



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA AMBIENTAL

Modelo Estocástico para el pronóstico de la concentración del Material Particulado (PM₁₀)
en los distritos de Jesús María y Ate Vitarte, 2019

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:
Ingeniero Ambiental

AUTOR:

Br. Merino Navarro, Erik Gregorio (ORCID: 0000-0003-1733-2629)

ASESOR:

Dr. Juan Julio Ordoñez Gálvez (ORCID: 0000-0002-3419-7361)

LÍNEA DE INVESTIGACIÓN:

Calidad y Gestión de los Recursos Naturales

LIMA - PERÚ
2019

DEDICATORIA

Dedico este trabajo investigativo a mi madre, que estima siempre en apoyarme. Ella, de buenos consejos y de regalarme los ánimos para ser mejor cada día.

A mis familiares y amigos más allegados que buscan en mí un ejemplo a seguir

AGRADECIMIENTO

Agradezco a Dios, por las oportunidades que me brinda; a los colaboradores del Senamhi, sobre todo al ingeniero Jhojan Rojas Quincho por su apoyo y alta estima personal y a la universidad Cesar Vallejo por impartir en mi formación profesional.

ÍNDICE

DEDICATORIA	i
AGRADECIMIENTO	ii
RESUMEN	
ABSTRACT	
I. INTRODUCCIÓN	1
1.1. Realidad problemática	1
1.2. Trabajos previos	2
1.3. Teorías relacionadas al tema.....	9
1.4. Formulación del problema.....	37
14.1. Problema General.....	37
14.2. Problemas Específicos	37
1.5. Justificación del estudio	38
1.6. Hipótesis	38
16.1. Hipótesis General.....	38
16.2. Hipótesis Específicas.....	38
1.7. Objetivos	39
17.1. Objetivo general	39
17.2. Objetivos Específicos	39
II. MÉTODO	39
2.1. Tipo y Diseño de investigación.....	40
2.2. Variables y Operacionalización	41
2.3. Población, muestra y muestreo.....	41
23.1. Población.....	41
23.2. Muestra	41
23.3. Muestreo.....	41
2.4. Técnicas e instrumentos de recolección de datos, Validez y Confiabilidad	41
24.1. Técnica e instrumentos de recolección de datos	41
24.2. Validez.....	41
24.3. Confiabilidad	41
2.5. Procedimiento	42
25.1. Procedimiento para la obtención de resultados del objetivo de estudio.....	42
25.2. Preprocesamiento de la base de datos	45

253. Desarrollo del modelo matemático de predicción para el Material Particulado PM ₁₀	45
2.5.4. Evaluación de los errores de predicción	45
2.6. Métodos de análisis de datos	47
2.7. Aspectos éticos	47
III. RESULTADOS	49
3.1. Área de estudio	49
3.2. Capa mezcla	50
3.3. Temperatura	50
3.4. Humedad relativa	52
3.5. Vientos	54
3.6. Concentración de Material Particulado (PM ₁₀)	57
3.7. Correlación de variables meteorológicas vs PM ₁₀	59
3.8. Base de datos para la entrada de la RNA	60
3.9. Agrupamiento de los datos para la RNA	61
3.10. Normalización de los datos	64
3.11. Arquitectura de la red neuronal artificial para el modelo de estimación del PM ₁₀	65
IV. DISCUSIÓN	71
V. CONCLUSIONES	72
VI. RECOMENDACIONES	73
REFERENCIAS	74
ANEXOS	78

ÍNDICE DE FIGURAS

Figura.1 Sistema sinóptico del tiempo	9
Figura.2 Capa de Mezcla	10
Figura.3 Comportamiento de los contaminantes emitidos a la atmósfera	11
Figura.4 Curvas típicas de la radiación y la temperatura del aire	14
Figura.5 Fenómeno de inversión térmica	15
Figura.6 Gradiente de presión y el viento	15
Figura.7 Rosa de vientos del área metropolitana del Lima- Callao	16
Figura.8 Perfil de velocidad de los vientos en relación a la rugosidad de la superficie	17
Figura.9 Condiciones de estabilidad atmosférica	19
Figura.10 Efectos aerodinámicos de los edificios y topografía irregular	19
Figura.11 Orografía del terreno	19
Figura.12 Ilustración del comportamiento de los contaminantes del aire por incidencia de la meteorología y topografía	19
Figura.13 Modelos matemáticos lineales utilizados en ciencias ambientales	20
Figura.14 Modelo de caja para el análisis del comportamiento atmosférico	21

Figura.15 Flujograma de análisis de datos de WRF.....	22
Figura.16 Diferentes técnicas de aprendizaje automático.....	23
Figura.17 Diferentes técnicas de aprendizaje automático.....	24
Figura.18 Forma principal de una neurona.....	25
Figura. 19 Forma de una neurona artificial y la similitud con una neurona biológica	26
Figura.20 Modelo neuronal de McCulloch-Pitts	27
Figura.21 Arquitectura unidireccional con tres capas de neuronas: Capa de entrada, Capa oculta, Capa de salida	28
Figura.22 Funciones de activación utilizadas.....	29
Figura.23 Orden de las redes neuronales artificiales	30
Figura.24 Tipos de aprendizaje	31
Figura.25 Arquitectura de un perceptrón multicapa.....	34
Figura.26 Arquitectura del algoritmo de retropropagación.....	36
Figura.27 Flujograma de procedimiento para los resultados del estudio	42
Figura.28 Mapa de ubicación de la Cuenca Atmosférica del AMLC.....	48
Figura.29 Mapa de ubicación de los distritos de Jesús María y Ate vitarte.....	49
Figura.30 Variación mensual de la Capa de mezcla-Estación meteorológica automática-Senamhi, 2018.....	50
Figura.31 Variación mensual de la temperatura del aire para los distritos de Ate, meteorológica automática-Senamhi, 2018	51
Figura. 32 Variación mensual de la temperatura del aire para los distritos de Jesús María, meteorológica automática-Senamhi, 2018	52
Figura.33 Variación del promedio mensual de la humedad relativa del distrito de Ate - Estación meteorológica automática-Senamhi, 2018	53
Figura.34 Variación del promedio mensual de la humedad relativa del distrito de Jesús María-Estación meteorológica automática-Senamhi, 2018	53
Figura.35 Rosa de vientos-Estación meteorológica automática ATE-Senamhi, 2018.....	54
Figura.36 Rosa de vientos-Estación meteorológica automática de Jesús María-Senmahi, 2018.	
Figura.37 Rosa de vientos-Estación meteorológica automática Ate y Jesús María-Senamhi, 2018	55
Figura.38 Comportamientos del Material Particulado PM ₁₀ en Ate Vitarte (Plaza municipal), 2018	57
Figura.39 Comportamientos del Material Particulado PM ₁₀ en Jesús María (Campo de marte), 2018.....	58
Figura.40 Correlación de variables meteorológicas vs PM ₁₀	59
Figura.41 Variables de salida del WRF-Chem, 2018	60
Figura.42 Agrupamiento de las variables de entrada para la RNA (k=3)	61
Figura.43 Agrupamiento de las variables de entrada para la RNA (k=5).....	61
Figura.44 Número óptimo de Clúster con el paquete NbClust para la RNA.....	62
Figura.45 Número óptimo de Clúster con el paquete factoextra para la RNA.....	62
Figura.46 Segmentos de datos analizados del agrupamiento para la RNA.....	63
Figura.47 Normalización de los datos	64
Figura.48 Arquitectura de la RNA para el distrito de Ate Vitarte (13-4-1).....	65
Figura.49 Similitud de los valores Reales vs los pronosticados para el distrito de Ate Vitarte.....	66
Figura.50Arquitectura de la RNA para el distrito de JesúsMaría (7-2-1)	67
Figura.51 Similitud de los valores Reales vs los pronosticados para el distrito de Jesús María..	68

Figura.52 Correlación de los datos Reales vs Pronosticados del distrito de Ate Vitarte	69
Figura.53 Correlación de los datos Reales vs Pronosticados del distrito de Jesús María	69

ÍNDICE DE TABLAS

Tabla N°1 Estándares de Calidad Ambiental para Aire.....	13
Tabla N°2 Turbulencia y condiciones del tiempo	17
Tabla N°3 Valores de los coeficientes de correlación (R)	47
Tabla N°4 Velocidades de vientos.....	56
Tabla N°5 Ubicación geográfica de las estaciones de Calidad del Aire-Senamhi.....	60
Tabla N°6 Valores mínimos y máximos permitidos	64
Tabla N°7 Arquitectura de la R NA.....	70
Tabla N°8 Arquitectura óptima para la R N A	70

RESUMEN

La contaminación atmosférica propicia el deterioro de la calidad del aire. Los efectos adversos que genera la concentración de los agentes contaminantes en los ecosistemas son severos. Hoy se registra información de personas con problemas respiratorios, cardiovasculares y cáncer de pulmón que han llevado a la muerte a miles de ellas, sobre todo por el material particulado que se considera en mayor proporción y distribución en las ciudades más grandes y pobladas. En cierta medida, las autoridades competentes responden a este problema creando una red de monitoreo de contaminantes primarios en las capitales más afectadas. Sin embargo, la medida no es completa debido a que la gestión no contempla un sistema de alerta temprana que pueda prever que las personas estén expuestas a la concentración del contaminante. Es por ello que, en el presente estudio, se realizó un modelo estocástico, que contempla los factores climáticos secundarios, para el pronóstico de la concentración del material particulado (PM_{10}) en los distritos de Ate Vitarte y Jesús María, para lo cual se utilizó la técnica de retropropagación de una red neuronal artificial multicapa, con lo cual se obtuvo como mejor algoritmo 4 neuronas en la capa oculta para el distrito de Ate vitarte y 2 neuronas, para el distrito de Jesús María, con error cuadrático medio de 1.59 y 0.01 respectivamente. Se verificó el grado de bondad de ajuste con una correlación de Pearson con lo que se obtuvo 0.7 y 0.9 para los distritos de Ate Vitarte y Jesús María. De modo que el modelo puede responder a estimar la concentración del material particulado (PM_{10}) en ambos distritos. Cabe resaltar que se debe realizar un ajuste mejor para los datos del distrito de Ate Vitarte.

Palabras clave: Contaminación atmosférica, agentes contaminantes, contaminantes primarios, Material particulado (PM_{10}), RNA, retropropagación.

ABSTRACT

Air pollution leads to deterioration of air quality. The adverse effects generated by the concentration of pollutants in ecosystems are severe. Today there is information on people with respiratory, cardiovascular and lung cancer problems that have killed thousands of them, especially because of the particulate material that is considered in greater proportion and distribution in the largest and most populated cities. To some extent, the competent authorities respond to this problem by creating a monitoring network of primary pollutants in the most affected capitals. However, the measure is not complete because the management does not contemplate an early warning system that can provide that people are exposed to the concentration of the contaminant. That is why, in the present study, a stochastic model was carried out, which contemplates secondary climatic factors, for the forecast of the concentration of particulate material (PM₁₀) in the districts of Ate Vitarte and Jesús María, for which it was used the retropropagation technique of a multilayer artificial neural network, which resulted in the best algorithm 4 neurons in the hidden layer for the district of Ate vitarte and 2 neurons, for the district of Jesús María, with a mean square error of 1.59 and 0.01 respectively. The degree of goodness of fit was verified with a Pearson correlation, obtaining 0.7 and 0.9 for the districts of Ate Vitarte and Jesús María. So the model can respond to estimate the concentration of particulate material (PM₁₀) in both districts. It should be noted that a better adjustment should be made for the data of the Ate Vitarte district.

Keywords: Air pollution, pollutants, primary pollutants, Particulate matter (PM₁₀), RNA, back propagation.

I. INTRODUCCIÓN

1.1. Realidad problemática

La Organización Mundial de la Salud (OMS) estimó que una de cada nueve muertes en todo el mundo es consecuencia del deterioro de la calidad del aire (WHO, GBoD 2016). A nivel mundial se estima que en promedio 1,3 millones de personas mueren cada año a causa de la contaminación atmosférica urbana; más de la mitad de las muertes ocurren en los países en desarrollo. En las ciudades que mantienen niveles de contaminantes atmosféricos elevados, diversos son los efectos a corto y largo plazo que esta puede ejercer sobre la salud de las personas: el riesgo de padecer enfermedades respiratorias y cardiovasculares, así como neumonía y cáncer del pulmón (OMS, 2016) generándose así pérdidas de bienestar económico. (BM & IHME, 2016).

En América Latina y el Caribe (ALC), se tiene una constante preocupación por el deterioro de la calidad del aire debido a que en las ciudades de Santiago, Bogotá, Sao Paulo, México, entre otras, han generado impactos gravísimos en la salud de las personas, sobre todo en los en vías con alto tráfico vehicular (Franco et al., 2009) y el Perú no es ajeno a estos efectos negativos mencionados, pues ya es un asunto de índole político; es así que dentro de los principales problemas ambientales *políticamente relevantes en el Perú*, se tiene como prioridad al “caos ambiental urbano” por la emisión de contaminantes primarios si se desea precisar. (Ráez & Dourojeanni, 2016).

Las tendencias temporales y espaciales en la variabilidad de PM_{10} y $PM_{2.5}$ desde 2010 hasta 2015 en el área metropolitana de Lima-Callao (AMLC) muestran que las concentraciones medias anuales de PM_{10} y $PM_{2.5}$ tuvieron, rangos (promedios) de 133 a $45 \mu\text{g}/\text{m}^3$ ($84 \mu\text{g}/\text{m}^3$) y $35\text{--}16 \mu\text{g}/\text{m}^3$ ($26 \mu\text{g}/\text{m}^3$), mostrándose así las concentraciones anuales más altas en la parte oriental de la ciudad, como resultado del patrón de vientos locales persistentes que ingresan desde la costa en dirección suroeste, la cual responde también a un patrón diurno que corresponde a la Tráfico máximo de un total de 9 millones de viajes al día. (Silva et al., 2017). Esto se puede verificar bajo la Red Automática de Monitoreo de la Calidad del Aire (RAMCA) del AMLC, pues la estación que registró las mayores concentraciones de PM_{10} principalmente durante los meses de verano y otoño (enero-mayo) es la estación “ATE” ubicada en el distrito de Ate, al este del AMLC (Senamhi, 2013; Senamhi, 2014; Senamhi, 2015).

Para este último año, 2018, el Servicio Nacional de Meteorología e Hidrología del Perú - Senamhi- registró el comportamiento de Material Particulado (PM₁₀) que superó el ECA en los distritos de: Villa María del Triunfo (27 días), Ate (19 días), San Juan de Lurigancho (7 días) y Carabaylo (7 días) para el mes de marzo. Para el mes de abril, los distritos de Ate (29 días), Villa María del Triunfo (16 días), Huachipa (16 días), San Juan de Lurigancho (13 días) y Carabaylo (12 días), Así mismo, para el mes de julio el Material Particulado (PM₁₀) tuvo menos días de polución en los distritos de: Puente Santa Anita (1 día), Villa María del Triunfo (3 días), Carabaylo (3 días), Piedra (4 días), San Juan de Lurigancho (5 días), y Huachipa (8 días), exceptuando el distrito de Ate (23 días). (Senamhi, 2018). Para el presente año, 2019, no se ha podido recoger datos del contaminante en mención, debido a que las estaciones de monitoreo, de la Red de Monitoreo de Calidad del Aire, se han encontrado en mantenimiento. (Senamhi, 2019). Es así que el presente trabajo pretende abordar una alternativa nueva para la solución de estos problemas técnicos. En la actualidad el AMLC es considerado como una de las ciudades más contaminadas por material particulado en Latinoamérica, por lo que se debería trabajar en acción rápida. (OMS, 2015; Silva et al., 2017).

De modo que es importante trabajar en la mitigación y plan de manejo del impacto que se generan por estas emisiones; como consecuencia se puede limitar la morbilidad y la mortalidad de las personas más expuestas derivadas a estos contaminantes, teniendo en cuenta que el nivel y el tipo de contaminantes varían en diferentes entornos urbanos y rurales por lo que se debe trabajar de manera específica y conjunta. (Sharma; et al.; 2018).

1.2. Trabajos previos

Memarianfard (2017) comenta que la mayor parte de las áreas urbanas enfrentan el problema de la flotación o suspensión de partículas finas. Por ello es crucial estimar las cantidades de partículas finas en la atmósfera urbana. En la investigación, Hatami utilizó la técnica de red neuronal artificial (RNA) para modelar la dispersión PM_{2.5} en la ciudad de Teherán. Los factores que influyen en el valor predicho consisten en datos relacionados con el clima y la contaminación del aire, es decir, la velocidad del viento, la humedad, la temperatura, SO₂, CO, NO₂ y PM_{2.5} como valores objetivo.

Estos factores se han considerado en 19 estaciones de medición (zonas) en áreas urbanas de la ciudad de Teherán durante cuatro años, desde marzo de 2011 hasta marzo de 2015. Los resultados indican que la red con una capa oculta, de seis neuronas en la etapa de entrenamiento a 113, tiene el mejor rendimiento con el valor de error más bajo ($MSE = 0.049438$) al considerar las concentraciones de $PM_{2.5}$ en las áreas metropolitanas de Teherán. Además, el valor "R" para el análisis de regresión del entrenamiento, la validación, la prueba y todos los datos son 0.65898, 0.6419, 0.54027 y 0.62331, respectivamente. Este estudio también representa las redes neuronales artificiales que se han implementado satisfactoriamente para resolver patrones complejos en el campo de la contaminación del aire.

Para Akbar (2017), los modelos estadísticos basados en redes neuronales artificiales (ANN) se han aplicado y evaluado cada vez más para el pronóstico de la calidad del aire. Es así que el desarrollo de ANN y las regresiones lineales múltiples (MLR) se han aplicado a la predicción a corto plazo de las concentraciones de NO_2 y NO_x en función de las condiciones meteorológicas. La estructura óptima de ANN se determinó mediante un método de prueba y error. Utilizando concentraciones de NO_x y NO_2 por hora y parámetros meteorológicos, red de monitoreo automático durante octubre y noviembre de 2012 para dos sitios de monitoreo (Abrasan y sitios de Farmandari) en Tabriz, Irán.

El diseño de la arquitectura de la red se basa en la teoría de la aproximación de Kolmogorov, y la estructura de ANN con 30 neuronas tuvo el mejor rendimiento. ANN entrenado por el algoritmo de entrenamiento escalado-gradiente conjugado (trainsecg) se ha implementado para modelar. También demuestra que las redes neuronales de MLP ofrecen varias ventajas sobre los modelos lineales de MLR. Los resultados muestran que los valores del coeficiente de correlación (R^2) son 0.92 y 0.94 para las concentraciones de NO_2 y NO_x , respectivamente. Pero en el modelo MLR, los valores de R^2 fueron 0.41 y 0.44 para las concentraciones de NO_2 y NO_x , respectivamente. De modo que el trabajo muestra que las redes neuronales de MLP pueden modelar con precisión la relación entre los datos meteorológicos locales y las concentraciones de NO_2 y NO_x en un entorno urbano en comparación con los modelos lineales.

Delavar, et al. (2019) realizaron un estudio para determinar los modelos de predicción en pro

de ver el comportamiento de contaminantes como PM_{10} y $PM_{2.5}$. Para predecir la contaminación del aire, fueron utilizados datos relacionados con el día de la semana, el mes del año, la topografía, la meteorología y la tasa de contaminantes de dos vecinos más cercanos como parámetros de entrada y métodos de aprendizaje automático. Estos métodos incluyen una máquina de vectores de apoyo a la regresión, una regresión ponderada geográficamente, una red neuronal artificial y una red neuronal no lineal auto-regresiva con una entrada externa como método de aprendizaje automático para la predicción de la contaminación del aire.

Luego se propuso un modelo de predicción para mejorar los métodos antes mencionados, por lo que el porcentaje de error se ha reducido y mejorado en un 57%, 47%, 47% y 94%, respectivamente. El algoritmo más confiable para la predicción de la contaminación del aire fue la red neuronal no lineal autorregresiva con entrada externa que utiliza el modelo de predicción propuesto, donde se alcanzó el error de predicción de un día. Finalmente, utilizando el algoritmo neuronal, los datos para el día de la semana, el mes del año, la topografía, la dirección del viento, la temperatura máxima y la tasa de contaminantes de los dos vecinos más cercanos se identificaron como los parámetros más efectivos en la predicción de la contaminación del aire.

Para Fadish, et al. (2015) El valor de AQI se puede predecir utilizando el modelo de red neuronal para un día antes, según las variables meteorológicas y el comportamiento de autocorrelación del índice en Kermanshah, una ciudad en el oeste de Irán. Los datos para la contaminación del aire y las variables meteorológicas, recopilados durante tres años, se retrasaron durante dos días. El AQI para el día siguiente se consideró como variable dependiente y otros se usaron como predictores. El desempeño del modelo se evaluó con el coeficiente de correlación (r). Las variables más importantes para predecir el AQI se identificaron mediante análisis de sensibilidad. El coeficiente r para la capacitación, validación y prueba del modelo fue de 0.75. Entre las variables meteorológicas, la vista horizontal y la precipitación tuvieron un mayor impacto en el AQI. La precipitación de un día puede reducir significativamente la cantidad de AQI para el día siguiente. Se encontró una relación inversa entre AQI y la vista horizontal.

Por consiguiente, el modelo propuesto puede usarse para predecir el índice AQI de la ciudad. Con respecto al tema de la contaminación del aire en esta ciudad, especialmente la contaminación por partículas finas; dicho modelo se usa dinámicamente para predecir el AQI (Índice de Calidad del Aire, en español), puede ser una herramienta útil para la declaración de una alerta de contaminación del aire. La preparación de un sistema en línea basado en modelos para la predicción del índice AQI para el distrito se sugiere para estudios futuros.

Capilla (2015) presentó la aplicación de las redes de perceptrón multicapa (MLP) de avance para predecir los niveles de óxidos de nitrógeno por hora con 24 h de anticipación. Los datos de entrada fueron variables meteorológicas, tráfico por hora promedio y niveles por hora de óxidos de nitrógeno. La introducción de cuatro componentes periódicos (términos de seno y coseno para los ciclos diarios y semanales) se analizó para mejorar los poderes de predicción de los modelos. Los datos se midieron durante 3 años en estaciones de monitoreo en Valencia (España) en dos lugares con alta densidad de tráfico. Los criterios de evaluación de los modelos fueron error medio absoluto, error cuadrático medio (RMSE), error porcentual absoluto medio y coeficiente de correlación entre observaciones y predicciones.

Las comparaciones de los modelos basados en MP demostraron que la inserción de cuatro variables de entrada estacionales adicionales mejoró la capacidad de obtener predicciones más precisas, lo que enfatiza la importancia de tener en cuenta el carácter estacional de los óxidos de nitrógeno. Cuando se usan componentes estacionales como factores predictivos, el error cuadrático medio mejora de 20.29 a 19.35 al predecir dióxido de nitrógeno y de 45.07 a 42.37 al pronosticar óxidos nítricos si el modelo incluye componentes estacionales en una ubicación de estudio. En la otra ubicación, RMSE cambia de 23.76 a 23.05 cuando predice dióxido de nitrógeno y de 33.94 a 33.10 para los pronósticos de otros contaminantes.

Las redes neuronales no requieren información muy exhaustiva sobre los contaminantes del aire, los mecanismos de reacción, los parámetros meteorológicos o las características del tráfico, y tenían la capacidad de permitir relaciones complejas y no lineales entre variables predictores muy diferentes en un entorno urbano.

Lauret, et al. (2017) mencionan que se requiere pericia para pronosticar la dispersión de contaminantes atmosféricos en cualquier industria que procese material peligroso. Aseguran que el comportamiento de la dispersión de polvo es difícil por el diámetro, dispersión, aglomeración, sedimentación, rango de densidades. Además, los sitios de producción pueden ubicarse dentro de un entorno complejo como las áreas urbanas, donde la precisión de los modelos de dispersión clásicos es baja. Por ello el trabajo que realizaron tuvo como objetivo evaluar la eficiencia de un modelo de redes neuronales artificiales (ANN) para predecir la dispersión de polvo en un área urbana sin conocimiento previo de la fuente. La base de datos experimental consta de 290 mediciones diarias de concentración media en un sitio ubicado a 500 m de la fuente de emisión. Las entradas se seleccionan a partir de datos meteorológicos. La fase de entrenamiento se realizó a través de la aplicación de *parada temprana*. La selección de modelo ANN se realiza sobre el mejor coeficiente de valor de determinación luego el rendimiento del modelo se evaluó utilizando criterios clásicos de calidad del aire mostrando así buenos resultados. Los resultados se incluyen dentro del rango aceptable. Se pueden lograr mejoras agregando información del término fuente como una entrada para el modelo ANN.

Así mismo Medina (2016) presentó un trabajo investigativo que tuvo como objetivo diseñar una Red Neuronal Artificial para predecir la concentración de PM_{10} en el distrito de ATE utilizando el tipo de aprendizaje supervisado bajo redes neuronales multicapa con algoritmo de aprendizaje backpropagation. En el estudio se tomó en cuenta los factores meteorológicos y topográficos de la zona; las fuentes de emisión de la zona como: fuente de área, fuentes móviles, etc. Se realizó el análisis del comportamiento temporal del Material Particulado del periodo 2010-2015. En los resultados se observa una tendencia de aumentar las concentraciones en los meses de primavera y verano, para que posteriormente disminuya y alcance sus valores mínimos en la estación invernal. Los picos más representativos se observan en los últimos meses de los años. El desempeño histórico es bueno pues se calculó un equivalente a 83% de captura de datos de todo el periodo en estudio. Así mismo presenta la validación del diseño de la red mediante el coeficiente de correlación R de [0.994-0.999], con una correspondencia de los parámetros de dispersión bajo el factor centesimal de correlación de Pearson, el cual fue un factor de formación del PM_{10} .

Por otro lado, Rojas & Silva (2016) Realizaron un trabajo de investigación en el que utilizaron las redes neuronales artificiales (RNA) para evaluar series temporales de la temperatura utilizando información de 44 años. En el estudio se tomaron como datos de entrada temperatura diaria máxima y mínima, horas d sol diaria, radiación extraterrestre, entre otras; para obtener como resultado temperatura diaria. Se tomó en cuenta los periodos que presentaron Fenómeno del Niño y la Niña. El tipo de entrenamiento con la red neuronal fue de tipo multicapa con algoritmo de entrenamiento “backpropagation”. Como resultados se obtuvo un mejor modelamiento en el patrón 4, que tuvo como coeficiente de correlación de Pearson (r) de 0.9; error medio (EM) de 0; error cuadrático (ECM) de 0.2 y una raíz del error cuadrático medio (RECM) de 0.5; mostrando que estos valores corroboran los datos reales en el modelamiento.

Para Paas, et al. (2017) la evaluación de la calidad del aire es una tarea importante para las autoridades locales debido a varios efectos adversos para la salud asociados con la exposición, por ejemplo, a las concentraciones de partículas urbanas en todo el mundo. En función del consumo de costos y el tiempo relacionado con los trabajos experimentales requeridos para las mediciones de la concentración de partículas en la atmósfera, surgen como opciones integradoras, la aplicación de modelos matemáticos. De modo que el estudio presenta un enfoque de la Red Neural Artificial (ANN) para predecir las concentraciones atmosféricas de masa de partículas considerando partículas con un diámetro aerodinámico de 0.25–1 μm (PM (0.25–1)), 0.25–2.5- μm (PM (0.25–2.5)), 0,25–10 μm (PM (0,25–10)), así como concentraciones en el número de partículas de partículas con un diámetro aerodinámico de 0,25–2,5 μm (PNC (0,25–2,5)).

Las variables de entrada del modelo ANN se definieron utilizando datos de mediciones de sonido locales, concentraciones de transporte de partículas de fondo y datos meteorológicos estándar. Se propone una metodología que incluye la selección de variables de entrada, la división de datos y una evaluación de su desempeño. Los modelos ANN se aplicaron a un sitio que presenta un parque en el interior de la ciudad para probar la capacidad del enfoque para recopilar información espacial de las concentraciones de aerosol. Se observó que ANN modela predicciones de PM (0.25–10) y PNC (0.25–2.5) en el caso del cañón de la calle, así como predicciones de PM (0.25–2.5), PM (0.25–10) y PNC (0.25–2.5) dentro del estudio de caso del área del parque muestra un buen acuerdo con las observaciones y cumple con los

estándares de calidad propuestos por la Comisión Europea con respecto a la predicción del valor medio. Los resultados indican que los modelos ANN propuestos pueden ser una herramienta bastante precisa para evaluar la predicción de concentraciones de partículas no solo en el tiempo sino también en el espacio.

