



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

Estimación de la distribución espacio-temporal de concentraciones de material particulado PM2.5 con el uso de redes neuronales artificiales e información geoespacial

TESIS

Que para obtener el título de

Ingeniero Civil

P R E S E N T A

Pablo de Jesús Ángeles Salto

DIRECTOR DE TESIS

M.I. Rodrigo Takashi Sepúlveda Hirose



Ciudad Universitaria, Cd. Mx., 2019

AGRADECIMIENTOS:

A toda mi familia. En especial a mi mamá por apoyarme incondicionalmente, y siempre recibirme con una sonrisa, a mi papá por sus consejos y apoyo durante toda la carrera, ambos por la confianza depositada en mí. A mis abuelos, ya que, sin su arduo trabajo, yo no podría estar aquí.

A mi maestro, Rodrigo Takashi Sepúlveda Hirose por brindarme la oportunidad de integrarme a su equipo de trabajo, sus enseñanzas y consejos que fueron fundamentales para la realización de este trabajo.

AGRADECIMIENTO ESPECIAL:

Agradecimiento especial a la Secretaría de Ciencia, Tecnología e Innovación de la Ciudad de México; por los fondos, asesorías y apoyos otorgados durante el desarrollo de esta investigación que formará parte del proyecto SECITI/089/2017 “Desarrollo de modelos para la estimación de material particulado suspendido de la Zona Metropolitana de la Ciudad de México mediante el uso de imágenes satelitales e información geoespacial”

Resumen

La presencia de PM2.5 en la atmósfera ha sido relacionado con problemas de salud pública. Se puede medir con precisión las concentraciones de este contaminante con equipos en tierra, pero estos están limitados debido a su alto costo, por ello se propone estimar las concentraciones por medio de imágenes satelitales Landsat 8, mediciones en tierra y modelos de regresión de redes neuronales artificiales.

El modelo aplicado cuenta con 13 variables de entrada, dos capas ocultas la primera con 100 nodos y la segunda con 60 nodos, ambas activadas con la función ReLU. Se obtuvo como resultado un modelo con coeficiente de correlación de 0.85 y raíz del error cuadrático medio de 7.16. Este modelo fue aplicado a imágenes satelitales de la Ciudad de México, logrando mapas de concentraciones de PM2.5.

Abstract

The presence of PM2.5 in the atmosphere has been related to public health problems. The concentrations of this pollutant can be accurately measured with ground equipment, but these are limited due to their high cost, for this reason it is proposed to estimate the concentrations using Landsat 8 satellite images, ground measurements and regression models of neural networks artificial.

The applied model has 13 input variables, two hidden layers the first with 100 nodes and the second with 60 nodes, both activated with the ReLU function. Resulting in a correlation coefficient of 0.85 and root mean square error of 7.16. This model was applied to satellite images of Mexico City, achieving maps of PM2.5 concentration.

Índice

Introducción	1
Justificación	2
Objetivos	3
Capítulo 1 Antecedentes	4
1.1 Percepción remota	4
1.1.1 Landsat 8	5
1.1.2 Corrección atmosférica	9
1.2 Índices Espectrales	11
1.3 Datos de Contaminantes en la Ciudad de México	14
1.4 Redes Neuronales Artificiales.....	15
1.4.1 Funciones de Activación.....	17
1.4.2 Red neuronal multicapa	20
1.4.3 Funciones de error	23
1.4.4 Propagación hacia atrás	24
Capítulo 2 Estado del arte	26
Capítulo 3 Metodología	29
3.1 Descarga de imágenes satelitales	29
3.2 Datos de contaminantes	30
3.3 Armado de la base de datos.....	34
3.3.1 Extracción de datos de imágenes satelitales	34
3.3.2 Agregar datos PM, humedad, temperatura y precipitación	36
3.4 Variables adicionales.....	36
3.5 Base de datos adicionales	37
3.6 Análisis de Variables.....	37
3.7 Creación de la Red Neuronal.....	41
3.7.1 Tensorflow y Keras	41
3.7.2 Configuración	42
3.8 Selección de variables predictoras.....	48
3.9 Aplicación de modelo.....	51
Capítulo 4 Análisis de resultados	54
4.1 Análisis Temporal	58
4.2 Comprobación del modelo.....	67

