Colombia, 2016, 33P. Disponible en: <http://hdl.handle.net/10495/5738>
- [19] Tecnol, C., Método propuesto para estimar la altura de capa de mezcla en la atmósfera, con ayuda de la aviación, 2011.
- [20] Nawrot, T.S. et al., Stronger associations between daily mortality and fine particulate air pollution in summer than in winter: evidence from a heavily polluted region in western Europe., *J. Epidemiol. Community Health*, 61(2), pp. 146-149, 2007. DOI: 10.1136/jech.2005.044263

- [21] Kim-Oanh, N.T. et al., Particulate air pollution in six Asian cities: spatial and temporal distributions, and associated sources, *Atmos. Environ.*, 40(18), pp. 3367-3380, 2006. DOI: 10.1016/j.atmosenv.2006.01.050
- [22] Cheng, Y., He, K., Du, Z., Zheng, M., Duan, F. and Ma, Y., Humidity plays an important role in the PM_{2.5} pollution in Beijing., *Environ. Pollut.*, 197, pp. 68-75, 2015. DOI: 10.1016/j.envpol.2014.11.028
- [23] Ministerio del Medio Ambiente de Chile, Contaminación del Aire. Capítulo 1, Inf. Estado del Medio Ambient. 2011, 68 P.
- [24] Wild, M. et al., From dimming to brightening: decadal changes in solar radiation at Earth's surface., *Science*, 308(5723), pp. 847-50, 2005. DOI: 10.1126/science.1103215
- [25] Kogawa, T., Okajima, J., Sakurai, A., Komiya, A. and Maruyama, S., Influence of radiation effect on turbulent natural convection in cubic cavity at normal temperature atmospheric gas, *Int. J. Heat Mass Transf.*, 104, pp. 456-466, 2017. DOI: 10.1016/j.ijheatmasstransfer.2016.08.059
- [26] Silva, P.P.J., Vawdrey, E.L., Corbett, M. and Erupe, M., Fine particle concentrations and composition during wintertime inversions in Logan, Utah, USA, *Atmos. Environ.*, 41(26), pp. 5410-5422, 2007. DOI: 10.1016/j.atmosenv.2007.02.016
- [27] Hermelin, M., Valle de Aburrá, Gob. Antioquia, no. 2, pp. 7-16, 2007.
- [28] Galatioto, F. and Bell, M.C., Exploring the processes governing roadside pollutant concentrations in urban street canyon, *Environ. Sci. Pollut. Res.*, 20(7), pp. 4750-4765, 2013. DOI: 10.1007/s11356-012-1428-5
- [29] Blocken, B., Stathopoulos, T., Carmeliet, J. and Hensen, J.L.M., Application of computational fluid dynamics in building performance simulation for the outdoor environment: an overview, *J. Build. Perform. Simul.*, 4(2), pp. 157-184, 2011. DOI: 10.1080/19401493.2010.513740
- [30] Michioka, T., Sato, A., Takimoto, H. and Kanda, M., Large-Eddy simulation for the mechanism of pollutant removal from a two-dimensional street Canyon, *Boundary-Layer Meteorol.*, 138(2), pp. 195-213, 2011. DOI: 10.1007/s10546-010-9556-2
- [31] Murena, F., Monitoring and modelling carbon monoxide concentrations in a deep street canyon: application of a two-box model, *Atmos. Pollut. Res.*, 3(3), pp. 311-316, 2012. DOI: 10.5094/APR.2012.034
- [32] Soulhac, L., Perkins, R.J. and Salizzoni, P.P., Flow in a street canyon for any external wind direction, *Boundary-Layer Meteorol.*, 126(3), pp. 365-388, 2008. DOI: 10.1007/s10546-007-9238-x
- [33] Cui, P.P.Y., Li, Z. and Tao, W.Q., Investigation of re-independence of turbulent flow and pollutant dispersion in urban street canyon using numerical wind tunnel (NWT) models, *Int. J. Heat Mass Transf.*, 79, pp. 176-188, 2014. DOI: 10.1016/j.ijheatmasstransfer.2014.07.096
- [34] Kikumoto, H. and Ooka, R., A numerical study of air pollutant dispersion with bimolecular chemical reactions in an urban street canyon using large-eddy simulation, *Atmos. Environ.*, 54, pp. 456-464, 2012. DOI: 10.1016/j.atmosenv.2012.02.039
- [35] Takano, Y. and Moonen, P.P., On the influence of roof shape on flow and dispersion in an urban street canyon, *J. Wind Eng. Ind. Aerodyn.*, 123, pp. 107-120, 2013. DOI: 10.1016/j.jweia.2013.10.006

D. Baena-Salazar, recibió su título de Ing. Ambiental en el año 2016. Con más de tres años y medio de experiencia en ciencias atmosféricas, en particular en el análisis de datos de calidad del aire, ruido, redes de monitoreo y planes de manejo de emisiones atmosféricas. Dos años de experiencia y práctica en consultoría para temas de evaluación de impactos ambientales, educación ambiental, planes de manejo, gestión ambiental y solicitud de. Es la fundadora de la organización sin ánimo de lucro Wom-en.
ORCID: 0000-0003-3937-1047

J.F. Jiménez, terminó el pregrado en Ing. Civil en el año de 1989, la MSc. en Aprovechamiento en Recursos Hidráulicos en 1993 y es Dr. en Ingeniería en el 2016, todos estos de la Universidad Nacional de Colombia. Ha realizado investigaciones en equilibrio fluvial, métodos estadísticos, modelamiento matemático, hidrodinámica, climatología, entre otros. Actualmente es docente titular de la Facultad de Minas, Universidad Nacional de Colombia, Sede Medellín. Ha dirigido más de 16 tesis de posgrados y más de 22 publicaciones.

ORCID: 0000-0003-2109-2373

C.E. Zapata, es Ing. de Petróleos, en el año 1986 y MSc. en Water Environment en 1995. Profesora de Ingeniería de Petróleos hasta 2002, de la Especialización en Gestión Ambiental desde 1998 hasta la actualidad, docente de la Maestría en Medioambiente y Desarrollo desde el 2002. Directora de la Oficina de Gestión Ambiental de la Universidad Nacional de Colombia, sede Medellín desde agosto del 2015.
ORCID: 0000-0002-2936-5973

Á. Ramírez-Cardona, es Ing. Ambiental egresado en el año 2016. Su experiencia profesional predomina en las áreas de la consultoría ambiental y estudios técnicos ambientales relacionados con emisiones atmosféricas, calidad del aire, calidad del agua y evaluación ambiental. Entre sus intereses profesionales: programación, modelación y análisis de datos lenguajes como Python, plataformas SIG y otros modelos de predicción numérica de estado del tiempo como es el WRF.
ORCID: 0000-0003-2887-7763



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Medio Ambiente

Oferta de Posgrados

Doctorado en Ingeniería - Recursos Hidráulicos
Maestría en Ingeniería - Recursos Hidráulicos
Maestría en Medio Ambiente y Desarrollo
Especialización en Aprovechamiento de
Recursos Hidráulicos
Especialización en Gestión Ambiental

Mayor información:

E-mail: acma_med@unal.edu.co
Teléfono: (57-4) 425 5105