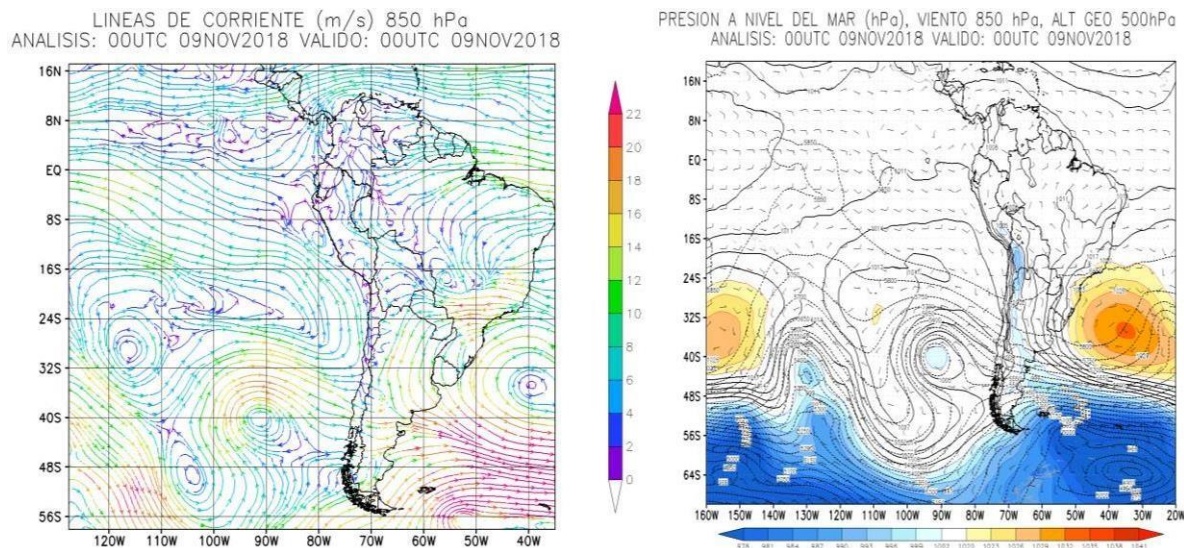
Feng, et al. (2015) presentan un nuevo modelo híbrido que combina el análisis de la trayectoria de la masa de aire y la transformación de agua para mejorar la precisión de la red neuronal artificial (ANN) de las concentraciones promedio diarias de $PM_{2.5}$ con dos días de antelación. El modelo se desarrolló a partir de 13 estaciones diferentes de monitoreo de la contaminación del aire en Beijing, Tianjin y la provincia de Hebei (área de Jing-Jin-Ji). Se utilizó la trayectoria de la masa de aire para reconocer distintos flujos que propician el transporte de aire “sucio” y aire “limpio” a las estaciones seleccionadas. Con cada corredor, se construyó una red de estación triangular basada en las trayectorias de masas de aire y las distancias entre los sitios vecinos.

La velocidad y la dirección del viento también se consideraron como parámetros al calcular el valor de un indicador de contaminación del aire basado en la trayectoria. Por otra parte, la serie de tiempo original de la concentración de $PM_{2.5}$ se descompuso por transformación wavelet en unas pocas subseries con menor variabilidad. La estrategia de predicción aplicada a cada uno de ellos y luego resumió los resultados de la predicción individual. Las variables de pronóstico meteorológico diario, así como los respectivos predictores de contaminantes, se utilizaron como entrada a un tipo de red de neuronas de propagación múltiple de perceptrón (MLP). La verificación experimental del modelo propuesto se llevó a cabo durante un período de más de un año (entre septiembre de 2013 y octubre de 2014). Se ha encontrado que el modelo geográfico basado en la trayectoria y la transformación de la información pueden ser herramientas efectivas para mejorar la precisión de pronóstico de $PM_{2.5}$. El error cuadrático medio (RMSE) del modelo híbrido se puede reducir, en promedio, hasta en un 40 por ciento. De modo que las altas concentraciones de $PM_{2.5}$ se pueden anticipar usando la descomposición del agua y la tasa de detección (DR) para un umbral de alerta dado que el modelo híbrido puede alcanzar un promedio del 90% de eficiencia. Este enfoque muestra el potencial para ser aplicado en los sistemas de pronóstico de calidad del aire de otros países.

1.3. Teorías relacionadas al tema

Dinámica atmosférica

Para entender la dinámica atmosférica es preferente conocer las teorías relacionadas a ella, verbigracia la *teoría cuasi-geostrófica*, que permite entender el comportamiento de la atmosfera a través de las leyes de la física, las cuales integran fenómenos de *conservación de momento*: masa y energía con la presión, temperatura y régimen de velocidad de vientos. Con ecuaciones, se intenta predecir los movimientos de una cuenca atmosférica. Estas predicciones estadísticas son sustentadas para una atmósfera adiabática, una relación directa entre el régimen de viento y la altura, así como la fuerza de fricción de flujos atmosféricos de escala sinóptica. Así mismos conceptos como termodinámica, y vorticidad se encuentran en la dinámica atmosférica. (Febre, 2017, p. 33- 34).

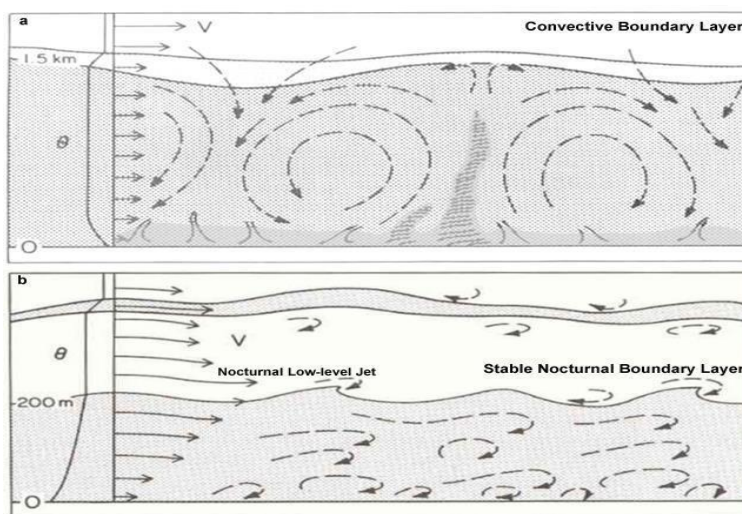


Fuente: Senamhi, 2018.

Figura.1 Sistema sinóptico del tiempo.

Los sistemas sinópticos de tiempo son grandes áreas con masas de aire regularmente homogéneas de escala de tiempo de semanales, contenidas dentro de los sistemas macroescalares. Como se presenta en la figura. 1, los vientos que ingresan al litoral peruano son influenciados por el Anticiclón del Pacífico subtropical (APS). Este se extiende sobre toda la cuenca oceánica del Pacífico Sur y es el que predomina en el sistema de las corrientes de Humboldt. Esencialmente en las costas de Perú durante el invierno los vientos favorables a (hacia el norte) incrementan su velocidad respecto al verano. Es así que tanto este fenómeno, como la topografía, acentúan la micro meteorología del lugar. Por otro lado, se verifica también *La capa límite atmosférica (CLA)*.

Esta es una porción de la troposfera que es influenciada por la superficie terrestre. Su estructura cambia a lo largo del día siendo mucho mayor este cambio sobre tierra que sobre los océanos. Dentro de ella, suceden procesos de intercambio de energía involucrando un sistema de relación sol-atmósfera-tierra, la distribución de las variables meteorológicas como el viento, temperatura y humedad, y las concentraciones de contaminantes primarios. Existen diferentes tipos de fuentes de polución al aire, entre ellas se clasifican fuentes primarias y las fuentes secundarias que son los temas relevantes para el presente estudio.



Recuperado de: http://ftp.comet.ucar.edu/ootw/tropical/textbook_2nd_edition/print_6.htm

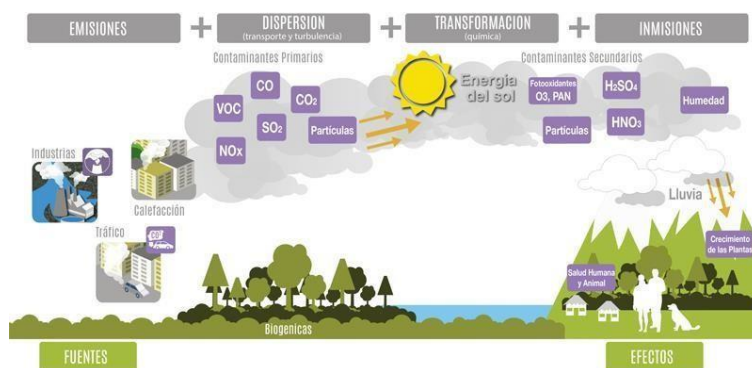
Figura.2 Capa de Mezcla

Existen factores que intervienen en el perfil de la temperatura del aire en la CLA: tipos de masas de aire que están por encima de la capa límite planetaria (CLP), que dependen de las condiciones sinópticas y del patrón de circulación a gran escala; características térmicas de la superficie (intervalo de temperaturas de superficie diurna); intercambios de calor, humedad específica de masa de aire, la advección, etc. Así mismo de los factores que influyen en los patrones de vientos locales son: Las gradientes de presión y temperatura horizontal de grandes escalas, el ciclo diurno del calentamiento o enfriamiento de la superficie, convergencia o divergencia horizontal a gran escala, presión de nubes y precipitación, características topográficas de la superficie, características de rugosidad, entre otros. (Vara, 2009, p. 22- 25).

Contaminación Atmosférica

Se entiende que la Contaminación atmosférica está definida por circunstancias determinadas en que existe presencia de sustancias no deseables, en concentraciones altas que afectan

significativamente el confort, bienestar y salud de las personas y del medio que les rodea. [...] Las sustancias no deseables son consideradas como agentes contaminantes y por su origen se clasifican en fuentes de contaminación primaria y secundaria. Con el fin de poder controlar estos agentes, se clasificaron según su importancia cuantitativa, a estas se les determinaron como contaminantes criterio. (Contreras, 2014, p. 67, 69).



Recuperado de: <http://www.siac.gov.co/aire>

Figura.3 Comportamiento de los contaminantes emitidos a la atmósfera

Contaminantes primarios

Son contaminantes atmosféricos producidos por acción directa de una reacción de fuente antropogénica. Es relevante por el efecto cuantitativo en el aire, referente al nivel de inmisión. (Contreras & Molero, 2014, p.71). Como lo son: Material Particulado (PM), (Dióxido de Azufre (SO₂), Monóxido de Carbono (CO), Dióxido de Nitrógeno (NO₂) y Ozono (O₃) principalmente.

Material Particulado (PM)

El Material Particulado es generado por diversas actividades antropogénicas y tiene efecto nocivo en los seres vivos, ya que afectan directamente a las vías respiratorias generando desde carrasperas hasta cáncer pulmonar. Las partículas pueden presentar grados de toxicidad debido a su composición física o química, o al actuar como conductor de sustancias en las vías de ingreso respiratorio de los seres vivos.

Por otro lado, es complicado obtener una relación entre la dosis a diferentes concentraciones de partículas y los efectos en la salud de las personas. Por ello el resultado que se tiene del análisis de los efectos adversos son debido al número de visitas médicas por personas con problemas respiratorios. (Wark & Warner, 2014, p. 39). Los contaminantes que pertenecen a esta categoría y que son registrados como contaminantes criterio son (PM₁₀) y (PM_{2.5}).

Dióxido de azufre (SO₂)

Las fuentes principales de SO₂ que son emitidas al aire provienen de la combustión de quema de carbón y derivados del petróleo, así como también de elementos no ferrosos. (Contreras & Molero, 2014, p.71). Este contaminante tiene efectos negativos en los seres vivos si son asimilados por acción directa o indirecta, pues pueden generar deterioro en la calidad de vida de las plantas y animales. Así también en la salud de las personas puede generar broncoconstricción: efecto en los bronquios por la cual hay una retención ligera en los conductos del aire. (Wark & Warner, 2014, p. 46).

Monóxido de Carbono (CO)

Generada de una combustión incompleta, el Monóxido de Carbono (CO), es un contaminante que deteriora la salud de las personas, es considerado un asfixiante; pues interfiere en el fluido sanguíneo, en este, no se completa el transporte del oxígeno a los tejidos del cuerpo y al inhalarse se inhibe la hemoglobina para formar la carboxihemoglobina, perjudicial para el corazón porque intentará bombear para compensar la cantidad de oxígeno faltante, como consecuencia la persona puede fallecer de paro cardíaco. (Masters & Ela; 2008, p. 395).

Dióxido de Nitrógeno (NO₂)

De los Óxidos de Nitrógenos (NO_x), el más importante es el Dióxido de Nitrógeno (NO₂) puesto que este compuesto puede irritar los pulmones y causar bronquitis y pulmonías, así mismo tiene otras consecuencias ambientales como contribuir a causar lluvia ácida y deteriorar a la vida de las plantas. (Masters & Ela; 2008, p. 397)

Ozono (O₃)

El Ozono es el oxidante fotoquímico más abundante en la atmósfera, pues es responsable de generar irritación en la membrana mucosa y en los ojos de las personas. Como efecto secundario propicia los compuestos orgánicos volátiles (COVs) en el ambiente. (Masters & Ela; 2008, p. 399).

Normativa ambiental para la calidad del aire

La normativa peruana en el Decreto Supremo N° 003 aprobado en el 2017 del Ministerio de Ambiente (D.S. N° 003-2017-MINAM) establece los parámetros de contaminantes

primarios con el fin de brindar un alcance de las regulaciones en cuanto a la gestión en pos de la buena salud de las personas y de los ecosistemas.

Tabla N°1 Estándares de Calidad Ambiental para Aire

Parámetros	Período	Valor [µg/m ³]	Criterios de evaluación	Método de análisis ^[1]
Benceno (C ₆ H ₆)	Anual	2	Media aritmética anual	Cromatografía de gases
Dióxido de Azufre (SO ₂)	24 horas	250	NE más de 7 veces al año	Fluorescencia ultravioleta (Método automático)
Dióxido de Nitrógeno (NO ₂)	1 hora	200	NE más de 24 veces al año	Quimioluminiscencia (Método automático)
	Anual	100	Media aritmética anual	
Material Particulado con diámetro menor a 2,5 micras (PM _{2.5})	24 horas	50	NE más de 7 veces al año	Separación inercial/filtración (Gravimetría)
	Anual	25	Media aritmética anual	
Material Particulado con diámetro menor a 10 micras (PM ₁₀)	24 horas	100	NE más de 7 veces al año	Separación inercial/filtración (Gravimetría)
	Anual	50	Media aritmética anual	
Mercurio Gaseoso Total (Hg) ^[2]	24 horas	2	No exceder	Espectrometría de absorción atómica de vapor frío (CVAAS) o Espectrometría de fluorescencia atómica de vapor frío (CVAFS) o Espectrometría de absorción atómica Zeeman. (Métodos automáticos)
Monóxido de Carbono (CO)	1 hora	30000	NE más de 1 vez al año	Infrarrojo no dispersivo (NDIR) (Método automático)
	8 horas	10000	Media aritmética móvil	
Ozono (O ₃)	8 horas	100	Máxima media diaria NE más de 24 veces al año	Fotometría de absorción ultravioleta (Método automático)
Plomo (Pb) en PM ₁₀	Mensual	1,5	NE más de 4 veces al año	Método para PM ₁₀ (Espectrofotometría de absorción atómica)
	Anual	0,5	Media aritmética de los valores mensuales	
Sulfuro de Hidrógeno (H ₂ S)	24 horas	150	Media aritmética	Fluorescencia ultravioleta (Método automático)

NE: No Exceder.

^[1] o método equivalente aprobado.

^[2] El estándar de calidad ambiental para Mercurio Gaseoso Total entrará en vigencia al día siguiente de la publicación del Protocolo Nacional de Monitoreo de la Calidad Ambiental del Aire, de conformidad con lo establecido en la Sétima Disposición Complementaria Final del presente Decreto Supremo.

Factores que inciden en el comportamiento de los contaminantes del aire

Los factores meteorológicos y topográficos son aquellos que inciden en la concentración de contaminantes atmosféricos; estos son: temperatura, dirección y velocidad de los vientos, así como las características de la superficie del suelo o aspereza de la superficie, es decir, de los elementos que se registran en una zona determinada. (Parker, 2014, p. 131). En algunas zonas los factores meteorológicos y topográficos inducen a la acumulación y concentración de los contaminantes atmosféricos, entre otros, que aceleraron el deterioro de la salud pública, así como la vegetación del lugar. (Wark.Warner, 2014). Para Arrieta (2016) En cuanto a la dispersión de PM₁₀ depende mucho de la topografía simple y compleja del territorio, pues la dirección y velocidad del viento está condicionada por la misma. Ésta

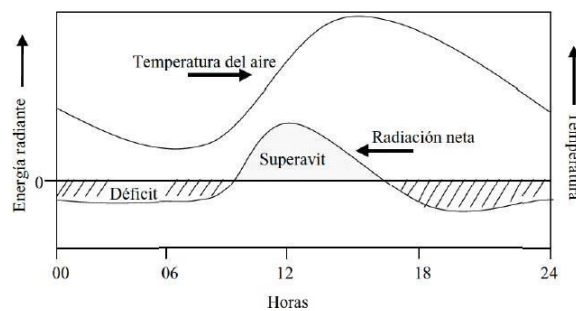
permite un flujo libre en la dirección predominante de los vientos para la zona de topografía simple y un flujo turbulento en las zonas complejas.

La Meteorología

Ciencia que trata sobre el estudio de la atmósfera, de sus propiedades o características y de los fenómenos que se manifiestan en ella. Es estudiada en un tiempo y espacio determinado. (Rodríguez, et al., 2004, p.9).

Temperatura

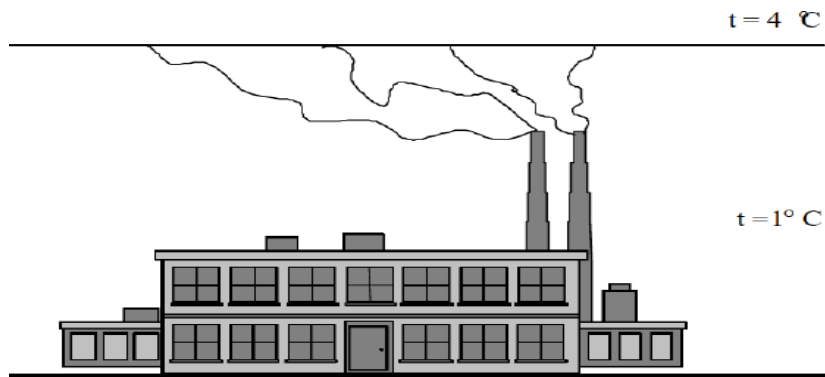
La temperatura es considerada un indicador que permite verificar el nivel de calor de un cuerpo; calor que, según el segundo principio de la termodinámica, se puede transmitir de un cuerpo con mayor temperatura a uno con menor temperatura; estas masas de aire (figura. 3) pueden ser moverse por: conducción, convección y radiación. (Andrades & Muñoz, 2012, p. 19)



Fuente: Fundamentos de Climatología, 2012

Figura.4 Curvas típicas de la radiación y la temperatura del aire

La temperatura varía en la troposfera según la altura, en promedio esta variación se estima en $-6.5\text{ }^{\circ}\text{C/km}$, de modo que no es un valor constante puesto que depende del contenido de la humedad del aire, de las estaciones, etc. Sin embargo, en ocasiones se presenta el fenómeno llamado inversión térmica, pues la temperatura aumenta en el perfil atmosférico. La inversión térmica (figura. 4) también puede influir en la contaminación atmosférica, ya que propicia la dispersión del contaminante en la periferia. (Andrades & Muñoz, 2012, p. 21, 22).

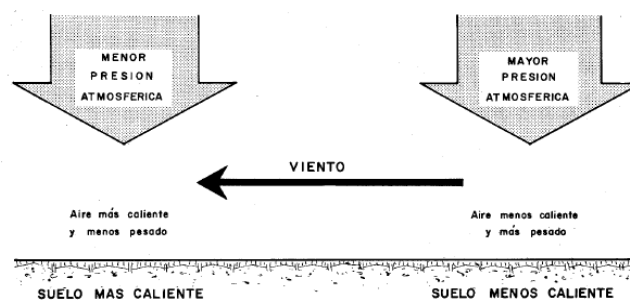


Fuente: Fundamentos de Climatología, 2012.

Figura.5 Fenómeno de inversión térmica

Vientos

Los vientos son originados por gradientes de presión. En la atmosfera la presión y la temperatura oscilan debido a la radiación solar que es emitida al planea tierra y esto propicia que el aire se caliente en algunas zonas, luego el aire más caliente se vuelve más denso y las masas de aire se desplazan en forma horizontal en la superficie de la tierra. (Andrades & Muñoz, 2012, p. 31).

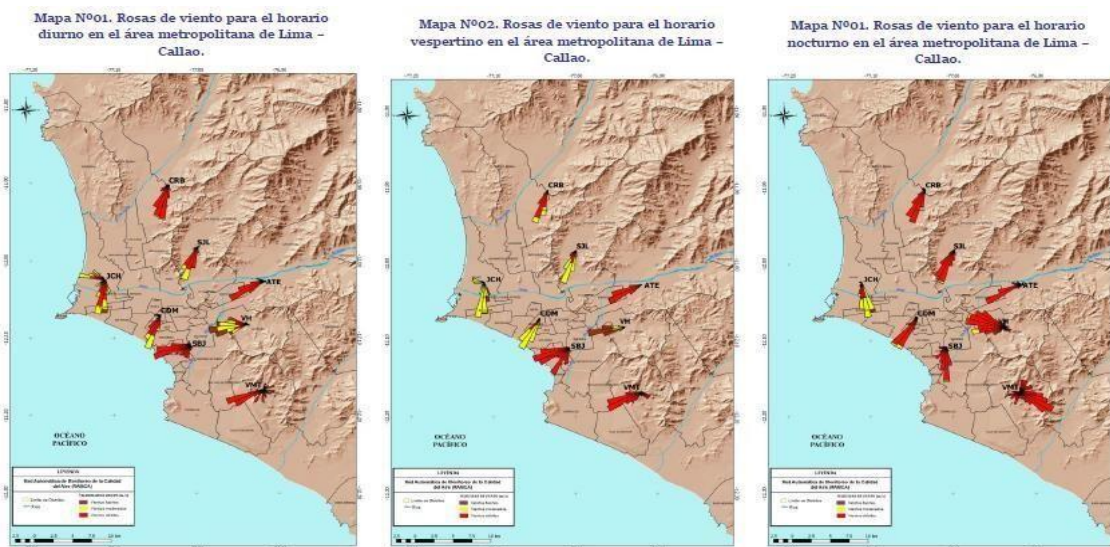


Fuente: Fundamentos de Climatología, 2012.

Figura.6 Gradiente de presión y el viento

Rosa de Vientos

Los vientos se pueden calcular bajo dos características: dirección y velocidad. La dirección de los vientos se estima con la rosa de los vientos; mientras que la velocidad se estima con el cálculo medio y/o promedio de los mismos. Generalmente para su representación se emplea la terminología marina “Rosa de vientos”, como se ve en la figura. 6. Los vientos son el condicionante mayor en la dispersión del contaminante, puesto según su intensidad la concentración del contaminante emitido a la atmosfera variará con respecto a la distancia.



Fuente: Senamhi, 2018.

Figura.7 Rosa de vientos del área metropolitana del Lima- Callao

Tasa de lapso adiabático

Una de las cualidades más relevantes de la atmosfera es la estabilidad que presenta, es decir, la tendencia que tiene para resistir o suprimir la turbulencia del lugar. Tiene que ver mucho con la dispersión de los contaminantes de manera vertical, pues el cambio de temperatura se da por la expansión adiabática y permite dispersar la polución emitida de manera ascendente. (Wark & Warner, 2014, p. 109).

Turbulencia

La turbulencia atmosférica puede generarse de dos maneras convectiva o mecánica. Estas se dan a través del calor emitido de la superficie terrestre y depende de la estabilidad térmica de la atmosfera y del movimiento de masas de aire sobre la superficie respectivamente.

Este último tiene relación con la rugosidad de la superficie: a mayor rugosidad, mayor turbulencia. (Parker, 2013, p. 133).

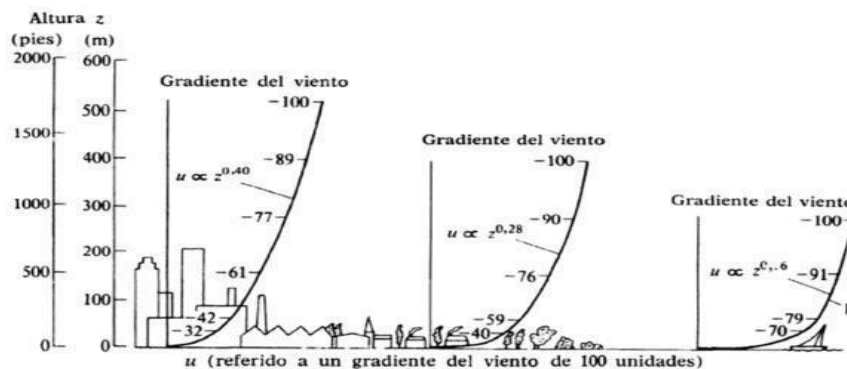


FIG. 5.2. Perfiles de velocidad media del viento sobre terrenos de diferente rugosidad.

Fuente: Contaminación del aire por la industria, 2014.

Figura.8 Perfil de velocidad de los vientos en relación a la rugosidad de la superficie.

De modo que la turbulencia es el efecto producido por remolinos de aire de distintos tamaños coexistiendo unos dentro de otros que aumentan la mezcla y la difusión. Por lo tanto, es evidente que los elementos que hay en un territorio condicionan diferentes gradientes de velocidad del viento en el perfil vertical de la zona. En el perfil establecido de velocidad del viento se busca diferenciar el comportamiento del mismo durante el día y la noche, por ello se percibe mayor velocidad de vientos en la noche que en el día.

Condiciones de estabilidad

Las condiciones estables de la atmósfera se establecieron a través de una tabla que permite diferenciar la dinámica y el enfriamiento de las masas de aire en la zona, es así que se pueden presentar condiciones estables o condiciones inestables –a manera de ejemplo- por la relación que existe entre la turbulencia y las condiciones del tiempo como se muestra en la Tabla.2 y figura. 9.

Tabla N°2 Turbulencia y condiciones del tiempo

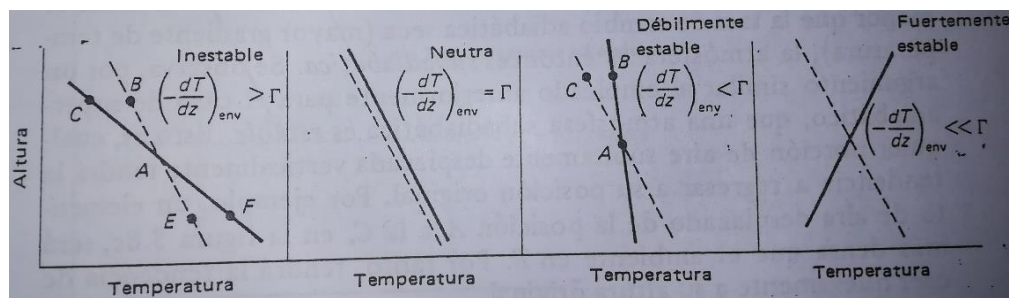
Relación de los tipos de turbulencia y las condiciones del tiempo

Velocidad del viento superficial ms ⁻¹	Insolación diaria			Condiciones nocturnas	
	Fuerte	Moderada	Ligera	Nubosidad ≥ 4/8 ‡	Nubosidad ≤ 3/8
< 2	A	A-B	B	E	F
2	A-B	B	C	D	E
4	B	B-C	C	D	E
6	C	C-D	D	D	D
> 6	C	D	D	D	D

† Aplicable a un nublado espeso, día y noche.

‡ El grado de nubosidad se define como la fracción del firmamento por encima del horizonte aparente local que está cubierto por nubes.

Fuente: Contaminación del aire por la industria, 2014



Fuente: Contaminación del aire Origen y control, 2014.

Figura.9 Condiciones de estabilidad atmosférica.

La estratificación térmica determina, a nivel local, la forma de una pluma procedente de una chimenea industrial. Por otro lado, también debe tomarse en cuenta a la capa de mezcla, pues es la capa en la que la atmósfera se mezcla homogéneamente y varía a lo largo del día.

El desplazamiento de las masas de aire, al tener mayor velocidad que la velocidad de salida de los gases emitidos por ejemplo el de las chimeneas, la pluma tenderá a depositarse a su alrededor debido a la presión atmosférica que se ejerce en el lugar. Por otro lado, si en el lugar se encuentra infraestructura urbana, la pluma tenderá a distorsionarse debido a las condiciones de viento que se verificaban antes de la dispersión del contaminante en la zona. En la figura. 12 se puede percibir el efecto que generan los edificios y la chimenea en la concentración del contaminante. (Parker, 2013, p. 138).