Capítulo 5 Conclusiones y recomendaciones	68
Anexo	70
Bibliografía.....	72

Índice de Figuras

Figura 1.1 Diagrama de los elementos involucrados en la percepción remota (Servicio Geológico Minero Argentino, 2018).....	5
Figura 1.2 Espectro electromagnético.....	7
Figura 1.3 Diagrama de perceptrón simple	16
Figura 1.4 Diagrama de un perceptrón con función de activación sigmoide.....	17
Figura 1.5 Gráfica de la función sigmoide	18
Figura 1.6 Gráfica de la función tangente hiperbólico	19
Figura 1.7 Gráfica de la función lineal rectificadora.....	20
Figura 1.8 Diferentes arquitecturas de Redes Neuronales Artificiales.....	20
Figura 1.9 Arquitectura multicapa de Perceptrones	21
Figura 1.10 Esquema método del descenso por gradiente	23
Figura 2.1 Red Neuronal entrenada para predecir Concentraciones de PM2.5 con una malla de celdas de 1kmx 1km en Estados Unidos de América.	26
Figura 2.2 Gráfica IMECA mensual, real vs pronostico con RNA.(V. H. Torres, 2017).....	27
Figura 2.3 Gráfica de predicción concentraciones de ozono horario.(Arrellano Monreal, 2019)	27
Figura 2.4 Gráfica de predicción concentraciones de ozono promedio diario.(Camacho Gonzáles, 2018)	28
Figura 3.1 Mapa de ubicación de estaciones de medición PM2.5.....	30
Figura 3.2 El lado izquierdo está el formato del archivo descargado y a la derecha con el formato después de reacomodar datos.....	31
Figura 3.3 Gráfico de caja concentraciones promedio diarias PM2.5, 2013-2018.	31
Figura 3.4 Gráfica de precipitación contra concentración de PM promedio.....	33
Figura 3.5 Gráfica de temperatura contra concentración de PM promedio.....	33
Figura 3.6 Ejemplo formato de extracciones	36
Figura 3.7 Gráficas de correlación entre reflectancias atmosféricas y PM2.5	38
Figura 3.8 Gráficas de correlación entre índices espectrales y PM2.5	39
Figura 3.9 Matriz de correlaciones entre variables independientes	40
Figura 3.10 Ejemplo de código de Redes Neuronales en R usando Keras.	42
Figura 3.11 Gráficas de dispersión valores predichos y valores reales, conjunto de datos solo estaciones número de datos=834.....	44
Figura 3.12 Gráfica de dispersión modelo de RNA fallido	45

Figura 3.13 Gráficas de dispersión valores predichos y valores reales, modelo 2 número de datos=834.....	46
Figura 3.14 Gráficas de dispersión valores predichos y valores reales, modelo 3, conjunto de datos con pixeles vecinos, número de datos=7781.....	47
Figura 3.15 Gráfico de barras R^2 / desviación estándar de modelos con remoción de una variable.....	49
Figura 3.16 Gráfico de barras R^2 / desviación estándar de modelos una vez removidas 15 variables.....	50
Figura 3.17 Arquitectura del modelo de Red Neuronal Artificial.....	52
Figura 3.18 Gráficas de dispersión del primer modelo aplicado.....	53
Figura 4.1 Mapa de concentraciones aplicando el 1er modelo de Redes Neuronales Generado.....	54
Figura 4.2 Mapa de concentraciones PM2.5 modelo 2(sin NDVI).....	55
Figura 4.3 Distribución de valores predichos de PM2.5 tomando en cuenta toda la CDMX.....	56
Figura 4.4 Concentraciones de PM2.5 en alrededores Bosque de Chapultepec.....	57
Figura 4.5 Concentraciones de PM2.5 sobre naves industriales, Hacienda del parque, Cuautitlán Izcalli, 2018-01-10.....	58
Figura 4.6 Grafica de caja concentraciones PM2.5.....	59
Figura 4.7 Mapa concentraciones PM2.5, 2015-02-19.....	60
Figura 4.8 Mapa concentraciones PM2.5, 2015-03-07.....	61
Figura 4.9 Mapa concentraciones PM2.5, 2015-04-24.....	62
Figura 4.10 Mapa concentraciones PM2.5, 2015-05-10.....	63
Figura 4.11 Mapa concentraciones PM2.5, 2015-06-11.....	64
Figura 4.12 Mapa concentraciones PM2.5, 2015-07-13.....	65
Figura 4.13 Mapa concentraciones PM2.5, 2015-12-20.....	66
Figura 4.14 Mapa concentraciones PM2.5 2019-01-29.....	67

Introducción

En los últimos años han aumentado la preocupación de la calidad del aire en la comunidad científica, a la luz de investigaciones que relacionan contaminantes atmosféricos con problemas de salud y efectos atmosféricos como el deterioro en la capa estratosférica de ozono, o el efecto invernadero.

Aunque generalmente la mayoría de los contaminantes atmosféricos siempre se encuentran presentes debido a efectos de la naturaleza, la acción del hombre a través de la quema de combustibles fósiles y otras actividades ha provocado un incremento en las concentraciones de estos contaminantes. Hay dos factores que intervienen para que un contaminante produzca efectos adversos a la salud humana, el primero, como se había mencionado, es la concentración del contaminante, y el segundo, el tiempo de exposición. Es por ello que en las normas mexicanas se mencionan dos valores límites: el de exposición aguda (o de tiempos cortos), generalmente asociado a la evaluación mediante el promedio diario, y el de exposición crónica, que se cuantifica a través del promedio anual de las concentraciones.

El entender la distribución espacial y temporal de los contaminantes es el primer paso para la toma de decisiones dirigidas a mitigar riesgos a la población, pero esto no es tarea sencilla ya que existen muchos factores que modifican estas concentraciones, como por ejemplo viento, temperatura, lluvia, humedad, ubicación de industrias, ubicación de vialidades principales, entre otras.

En el primer capítulo de antecedentes, se describen los tres componentes básicos para el desarrollo de este trabajo que son las imágenes satelitales del Landsat 8, las mediciones en campo de SEDEMA y las redes neuronales artificiales. En el segundo capítulo relacionado con el estado del arte se investigó información acerca de los trabajos realizados en los últimos años, principalmente en la Ciudad de México. En el capítulo tercero se detalla el procedimiento realizado para obtener el modelo. En el cuarto capítulo se analizan los resultados y se aplica el modelo a una escena diferente a las escenas ocupadas para realizar el modelo. En el último capítulo se habla de las oportunidades para implementar este tipo de modelos.

Justificación

El material particulado (PM) es una mezcla compleja de sustancias en estado líquido o sólido, que permanece suspendida en la atmósfera por periodos variables de tiempo. Por su origen, las partículas pueden definirse como primarias (aquellas producidas directamente por alguna fuente contaminante) o secundarias (las que se forman en la atmósfera, como resultado de la interacción química entre gases y partículas primarias) (NOM-251-SSA1-2009, 2009).

El PM 2.5 es el material particulado cuyo diámetro aerodinámico es menor a 2.5 mm, debido a este tamaño muchas partículas pasan fácilmente las barreras naturales del sistema respiratorio del cuerpo humano, como son bellos nasales o la mucosa, depositándose en los alveolos de los pulmones y pasando directamente a la sangre. Por ello han surgido muchos estudios que analizan los efectos en la salud humana por la exposición a estas partículas, (Anenberg, Horowitz, Tong, & West, 2010) obtuvieron que 3.5 millones de muertes cardiopulmonares y 220,000 muertes de cáncer pulmonar pueden estar relacionadas con PM2.5, esto a nivel mundial.

Por la alta mortalidad que existe al estar en contacto con estas partículas es necesario tener el conocimiento de su distribución espacial para poder informar y prevenir a la población de altos niveles de concentración, evaluar estrategias de control enfocadas a zonas específicas, entre otros aspectos. Y es la distribución espacial en la que se centra este trabajo al intentar determinar la concentración de los contaminantes y su distribución espacial, en específico PM2.5, de forma indirecta a través de imágenes satelitales.