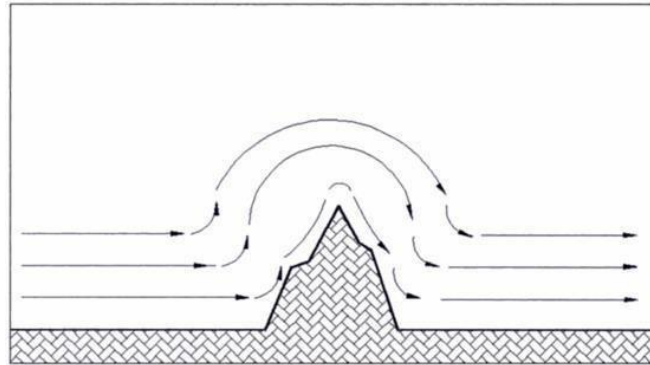


Fuente: Contaminación del aire por la Industria, 2014.

Figura.10 Efectos aerodinámicos de los edificios y topografía irregular.

Topografía

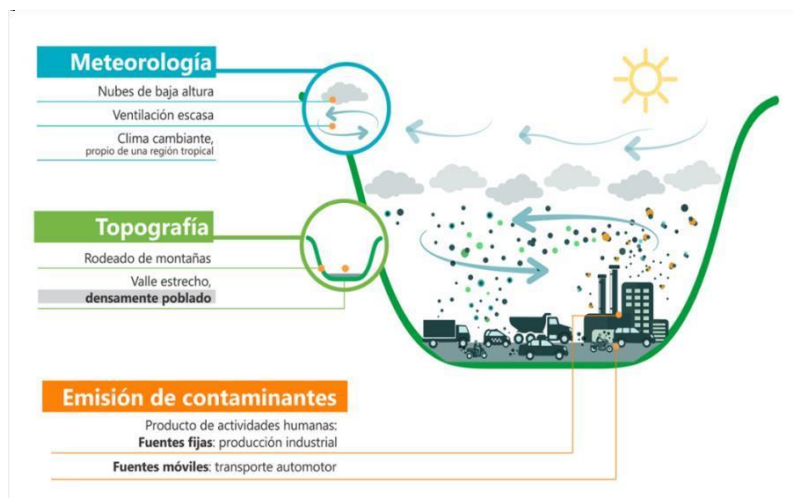
La topografía es una ciencia que representa gráficamente el relieve, la superficie terrestre o las deformaciones que tiene el suelo en un territorio. Es la orografía, o específicamente el relieve un de los factores topográficos que hace variar la concentración de los contaminantes atmosféricos de la zona, pues este incide por medio de dos componentes: la altitud y latitud. (Quereda, 2005, p. 41).



Fuente: Curso de Climatología general, 2004.

Figura.11 Orografía del terreno

Las montañas son importantes en el ciclo del agua porque captan la humedad de las masas de aire; así mismo en ellas, ocurre el efecto Fohen, este fenómeno determina variaciones en la temperatura a Barlovento y Sotavento del sistema montañoso en una misma altitud.



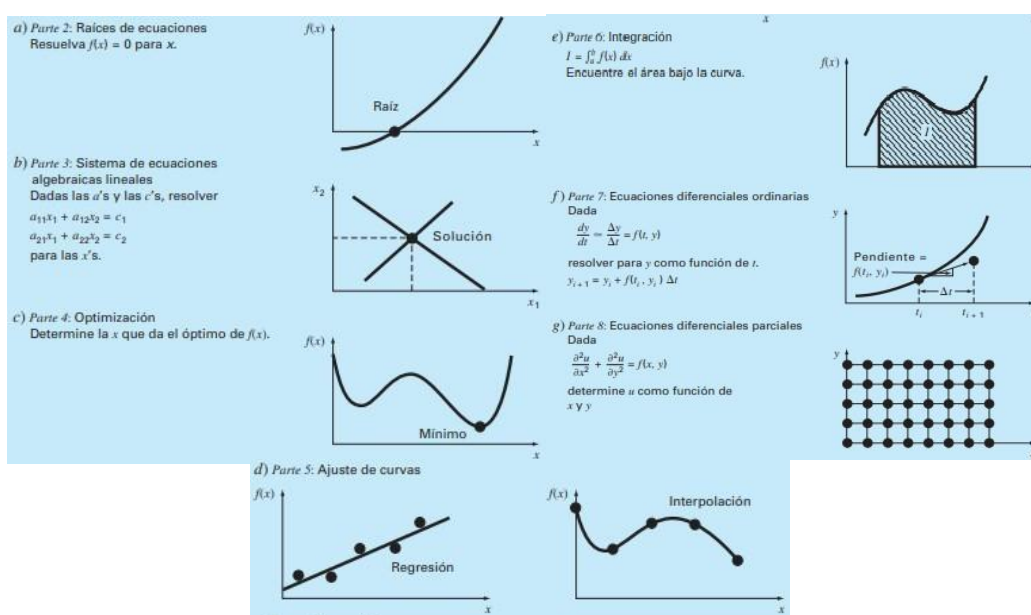
Recuperado de: <http://www.calidaddel aire.co/preguntas-frecuentes.php>

Figura.12 Ilustración del comportamiento de los contaminantes del aire por incidencia de la meteorología y topografía.

Las circulaciones locales son circulaciones de aire inducidas térmicamente por los procesos de superficie en regiones heterogéneas y estas influyen sobre la dispersión de contaminantes, estas pueden ser convencionales y no convencionales: Brisa marítima, brisa terrestre, variabilidad de la humedad del suelo, la vegetación, valle de montañas entre otros. (Chávez, 2011, p. 33).

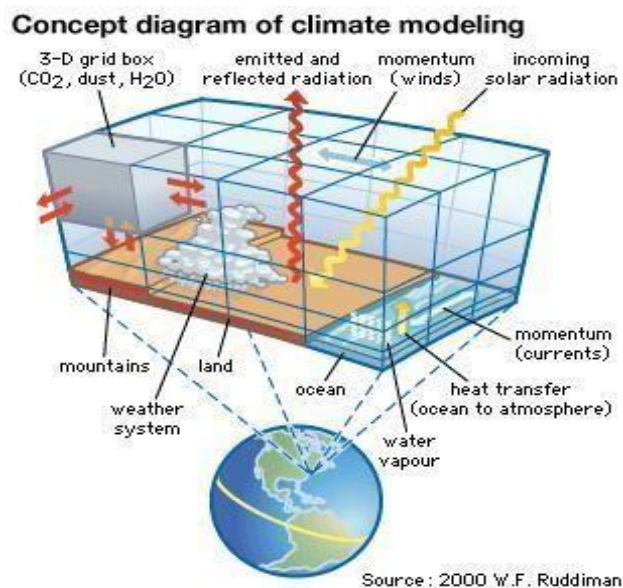
Modelamiento Numérico aplicado a las ciencias ambientales

Un modelo numérico es un conjunto de expresiones matemáticas que describen el comportamiento de un sistema físico-químico. Estas ecuaciones matemáticas son resueltas en un entorno de cálculo computacional y están basada en una serie de principios de conservación (de la masa, de la energía, del movimiento, del agua y en su caso de otros materiales gaseosos y aerosoles) tanto a nivel dinámico como termodinámico. Así mismo estos modelos pueden proporcionar información necesaria para pronosticar la evolución y comportamiento atmosférico en una zona determinada. [...] El uso de este método tiene su campo de aplicación en diversas áreas: 1) en la predicción y estudios del transporte y dispersión de contaminantes atmosféricos, 2) Un análisis más detallado de la predicción a escala regional y/o local, 3) Análisis detallado de fenómenos meteorológicos, 4) Estudios detallados para verificar zonas con mayores flujos de vientos para instalar parques eólicos. (Gaertner, 2002, p.10, 11).



Fuente: Métodos numéricos para ingenieros, 2007.

Figura.13 Modelos matemáticos lineales utilizados en ciencias ambientales.



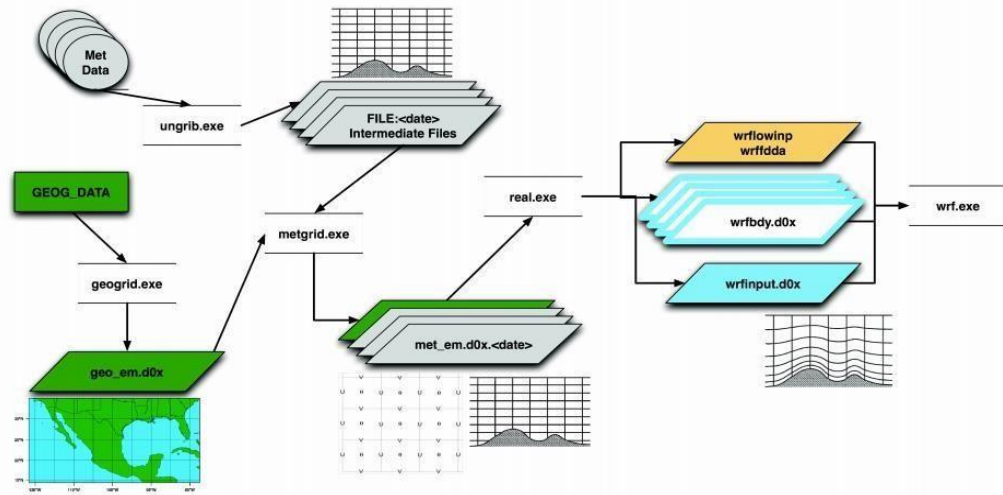
Fuente: <https://www.britannica.com/science/scientific-modeling>

Figura.14 Modelo de caja para el análisis del comportamiento atmosférico

Cabe resaltar que la figura 14 responde a un modelo de escala mayor, donde los sistemas meteorológicos abarcan grandes áreas en el planeta. De modo que para cada zona de estudio se realiza un ajuste de los parámetros para mayor precisión de información en la zona.

Weather Research and Forecast (WRF) Modeling System

El WRF es un modelo de numérico de área limitada (LAM) desarrollado por diversas entidades NCAR, NCAR, NOAA / ESRL, NOAA / NCEP / EMC entre otros; diseñado tanto para evaluar tanto el comportamiento atmosférico como para resolver las necesidades del pronóstico del tiempo. (User's Guide WRF, 2010, p.3). Esta comprende dos núcleos dinámicos: a) una propia de la investigación avanzada WRF (ARW) y b) un modelo de mesoescala no hidrostático (NMM) propia de la parte operativa. Con investigación, WRF, puede realizar el modelamiento y mejorar la resolución del lugar estudiando la meteorología. (WRF basics, 2012, p. 13).



Fuente: WRF Basics, 2012.

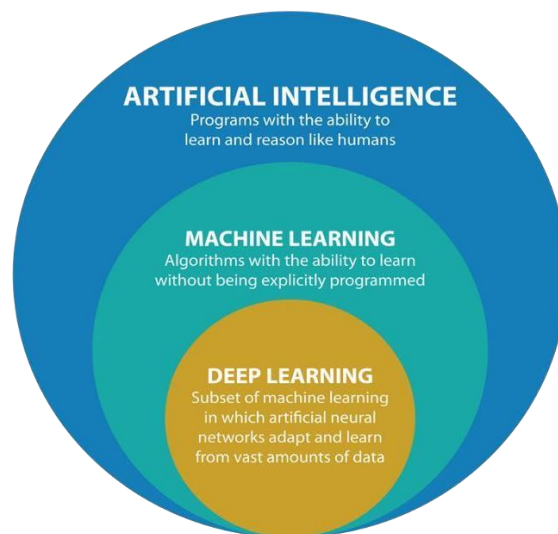
Figura.15 Flujograma de análisis de datos de WRF.

En la figura.15 se visualiza los datos de entrada, parte izquierda, necesarias para la modelación que se requiera (Datos meteorológicos, información topográfica georreferenciada, así como datos grillados); en la parte media se visualiza los parámetros que se toman en cuenta para modelar los datos ingresados, como una mejora en la resolución de los datos grillados y resolver el problema de las variables meteorológicas en los puntos específicos para el pronóstico. Por último, en la parte derecha se tienen las salidas que el modelo WRF puede generar: .CTL, .exe, .grib, etc. Para el presente estudio los datos son en formato .txt.

Inteligencia Artificial

La Inteligencia Artificial es definida por John Mc Carthy, (1956) como ‘la ciencia y la ingeniería de hacer inteligentes a las máquinas.’ De modo que la inteligencia artificial (IA) puede definirse como el medio por el cual las computadoras, los robots y otros dispositivos realizan tareas que normalmente requieren de la inteligencia humana. Por ejemplo, la resolución de cierto tipo de problemas, la capacidad de discriminar entre distintos objetos o el responder a órdenes verbales. La IA agrupa un conjunto de técnicas que, mediante circuitos electrónicos y programas avanzados de computadora, buscan imitar procedimientos similares a los procesos inductivos y deductivos del cerebro humano.

Por lo tanto, la IA es el desarrollo de algoritmos y métodos que permitan a las computadoras comportarse de modo inteligente.



Fuente: <https://medium.com/@experiencia18/diferencias-entre-la-inteligencia-artificial-y-el-machine-learning>

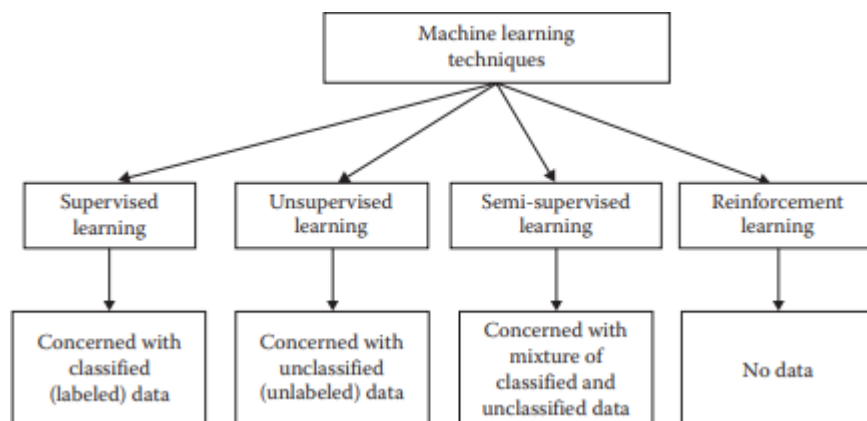
Figura.16 Diferentes técnicas de aprendizaje automático.

Existen diferentes modelos de inteligencia artificial, una clasificación de los modelos que se basan en el objetivo y la forma que trabaja un sistema. Entre ellos tenemos:

- Sistemas que Piensan como Humanos
- Sistemas que Actúan como Humanos
- Sistemas que piensan Racionalmente.
- Sistemas actuantes racionales.

Aprendizaje automático (Machine learning)

Las técnicas de aprendizaje automático (o machine learning) pueden definirse como un conjunto de métodos capaces de detectar automáticamente patrones de datos a analizar. En esta técnica se utilizan diversos métodos estadísticos con algoritmos simples, circunscritos casi exclusivamente al ámbito académico. Desde entonces el uso de los patrones detectados para realizar predicciones, o ara tomar otros tipos de decisiones en entornos de incertidumbre.



Fuente: Neural Network, 2012.

Figura.17 Diferentes técnicas de aprendizaje automático.

Aprendizaje profundo (Deep Learning)

El aprendizaje profundo es una técnica que se desarrolla en el ámbito del machine learning; utiliza métodos y algoritmos para resolver problemas, con conocimiento de la información proporcionada. Lleva a cabo el proceso de red neuronal artificial que se compone de un número de niveles jerárquicos. Entre ellos tenemos:

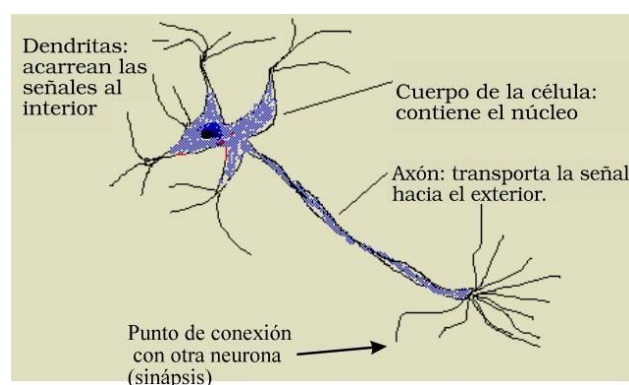
- Pre-procesamiento.
- Filtrados.

Redes neuronales Artificiales

Una Red Neuronal Artificial (RNA) se puede definir como: a) Un sistema computacional, inspirado en modelos biológicos (de ahí el concepto de una neurona) y/o b) Unos modelos matemáticos compuestos por un gran número de variables procesales organizadas en distintos niveles. [...] Cabe resaltar que las RNA responden a procesos de instrucciones u operaciones complejas [...], estas son el resultado de abundantes interconexiones de realimentación junto con procesos no lineales, pues los parámetros pueden ser incluso fenómenos dinámicos muy complejos. [...] La ventaja de utilizar una RNA es que los resultados son óptimos debido a las características tal cual está constituido la red, estas son: como: Aprendizaje adaptativo, auto organización, tolerancia a los fallos, operación en tiempo real y además fácil inserción a la tecnología vigente. (RNA, 1995, p. 17, 18). Por lo tanto, se puede entender como un modelo matemático compuesto por un gran número de elementos procesales organizados en niveles.

Concepto biológico neuronal

Para entender cuáles con los elementos principales que componen una red neuronal, es preciso conocer el funcionamiento de una neurona. Las neuronas tienen tres componentes principales: *Dendritas*, *Cuerpo de la célula o Soma* y el *Axón*. La conexión que realiza el axón de una célula con una dendrita de otra célula se llama *sinapsis*, en esta última, la interacción se concreta mediante señales eléctricas. Las dendritas, son las receptoras que cargan, de señales eléctricas, al Cuerpo de la célula. El cuerpo de la célula, procesa las señales de entrada y el Axón se encarga de llevar la señal desde el Cuerpo de la célula hasta otras neuronas (Acevedo, 2017, p.3). Las redes neuronales del conocimiento debido al hecho que se forman por asociación, comparten células y/o grupos celulares, pues un grupo puede ser parte de muchas redes neuronales. Ellas son un código relacional, y no se puede reducir a sus partes mínimas.



Fuente: RNA: fundamentos, modelos y aplicaciones 1995.

Figura.18 Forma principal de una neurona.

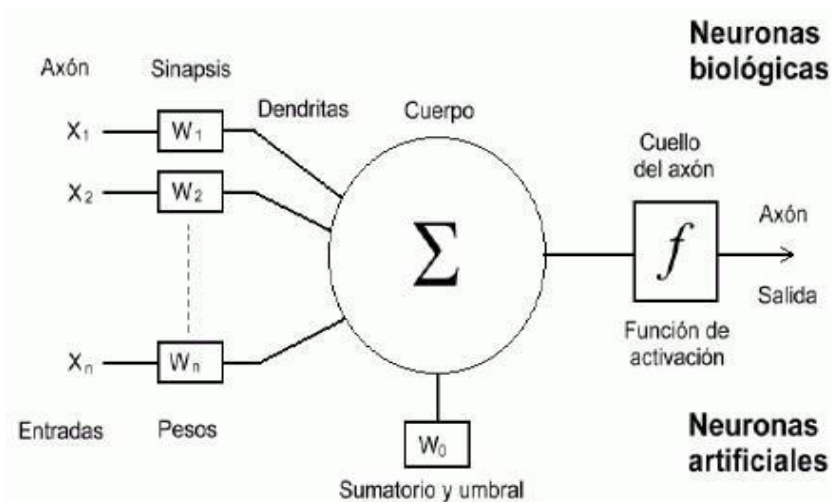
“La inmensa capacidad del almacén cognitivo del cerebro humano es fruto directo de dos hechos afines. Uno es el código relacional [...] el otro hecho es que estas relaciones constan de asociaciones entre células nerviosas. En la corteza humana, incluso un número limitado de neuronas –entre 10 mil y 20 mil millones- se puede asociar de maneras infinitamente distintas.” (Fuster, 2014, p.99).

De este modo, en términos computacionales, las dendritas, son las receptoras de la red, que cargan de señales eléctricas el cuerpo de la célula. El cuerpo de la célula, realiza la suma de esas señales de entrada y el axón proporciona la información desde el cuerpo de la célula hasta otras neuronas. [...] Por consiguiente una red neuronal está constituida por neuronas que se encuentran interconectadas y organizadas en tres capas.

Si podemos precisar la red neuronal puede entenderse como un proceso de entradas, procesamiento y salidas *capa de entrada* (input), que pasan a través de la *capa oculta* y salen por la *capa de salida* (output). (Acevedo, 2017, p.3).

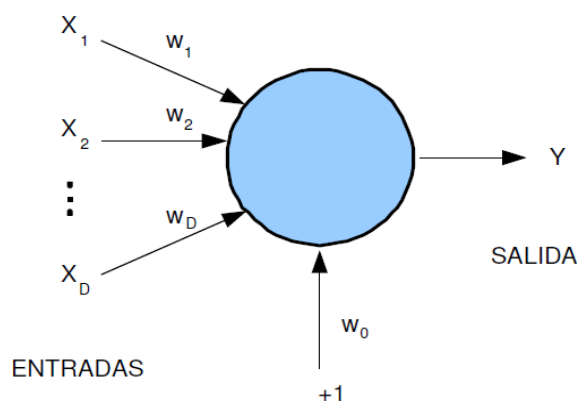
Arquitectura y funcionamiento de una neurona artificial

Es cierto que el modelo de neurona artificial, tiene su concepto biológico; sin embargo, la base de su diseñado lo conceptualizó McCulloch y Pitts, en 1943. “*Se denomina neurona a un modelo simple de cálculo que a partir de un vector de entrada $\{X_1, X_2, \dots, X_D\}$ que procede del exterior o de otras neuronas, que proporciona una única respuesta o salida Y .*” Aligual que las conexiones neuronales biológicas, las neuronas artificiales también establecen mayor o menor intensidad en sus procesos debido a lo pesos sinópticos que integran el modelo matemático; de modo que toda entrada X_i es afectada por un peso W_i . (Cortina, 2012, p. 57, 58). De este modo se tiene una estructura de neurona artificial que compone tres campos:a) Capa de Entrada: donde la información del exterior es recibida por este espacio, (aquí ingresan todas las variables que se toman en cuenta para el estudio). En las redes biológicas su símil son las *dendritas*. b) Capa oculta: donde se realizar el proceso de aprendizaje de la red. Esta se asemeja al *soma* en las redes biológicas. c) *Capa de Salida*: donde se brindan los resultados del proceso de la red al exterior y a otras neuronas. En las redes biológicas, esta sería una actividad realizada por el *axón*. (Acevedo, 2017, p.3). En las siguientes figuras: 19 y 20 se visualiza el concepto, la estructura de una neurona artificial y la similitud que tiene con la neurona biológica.



Fuente: Simulación de una Red Neuronal con mapas auto-organizados (SOM) de kohonen, 2017.

Figura. 19 Forma de una neurona artificial y la similitud con una neurona biológica.



Fuente: Aplicación de técnicas de inteligencia artificial a la predicción de contaminantes atmosféricos, 2012.

Figura.20 Modelo neuronal de McCulloch-Pitts.

Las salidas de la red neuronal son generadas bajo el cálculo de la suma ponderada a de la información de entradas que se tiene. El cálculo se realiza con la siguiente formula:

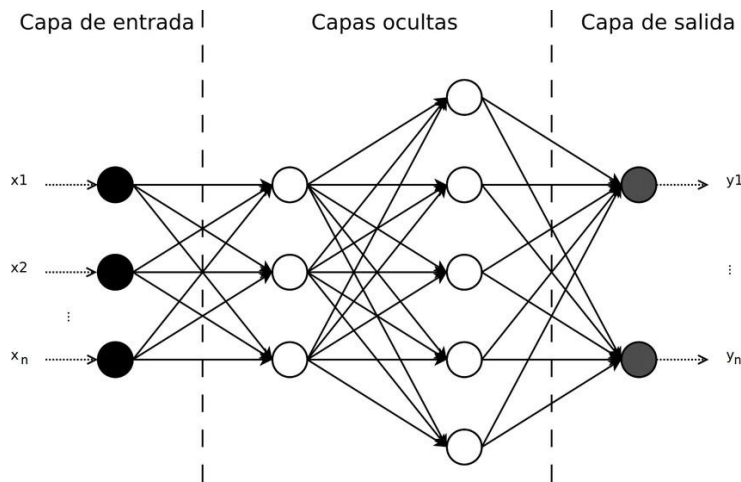
$$a = \sum_{i=1}^D w_i x_i + w_0 \quad \dots\dots\dots(1)$$

Donde w un sesgo que se utiliza para compensar la diferencia entre el valor medio de las entradas sobre el conjunto de entrenamiento, y el correspondiente valor medio de las salidas deseadas. Con esto tenemos también la función de activación o transferencia $g(a)$ que permite obtener la salida y mediante la ecuación:

$$y = g(a) = g \left(\sum_{i=1}^D w_i x_i + w_0 \right) = g \left(\sum_{i=0}^D w_i x_i \right) \quad \dots\dots\dots(2)$$

“Las neuronas artificiales se pueden clasificar de acuerdo a los valores que pueden tomar. Se pueden identificar dos tipos: 1) Neuronas binarias y 2) Neuronas reales. En el caso de las neuronas binarias, únicamente pueden tomar valores que se encuentren dentro del intervalo $\{-1, 1\}$ o $\{0, 1\}$. En el caso de las neuronas reales, estas pueden tomar valores que se encuentren dentro de los intervalos $[0, 1]$ o $[-1, 1]$. Generalmente, los pesos no se encuentran restringidos a intervalos específicos, aunque para aplicaciones específicas puede ser esto necesario” (Acevedo, 2017, p. 4).

Por otro lado, cabe resaltar que la topología (número de neuronas conectadas entre sí) es conocida como arquitectura de la red neuronal, y es allí donde se distinguen los tres tipos de capas mencionadas anteriormente. En la figura 21 se visualiza las capas y las conexiones que tienen las neuronas y como estas forman la RNA.



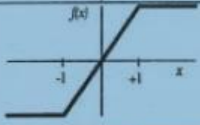
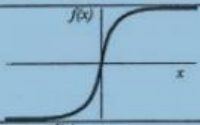
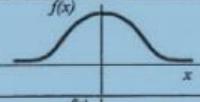
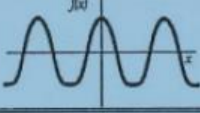


Fuente: Pronóstico del tipo de cambio USD/MXN con redes neuronales de retropropagación, 2017.
Figura.21 Arquitectura unidireccional con tres capas de neuronas: Capa de entrada, Capa oculta, Capa de salida.

Las redes pueden ser monocapa o multicapas. La red monocapa es aquella que está compuestas por una capa de neuronas; las multicapas, por varias capas neuronales. Teniendo en cuenta la dirección que los datos de la red neuronal, estas pueden distinguirse en: redes unidireccionales (*feedforward*) y redes recurrentes (*feedback*). Como su mismo término lo dice: las redes unidireccionales tienen un solo sentido; mientras que, para las redes recurrentes, la información puede circular entre las capas, ya sea el de salida o entrada. (Cortina, 2012, p. 62).

Funcionamiento de una RNA

Las redes neuronales pueden resolver problemas propuestos después de un entrenamiento, pues estas responden al peso sinóptico que la red establece a través de la información brindada y a la regla de propagación que se establece debido al patrón de conexión de las entradas. De modo que es preciso resaltar que la RNA responde a funciones de activación por cada valor que se obtiene con la regla de propagación. Las funciones de activación se eligen dependiendo el objetivo de entrenamiento de la red. En la figura 22 se muestran algunas de las funciones de activación más utilizadas.

	Función	Rango	Gráfica
Identidad	$y = x$	$[-\infty, +\infty]$	
Escalón	$y = \text{sign}(x)$ $y = H(x)$	$\{-1, +1\}$ $\{0, +1\}$	
Lineal a tramos	$y = \begin{cases} -1, & \text{si } x < -l \\ x, & \text{si } -l \leq x \leq +l \\ +1, & \text{si } x > +l \end{cases}$	$[-1, +1]$	
Sigmoidea	$y = \frac{1}{1+e^{-x}}$ $y = \text{tgh}(x)$	$[0, +1]$ $[-1, +1]$	
Gaussiana	$y = Ae^{-bx^2}$	$[0, +1]$	
Sinusoidal	$y = A \text{sen}(\omega x + \varphi)$	$[-1, +1]$	

Fuente: Principles and characteristics of artificial neural networks, 2017.

Figura. 22 Funciones de activación utilizadas.