Objetivos

Objetivo general

Realizar un modelo de regresión de PM2.5 mediante Redes Neuronales usando imágenes satelitales y mediciones de la Red Automática de Monitoreo Atmosférico en la Ciudad de México.

Objetivos Específicos

1. Estudiar las características de las imágenes satelitales del Landsat 8 y recopilar las que estén disponibles en Ciudad de México
2. Extraer la información de las imágenes en las coordenadas de las estaciones de monitoreo.
3. Analizar las relaciones entre la información de las imágenes y los datos de PM2.5 de la Ciudad de México.
4. Aplicar el mejor modelo de regresión generado en toda el área de la CDMX para obtener un mapa de concentraciones de PM2.5.

Capítulo 1 Antecedentes

1.1 Percepción remota

“La percepción remota es la ciencia de adquirir información acerca de la superficie de la Tierra sin estar en contacto directo con ella. Esto es captando y grabando la energía que se emite o refleja, procesando, analizando y aplicando esta información” (Canada Centre for Remote Sensing, 2016).

La información adquirida a través de la percepción remota nos brinda conocimiento de los recursos naturales, como son bosques, cultivos, minas, cuerpos de agua, etc. También nos puede ayudar en entender el comportamiento de fenómenos naturales, así como intentar predecirlos, lo que implica desarrollo tecnológico, económico e industrial para un país.

Los elementos involucrados en el proceso de adquirir información de la Tierra por medio de sensores remotos son los siguientes, estos se pueden observar en la figura 1.1.

1. Fuente de Energía. El primer elemento de la percepción remota es una fuente de energía que ilumine el objeto de interés, en la mayoría de los satélites esta fuente es el Sol, aunque existen satélites equipados con fuentes propias a los cuales se les denomina sistemas activos. En teoría todos los objetos que tengan una temperatura por encima del cero absoluto emiten energía, pero esta es muy pequeña, por lo que los sensores convencionales de los satélites no logran medir, aun así, existen imágenes tomadas durante la noche donde no se tiene ninguna fuente de energía más que la emitida por los propios objetivos, la mayoría creado por el hombre.
2. Radiación y atmósfera. La energía viaja desde la fuente hasta el objetivo, en este viaje la energía entra en contacto con la atmósfera donde una parte es absorbida, otra dispersada y otra pasa a través de ella. Esto ocurre por segunda vez cuando la energía reflejada viaja de la superficie terrestre al satélite. Este es el elemento más importante en este trabajo ya que nos interesa más la información de la interacción con la atmosfera que la información de la superficie de la Tierra.
3. Interacción con el objetivo. De igual forma en cuando la energía toca la superficie de la Tierra esta es absorbida y reflejada en varias direcciones, dependiendo las propiedades del objeto.

4. Almacenamiento de la información que llega al sensor. Una vez que la energía llega al sensor, después de los pasos anteriores, ésta es almacenada en forma digital en el satélite.
5. Transmisión, recepción y procesamiento. Los datos almacenados en el sensor son enviados a una estación de procesamiento, generalmente en tierra, en esta estación son procesados en una imagen digital usualmente ya georreferenciada.
6. Interpretación y análisis. La imagen procesada es visualizada para extraer información acerca de los objetos de interés.
7. Aplicación. El último paso del proceso de percepción remota es cuando se aplica la información extraída de las imágenes, esto para entender mejor los objetos captados, revelar nueva información o ayudar en resolver problemas.

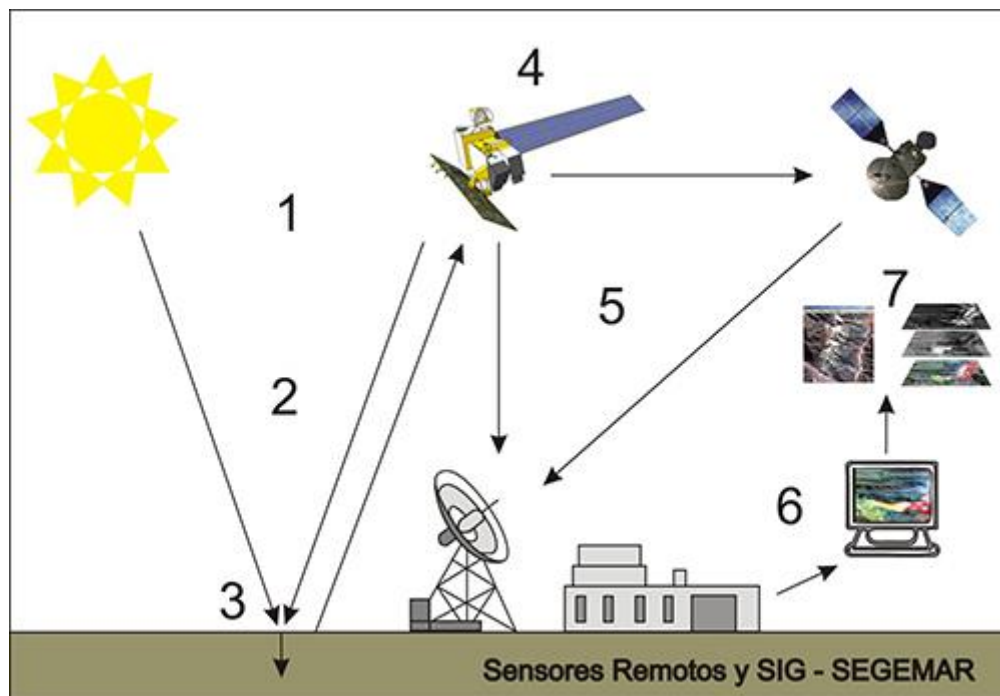


Figura 1.1 Diagrama de los elementos involucrados en la percepción remota (Servicio Geológico Minero Argentino, 2018)

1.1.1 Landsat 8

Como parte del proyecto del cual surge este trabajo, se decidió el uso del satélite Landsat 8 para la obtención de imágenes entre las ventajas que presentan respecto otros satélites se encuentra su alta resolución espacial, una banda destinada a la medición de aerosoles (Banda1), la gratuidad de las imágenes, entre otras características que se describen a continuación.

El 21 de septiembre de 1966 el secretario de interior de los Estados Unidos, Stewart Udall anuncio el proyecto "Project EROS (Earth Resources Observation Satellites)". La visión de Udall fue observar la Tierra para el beneficio de todos. Declaró: "*el programa proveerá la oportunidad de coleccionar información valiosa que se usará para mejorar la calidad de nuestro ambiente*" (United States Geological Survey, 2018b).

Este proyecto marcó la pauta para desarrollar el satélite ERTS-1, que posteriormente fue renombrado como Landsat 1, lanzado en el año de 1972 que fue el primer satélite lanzado para monitorear y estudiar nuestro planeta.

La función del Landsat 8 es coleccionar, archivar, procesar y distribuir datos de manera consistente con el satélite Landsat 7. Este opera a una altitud de 705 km del ecuador, el paso del satélite es cíclico y pasa cada 16 días por el mismo punto (Department of the Interior U.S., 2018).

Todos los objetos con temperatura mayor al cero absoluto emiten y reflejan energía que viaja por el espacio, esta energía tiene efectos eléctricos y magnéticos, por lo que se denomina radiación electromagnética. Ésta viaja en forma de ondas las cuales tienen diferentes longitudes, el espectro electromagnético está constituido por todas las longitudes de onda, desde las más grandes hasta las más pequeñas (rayos gamma), y todas las que están entre ellas.

Todas estas ondas electromagnéticas proporcionan información, pero nosotros con nuestros ojos sólo podemos percibir una pequeña fracción de ellas en lo que llamamos luz visible, en la figura 1.2 observamos el espectro electromagnético en escala gráfica y la pequeña porción que constituye la luz visible. Sin embargo, se han creado sensores capaces de percibir y emitir en otras longitudes de onda, un primer ejemplo de estos es el radio que tiene longitudes de onda muy grandes a comparación de la luz visible, gracias a este invento se pudo captar y enviar información a grandes distancias sin necesidad de cables.