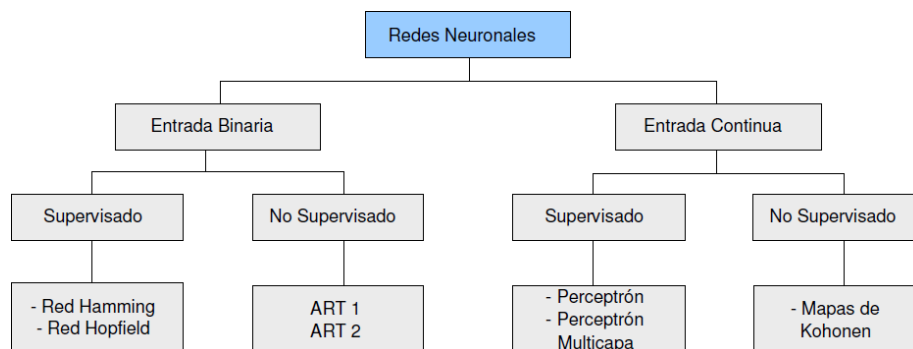
Para ello, Cortina (2012), nos hace un resumen de las siguientes fases por la que pasa la RNA:

- ✓ Fase de conceptualización del modelo para el estudio de un problema concreto. En esta fase se deben señalar las entradas, salidas y la información con la que se dispone.
- ✓ Fase de organización de los datos que se dispone a la estructura de la red a crear. Para esta fase se requiere construir los patrones de aprendizaje, es aquí que parte de la información es validada por la red.
- ✓ Fase de aprendizaje. Donde la red organiza la información y genera patrones para luego proporcionar salidas, estas salidas se repiten un cierto número de etapas, luego estas son comparadas con las salidas que se desean tener y los diversos algoritmos de aprendizaje para luego ver los patrones de validación.

- ✓ Fase de validación. En esta fase se presenta los resultados de la red entrenada al conjunto de valores que se desean y se ve el error cometido por la red en su conjunto. La validación es dada por la misma red cuando se procesa de nuevo los datos corregidos.
- ✓ Fase de generalización. Es la fase donde ya se logró conseguir la red adecuada para poder utilizar el modelo predictor, aportándole una nueva entrada y se procesará para dar una nueva salida.

Proceso de aprendizaje

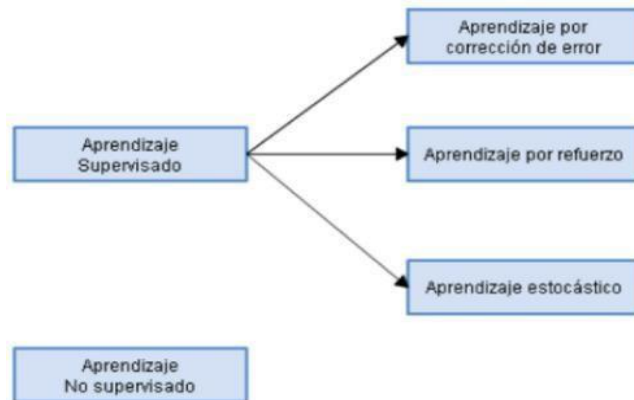
El proceso de aprendizaje de una neurona se determina a través del entrenamiento o de la experiencia con la que la RNA ha sido alimentada. Por consiguiente, dependiendo de cómo este constituida el mecanismo de aprendizaje, estas se pueden clasificar como: a) redes supervisadas y redes no supervisadas. Para mayor alcance se muestra la figura 21, donde se puede ver el orden de estas.



Fuente: Aplicación de técnicas de inteligencia artificial a la predicción de contaminantes atmosféricos, 2012.

Figura.23 Orden de las redes neuronales artificiales.

En cuanto a la red neuronal supervisada y no supervisada, la diferencia radica en que los resultados por el primer tipo de aprendizaje son controlados por el modelador, es decir, bajo control de calidad o ajuste se cambian los pesos, variables o parámetros, para obtener el resultado determinado, mientras que el otro tipo no.



Fuente: Principles and characteristics of artificial neural networks, 2017.

Figura.24 Tipos de aprendizaje.

Cuando se usa el tipo de aprendizaje supervisado se toma en cuenta todas las variables de entradas y salidas deseadas que la red debería asumir, puesto que el objetivo es aprender la relación entre ellas, es decir, se busca visualizar todo en general (entradas y salidas) que se tiene en la red. La notación matemática se representa de la siguiente manera:

$$X = \{(x^n, t^n)\}_{n=1}^N \dots\dots\dots(3)$$

X es el valor de entrada, t el valor de salida deseada y N es el tamaño del conjunto, es así que el entrenamiento consiste en que la red pueda repetir estos patrones con el menor error posible. De este modo se considera una función genérica de regresión lineal para poder clasificar y ajustar los datos. La función responde al criterio de tener una buena aproximación de t:

$$y = f(x|w) \dots\dots\dots(4)$$

Y representa la salida proporcionada por la red y está en función a las variables de entrada y el peso que se les designa según la relación que se determine en la modelación. Por otro lado, lo que se pretende es tener el menor error posible, por ello en cada fase de aprendizaje se debe indicar que tan cerca se está de la solución. La notación matemática utilizada para minimizar el error es el Error Cuadrático Medio (MSE-Mean Square Error):

$$MSE = \frac{1}{N} \sum_{n=1}^N (y^n - t^n)^2 \dots\dots\dots(5)^*$$

Donde: n es la cantidad de patrones, Y^n es la salida obtenida por la red, y t^n la salida que debería obtenerse.

Para el presente estudio de investigación se considera la red neuronal perceptrón multicapa, de tipo de aprendizaje supervisado de entrada continua. Por lo que es conveniente realizar una introducción al perceptrón simple y multicapa.

Perceptrón simple y Perceptrón Multicapa

Un perceptrón es la RNA más simple, tiene su origen bajo la regla de Hebb, pues basándose en el hecho biológico, cuando dos neuronas se activan simultáneamente, su conexión se refuerza, de modo que el modelo matemático para el perceptrón simple responde a la siguiente ecuación:

$$w_i(\tau + 1) = w_i(\tau) + \eta x_i y \dots\dots\dots(6)$$

Donde la entrada X_i es la salida de otra neurona Y , W_i es el peso en el momento $(T+1)$ y η es la tasa de aprendizaje.

De este modo se analiza las salidas del perceptrón simple compuesto por dos capas. La ecuación matemática utilizada se representa como:

$$y = g(a) = g \left(\sum_{j=0}^M w_j \phi_j(\mathbf{x}) \right) = g(\mathbf{w}^T \phi) \dots\dots\dots(7)$$

Donde j número de neuronas en una capa de entrada con funciones ϕ_j que realiza la transformación de los datos de entrada. Así mismo ϕ es el vector generado por las funciones de activación y g es la función de escala.

Teniendo en cuenta que el algoritmo (conjunto de operaciones sistemáticas) de aprendizaje de un perceptrón sencillo se fundamenta más en la corrección de error, es menester que se ajusten los pesos para que se actualice la nueva red que se proporciona y así obtener las salidas deseadas. El criterio que sigue el perceptrón simple es el siguiente:

$$w_j(\tau + 1) = w_j(\tau) + \eta \phi_j^n t^n \dots\dots\dots(8)$$

Donde n es la tasa de aprendizaje.

Así mismo, para el perceptrón simple, es necesario visualizar los procesos de activación debido a que cada entrada genera uno nuevo. La fórmula que responde al control de estas se visualiza bajo la siguiente notación:

$$E^{perc}(\mathbf{w}) = - \sum_{\phi^n \in M} (\mathbf{w}^T \phi^n) t^n \dots\dots\dots(9)$$

Donde M es el conjunto de vectores ϕ^n que han sido clasificados incorrectamente, por lo tanto, E^{perc} tendrá el valor de 0 cuando los vectores estén debidamente clasificados.

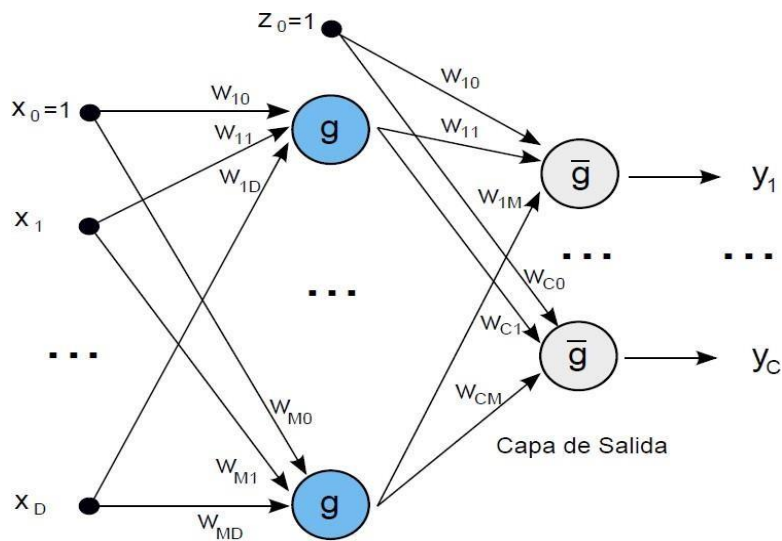
La arquitectura de una red de perceptrón multicapa constituye algoritmos de retropropagación de errores. De modo que la fórmula para ver el control de estas se determina bajo la siguiente formula:

$$z_j = g(a_j) = g \left(\sum_{i=1}^D w_{ji} x_i \right); \quad k = 1, 2, \dots, M \dots\dots\dots(10)$$

Donde D es el número de entradas, M número de neuronas en la capa oculta y C el número de neuronas en la capa de salida. Por otro lado, a_j es el nivel de activación de la neurona j de la capa oculta; así mismo W_{ji} el peso asociado a la neurona mencionada. En general, se calcula como una combinación lineal de las D entras X_i luego de aplicar la función de transferencia g se obtiene la salida Z_j de las neuronas. La salida de la red neuronal es obtenida como la sumatoria ponderada de las salidas de las neuronas de la capa oculta y sobre ella es menester aplicar una función de transferencia de modo que la notación matemática tiene la forma:

$$y_k = \tilde{g} \left(\sum_{j=0}^M w_{kj} z_j \right) = \left(\sum_{j=0}^M w_{kj} g \left(\sum_{i=1}^D w_{ji} x_i \right) \right); \quad k = 1, 2, \dots, C \quad \dots\dots\dots(11)$$

En la figura 23 se presenta la arquitectura de un perceptrón multicapa, donde se visualiza una red alterna que sirve para comparar los resultados que se desean tener y alimentar de la misma forma a la entrada de la red de perceptrón multicapa.



Fuente: Aplicación de técnicas de inteligencia artificial a la predicción de contaminantes atmosféricos, 2012.

Figura.25 Arquitectura de un perceptrón multicapa.

Algoritmo de Retropropagación

El perceptrón multicapa sustenta el aprendizaje de sus pesos primordialmente en la regla de ajustes de error. Para ello el Perceptrón multicapa utiliza el método de descenso del gradiente para ajustar esos pesos de la red con la que fue alimentada. Cortina (2012) nos ilustra en dos fases como se entiende este proceso de retropropagación. Primea fase: En esta fase se introduce un patrón de datos a la entrada de la red para que el proceso de la función del algoritmo se propague capa tras capa (paso hacia adelante) teniendo como resultado un conjunto de salidas con determinados pesos como resultado a los patrones de entrada. Segunda fase: Consiste en realizar la propagación de lo aprendido, teniendo en cuenta los

pesos de la red, se recalculan según a la regla de ajuste de error, aquí es donde se compara la respuesta actual de la red con la respuesta deseada, propagando el error hacia atrás.

La notación matemática que responde sobre el error dentro de la retropropagación es el error cuadrático para cada patrón, se determina de la siguiente forma:

$$E^n = \frac{1}{2} \sum_{k=0}^C (e_k^n)^2 \quad \text{donde} \quad e_k^n = y_k^n - t_k^n \quad \dots\dots\dots(12)$$

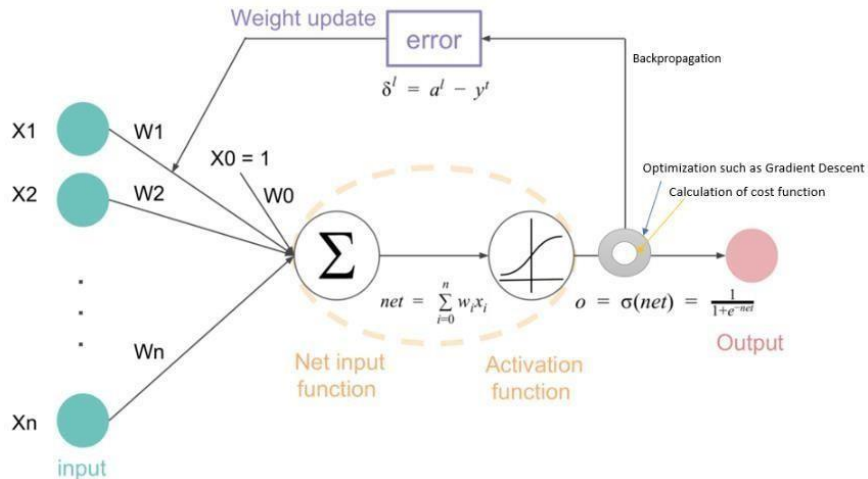
El mismo proceso de retropropagación corrige el peso sinóptico que se le asigna W_{kj} , y se añade un incremento diferencial Δw_{kj} que es proporcional al gradiente mencionado anteriormente. La ecuación que sirve para actualizar los pesos en el algoritmo generado está representada de la siguiente manera:

$$\Delta w_{ji} = -\eta \delta_j x_i \quad \dots\dots\dots(13)$$

Donde representa el gradiente local de neuronas en la capa oculta, estimada con la siguiente fórmula:

$$\delta_j = g'(a_j) \sum_k \delta_k w_{kj} \quad \text{donde} \quad g'(a_j) = \frac{\delta z_j}{\delta a_j} \quad \dots\dots\dots(14)$$

Este procedimiento permite obtener los datos más precisos para el aprendizaje de la red neuronal y así aplicar retropropagación es aplicar capa tras capa moviendo el erro hacia atrás, gracias a esto cuando llegemos a la primera capa, se habrá obtenido cual es el error para cada neurona y para cada uno en sus parámetros, solamente propagando una sola vez el error hacia atrás.



Fuente: <https://datascience.stackexchange.com/questions/44703/how-does-gradient-descent-and-backpropagation-work-together>

Figura. 26 Arquitectura del algoritmo de retropropagación.

Criterio de ajuste de los parámetros de entrada de la RNA

El objetivo de analizar la base de datos es poder generalizar un modelo acorde al comportamiento de las variables incluidas en la base de datos. Es así que un modelo que no se ajusta al comportamiento de las variables no permitirá realizar buenas predicciones, a este tipo de modelo se le determina como *underfitting* lo que puede estar ocurriendo, en este modelo, es que el mismo modelo no tiene la flexibilidad suficiente como para poder adaptarse a los puntos de la nube de datos. Por otro lado, ajustar demasiado un modelo presenta *overfitting* pues existe demasiada flexibilidad en el modelo y solo respondería frente a las determinadas características a la que se le ajusta. De modo que un modelo bien ajustado es aquel que se encuentra en un punto intermedio, que sea capaz de aprender en el entrenamiento de la RNA, y que también sea capaz de generalizar su conocimiento.

El criterio para no aterrizar en los modelos ineficientes y tener modelos que generalicen, consiste en separar la base de datos en porcentajes para el entrenamiento del aprendizaje (80% de datos de la base original de variables) y lo otro para la prueba de predicción de los comportamientos en función a la relación que tienen cada variable (20% datos de la base original de variables), de este modo se asegura la calibración de un modelo acorde a la realidad. El método que apoya esta adecuada función es el aprendizaje supervisado, pues hay una variable independiente dentro de todos los datos y estas se pueden agrupar de manera beneficiosa para un modelo adecuado y preciso.

Planteamiento del problema

El problema de la polución del aire en el Área de Lima metropolitana y Callao (AMLC) se hace relevante en la medida que se trate de gestionar y/o controlar las fuentes generadoras de los agentes adversos a la salud y al ambiente. Hoy se cuenta con una Red de Monitoreo de Calidad del Aire (RAMCA) que vigilia las concentraciones de los contaminantes en diez distritos, los resultados de este monitoreo registra a los distritos de Villa María del Triunfo, Carabayllo y Ate como los más afectados por el Material Particulado (PM_{10}). Por otro lado, los distritos de San Borja y Jesús María presentan concentraciones bajas. Por ello es importante seguir con las medidas de control orientadas a reducir estas emisiones, sin embargo, no se verifica si las medidas puedan aplicarse a mediano o largo plazo, debido a que no se cuenta con información prospectiva sobre este hecho. Por lo que es menester que se cuente con una herramienta que permita visualizar con anticipación el comportamiento de contaminantes atmosféricos -teniendo en cuenta la meteorología y física de la atmósfera- sobre todo del PM_{10} que es un contaminante que supera, en la mayoría de los distritos de Lima, el ECA- aire. En el Perú ya se trabaja con modelos para el pronóstico de calidad del aire, sin embargo, no se verifica una proyección eficiente y esto es por la cantidad de información que se necesita analizar y manipular. De modo que la presente investigación pretende realizar un análisis de datos de los distritos de Ate Vitarte (Plaza municipal) y Jesús María (Campo de Marte) sobre el comportamiento y pronóstico del PM_{10} intentando diseñar un modelo matemático basado en Redes Neuronales Artificiales (RNA), los cuales presentan un aprendizaje adaptativo, capaz de aprender a realizar tareas basadas en un entrenamiento o experiencia previa. Pues el resultado de esto hace capaz al sistema, de entender el comportamiento del PM_{10} a partir de información histórica del mismo contaminante y de las variables que puedan influenciar su concentración.

1.4. Formulación del problema

1.4.1. Problema General

¿De qué manera el modelo estocástico permitirá estimar la concentración de Material Particulado (PM_{10}) en los distritos de Jesús María y Ate Vitarte?

1.4.2. Problemas Específicos

¿En qué medida el modelo estocástico permitirá estimar la concentración de Material Particulado (PM_{10}) en el distrito de Jesús María (Campo de Marte)?

¿En qué medida el modelo estocástico permitirá estimar la concentración de Material Particulado (PM_{10}) en el distrito de Ate Vitarte (Plaza municipal)?

1.5. Justificación del estudio

El uso de un modelo matemático basado en la RNA del presente estudio tiene justificación ambiental puesto que permitirá implementar un sistema de pronóstico en el comportamiento del Material Particulado PM_{10} esperando así, se establezca una red de monitoreo acorde a la realidad acontecida; teniendo en cuenta los factores que pueden intervenir en sudispersión.

Así mismo el estudio se justifica por su relevancia social debido a que el modelo permitirá diagnosticar y alertar, de manera anticipada, la concentración de PM_{10} el cual la población puede estar afectada y así reforzar las políticas públicas que intentan menguar el problema que acontecerá.

Finalmente se justifica por su valor teórico-metodológico, pues el modelo tiene importancia de uso global ya que permite establecer controles y de redes de monitoreo a nivel local realizando algoritmos de modelación y viendo así la incidencia de los factores atmosféricos con mayor resolución y precisión.

1.6. Hipótesis

1.6.1. Hipótesis General

El modelo estocástico permitirá estimar la concentración de Material Particulado (PM_{10}) en los distritos de Jesús María y Ate Vitarte.

1.6.2. Hipótesis Específicas

El modelo estocástico permitirá estimar la concentración de Material Particulado (PM_{10}) en los distritos de Jesús María (Campo de Marte).

El modelo estocástico permitirá estimar la concentración de Material Particulado (PM_{10}) en el distrito de Ate Vitarte (Plaza municipal)

1.7. Objetivos

1.7.1. Objetivo general

Desarrollar un modelo estocástico que permita estimar la concentración de Material Particulado (PM_{10}) en los distritos de Jesús María y Ate Vitarte., 2019

1.7.2. Objetivos Específicos

Elaborar un modelo estocástico que permita estimar la concentración de Material Particulado (PM_{10}) en el distrito de Jesús María (Campo de Marte).

Elaborar un modelo estocástico que permita estimar la concentración de Material Particulado (PM_{10}) en el distrito de Ate Vitarte (Plaza municipal).

II. MÉTODO

2.1. Tipo y Diseño de investigación

El presente estudio tiene un alcance cuantitativo y como tipo de investigación es descriptiva, pues identifica el comportamiento del fenómeno estudiado; correlacional porque existe una asociación de las variables presentadas, así mismo, permite realizar predicciones del comportamiento de la variable dependiente. El diseño de investigación es no experimental (longitudinal) ya que se analizará los cambios de la variable a través del paso del tiempo; es a su vez de nivel aplicativo puesto que sus variables son medibles y se emplearán herramientas para generar estrategias en la investigación del proyecto respectivamente.

2.2. Operacionalización de variables

	VARIABLE	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADOR	UNIDAD
INDEPENDIENTE	Factores Climáticos secundarios	Los factores climáticos secundarios son aquellos que tienen incidencia en los seres vivos pero que su acción es menos relevante, debido a su permanencia o poca frecuencia. (Pulido, 2013, p. 71)	Los factores climáticos secundarios son registrados mediante monitoreos de la estación meteorológica ubicada en ATE y Jesús María	Factores meteorológicos.	Temperatura	° C
					Velocidad de los vientos	m/s
					Dirección de los vientos	(°)
					Radiación solar	w/ m3
					Humedad relativa	%
					Presión atmosférica	Pha
	Contaminantes primarios	Son contaminantes atmosféricos producidos por acción directa de una reacción de fuente antropogénica. Es relevante por el efecto cuantitativo en el aire, referente al nivel de inmisión. (Contreras & Molero, 2014, p.71)	Las concentraciones de los contaminantes atmosféricos son simulados mediante el software WRF	Estabilidad espacial de los contaminantes primarios	Concntración de contaminantes primarios	ug/m3
DEPENDIENTE	Material Particulado (PM 10)	Son partículas pequeñas que pueden presentar grados de toxicidad debido a su composición física o química en el ambiente y en el aparato respiratorio de los seres vivos (Wark & Warner, 2014, p. 39)	El pronóstico de la concentración del Material Particulado (PM10), se estima mediante el modelo de Red Neuronal Artificial (RNA)	Estabilidad espacial del Material Particulado (PM10)	Concentración del Material Particulado (PM10)	ug/m3

2.3. Población, muestra y muestreo

2.3.1. Población

Data histórica de la concentración de Material Particulado de los distritos de Jesús María y Ate Vitarte.

2.3.2. Muestra

Registro de datos del año 2018 de la concentración de Material Particulado (PM₁₀) de los distritos de Jesús María y Ate Vitarte con los respectivos datos meteorológicos de la zona.

2.3.3. Muestreo

El muestreo presentado en el trabajo investigativo es de carácter aleatorio-simple, puesto que cada miembro de la población tiene la misma probabilidad de ser seleccionado como sujeto y se toma la muestra por interés del investigador, preferentemente de las dos zonas del distrito Jesús María y Ate Vitarte.

2.4. Técnicas e instrumentos de recolección de datos, Validez y Confiabilidad

2.4.1. Técnica e instrumentos de recolección de datos

El presente trabajo de investigación se sustenta mediante la recolección y análisis de la información de datos meteorológicos y la concentración del Material Particulado (PM₁₀) en las zonas específicas. Es así que la técnica a utilizar es presentada mediante los instrumentos de caracterización de la zona y la toma de parámetros del aire en un periodo determinado. Entre los instrumentos utilizados se presentará el formato de campo o de registro de datos de la zona.

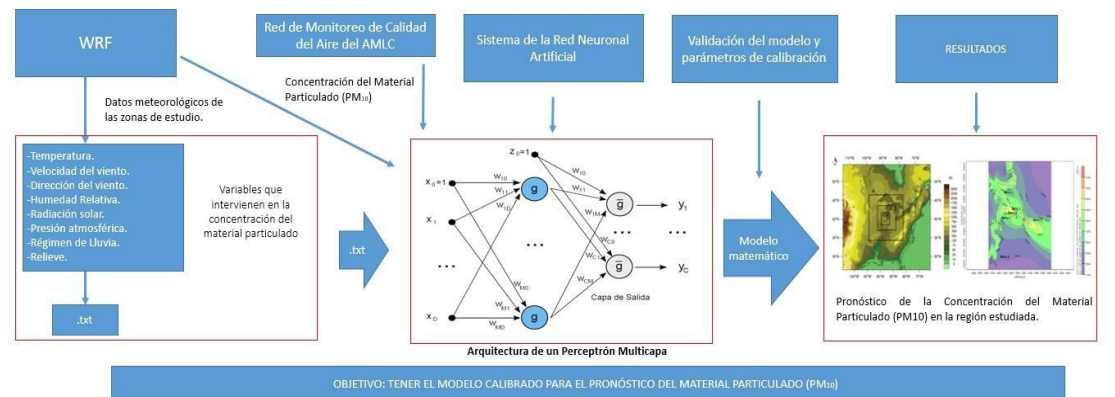
2.4.2. Validez

La validez es el grado en que un instrumento puede medir las variables y refleja un dominio específico del contenido. De modo que los instrumentos de medición presentados en el estudio, serán validados por el juicio de tres especialistas.

2.4.3. Confiabilidad

Los datos obtenidos, en el proceso de recolección de información, para el presente estudio son brindados por el Senamhi, bajo previa supervisión de la Subdirección de Evaluación del Ambiente Atmosférico (SEA).

2.5. Procedimiento



Fuente: Elaboración propia.

Figura.27 Flujograma de procedimiento para los resultados del estudio.

La metodología para obtener los objetivos de estudio se distingue en las siguientes fases:

- ✓ Identificación del área de estudio.
- ✓ Procesamiento de la base de datos proporcionados por la RAMCA y el modelamiento del software WRF.
- ✓ Desarrollo del modelo matemático de predicción para el Material Particulado PM₁₀.
- ✓ Evaluación de los errores de predicción.

2.5.1. Preprocesamiento de la base de datos

La base de datos explorada para la entrada de la RNA debe ser estandarizada y agrupada debidamente puesto que la información está sujeta a diferentes factores que provocan el error de los mismos, como las fallas en el sistema de comunicación, instrumentos no calibrados, falta de mantenimiento, cortes de energía eléctrica, horas repetidas, entre otros. De modo que, para la validación de la información de datos horarios, se le estima criterios como descartar la información que muestren valores negativos o nulos; descartar los valores que son anómalos o niveles altos en los datos.

Para el descarte de los niveles altos en las mediciones de los equipos tanto del WRF y de la estación automática de calidad del aire, se propuso no superar los siguiente:

- Material Particulado (PM_{10}), valores no mayores a $400 \mu\text{g}/\text{m}^3$.
- Velocidad de vientos (VV), valores no mayores a 20 m/s.
- Dirección del viento (DV), valores no mayores a 360 grados.
- Temperatura (T), no mayores a los 50 grados Celsius.
- Humedad relativa (HR), no mayor a 100 %.

El resto de datos que se suman a las variables de entradas para la RNA, mantienen sus valores normales. Además, es aconsejable que la cantidad tenga una suficiencia de 75%.

Por otro lado, para el análisis de las variables de entrada a la RNA es menester realizar la extracción de conocimiento debido a la gran cantidad de datos. En inteligencia artificial los datos son reconocidos a través del *reconocimiento de patrones* de modo que la eficiencia depende de cómo se encuentran los datos debidamente agrupados. Para ello se usan técnicas estadísticas que emplean algoritmos para encontrar, correlaciones, asociaciones, anomalías y patrones ocultos de la base de datos. Por ello el proceso de agrupamiento consiste en el particionamiento de datos en diferentes *clusters* o grupos, de modo que se busca maximizar la similitud en el grupo de datos, que no sean muy distintos de las características o cualidades de los demás.

Es importante resaltar que, en contaminación del aire, los procesos que se realizan son complejos por diversos factores no lineales, pues son caóticos por las reacciones fisicoquímicas que se manifiestan en la atmósfera. Con la finalidad de encontrar la correlación entre los factores climáticos secundarios y la concentración del contaminante, se utilizan técnicas estadísticas que permiten ver el espacio que tiene los datos en una serie temporal.

La técnica para realizar la similaridad esperada en el agrupamiento (cluster) de los datos que se utilizan en el presente estudio es K-means. La técnica en mención, sigue un proceso de clasificación simple de un conjunto de datos de entrada. Esta técnica realiza la división del conjunto de vectores de entradas en una cantidad m (k para el código del script) de grupos y estableciendo la distancia entre los vectores de entrada, así como las cualidades de referencia para generar los grupos. Esta técnica es un algoritmo de agrupamiento no supervisado.

La técnica es representada mediante la siguiente expresión matemática:

$$J = \sum_{j=1}^m \sum_{i=1}^n \beta_j(x_i) \|x_i - c_j\|^2 \quad \dots\dots\dots(15)$$

Donde n es la cantidad de patrones de entrada, m es la cantidad de grupos y la función de pertenencia β , de patrón de entrada X_i al cluster C_j está definida en el rango de salida (0, 1):

$$\beta_{c_j}(x_i) = \begin{cases} 1 & \text{si } \|x_i - c_j\|^2 < \|x_i - c_l\|^2 \forall l \neq j, l=1, \dots, m \\ 0 & \text{en otro caso} \end{cases} \quad \dots\dots\dots(16)$$

Así mismo cada patrón de entrada se asigna un cluster dependiendo las cercanías de las demás variables.

$$c_j = \{x_i \in X : \beta_{c_j}(x_i) = 1\} \quad \dots\dots\dots(17)$$

De modo que se recalculan los cluster de la división, que asigna a cada uno, en el centroide de los vectores de entrada.

$$c_j = \frac{\sum_{i=1}^n \beta_{c_j}(x_i) x_i}{\sum_{i=1}^n \beta_{c_j}(x_i)} \quad \dots\dots\dots(18)$$

Así, los clusters, también se pueden detenerse en el algoritmo utilizado, por ello se utiliza un valor umbral como una medida de parada del algoritmo comparado con la distorsión de la división actual. La ecuación que determina la distorsión es la siguiente:

$$\varepsilon = \sum_{j=1}^m \varepsilon_j \quad \text{donde} \quad \varepsilon_j = \sum_{x_j \in c_j} \|x_i - c_j\| \quad \dots\dots\dots(19)$$

2.5.2. Desarrollo del modelo matemático de predicción para el Material Particulado

PM₁₀

El objetivo de analizar series temporales es usar modelos que intenten explicar la correlación de los parámetros o medida de tendencia de la evolución de las variables, usando su propia data histórica para implementar un modelo de regresión que permita predecir a corto y largo plazo lo que se desea identificar. Para poder predecir los valores futuros del contaminante de interés es necesario tener en cuenta los siguientes pasos: a) Normalización de datos; que consiste en normalizar los datos de entrada dentro del rango [0-1]. b) Guía de entrenamiento y prueba; son los patrones de la información que se utiliza, viendo las concentraciones pasadas y presentes para predecir las concentraciones del contaminante estudiado en corto tiempo. c) Arquitectura de la RNA y d) Ajuste de resultados de la predicción.

Para realizar el desarrollo del modelo se utilizó el software Rstudio que permite ejecutar trabajos estadísticos en forma rápida. En el diseño de la red neuronal se considera:

- La red de Perceptrón Multicapa tipo retropropagación que permite tener una cantidad apropiada de neuronas en la capa oculta.
- Función de activación escogida para este tipo de trabajo es la función Sigmoide.
- El entrenamiento se realiza con los ajustes del parámetro de error del algoritmo de retropropagación. Después de cada entrenamiento empleado se realiza una validación para comprobar si no se presenta sobre aprendizaje. Es así que se realizó la partición de dos grupos (Entrenamiento-test) con el fin de no presentar problemas de *overfitting* o *underfitting*

2.5.3. Evaluación de los errores de predicción

Consiste en ver las técnicas que se pueden utilizar para evitar la pérdida de generalización, de ello, proporcionar un esquema final de red neuronal y la validación de la red neuronal para la estimación de PM₁₀.

Para la validación de los resultados el primer paso es determinar la confiabilidad de los resultados comparando los datos reales con los datos pronosticados. Como bien lo dice Cortina (2012) las fórmulas que se utilizan para ver el porcentaje de error y los ajustes de confiabilidad son:

Error Cuadrático Medio (MSE por sus siglas en inglés), realiza las diferencias entre el valor real y el predicho.

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_p - Y_o)^2 \dots\dots\dots(5)^*$$

Dónde: n es la cantidad de patrones, Y^n es la salida obtenida por la red, y t^n la salida que debería obtenerse.

-Raíz del Error Cuadrático Medio (RMSE por sus siglas en inglés), que proporciona una idea global de la diferencia entre los valores predichos y reales.

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^N (Y_p - Y_o)^2 \right]^{1/2} \dots\dots\dots(20)$$

-Media del Error Absoluto (MAE por sus siglas en inglés), cuantifica los errores residuales.

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_p - Y_o| \dots\dots\dots(21)$$

-Correlación (R), indica una descripción global del modelo.

$$R = \frac{\mu((Y_o - \mu(Y_o))(Y_p - \mu(Y_p)))}{\sigma_{Y_o} \sigma_{Y_p}} \dots\dots\dots(22)$$

La tabla 3, muestra la interpretación de los valores de R. Esto es tomado como referencia de los criterios de estimados por Cortina.

Tabla N°3 Valores de los coeficientes de correlación (R)

Coeficiente	Interpretación
0	Relación nula
0-0,2	Relación muy baja
0,2-0,4	Relación baja
0,4-0,6	Relación moderada
0,6-0,8	Relación alta
0,8-1	Relación muy alta
1	Relación perfecta

d , propone la congruencia entre los datos estimados y datos registrados, donde Y_o indica los valores observados; Y_p , los valores pronosticados; N , el número total de observaciones; μ , la media aritmética; σY_o y σY_p la varianza de los valores observados y estimados.

$$d = 1 - \frac{\sum_{i=1}^n (Y_p - Y_o)^2}{\sum_{i=1}^n (|Y_p - \mu(Y_p)| + |Y_o - \mu(Y_o)|)^2} \dots\dots\dots(23)$$

2.6. Método de análisis de datos

El análisis de datos será corroborado y precisado mediante pruebas estadísticas de correlación, regresión lineal y porcentaje de error (por medio del algoritmo propuesto) pues los datos evaluados tendrán un valor y peso mediante el proceso de modelamiento de la misma.

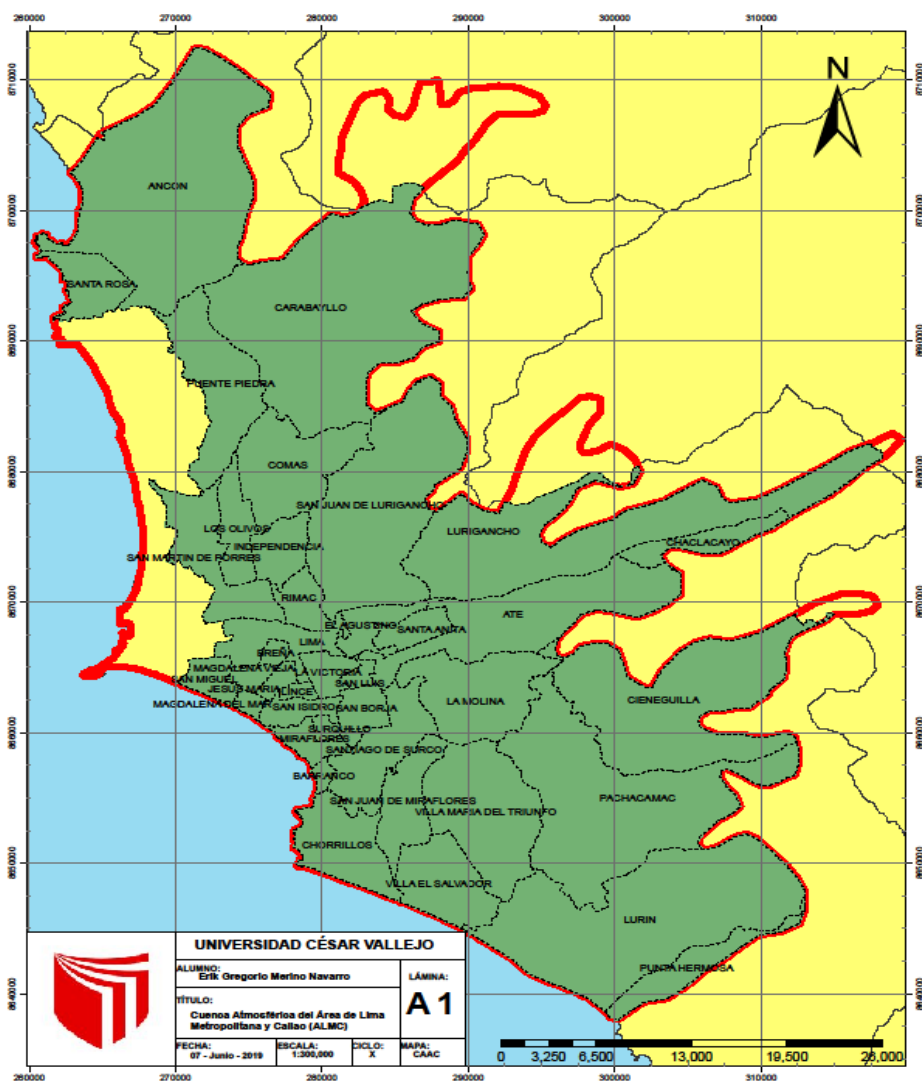
2.7. Aspectos éticos

Los resultados obtenidos del presente trabajo de investigación, dedico mi plena entrega para que estos sean los más verídicos y transparentes posibles. De modo que no perjudique al ambiente y a las personas en su desarrollo y toma de decisiones respectivamente.

III. RESULTADOS

3.1. Área de estudio

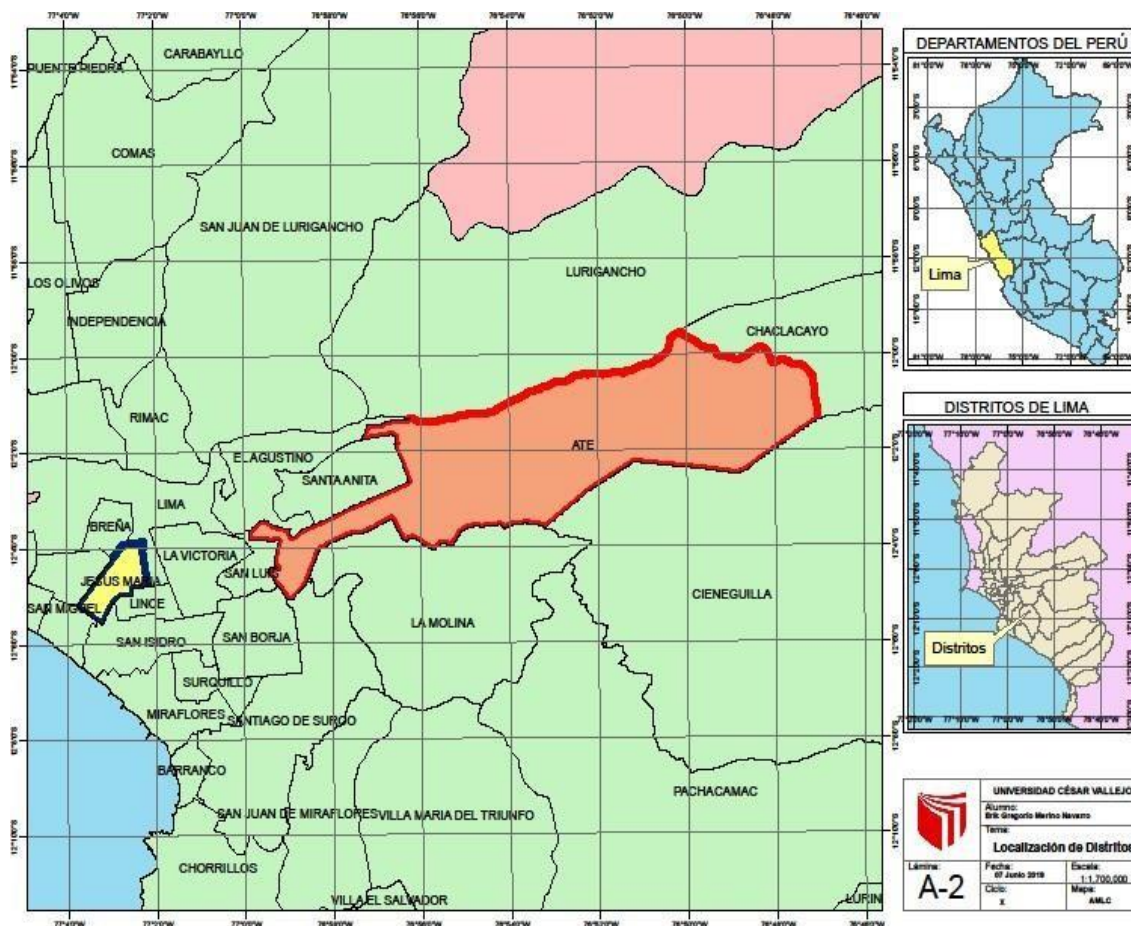
El Perú está influenciado por factores climáticos como: la Cordillera de los andes, el Anticiclón del pacífico sur, también por la corriente Oceánica peruana, y el fenómeno del niño. Estas determinan el clima en el país; específicamente en la costa central donde es notable la capa de inversión térmica entre los 600-800 metros de altura, la cual a su vez origina la estabilidad atmosférica con la formación de nubes estratiforme que predomina de 7 a 8 meses al año. El área metropolitana de Lima y Callao (AMLC) tiene una cuenca atmosférica que abarca las cinco zonas de la costa central del Perú (figura 28). Esto es relevante puesto que el fin es identificar la distribución espacio-temporal de la contaminación atmosférica.



Fuente: Elaboración Propia.

Figura.28 Mapa de ubicación de la Cuenca Atmosférica del AMLC.

Dentro de la cuenca atmosférica presentada en la figura 29 (contorno rojo en el mapa), se encuentra un predominante flujo de vientos rumbo norte. Así mismo, debido a la complejidad del suelo y características urbanas, se generan vientos turbulentos que a su vez permite que la dispersión o la sedimentación de contaminantes atmosféricos.



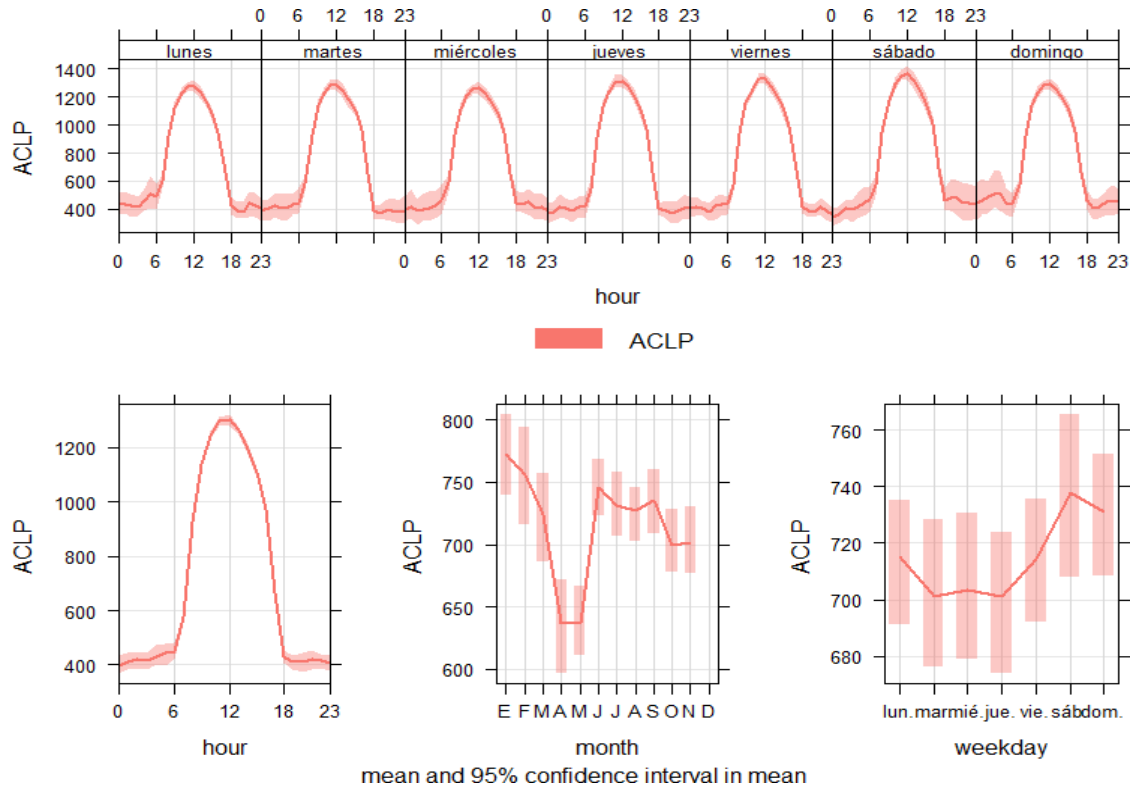
Fuente: Elaboración Propia.

Figura.29 Mapa de ubicación de los distritos de Jesús María y Ate vitarte.

3.2. Capa mezcla

Un parámetro relevante, como se dijo anteriormente, es la capa de mezcla (CLP) en la dispersión de contaminantes atmosféricos, pues representa el volumen de aire disponible para la concentración y transporte de los mismos. En la Figura 30, se muestra la variación promedio mensual de la altura de la capa de mezcla en el Aeropuerto Jorge Chávez, determinada con el sistema de radiosondeo de la Corporación Peruana de Aeropuertos y Aviación Comercial (CORPAC S. A.) del año en el 2018. La menor altura de la capa de mezcla se visualiza en los meses de abril y mayo, lo cual se debió a la disminución de la

intensidad del anticiclón del Pacífico sur. La altura de capa de mezcla durante el día, aumenta su volumen a las 12:00 horas del mediodía, en promedio 1.3 km, luego baja su altura durante el resto del día.

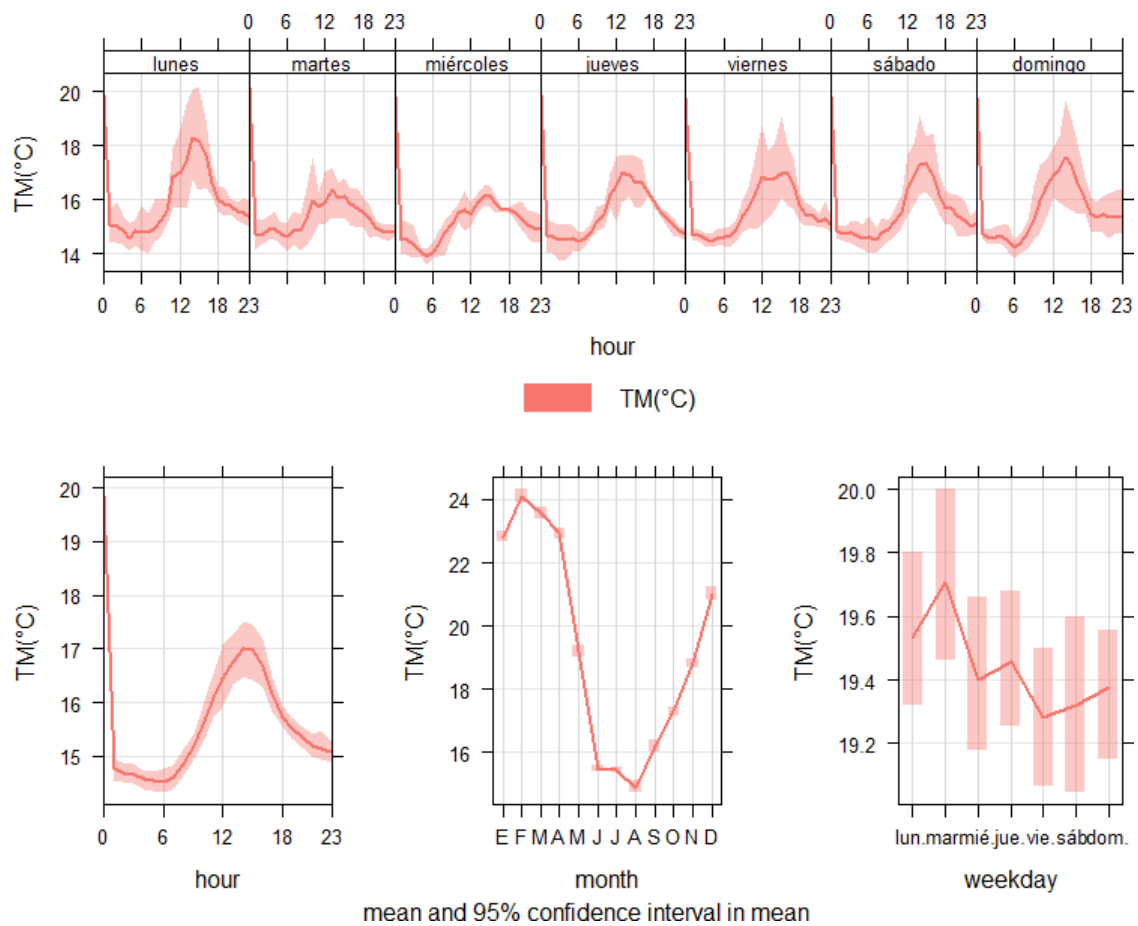


Fuente: Elaboración propia.

Figura.30 Variación mensual de la Capa de mezcla-Estación meteorológica automática-Senamhi, 2018.

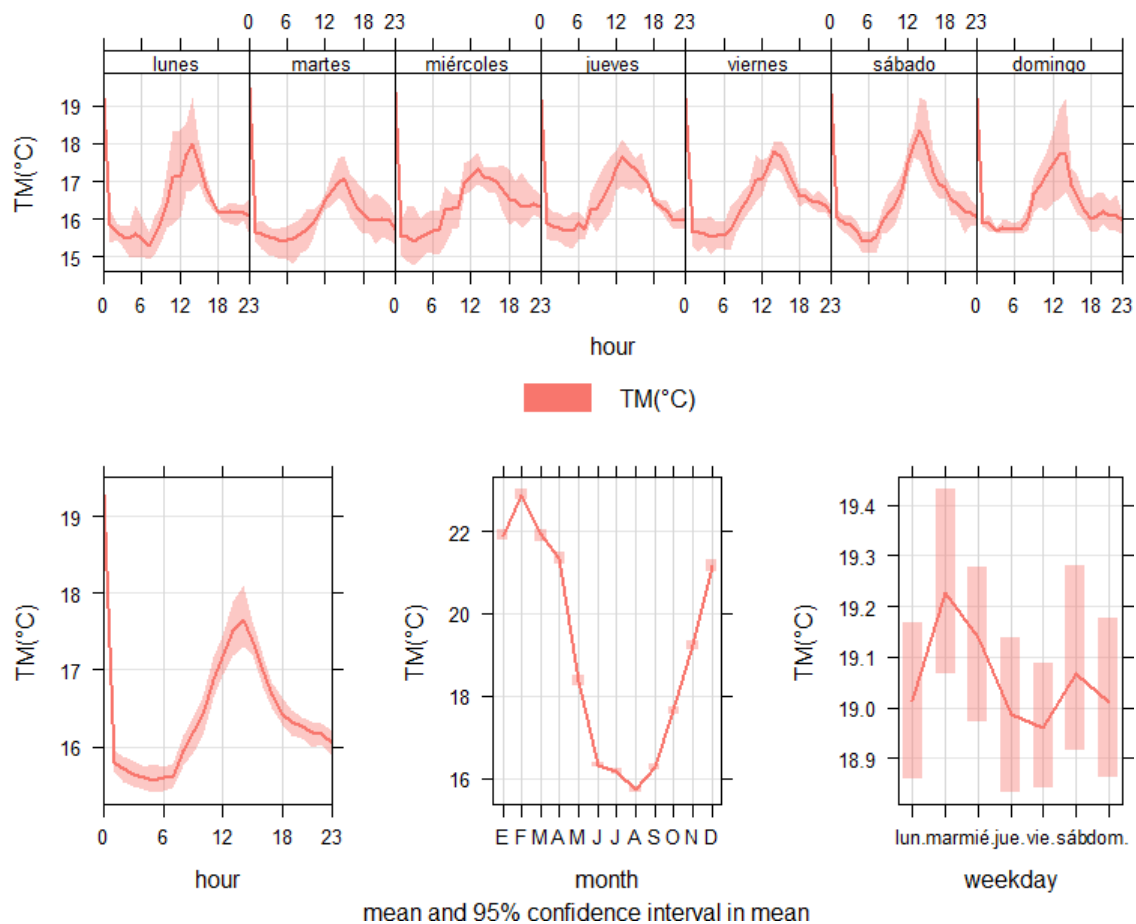
3.3. Temperatura

En la Figura 31 y 32, se muestra la variación mensual de la temperatura del aire en las estaciones meteorológicas de Plaza municipal (Ate Vitarte) y Campo de Marte (Jesús María). Se observa una variabilidad térmica baja en los meses de agosto de ambas estaciones, debido al alto calor específico que necesita el agua del mar, que permite la evaporación y actúa como regulador térmico en sectores colindantes a la zona marino costera, como el AMLC. El promedio mensual que contiene valores altos fue registrado en la estación de la Plaza municipal (Ate); el valor alto de la temperatura máxima fue de 24.6 °C y se registró en febrero; el valor más bajo de la temperatura fue de 15.3 °C en el mes de agosto.



Fuente: Elaboración propia.

Figura.31 Variación mensual de la temperatura del aire para los distritos de Ate, meteorológica automática-Senamhi, 2018.

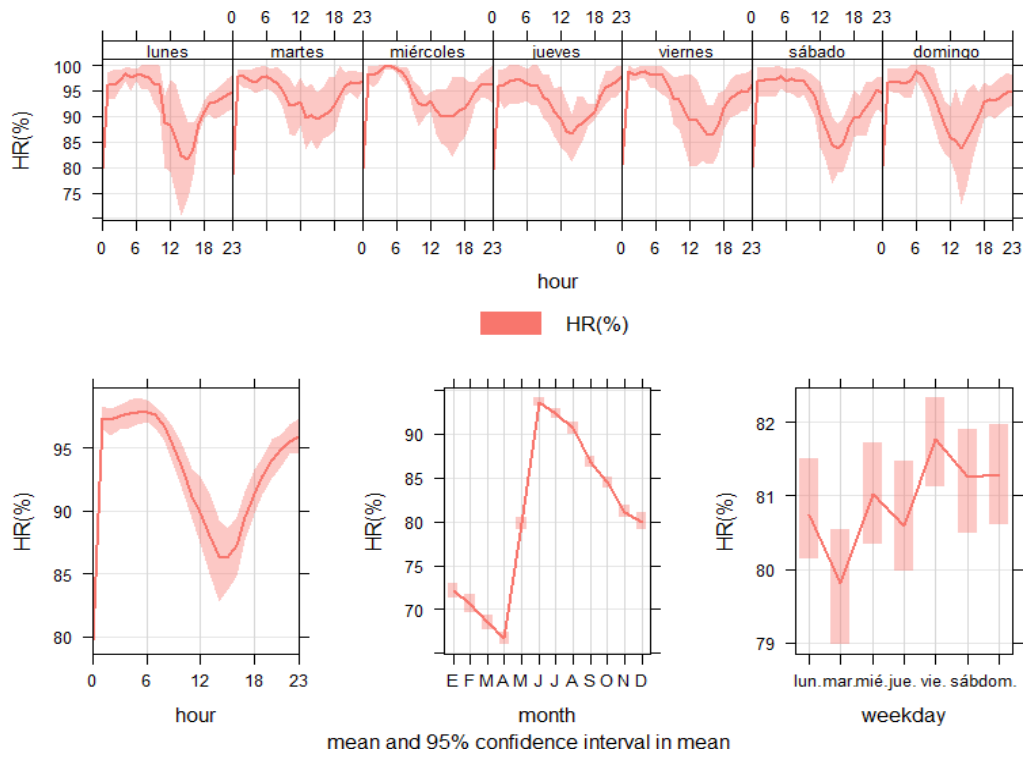


Fuente: Elaboración propia.

Figura. 32 Variación mensual de la temperatura del aire para los distritos de Jesús María, meteorológica automática-Senamhi, 2018.