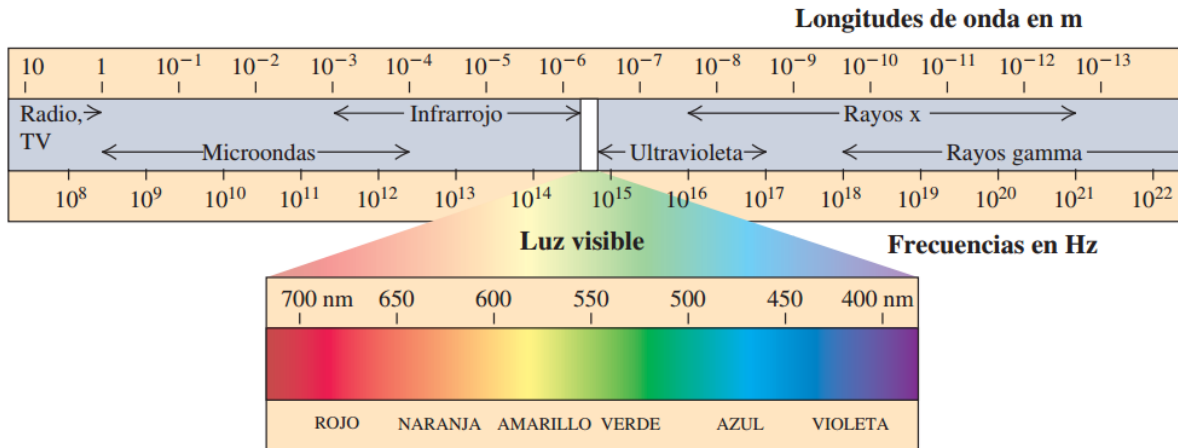


Figura 1.2 Espectro electromagnético

Desde la creación del radio los avances tecnológicos permitieron la creación de sensores capaces de captar distintas partes del espectro electromagnético, los sensores creados van desde las cámaras digitales que captan, al igual que nosotros, la luz visible, hasta sensores que captan longitudes de onda muy pequeños como son los rayos gamma. Los sensores del Landsat 8 perciben en un rango de luz que va de los infrarrojos hasta la primera parte de los ultravioleta.

El satélite Landsat 8 está equipado con 2 sensores, el Operational Land Imager (OLI) y el Thermal Infrared Sensor (TIRS), que operan de forma conjunta pero que pueden llegar a funcionar por separado si llega a ocurrir un fallo. Estos sensores miden la luz del sol reflejada en el planeta, en diferentes rangos del espectro de luz el cual almacenan en distintas “bandas”¹. En la tabla 1.1 se presentan los distintos rangos que miden tanto el sensor TIRS y OLI, y la denominación de la banda, además una comparación con su predecesor.

¹ Banda espectral se le denomina la información captada en un delimitado rango de valores a lo largo del espectro electromagnético.

Tabla 1.1 Comparación denominación de bandas en Landsat 7 y 8 (Department of the Interior U.S., 2018).

Landsat-7 ETM+ Bands (μm)			Landsat-8 OLI and TIRS Bands (μm)		
			30 m Coastal/Aerosol	0.435 - 0.451	Band 1
Band 1	30 m Blue	0.441 - 0.514	30 m Blue	0.452 - 0.512	Band 2
Band 2	30 m Green	0.519 - 0.601	30 m Green	0.533 - 0.590	Band 3
Band 3	30 m Red	0.631 - 0.692	30 m Red	0.636 - 0.673	Band 4
Band 4	30 m NIR	0.772 - 0.898	30 m NIR	0.851 - 0.879	Band 5
Band 5	30 m SWIR-1	1.547 - 1.749	30 m SWIR-1	1.566 - 1.651	Band 6
Band 6	60 m TIR	10.31 - 12.36	100 m TIR-1	10.60 - 11.19	Band 10
			100 m TIR-2	11.50 - 12.51	Band 11
Band 7	30 m SWIR-2	2.064 - 2.345	30 m SWIR-2	2.107 - 2.294	Band 7
Band 8	15 m Pan	0.515 - 0.896	15 m Pan	0.503 - 0.676	Band 8
			30 m Cirrus	1.363 - 1.384	Band 9

La resolución espacial del sensor OLI es de 30x30 m, el sensor TIRS tiene una resolución espacial de 100m. Una escena completa mide 185 x180 km, lo que equivale a una imagen con 36,996,000 pixeles con información guardada en 12 bits, esto es que cada pixel puede tomar un valor de entre 0-4096.