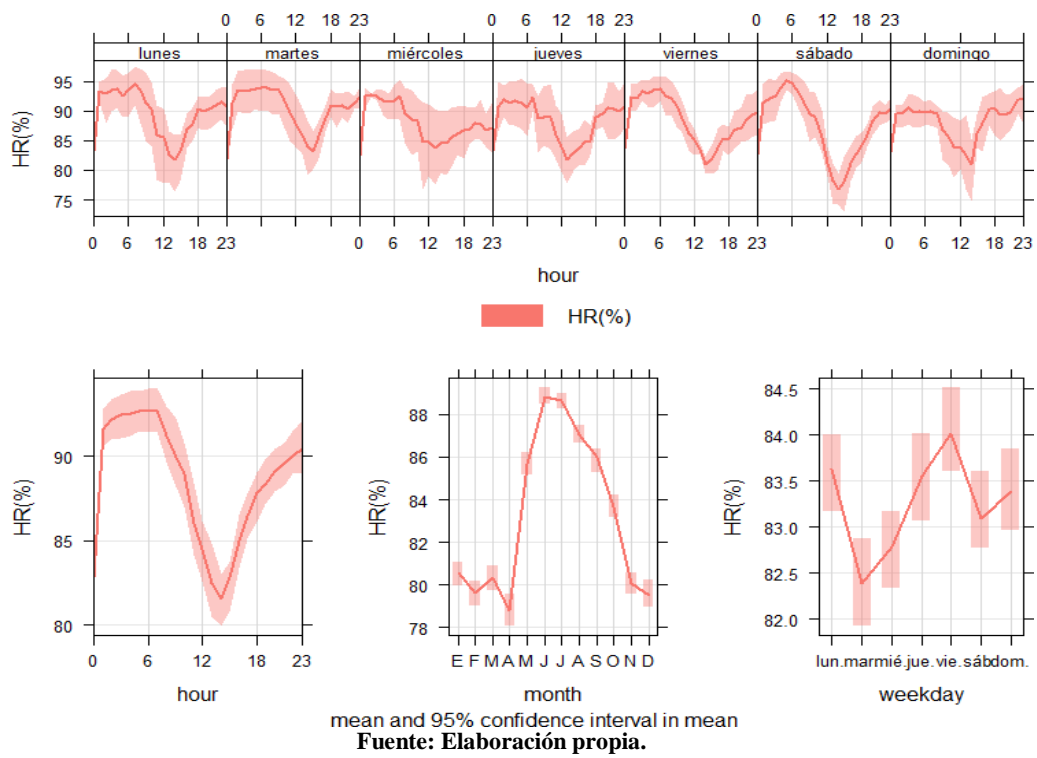
3.4. Humedad relativa

Por otro lado, en las figuras 33 y 34 se presenta la variación promedio mensual de la humedad relativa registrada en las estaciones meteorológicas de Plaza municipal (Ate) y Campo de Marte (Jesús María). El promedio anual para ambas estaciones fue de 81% y 83 % respectivamente. La estación de Ate presentó la mayor humedad relativa con 97 % en el mes de junio y una baja humedad relativa en el mes de abril. Mientras que la estación de Campo de marte presentó la mayor humedad relativa en el mes de junio con 89 % y una baja humedad relativa en el mes de abril.



Fuente: Elaboración propia.

Figura.33 Variación del promedio mensual de la humedad relativa del distrito de Ate -Estación meteorológica automática-Senamhi, 2018.

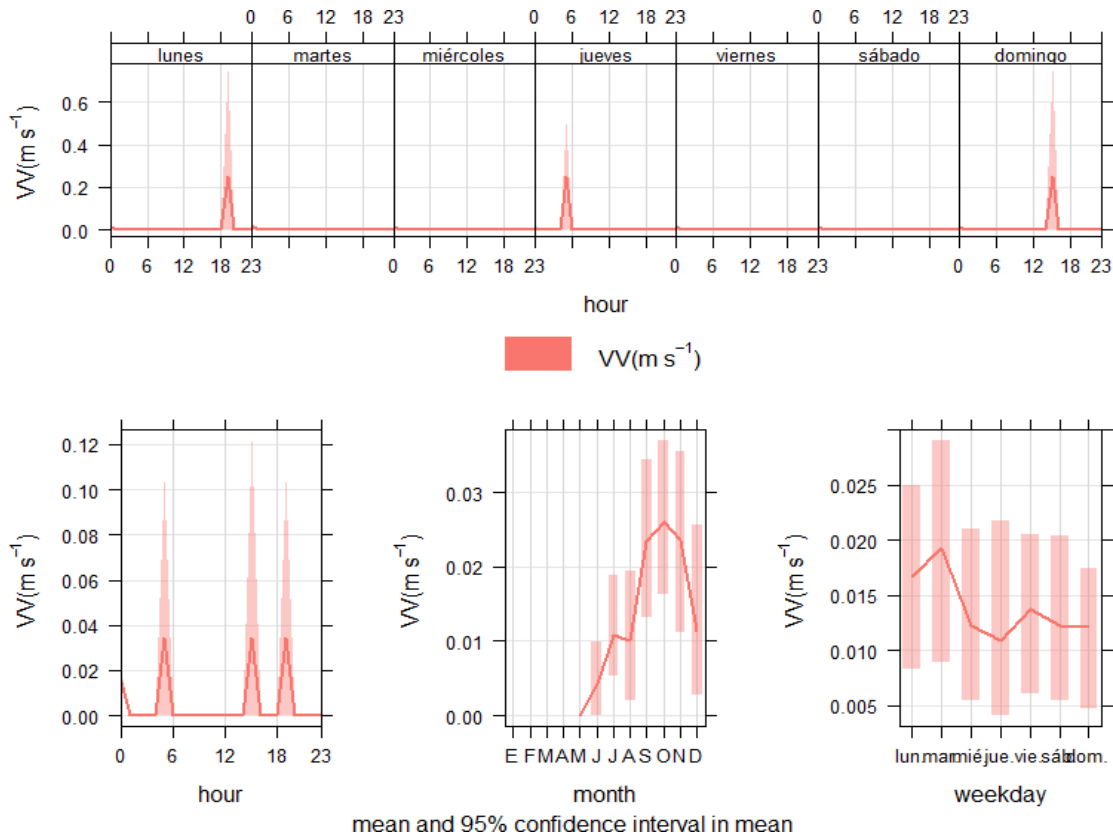


Fuente: Elaboración propia.

Figura.34 Variación del promedio mensual de la humedad relativa del distrito de Jesús María- Estación meteorológica automática-Senamhi, 2018.

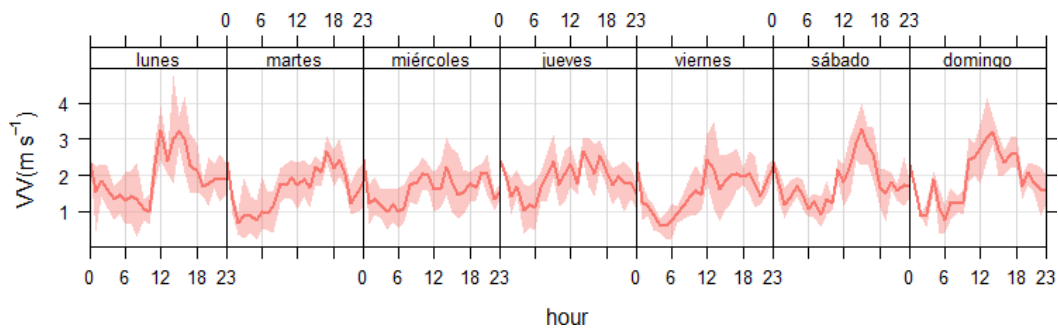
3.5. Vientos

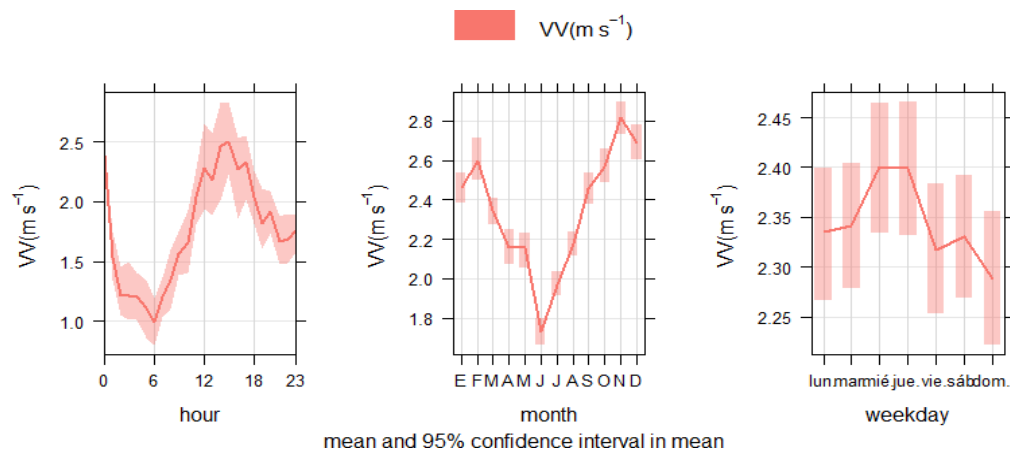
Las variaciones mensuales de la velocidad del viento de las estaciones meteorológicas de Campo de Marte y Plaza municipal de Ate presentan diferentes medias a partir del mes de junio, de modo que en la estación de Campo de Marte se registró velocidades entre 1.5 (junio) y 2.8 m/s (noviembre) y velocidades entre 0.0 1 m/s (mayo) y 0.0 2 m/s (agosto) para el distrito de Ate. (Figura 35 y 36).



Fuente: Elaboración propia.

Figura.35 Rosa de vientos-Estación meteorológica automática ATE-Senamhi, 2018.

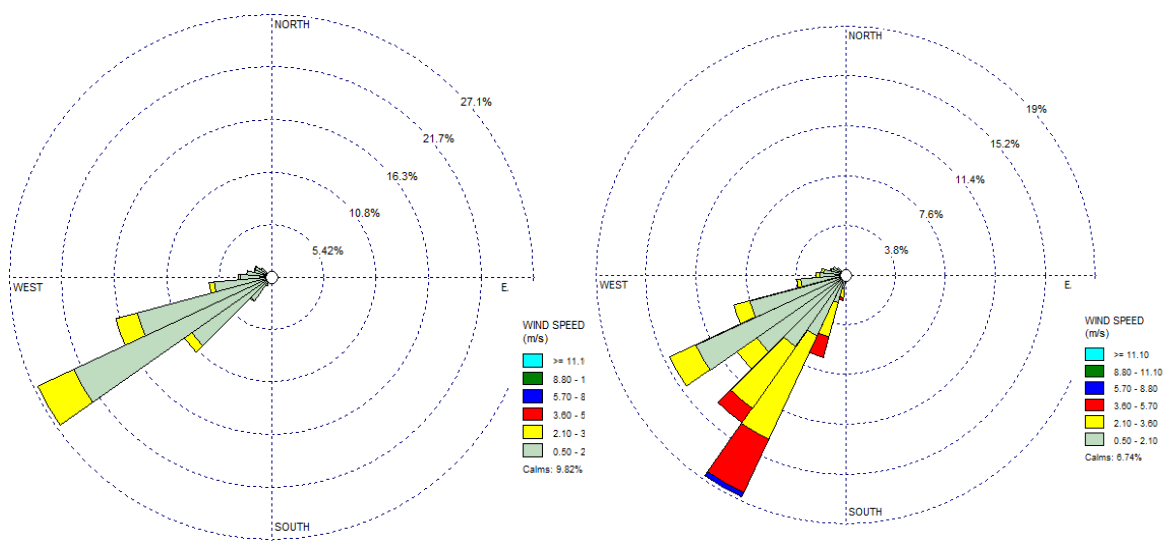




Fuente: Elaboración propia.

Figura.36 Rosa de vientos-Estación meteorológica automática de Jesús María-Senmahi, 2018.

El comportamiento de los vientos locales de Lima responde a un flujo estándar de los vientos con dirección NOR-ESTE. Sin embargo, en cada zona de los distritos se presentan diferentes flujos de vientos puesto que las pendientes, las zonas accidentadas, los componentes geográficos, y los edificios del lugar influyen en las salidas de los mismos. Como se puede apreciar, en la figura 37, hay un predominio de vientos provenientes SE que en ambas zonas se manifiestan.



Fuente: Elaboración propia.

Figura.37 Rosa de vientos-Estación meteorológica automática Ate y Jesús María-Senamhi, 2018.

Se observa, además, la rosa de vientos que permite identificar y analizar el comportamiento de los vientos locales respecto a la velocidad y dirección de los mismos.

Los valores promedios encontrados en el distrito de Ate y Jesús María presentan un rango de 0.01 – 0.02 m/s y 1.5. – 2.8 m/s respectivamente. La escala Beaufort puede clasificar los tipos de vientos que se presentan en los distritos.

Tabla N°4 Velocidades de vientos

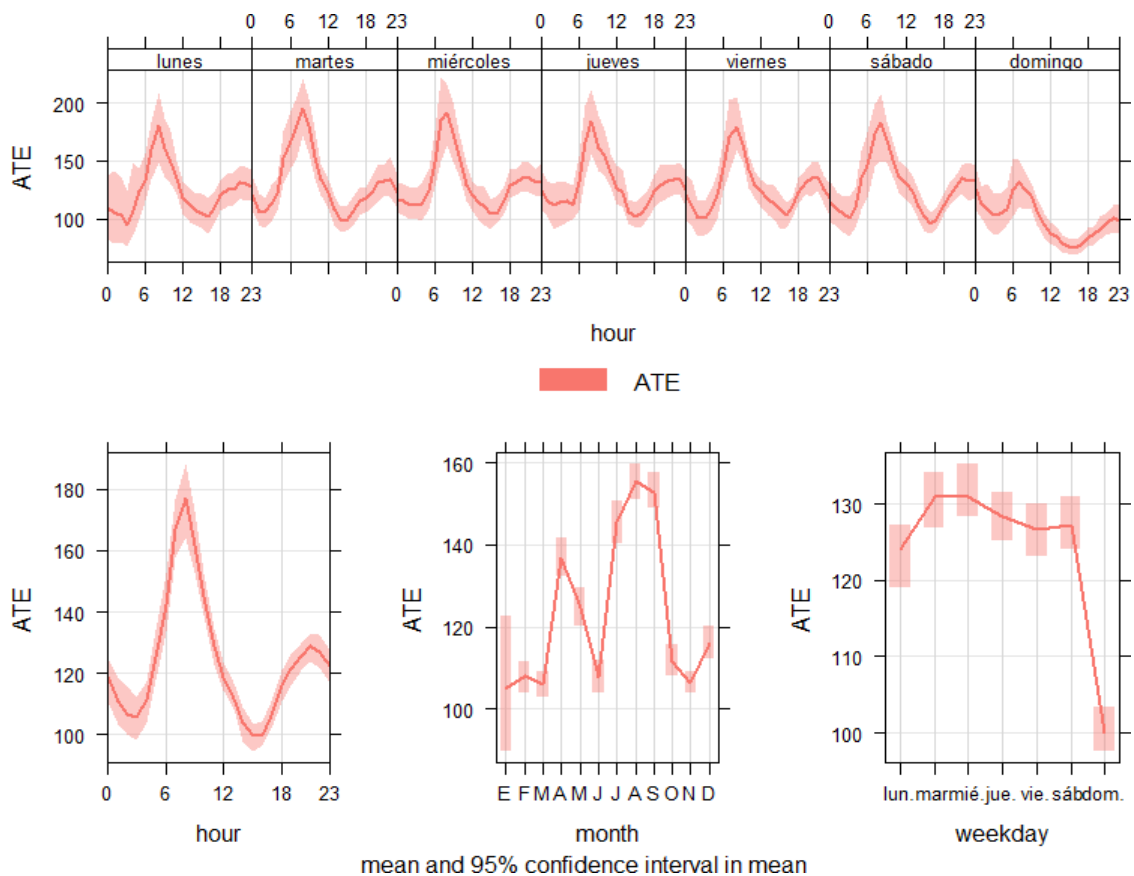
Fuerza	Tipo	velocidad (m/s)	Característica
0	Calma	0-0.2	Elevación de humo verticalmente
1	Brisa suave	0.3 - 1.6	Indica la dirección a donde se dirige el humo
3	Viento suave	1.9 - 3.3	Se siente el viento en la cara, mueve las ropas en las zonas urbanas.
4	Viento leve	3.6 - 5	Mueve constantemente las hojas.
5	Viento moderado	5.3 - 7.2	Levanta polvo, hojas y papeles.
6	Viento fuerte	7.5 - 9.8	Agita en menor proporción arbustos y árboles.
7	Viento muy fuerte	10 - 12.3	Mueve arboles pequeños.
8	Temporal	12 - 14.8	Puede desestabilizar instalaciones telegráficas y sacarlas de su lugar.
9	Temporal fuerte	15 - 18.2	Mueve arboles por completo; se hace difícil andar.
10	Temporal muy fuerte	18.6 - 21	Causa averías leves en las estructuras, y arranca los tejados de los techos.
11	Tempestad	21.6 - 25	Ocasiona grandes estragos.
12	Huracán	25 - a más	Vientos causantes de desastres.

Fuente: Escala Beaufort

De modo que, para el distrito de Ate Vitarte tenemos un tipo de viento denominado *calma* lo cual permite percibir la constante calma de flujo de vientos en la zona. Así mismo para el distrito de Jesús María que presenta *Viento suave* lo cual permite visualizar el flujo de vientos que tienen potencial de mover levemente los objetos pequeños.

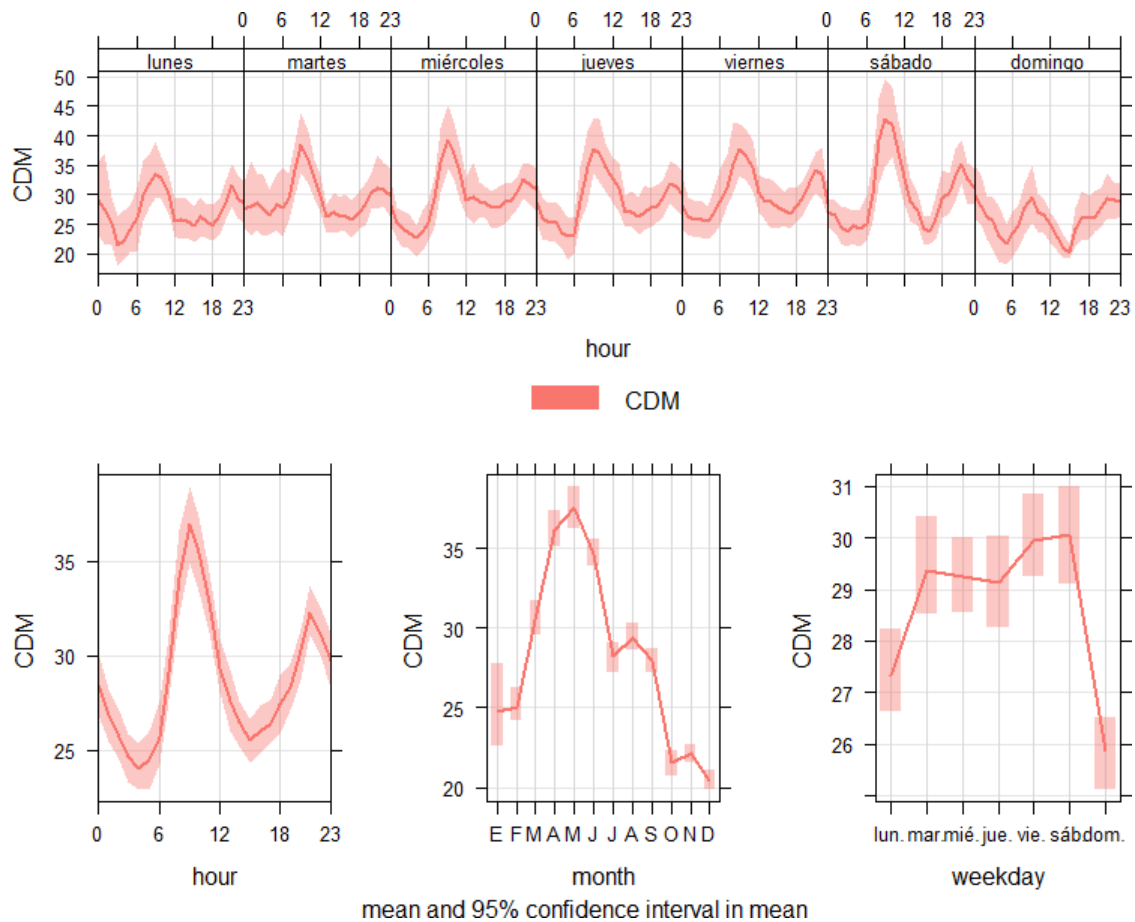
3.6. Concentración de Material Particulado (PM₁₀)

La implementación de la Red Automática de Monitoreo de Calidad del Aire (RAMCA) ha permitido que se pueda vigilar la concentración del contaminante PM₁₀ en el Área de Lima Metropolitana y Callao (AMLC). Además, obtener la tendencia del mismo y su relación con el tiempo y el clima locales. Los distritos que se presenta en el estudio, comprenden un comportamiento de contaminantes primarios distinto para cada uno, esto es, por la influencia de la topografía y la meteorología y la flota vehicular. A todo ello se puede realizar un estudio bajo la metodología del factor de enriquecimiento para ver el origen específico del contaminante en mención. En las figuras 38 y 39 se visualiza el comportamiento de la concentración del Material Particulado PM₁₀ para las zonas de Plaza municipal Ate y Campo de Marte respectivamente.



Fuente: Elaboración propia.

Figura.38 Comportamientos del Material Particulado PM₁₀ en Ate Vitarte (Plaza municipal), 2018.



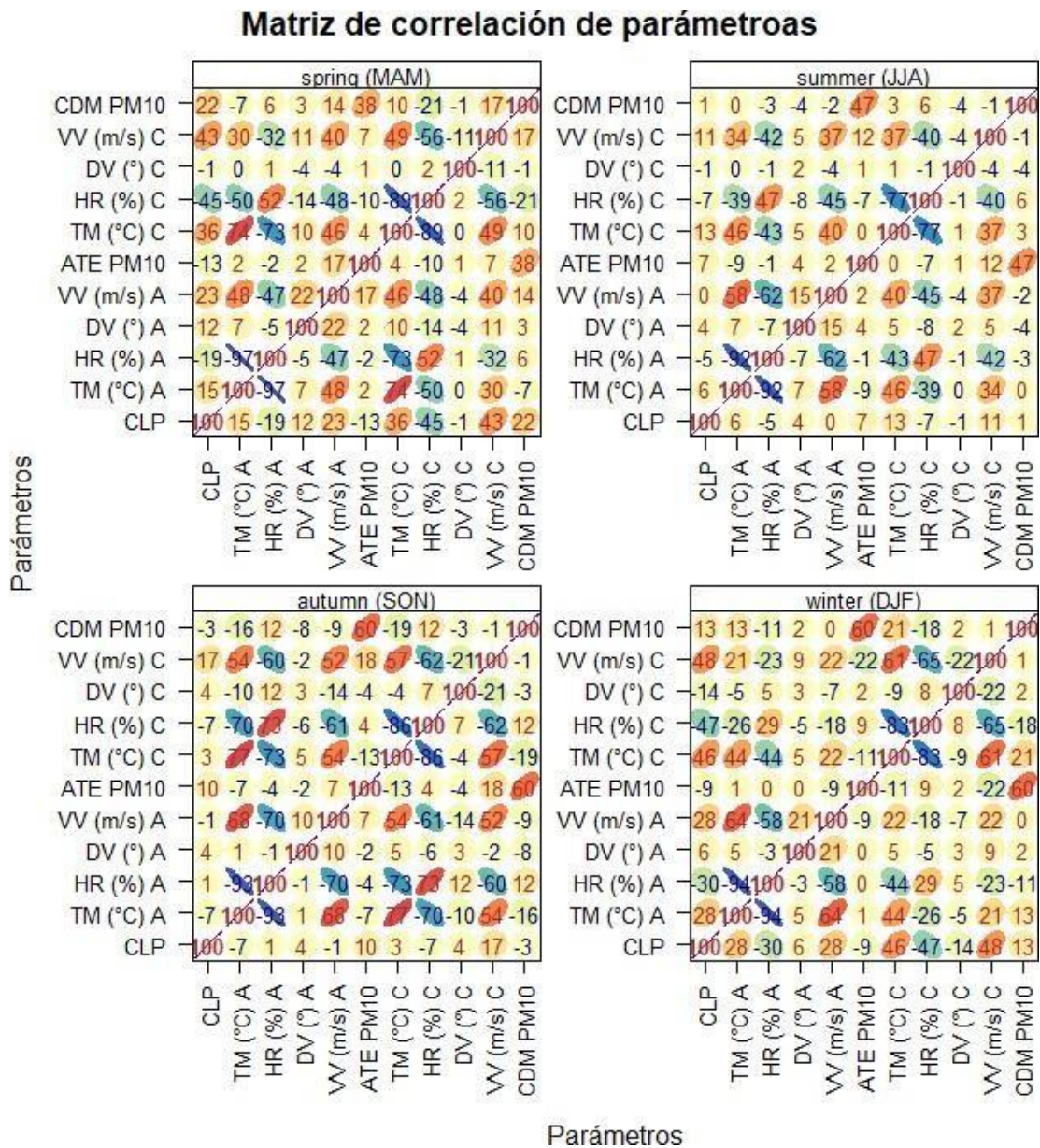
Fuente: Elaboración propia.

Figura.39 Comportamientos del Material Particulado PM₁₀ en Jesús María (Campo de marte), 2018.

El comportamiento del PM₁₀ en la estación de Ate muestran las concentraciones más altas para los meses de julio, agosto y setiembre, las más bajas en los meses de enero, febrero y marzo. Así mismo en la estación de Campo de Marte muestra las concentraciones más altas para los meses de abril, mayo y junio; con las más bajas en los meses de noviembre, diciembre, enero y febrero. La concentración en promedio para la estación de Ate es de 156 ug/m³ y para la estación de Campo de marte, 38 ug/m³. Esto pone en evidencia que para el distrito de Ate la concentración supera el ECA establecido de la normativa peruana. Por otro lado, se puede visualizar durante el día, en las horas *am*, la concentración aumenta por la dinámica antropogénica en los distritos.

3.7. Correlación de variables meteorológicas vs PM₁₀

El análisis de la correlación de dispersión de los datos meteorológicos vs PM₁₀ se realiza a través de una función de la librería *Openaire* llamada *CorPlot* aplicado en el software R. Este utiliza el coeficiente centesimal de correlación de Pearson, graficando así una matriz de doble entrada para los datos establecidos por cada estación o periodo del año. En la figura 40 se muestra la correlación de las variables meteorológicas que precisan los factores de formación del Material Particulado; con ello permite identificar cuál de las variables influye más en la formación del contaminante mencionado.



Fuente: Elaboración propia.

Figura.40 Correlación de variables meteorológicas vs PM₁₀.

Los números, la elipse y el color presentados en la gráfica son los coeficientes de relación de Pearson en escala centesimal, la forma y la intensidad de relación que se tienen entre las variables respectivamente. De modo que si mayor es la cercanía al valor 100, este tiene una relación directamente proporcional perfecta. Es así que en los cuatro periodos del año las variables no muestran una mayor intensidad en la formación de PM₁₀.

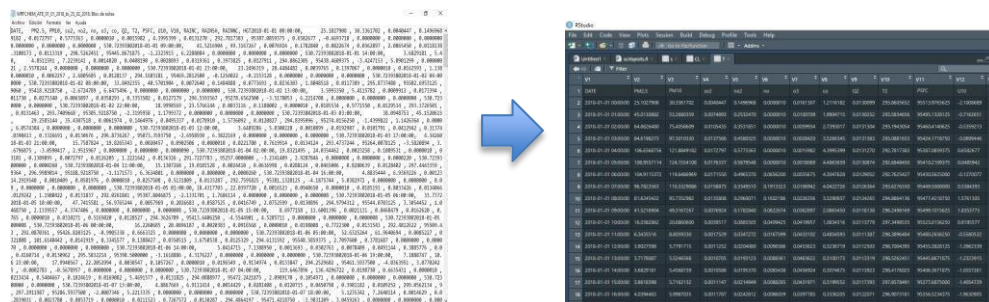
3.8. Base de datos para la entrada de la RNA

Los datos utilizados en el presente estudio fueron proporcionados por la Red Automática de Monitoreo de Calidad del Aire (Tabla N° 4) y del Software WRF-Chem encargado del modelamiento atmosférico para el pronóstico de las condiciones meteorológicas y de contaminantes del aire. La información de salida del WRF está contenida en formato .txt los cuales pueden ser visualizados en bloc de notas o Microsoft Excel. Los archivos contienen información horaria evaluada de todo el año 2018. Por otro lado, debido a la diversidad de factores que intervienen en la medición de los contaminantes del aire, es menester analizar y revisar los datos que se presentan. La figura 41 muestra las variables de salida del WRF que se procesan para los valores de entrada de la RNA; el proceso consiste en poder ordenar los datos de forma que se obtengan las columnas y filas (matriz).

Tabla N°5 Ubicación geográfica de las estaciones de Calidad del Aire-Senamhi

N°	ESTACIÓN	Coordenadas UTM(m)			Lugar
		Este	Norte	Altitud	
1	ATE	291 097	8 669 823	358	Plaza municipal de Ate
2	JESÚS MARÍA	277 601	8 664 893	159	Campo de Marte

Fuente: Senamhi, 2018

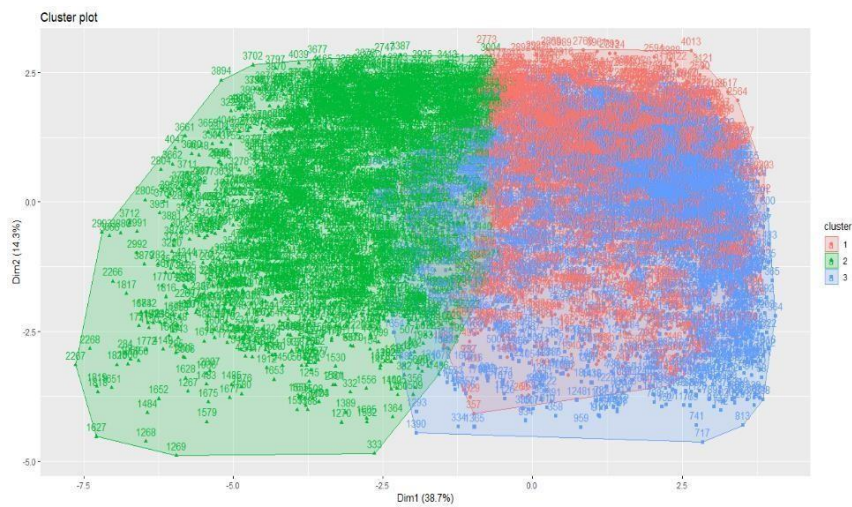


Fuente: WRF-Chem, Senamhi.