El producto que entrega Landsat escala los valores de 12 bits a 16 bits, además hace la georreferencia de las imágenes, a este producto lo denomina niveles digitales (DN) que físicamente no representa nada. Posteriormente se puede procesar estos niveles digitales a reflectancia o radiancia en el tope de la atmósfera (TOA por sus siglas en inglés), con ayuda de ecuaciones de transformación y los coeficientes de escalamiento radiométrico que se provee en los metadatos de la escena (United States Geological Survey, 2018a).

La ecuación 1.1 es para calcular la reflectancia TOA a partir de los valores digitales de la imagen.

$$\rho_{\lambda}' = M_{\rho} Q_{cal} + A_{\rho} \quad \dots 1.1$$

Donde:

ρ_{λ}' = Reflectancia planetaria TOA, sin corrección por el ángulo solar

M_{ρ} = Factor multiplicativo específico por banda, obtenido metadatos (REFLECTANCE_MULT_BAND)

A_{ρ} = Factor aditivo específico por banda, obtenido metadatos (REFLECTANCE_ADD_BAND)

Q_{cal} = Valores de píxeles de productos estándar cuantificados y calibrados (DN)

Para realizar la corrección por el ángulo solar se ocupa la ecuación 1.2

$$\rho_{\lambda} = \frac{\rho_{\lambda}'}{\sin(\theta_{SZ})} \quad \dots 1.2$$

Donde:

ρ_{λ} = Reflectancia planetaria TOA

θ_{SZ} = Elevación del ángulo solar local. El ángulo de elevación del sol del centro de la escena en grados se proporciona en los metadatos

Aunque se puede descargar el producto de niveles digitales para luego aplicarles las ecuaciones anteriores de forma manual, existe la descarga directa del producto reflectancia TOA la principal ventaja es que la corrección por el ángulo solar del producto de Landsat lo realiza pixel por pixel, en cambio al hacerlo de forma manual se realiza con el ángulo solar al centro de la imagen, lo que simplifica los cálculos, pero disminuye la exactitud de los valores en más o menos 3%.

1.1.2 Corrección atmosférica

En la corrección atmosférica se busca eliminar los efectos de los gases de la atmósfera para solo obtener la reflectancia de la superficie, en muchos casos el objeto de estudio son cuerpos de agua o masas de tierra por lo que se desea eliminar los efectos de la atmósfera, en este trabajo se desea obtener los efectos solo de la atmósfera por lo que se tiene que eliminar los efectos de la superficie, esto se logró utilizando aplicando la corrección

atmosférica a las 7 primeras bandas del landsat8, ya que si tenemos la reflectancia TOA y la reflectancia de superficie se puede inferir rápidamente la reflectancia atmosférica.

Existen distintos métodos para realizar esta corrección atmosférica la mayoría de estos métodos pueden ser catalogados en los siguientes grupos: objetos invariantes, comparación de histogramas, objetos oscuros y reducción de contraste (Liang, Fang, & Chen, 2001).

-Objetos invariantes es un método que asume que existen algunos píxeles que en cualquier escena son bastante constantes, por lo que las diferencias presentadas entre dos imágenes son causadas por efectos de la atmósfera, y con ello se puede hacer una corrección a toda la escena.

-El método de comparación de histogramas supone que los histogramas de reflectancia de regiones despejadas y de regiones brumosas deberían ser iguales, por lo que al identificar secciones despejadas se puede realizar una corrección al histograma de regiones brumosas.

-El método de objetos oscuros hace la suposición de que existen objetos que prácticamente no deberían reflejar nada de luz en ciertas bandas, por ejemplo, vegetación muy densa casi no refleja azul (banda 2), por lo que los valores medidos son debidos a la atmósfera, con estos valores se hace una corrección a toda la escena o fracción de la misma, este método funciona mejor en zonas pequeñas y con bajas concentraciones de aerosoles.