Figura.41 Variables de salida del WRF-Chem, 2018.

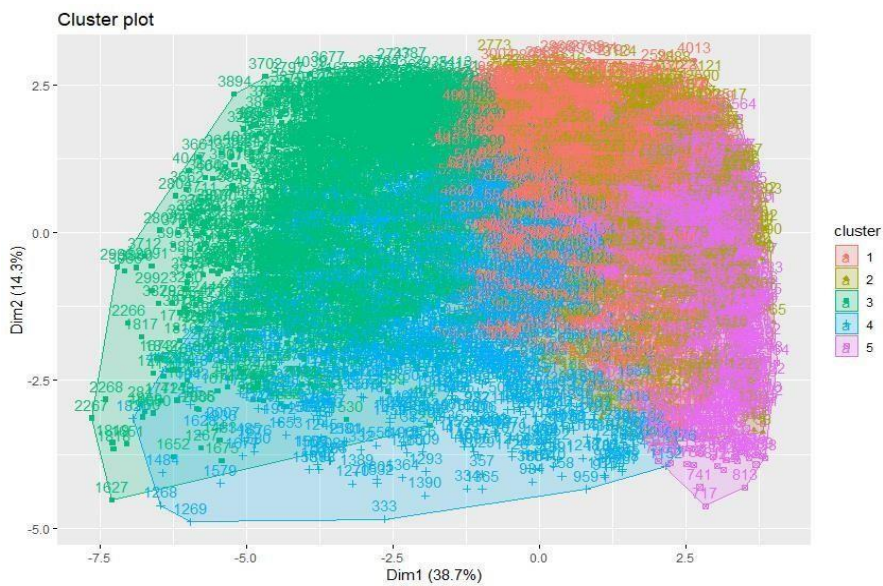
3.9. Agrupamiento de los datos para la RNA

Como se mencionó en la metodología que las técnicas de agrupamiento garantizan la relación adecuada de las variables y la formación de patrones que se necesiten de entrada a la RNA. La lógica difusa permite caracterizar relaciones funcionales no-lineales y sirven para ver los valores en series temporales. De este modo se procedió a ejecutar las funciones del script y aplicar la técnica mencionada. En la figura 42 y 43, se muestran el agrupamiento de las variables da mayor afinidad o similitud que respecta del orden de datos, es decir, partición de los grupos o cluster que identifican relación directa o indirecta en la formación de cualquiera de ellos.



Fuente: Elaboración propia.

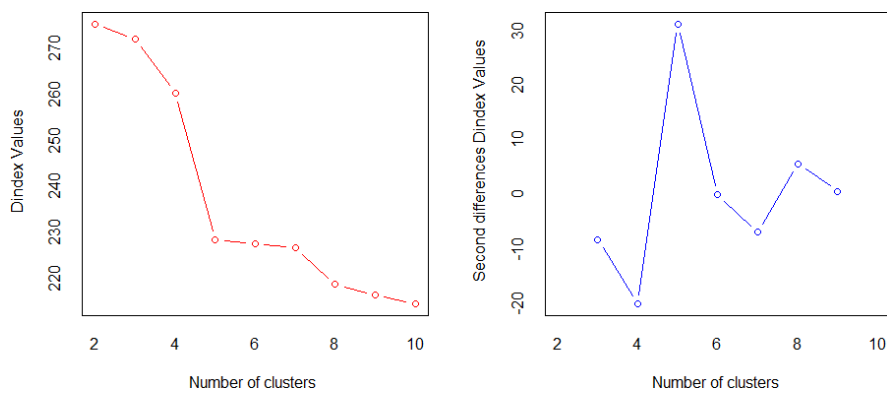
Figura.42 Agrupamiento de las variables de entrada para la RNA (k=3).



Fuente: Elaboración propia.

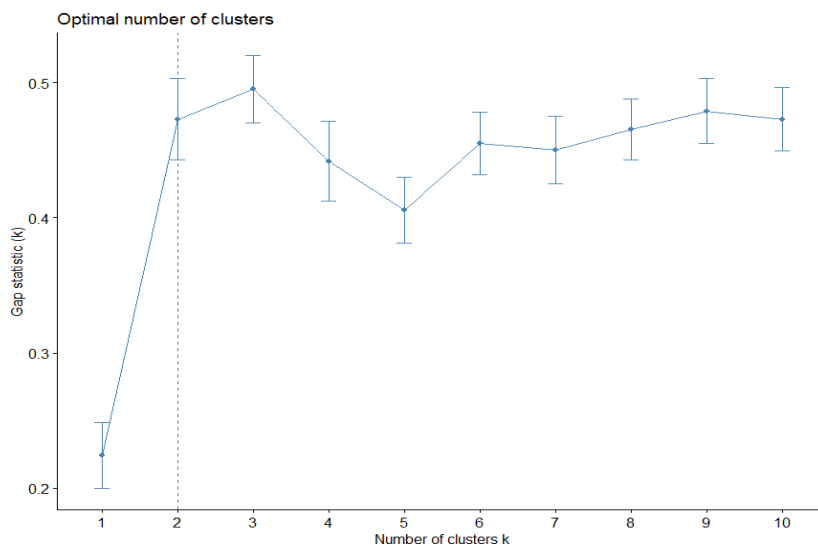
Figura.43 Agrupamiento de las variables de entrada para la RNA (k=5).

Así mismo realizar la partición o el agrupamiento adecuado de datos puede ser complicado puesto que los grupos que se pueden formar no representen lo que se desea determinar. Por ello la misma prueba estadística, Kmeans, ofrece realizar estas particiones y mostrara el numero óptimo de clúster con el que se puede identificar la relación entre variables; para ello se vale de las iteraciones que hace la técnica. Un número máximo de iteraciones propuesta para este trabajo es de 100 iteraciones por dato y el resultado se muestra en las siguientes figuras con diferentes paquetes de prueba en el software Rstudio.



Fuente: Elaboración propia.

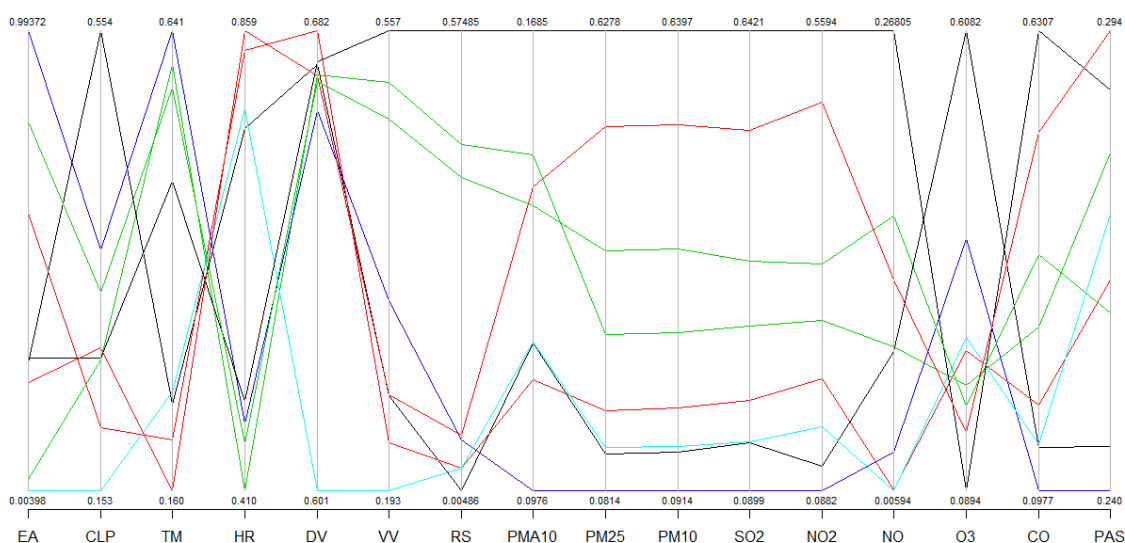
Figura.44 Número óptimo de Clúster con el paquete NbClust para la RNA.



Fuente: Elaboración propia.

Figura.45 Número óptimo de Clúster con el paquete factoextra para la RNA.

Importante es realizar la partición y esta sea notoria en el grupo de datos. En la figura 42 se puede ver bien definido los grupos y es corroborado con las pruebas del número óptimo de clústers que se visualizan en las figuras 44 y 45. Estas muestran el número óptimo de clúster (agrupamiento) para los datos registrados de una serie de tiempo ambientales. Se visualiza, entonces, la cantidad de bloques con los cuales se conforman los grupos es de tercer orden (número de grupos: 3). Para este efecto, se repitió el proceso varias para cada diferente clúster hasta ver si se mantenía constante el agrupamiento, el resultado es constante dado la cantidad de clúster del programa mostrando la siguiente figura.



Fuente: Elaboración propia.

Figura.46 Segmentos de datos analizados del agrupamiento para la RNA.

Cada línea muestra el segmento de grupos de clúster de las variables. En la figura 46 muestra los segmentos que tiene mayor afinidad en un particionamiento de 8 clúster. Mayormente se agrupan con la variable dependiente del presente estudio con la temperatura (TM), capa límite planetaria (CLP), velocidad de vientos (VV), dirección del viento (DV), material particulado (PM_{10}), material particulado ($PM_{2.5}$), material particulado (PM_{10}), dióxido de azufre (SO_2), monóxido de nitrógeno (NO), monóxido de carbono (CO) y la presión atmosférica (PAS). Para ello se puede confirmar los patrones de entrada a la RNA, para Ate y Jesús María.

$$P_{ATE}(PM_{10}) = [PMA_{10}, CPM_{2.5}, PM_{10}, VVA, DVA, TMA, CO, NO, SO_2, CLP, PAS]$$

$$P_{CMD}(PM_{10}) = [PMC_{10}, VV, DV, TM, CLP, PAS]$$

3.10. Normalización de los datos

Los datos registrados de la RAMCA y del Software WRF son normalizados mediante la siguiente ecuación:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

X' es el valor normalizado, X_{máx} y X_{mín} valores máximos y mínimos que son tomados de la base de datos. Cabe resaltar que los valores máximos son los valores permitidos para las variables a criterio de que no se registre valores muy altos. En la tabla 5 se visualiza los valores mínimos y máximos mencionados.

Tabla N°6 Valores mínimos y máximos permitidos

Variable	V. Min.	V. Máx.
PM ₁₀	0	400 ug/m ³
Velocidad del viento	0	20 m/s
Dirección del viento	0	360 °
Temperatura	0	50 °C
Humedad relativa	0	100%

Realizar el escalonamiento o normalización de los datos es conveniente puesto que permite realizar un adecuado análisis de datos y los pesos de que puede tener cada variable no sea imperante sobre otra por tener mayor número o cantidad dentro de los vectores de entrada a la red. Es así que en la figura 47 se observa los datos normalizados en los rangos propuestos.

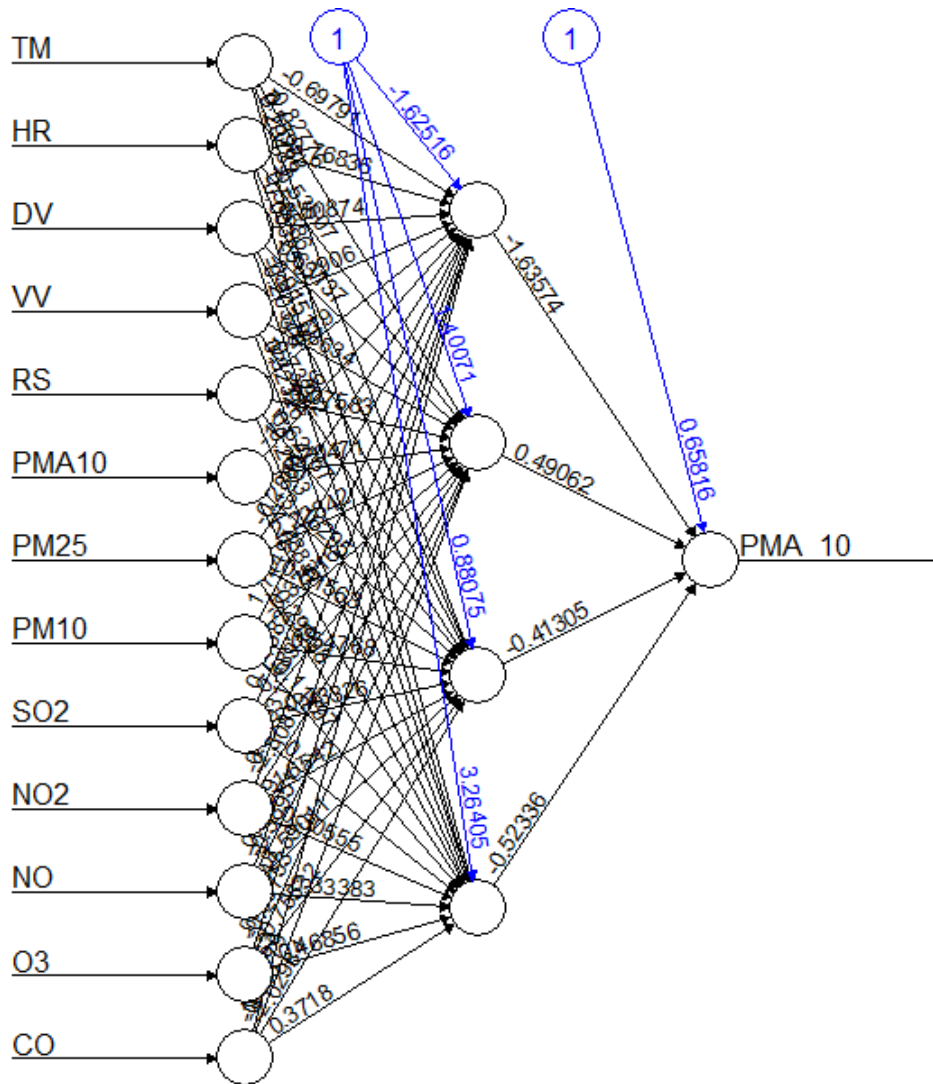
Showing 1 to 23 of 5,826 entries, 20 total columns

Fuente: Elaboración propia.

Figura.47 Normalización de los datos.

3.11. Arquitectura de la red neuronal artificial para el modelo de estimación del PM₁₀

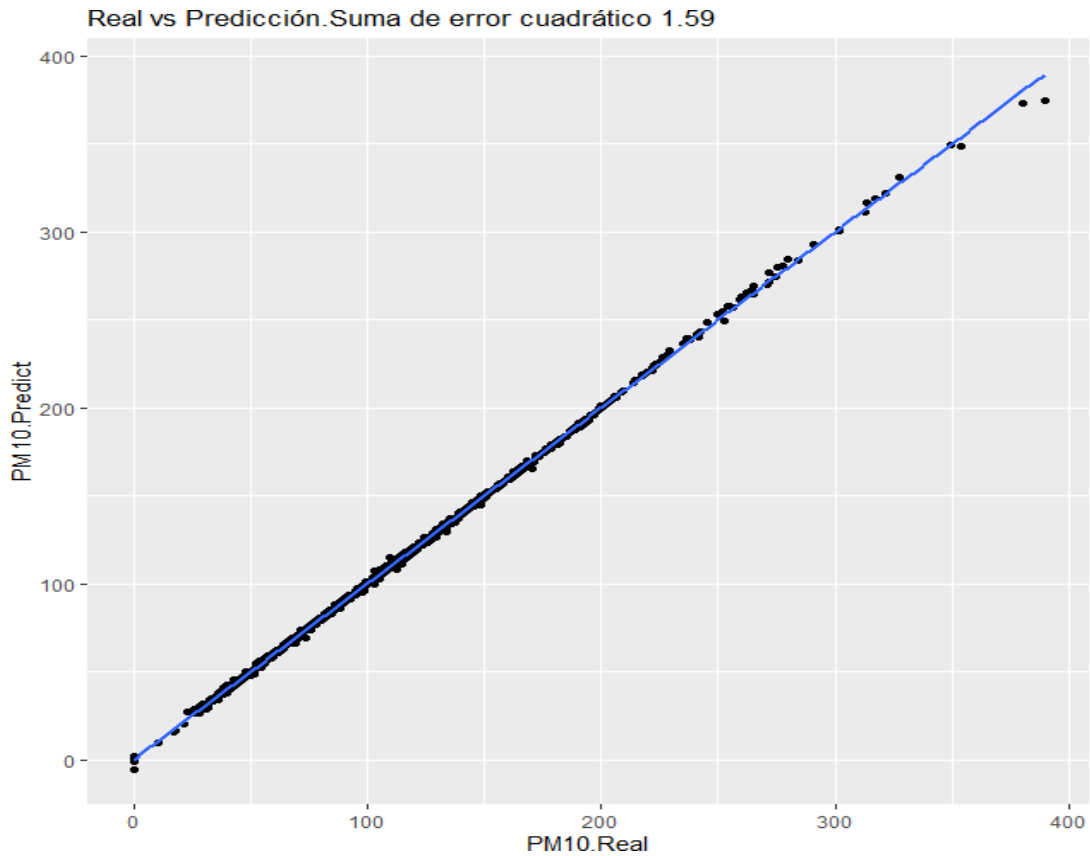
Para la arquitectura de la RNA es importante realizar varias pruebas que permitan identificar el número de neuronas en la capa oculta, esto es para alcanzar los valores estimados más próximos a los reales.



Fuente: Elaboración propia.

Figura.48 Arquitectura de la RNA con estructura óptima (13-4-1) para el distrito de Ate Vitarte.

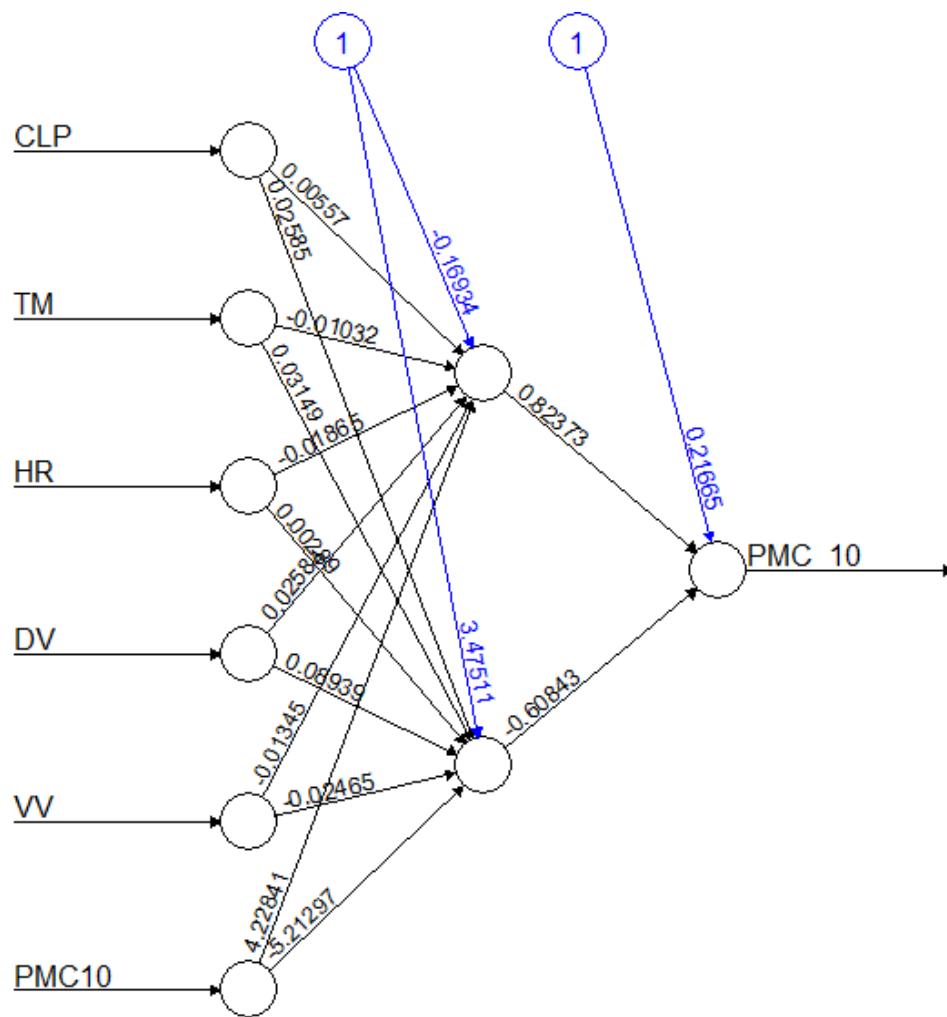
En la figura 48 se visualiza la arquitectura óptima con la cual se alcanzó el ajuste del modelo para el pronóstico del objetivo del estudio.



Fuente: Elaboración propia.

Figura.49 Similitud de los valores Reales vs los pronosticados para el distrito de Ate Vitarte.

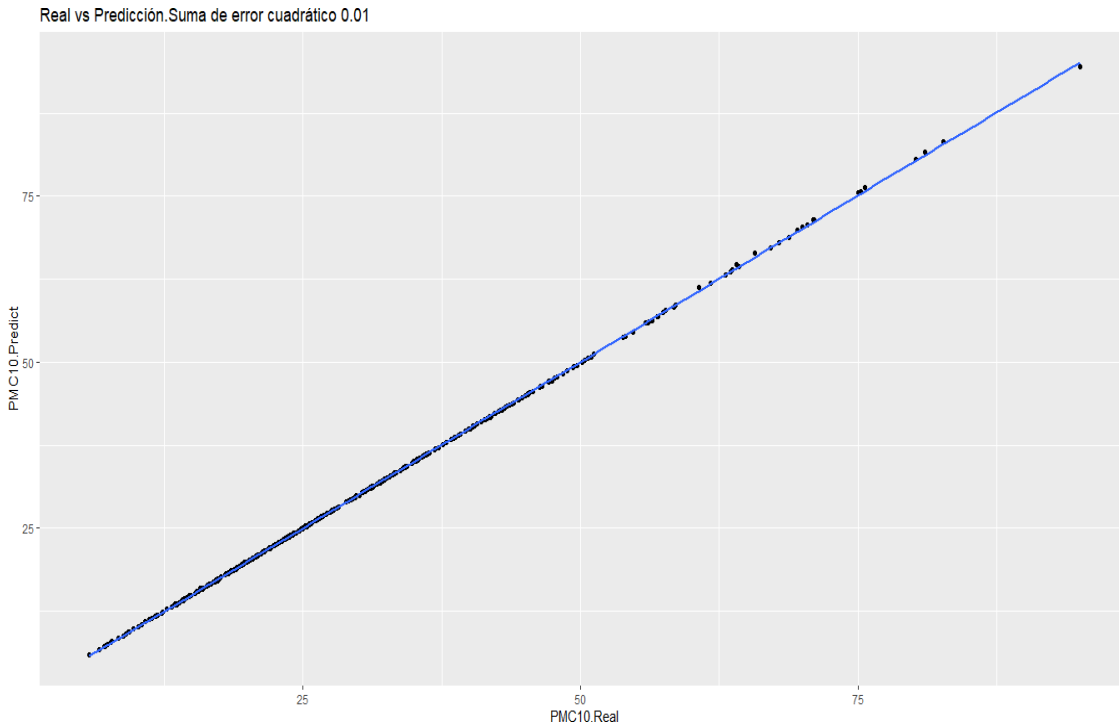
En la figura 49 se visualiza la precisión y similitud de valores entre el pronóstico y los valores reales aplicado por el algoritmo de retropropagación. Así mismo, gráficamente se puede visualizar el error que se tiene en la precisión de los valores altos. Así mismo, el Error Cuadrático Medio (MSE) tiene un valor de 1.59. Siendo este un error muy alto para la precisión de datos en general.



Fuente: Elaboración propia

Figura.50 Arquitectura de la RNA con estructura óptima (7-2-1) para el distrito de Jesús María.

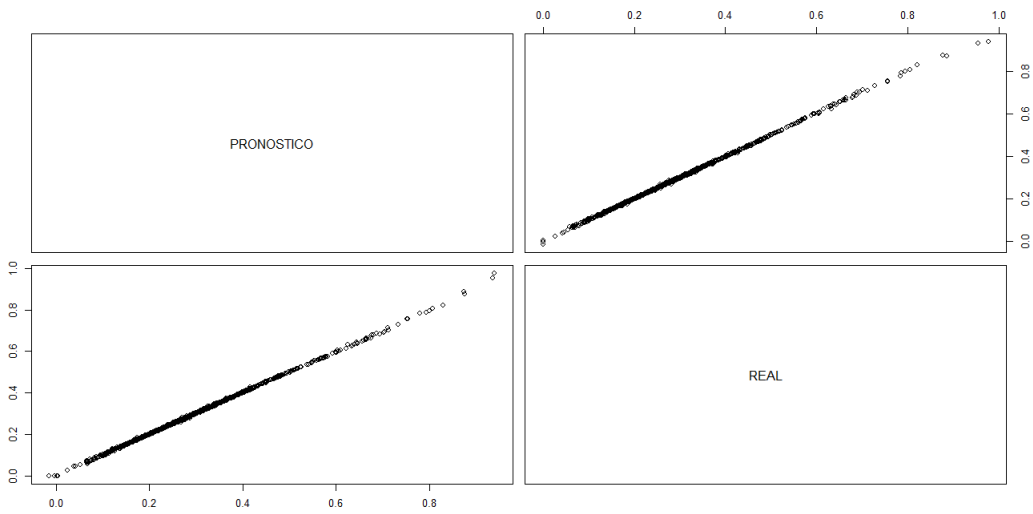
En la figura 50 se visualiza la arquitectura óptima con la cual se alcanzó el ajuste del modelo para el pronóstico del objetivo del estudio.



Fuente: Elaboración propia.

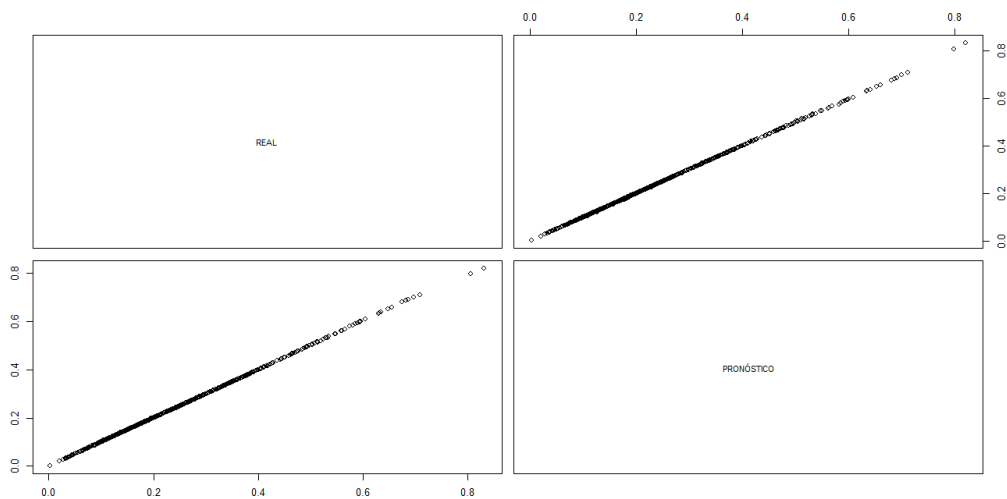
Figura 51. Similitud de los valores Reales vs los pronosticados para el distrito de Jesús María.

En la figura 51 se visualiza la precisión y similitud de valores entre el pronóstico y los valores reales aplicado por el algoritmo de retropropagación. Así mismo, gráficamente se puede visualizar el error que se tiene en la precisión de los valores altos. Por otro lado, el Error Cuadrático Medio (MSE) tiene un valor de 0.01. Siendo este un error aceptable para el pronóstico del contaminante del objetivo de estudio.



Fuente: Elaboración propia.

Figura.52 Correlación de los datos Reales vs Pronosticados del distrito de Ate Vitarte



Fuente: Elaboración propia.

Figura.53 Correlación de los datos Reales vs Pronosticados del distrito Jesús María.

La figura 52 y 53 muestran el grado de bondad del modelo que se estima en cuanto a la correlación de Pearson. El modelo propuesto tiene un valor de 0.7 para el distrito de Ate Vitarte y 0.9 para el distrito de Jesús María de correlación, lo cual indica una relación alta a los valores estimados.

Tabla N°7 Arquitectura de la R N A

Arquitectura de la Red Neuronal	
Tipo de red	Perceptrón Multicapa- Retropropagación
Tipo de Conexión Recurrente	
Capa de Entrada	Patrón de 14 y 7 entradas *
Capa oculta	En prueba y análisis de datos
Capa de Salida	Una neurona
Función de activación	Sigmoidea-capa oculta Lineal-cap de salida

*De Ate vitarte y Jesús María respectivamente.

Tabla N°8 Arquitectura óptima para la R N A

Zona	Prueba Óptima	Número de capas	Número de neuronas en la capa oculta	MSE	R
ATE	20	1	4	1.59	0.7
CDM	10	1	2	0.01	0.9

Las tablas N° 7 y N°8 muestran el resumen de la arquitectura óptima para y el error obtenido para cada proceso del modelo en relación a la predicción de los valores esperados.

IV. DISCUSIÓN

4.1. Preparación de datos de entrada a la RNA

Para el preprocesamiento de los datos según Medina (2016), la data debe dividirse en tres grupos distintos (Entrenamiento-test-validez) para poder entrenar a la red con la mayor precisión posible. Esto es coherente con los criterios de Memarianfard (2017), Rosales, *et. al* (2018) y Salazar (2019), debido a que se maneja grandes cantidades de datos. Sin embargo, también se puede agrupar los datos en dos grandes grupos (ya sea en [70-30]% u [80-20]%) debido a la similitud y homogeneidad de los valores de la data. Por lo que para el presente estudio resultó ser más eficiente utilizar los criterios de Cortina *et.al* (2015) en los vectores de entrada a la RNA (entrenamiento y validación).

4.2. Arquitectura óptima de la RNA para la estimación de la concentración de PM₁₀

Para el entrenamiento de la RNA lo sólito es tener mayor número de neuronas en la capa oculta que la cantidad de vectores o patrones de entrada, como se registra en los trabajos de Capilla (2015) y Medina (2016) 30 y 18 neuronas en la capa oculta respectivamente. No obstante, Fadish (2015), Feng (2015), Memarianfard (2017) y Paas (2017) utilizan menor número de neuronas en la capa oculta que el patrón de entradas para la red [27-8-6-5] neuronas, respectivamente. De modo que no es incongruente tener número de neuronas menores que el patrón de entradas, lo cual coincidió con los resultados obtenidos del presente estudio. Se coincidió también en la capa oculta igual a una en todos los trabajos mencionados. A saber, este tipo de pruebas del diseño de una RNA no es definido mediante una propia metodología, pues para cada análisis, se experimenta con las distintas arquitecturas variando la cantidad de neuronas en sus capas ocultas.

En la Tabla N° 7 se visualiza los resultados de la arquitectura óptima para el modelo esperado, así mismo, se coinciden con los trabajos anteriores de los antecedentes presentados en este estudio, en el nuerro de capas ocultas, más no en el número de neuronas dentro de la capa oculta. Se observa también, que el error cuadrado medio (MSE) es de 1.59 y 0.01 en los distritos de Ate y Jesús María. Esto es porque no se siguió la metodología de Medina (2016) que permite ver cuándo debe parar el algoritmo detectando el error generalizado (Técnica early-stopping); pues de lo contrario se tendría que realizar varias pruebas hasta encontrar un camino de algoritmo adecuado.

En cuanto al grado de bondad del modelo diseñado tiene una precisión de 0.9 para el distrito de Jesús María y 0.7 para el distrito de Ate Vitarte. Las funciones de activación utilizadas en el algoritmo de retropropagación es la sigmoideal de entrada a la capa oculta y lineal de salida de la capa, concordando con el trabajo realizado por Rojas & Medina (2016). Por consiguiente, el modelo diseñado denota una estructura única para la cantidad de datos y patrones de entradas. A comparación de los distintos trabajos realizados anteriormente, no se coincidió tanto en el número de neuronas en la capa oculta cuanto en el error cuadrático medio, debido al preprocesamiento de datos y como estos ingresan a la RNA.

V. CONCLUSIONES

1. La red diseñada es un perceptrón multicapa con una capa oculta, de 4 y 2 neuronas para los datos de Jesús María y Ate vitarte. Siendo el error más elevado para este último distrito. Esto se debe a la cantidad de datos y a la homogeneidad de las variables. A pesar de que el proceso de clúster es eficiente en cuanto a segmentar los grupos de instancia, se tiene que ver la correlación de las variables que sean altas. La medida en que el modelo puede resolver el problema del pronóstico del día siguiente en los distritos de Jesús María y Ate Vitarte están sujetos a diferentes análisis con el uso de técnicas para poder ajustar el preprocesamiento de los datos. De modo que:

- ✓ El modelo estocástico presentado para estimar la concentración de Material Particulado (PM_{10}) en el distrito de Jesús María es eficiente porque responde de manera rápida los resultados esperados.
- ✓ En cambio, el modelo estocástico presentado bajo un algoritmo diferente al utilizado para las variables del distrito de Jesús María, no permitió estimar la concentración de Material Particulado (PM_{10}) en el distrito de Ate Vitarte debido a el preprocesamiento de los datos en la metodología utilizada en el presente estudio.

2. La técnica de agrupamiento de datos, kmeans, es eficiente en cuanto a los procesos de partición, pues agrupa a las variables dándoles afinidad entre ellos. Por otro lado, ayuda también a verificar cuál es el número óptimo de grupos con los que se hace la partición y con ello trabajar.

3. Así mismo el algoritmo de aprendizaje de retropropagación en el paquete *neuralnet* de Rstudio es eficiente en cuanto a su programación debido a que se tiene establecido el argumento: "*rprop+*" que refiere a procesar el algoritmo con retroceso de peso y así asignar el error de manera automática.

VI. RECOMENDACIONES

- Generar un script que permita considerar la Técnica early-stopping para poder parar el procesamiento de corridas hasta llegar al error más bajo.
- Analizar y preparar adecuadamente los datos debido a que pueden existir picos altos que no permitan integrar mejor la distancia entre ellos. Sobre todo, si se realiza un método de agrupamiento como en el presente estudio.
- El software Rstudio permite realizar los procesos de backpropagation de manera eficaz, sin embargo, requiere de cdear o codificar nuevo script para el fin que se requiere. Para ello se puede sugerir ver la compilación de códigos que otros, estudios de la ciencia de datos, han realizado siguiendo la lógica de su propuesta.
- No solo la técnica *Kmeans* puede ayudar ver la afinidad de datos con los que se puede trabajar, las técnicas como *Fuzzy-kmeans* o metodos como *falsos vecinos* pueden aportar para el preprocesamiento de datos.
- Seguir realizando estudios de PM_{10} en los distritos de Jesús María y Ate Vitarte, para poder evaluar el comportamiento del contaminante de la zona.
- Implementar una guía de reanálisis de datos para que estos no se pierdan y puedan servir en la construcción de un buen pronóstico a las personas.

REFERENCIAS

- ACEVEDO, Edgar; SERNA, Aelxia; SERNA, Edgar. Principles and characteristics of artificial neural networks. Actas de ingeniería [en línea]. 2017. [Fecha de consulta: 23 de mayo de 2019]. Disponible en: <http://fundacioniai.org/actas/Actas3/Actas3.40.pdf> ISSN: ??
- AKBAR, Rahimi. Short-term prediction of NO₂ and NO_x concentrations using multilayer perceptron neural network: a case study of Tabriz, Iran. Springer Open. [en línea]. 2017. [Fecha de consulta: 23 de mayo de 2019]. Disponible en: <https://link.springer.com/article/10.1186/s13717-016-0069-x> ISSN: ??
- ALCALÁ, Tomás. Introducción a los modelos de redes neuronales artificiales: El Perceptron simple y multicapa. Universidad Zaragoza. Zaragoza: España, 2018.
- ANDRADES, Marisol & MUÑOZ, Carmen. Fundamentos de Climatología. Universidad de la Rioja, 2012. 64 pp. ISBN: 9788469527993
- ARMAS, Carlos & ARMAS, Carlos E. TECNOLOGÍA AMBIENTAL: En nuestro hogar la nave sideral. Trujillo: CONCYTEC, 2002. 690 pp. ISBN: 997294374
- ARRIETA, Alvaro. Dispersión de Material Particulado (PM₁₀), con interrelación de Factores Meteorológicos y Topográficos. Dialnet [en línea]. mayo 2016. [Fecha de consulta: 27 de abril de 2019]. Disponible en: <https://dialnet.unirioja.es/descarga/articulo/6096112.pdf> ISSN: 1900-771X
- CAPILLA, Carmen. Artificial Neural Network Approach for Forecasting Nitrogen Oxides Concentrations. Environmental engineering science [en línea]. Mayo 2015. [Fecha de consulta: 27 de mayo de 2019]. Disponible en https://www.researchgate.net/publication/281266520_Artificial_Neural_Network_Approach_for_Forecasting_Nitrogen_Oxides_Concentrations
- CONESA, Vicente. Guía Metodológica para la Evaluación del Impacto Ambiental. Madrid: MP, 2010. 864 pp. ISBN: 9788484763840
- CONTRERAS, Alfonso & MOLERO, Mariano. Ciencia y Tecnología del Medioambiente. MADRID: UNED, 2014. 384 pp. ISBN: 9788436252965
- CORTINA, Januchs; DOMINGUEZ, Joel; CORONA, Antonio; ANDINO, Diego. Development of a model for forecasting of PM₁₀ concentrations in Salamanca, Mexico. Environmental engineering science [en línea]. Junio 2015. [Fecha de consulta: 27 de mayo de 2019]. Disponible en https://www.researchgate.net/publication/281266520_Artificial_Neural_Network_Approach_for_Forecasting_Nitrogen_Oxides_Concentrations ISSN: ??
- CORTINA, Januchs. Aplicación de técnicas de inteligencia artificial a la predicción de contaminantes atmosféricos. universidad politécnica de Madrid: España, 2012.
- CUADRAT, José & PITA, Fernanda. CLIMATOLOGÍA. Madrid: Juan Ignacio Luca de Tena, 2014. 750 pp. ISBN: 9788437615318

-CHAVEZ, José. Evaluación del modelo brams/teb de calidad del aire y estudio de la circulación local en la ciudad de lima en los procesos de dispersión de contaminantes. Universidad Nacional Agraria la Molina. Lima: Perú, 2011.

-CHAPRA, Steven; CANALE, Raymond. Métodos numéricos para ingenieros. 5^{ta} ed. Madrid: The McGraw-Hill Companies, 2007. 30 pp. ISBN: 9789701061145

-DELAVAR, Mahmoud; GHOLAMI, Amin; SHIRAN, Gholam; RASHIDI, Yousef; NAKHAEIZADEH, Gholam; FEDRA, Kurt; AFSHAR, Smaeil. A Novel Method for Improving Air Pollution Prediction Based on Machine Learning Approaches: A Case Study Applied to the Capital City of Tehran. Isprs. Enero 2019 [Fecha de consulta: 22 de mayo de 2019]. Disponible en: <file:///C:/Users/pc1/Desktop/REDES%20NEURONALES%20PARA%20EL%20PRON%20C3%93STICO%20DE%20CALIADAD%20DEL%20AIRE/antedecentes/pronostico%20pm10.pdf>. ISBN: ??

-DURAN, Susana; CÓRDOVA, Teodoro. Simulación de una red neuronal con mapas auto-organizados (som) de kohonen. Rev. Jóvenes en la ciencia. Enero 2017 [Fecha de consulta: 22 de mayo de 2019]. Disponible en: <http://www.jovenesenlaciencia.ugto.mx/index.php/jovenesenlaciencia/article/view/2094/1589>

-FASIH, Zeinab; MESDAGHINIA, Alireza; NADAFI, Kazem; NODEHI, Ramin; MAHVI, Amir; HADI, Mahdi. Forecasting the air quality index based on meteorological variables and autocorrelation terms using artificial neural network. Razi Journal of Medical Sciences. [en línea]. November 2015 [Fecha de consulta: 22 de mayo de 2019]. Disponible en: <http://rjms.iums.ac.ir/article-1-4073-en.html> ISBN: ??

-FEBRE, Cristian. Estudio de la dinámica y patrones de circulación media del jet costero de bajos niveles en la costa central y sur del Perú. Universidad Nacional Agraria la Molina. Lima: Perú, 2019.

-FENG, Xiao, LI, Qi; ZHU, Yajie; HOU, Junxiong; JIN, Lingyan; WANG, Jingjie. Artificial neural networks forecasting of PM_{2.5} pollution using air mass trajectory based geographic model and wavelet transformation. ELSEVIER. November 2015 [Fecha de consulta: 22 de mayo de 2019]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1352231015001491> ISBN: ??

-FRANCO, Juan; ROJAS, Nestor; SARMIENTO, Olga; HERNANDEZ, Luis; ZAPATA, Elizabeht; MALDONADO, Andrea; MATIZ, Leonardo; BEHRENT, Eduardo. Niveles de material particulado en colegios distritales ubicados en vías con alto tráfico vehicular en la ciudad de Bogotá: estudio piloto. Rev. Fac. Ing. Univ. Antioquia. [en línea]. September 2009 [Fecha de consulta: 22 de abril de 2019]. Disponible en: <http://aprendeonline.udea.edu.co/revistas/index.php/ingenieria/article/view/15961/13833>. ISBN: ??

-FUSTER, Joaquín. Cerebro y Libertad: Los cimientos cerebrales de nuestra capacidad para elegir. 2014. Planeta, Barcelona: España. 99 pp. ISBN: 9788408153818

- GAERTNER, Miguel. Aplicación de un modelo numérico de predicción meteorológica a la simulación de flujos atmosféricos a mesoescala en la zona centro de la Península Ibérica. Universidad Complutense de Madrid. Madrid: España, 2002.
- GARCIA, Makel. WRF basics. 2012. Universidad de Cantabria, Santander: España.
- HILERA, José; MARTÍNEZ, Víctor. REDES NEURONALES ARTIFICIALES, Fundamentos, modelos y aplicaciones, 1995. RA-MA. Madrid: España.
- HOSSEINABAD, E.R. and MORAGA, R.J. Air Pollution Mitigation in Metropolitans using System Dynamics Approach. IIE Annual Conference.Proceedings, 2017. pp. 638-643 ProQuest Central.
- LAURET, Pierre; HEYMES, Frederic; FORESTIER, Serge; APRIN, Laurent; PEY, Alexis, PERRIN, Marcia. Forecasting powder dispersion in a complex environment using Artificial Neural Networks. ELSEVIER. [en línea]. Febrary 2017 [Fecha de consulta: 22 de mayo de 2019]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0957582017300356> ISBN: ??
- MAPFRE, fundación. Manual de Contaminación Ambiental. Fernández: ITSEMAP AMBIENTAL, 2000. 706 pp. ISBN: 8471008017
- MASTERS, Gilbert & ELA, Wendell. Introducción a la Ingeniería Medioambiental. 3ª ed. Madrid: PEARSON, 2008. 752 pp. ISBN: 9788483224441
- MEDINA, Elvis. Redes Neuronales Artificiales para la predicción de la concentración de PM₁₀ en el distrito de ATE. Universidad Agraria de la Selva. Tingo Maria: Perú. 2016.
- MEMARIAFARD, Hatami. Artificial neural network forecast application for fine particulate matter concentration using meteorological data. Short Comuncation. [en línea] summer 2017 [Fecha de consulta: 29 de mayo de 2019]. Disponible en: https://www.gjesm.net/article_23079.html ISSN: ??
- MÉNESES, Francisco; ALVARADO, Matías. Pronóstico del tipo USD/MXN con redes neuronales de retropropagación. ResearchGate [en línea] May 2017 [Fecha de consulta: 14 de febrero de 2019]. Disponible en: https://www.researchgate.net/publication/323985249_Pronostico_del_tipo_de_cambio_USDMXN_con_redes_neuronales_de_retropropagacion ISSN: 1870-4069
- MÉNDEZ, Juan; PINTO, Laura; GALVIS, Remolina, PACHÓN, Jorge. Estimación de factores de emisión de material particulado resuspendido antes, durante y después de la pavimentación de una vía en Bogotá. Ciencia e Ingeniería Neogranadina [en línea] Enero-Junio 2017 [Fecha de consulta: 14 de febrero de 2019]. Disponible en: <http://www.scielo.org.co/pdf/cein/v27n1/v27n1a03.pdf> ISSN: 0124-8170

-ORGANIZACIÓN Mundial de la Salud & OPS. Contaminación del Aire Ambiental. 10 de noviembre de 2016. Disponible en: https://www.paho.org/hq/index.php?option=com_content&view=article&id=12918%3A2017-ambient-air-pollution&catid=2619%3Aenvironmental-health&Itemid=42246&lang=es

-ORGANIZACIÓN Mundial de la Salud & OPS. Efectos sobre la salud por material particulado. 13 de diciembre de 2015. Disponible en: <https://www.paho.org/annual-report-2015/Espanol.html>

-PAAS, Bastian; STIENEN, Jonas; VORLANDER, Michael; SCHNEIDER, Christoph. Modelling of Urban Near-Road Atmospheric PM Concentrations Using an Artificial Neural Network Approach with Acoustic Data Input. *Environments*. [en línea] marzo 2017 [Fecha de consulta: 18 de febrero de 2019]. Disponible en: https://www.researchgate.net/publication/315663999_Modelling_of_Urban_NearRoad_Atmospheric_PM_Concentrations_Using_an_Artificial_Neural_Network_Approach_with_Acoustic_Data_Input ISSN: ??

-PARKER, Albert. Contaminación del aire por la industria. Londres: Editorial Reverté, 2014. 709 pp. ISBN: 8429174648

-PULIDO, Víctor. Ecología general y del Perú. 2013, UIGV Lima: Perú. 242 pp. ISBN: 9786124050626

-ROJAS, Jhojan & SILVA, José. Modelamiento de series temporales de la temperatura ambiental con redes neuronales artificiales, Tingo María - Huánuco. Setiembre 2016. [Fecha de consulta: 30 de abril de 2019]. Disponible en: https://www.academia.edu/29128388/time_series_modeling_of_environmental_temperature_with_artificial_neural_networks_tingo_maria_huanuco_modelamiento_de_series_temporales_de_la_temperatura_ambiental_con_redes_neu

-SALAZAR, Daniela; JIMÉNEZ, José; ZAPATA, Carmen; CARDONA, Álvaro. Artificial neural network applied for the forecast of critical PM2.5 events in the Aburra Valley. Universidad Nacional de Colombia Sede Medellín. Marzo 2019. [Fecha de consulta: 30 de abril de 2019] Disponible en: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0012-73532019000200347 ISSN: 0012-7353

-SENAMHI, Boletín Mensual de la Calidad del aire en Lima Metropolitana. Vol 19, N°3. (2019). 2 PP. Vol 18, N° 4. (2018). 14 pp.; Vol 20, N° 4. (2015). 13 pp.; Vol 15, N° 4. (2014) 14 pp.; Vol 14, N° 4. (2013). 16 pp.

-SENDIÑA, Irene & PÉREZ, Vicente. Fundamentos de Meteorología. Santiago de Compostela: Universidade Servizo de publicacións e Intercambio Científico, 2006. 194 pp. ISBN: 8497897506458

-SHARMA, A.K., BALIYAN, P. and KUMAR, P. Air Pollution and Public Health: The Challenges for Delhi, India. *Reviews on Environmental Health*, 2018, vol. 33, no. 1. pp. 77-86 ProQuest Central. ISSN 00487554. DOI <http://dx.doi.org/10.1515/revh-2017-0032>.

-SILVA, José; ROJAS, Jhojan; NORABUENA, Magdalena; MOLINA, Carolina; TORO, Richard. Particulate matter levels in a South American megacity: the metropolitan area of Lima-Callao, Peru. Springer. [en línea]. November 2017. [Fecha de consulta: 30 de abril de 2019]. Disponible en:

file:///c:/users/pc1/desktop/redes%20neuronales%20para%20el%20pron%c3%93stico%20de%20calidad%20del%20aire/realidad%20problemativa/j.silva_2017.pdf

- WANG P, Liu Y, QIN Z, ZHANG G . A novel hybrid forecasting model for PM10 and SO2 daily concentrations. Sci Total Environ. [en línea]. November 2017. [Fecha de consulta: 30 de abril de 2019]. Disponible en: <http://adsabs.harvard.edu/abs/2015ScTEen.505.1202W> ISBN: 505:1202-1212

-WARK, Kenneth & Warner, Cecil. Contaminación del Aire. LIMUSA, 2014. 650 pp. ISBN: 9681819543

-WHO. Ambient air Pollution. Library Cataloguing, 2016. 121 pp. ISBN: 978924151353

- WRF-Chem Version 3.9.1.1 User's Guide.

ANEXOS

SCRIPT DE PREPARACIÓN Y CORRIDA DE DATOS

```
1 DATE<-read_excel("C:/Users/Usuario/Desktop/DatosC_entrena_CLUD.xlsx")
2 View(DATE)
3
4 # ELIMINANDO LOS NA #
5 CLU <- na.omit(DATE)
6 as.data.frame(CLU)
7 View(CLU)
8 nrow(CLU)
9 set.seed(100)
10
11 ## EXTRAYENDO MUESTRA ALEATORIA-80% ##
12 muestr <- sample(1:nrow(CLU),3390)
13 varios <- CLU[muestr,]
14 nrow(varios)
15 View(varios)
16
17 ## MUESTRA DE 20% (809 datos restantes)##
18 restmues<- CLU[-muestr,]
19 nrow(restmues)
20 View(restmues)
21
22 # ESCALONAMIENTO DE DATOS #
23 esca <-c()
24 for (i in 1:7) {
25   esca[i] <-(CLU[,i]-min(CLU[,i]))/(max(CLU[,i])-min(CLU[,i]))
26 }
27 names(esca)<-names(CLU)
28 datestand<-as.data.frame(esca)
29 nrow(datestand)
30 View(datestand)
31
32 train <- datestand[muestr,]
33 test <- datestand[-muestr,]
34 nrow(train)
35 nrow(test)
36 View(train)
37 View(test)
38 test$PM10
39
40 #FORMULAS#
41
42 nms <- names(train)
43 nms
44
45 ecuación<- as.formula(paste("(PM_10)~", paste(nms[!nms%in%"PM_10"], collapse = "+")))
46
47 str(nms)
48 str(ecuación)
49
```

```

50 # MODELO
51 maxIteraciones=10 #Cantidad de corridas que harias manualmente
52 errorFactible=0.5 #
53 vectorerror<-c() #
54 vectorNeuronas<-c() #
55 vectorErrorNeurona<-c() #
56 posicionIteracionError<-c() #
57
58 #
59 for (j in 1:5){
60   neuronas = 2*j
61
62   for (i in 1:maxIteraciones){
63     modelo.nn<- neuralnet(ecuación, data = train, hidden = neuronas,lifesign.step = 300,
64       threshold = 0.05,algorithm = "rprop+")
65
66     prediccion<- compute(modelo.nn, within(test,rm(PMC_10)))
67     PMC10.Predict<- prediccion$net.result*(max(CLU$PMC_10)-min(CLU$PMC_10))+min(CLU$PMC_10)
68     PMC10.Real<- (test$PMC_10)*(max(CLU$PMC_10)-min(CLU$PMC_10))+min(CLU$PMC_10)
69     MSE<- sum((PMC10.Real-PMC10.Predict)^2)/nrow(test)
70     vectorerror[i]<-MSE
71   }
72
73   if(min(vectorerror) < errorFactible){
74     vectorErrorNeurona[j]<-min(vectorerror)
75     posicionIteracionError[j]<-which.min(vectorerror)
76     vectorNeuronas[j]<-neuronas
77   }
78
79   prediccion<- compute(modelo.nn, within(test,rm(PMC_10)))
80   PMC10.Predict<- prediccion$net.result*(max(CLU$PMC_10)-min(CLU$PMC_10))+min(CLU$PMC_10)
81   PMC10.Real<- (test$PMC_10)*(max(CLU$PMC_10)-min(CLU$PMC_10))+min(CLU$PMC_10)
82   MSE<- sum((PMC10.Real-PMC10.Predict)^2)/nrow(test)
83   vectorerror[i]<-MSE
84 }
85
86 if(min(vectorerror) < errorFactible){
87   vectorErrorNeurona[j]<-min(vectorerror)
88   posicionIteracionError[j]<-which.min(vectorerror)
89   vectorNeuronas[j]<-neuronas
90 }
91
92 vectorerror
93 posicionIteracionError
94
95 vectorNeuronas
96 vectorErrorNeurona

```



```

1
2 #LIBRERIAS#
3 install.packages(c('rattle',"corrplot","factoextra","mclust","FactoMineR"))
4
5 library(cValid);library(rattle);library(mclust);library(FactoMineR)
6 library(NbClust);library(ppclust);library(fclust);library(corrplot)
7 library(psych);library(ggcorrplot);library(neuralnet);library(ggplot2)
8 library(MASS);library(readxl);library(cluster);library(factoextra)
9 ###Corrida para kmeans###
10
11 clusATE<-read_excel("C:/Users/Usuario/Desktop/Datos_entrena_ATE.xlsx")
12 View(clusATE)
13 #Script para eliminar los NA#
14 CLU <- na.omit(clusATE)
15 View(CLU)
16 names(CLU)
17
18 ##verificando la correlación##
19 pairs.panels(CLU)
20 correlations <- cor(CLU) # correlation matrix
21 corrplot(correlations, method="circle")
22
23 #reducir datos en dos dígitos#
24 round(CLU, digits = 3)
25
26 #ESCALONAMIENTO DE DATOS#
27
28 esca <-c()
29
30 for (i in 1:17) {
31   esca[i] <- (CLU[,i]-min(CLU[,i]))/(max(CLU[,i])-min(CLU[,i]))
32 }
33 names(esca)<-names(CLU)
34 datnorm<-as.data.frame(esca)
35 View(datnorm)
36 apply(datnorm, 2, range)
37 ##script para kmeans##
38 set.seed(80)
39 kmdat <- kmeans(datnorm, centers = 8, iter.max = 200)
40 View(kmdat$cluster)
41 kmdat$cluster
42 write.csv(kmdat$cluster, file = "clusterx.csv")
43 kmdat$totss
44 kmdat$betweenss
45 kmdat$withinss
46 kmdat$tot.withinss
47 kmdat$centers
48 kmdat$ifault
49 #Mostrar la cantidad de datos en cada cluster#
50 table(kmdat$cluster)
51 #Creando graficos con distribucion de todos los centroides#
52 centr <- kmdat$centers
53 mini<-apply(datnorm,2,min)
54 maxi<-apply(datnorm,2,max)
55 maxi<-as.numeric(maxi)
56 mini<-as.numeric(mini)
57 datce<- t(mini+t(centr)*(maxi-mini))
58 dat_cent <- data.frame(datce)
59 parcoord(dat_cent, col= 1:5, var.label = TRUE)
60 par(xpd=TRUE)
61
62 #variables de entrada#
63 #EA, CLR, TM, HR, DV, VA, RS, PMA10,
64 #PM25, PM10, SO2, NO2, NO, O3, CO, PAS##
65 #Determinar número de cluster óptimo##
66
67 clusop <- kmeans(CLU, centers = 5)$betweenss
68 for (i in 1:17) clusop[i] <-kmeans(CLU, centers = i)$betweenss
69
70 plot(1:15, clusop, type = "b", xlab = "Número óptimo de cluster", ylab = "variación")

```



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA
ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA AMBIENTAL

Modelo Ecológico para el pronóstico de la concentración del Material Particulado (PM10) en los distritos de Jesús María y Ate Vitarte, 2019

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:
Ingeniero Ambiental

AUTOR:
Dr. Merino Navarro, Erik Gregorio (ORCID: 0000-0003-1735-2629)

ASESOR:
Dr. Juan Julio Ojeda Gámez (PHD) (ORCID: 0000-0002-3419-7361)

LÍNEA DE INVESTIGACIÓN:
Gestión y Conservación de los Recursos Naturales

LIMA - PERÚ
2019



Resumen de coincidencias

22 %

Se están viendo fuentes estándar
Ver fuentes en inglés (Beta)

Coincidencias	Porcentaje
1 cajunm.ee Fuente de Internet	4 %
2 Entregado a Universidad... Trabajo de tesis	3 %
3 fundaciona.org Fuente de Internet	2 %
4 .issuu.com Fuente de Internet	1 %
5 repositorio.unc.edu.pe Fuente de Internet	1 %
6 repositorio.uns.edu.pe Fuente de Internet	1 %
7 www.aeromil.gub.pe Fuente de Internet	1 %
8 vnufile.biogspat.com Fuente de Internet	<1 %
9 Entregado a Universidad... Trabajo de tesis	<1 %