Los métodos anteriores son semi empíricos, ya que no toman en cuenta todas las variables que influyen en el paso de la luz a través de la atmósfera, además las correcciones que hace son muchas veces constantes esto es que cada pixel se le disminuye un valor constante, y en nuestro caso es prácticamente inútil esta información. Existen otros métodos más complejos, que son los modelos de transferencia radiativa, que toman en cuenta los procesos que modifican la radiancia reflejada por un objeto mirado desde el espacio: la absorción por los gases, y la dispersión por aerosoles y moléculas. Las deducciones de fórmulas y los métodos para aplicar estos métodos escapan de los objetivos de esta tesis, pero la deducción de un método de este tipo puede ser consultada en el artículo "Second Simulation of the Satellite Signal in the Solar Spectrum, 6S"(Kraus, Szalay, Meyer, Kilian, & Schnettler, 2007), este método es uno de los más aplicados en software comercial.

Landsat 8 tiene su producto de reflectancia superficial (SR) donde se aplicó una corrección atmosférica, para obtener este producto Landsat usa un modelo de transferencia radiativa propio del Landsat 8, pero que es similar al 6S. En su hoja de especificaciones (U.S. Geological Survey, 2018) se mencionan los diferentes parámetros que ocupa su método de corrección atmosférica y algunos de estos parámetros no son medidos directamente por Landsat por lo que son tomados de otros satélites, por ejemplo: el vapor de agua, la temperatura del aire y AOT son calculados por MODIS CMA², el modelo digital de elevaciones es obtenido por el “Global Climate Model DEM”.

Debido a la dificultad de generar un modelo para obtener la reflectancia de superficie se optó por usar este producto del Landsat 8, el único inconveniente es que solo se encuentra disponible bajo demanda por lo que se tiene que pedir y esperar a que se procese la solicitud, y una vez procesada solo estará disponible durante una semana para su descarga.

1.2 Índices Espectrales

Los índices espectrales se definen como nuevas variables generadas por combinaciones matemáticas de dos o más bandas originales y espectrales escogidas como una forma en que el nuevo índice va a resaltar con mayor claridad parámetros biofísicos de interés (Jones & Vaughan, 2010).

El Landsat 8 entrega un producto con 7 índices que son los más ocupados en la percepción remota, se decidió descargar estos índices ya que en otros trabajos estos índices presentaron correlaciones con las concentraciones de PM, un ejemplo de esto es en el trabajo (N. Torres & Vivianco, 2018) ocuparon los índices espectrales en modelos lineales, con buenos resultados. A continuación, se describirán brevemente los índices que se ocupan usualmente y la forma de calcularlos, aunque como se dijo anteriormente es posible solicitarlos para su descarga y así reducir tiempo de procesamiento en caso de no tener un procesador muy rápido.

² Moderate Resolution Imaging Spectroradiometer (MODIS) y su producto “Aerosol Optical Thickness Daily”

Índice de Vegetación Normalizado (NDVI)

Es un índice que resalta la presencia de vegetación verde en la superficie y ayuda a delimitar su distribución espacial. Se calcula con la banda de color rojo (banda 4) y la banda del infrarrojo cercano (banda 5), se puede observar en la ecuación 1.3.

$$NDVI = \frac{\text{Banda 5} - \text{Banda 4}}{\text{Banda 5} + \text{Banda 4}} \quad \dots 1.3$$

Índice de Vegetación Mejorado (EVI)

Este índice incorpora dos coeficientes, “L” que realiza un ajuste en relación al canopeo³ y el coeficiente “C” como coeficiente de resistencia de la atmósfera, y un valor correspondiente a la banda azul. Estas las mejoras permiten el cálculo del índice como una relación entre los valores Rojo (R) e Infrarrojo cercano (NIR -del inglés *Near InfraRed*), al tiempo que reduce el ruido de fondo, el ruido atmosférico y la saturación en la mayoría de los casos (U.S. Geological Survey, 2018).

$$EVI = \frac{G * (NIR - R)}{NIR + C1 * R - C2 * B + L}$$

Dado que el infrarrojo cercano (NIR) corresponde a la banda 5, el rojo (R) a la banda 4 y el azul (B) a la banda 2, la ecuación para el Landsat 8 se muestra en la ecuación 1.4.

$$EVI = \frac{2.5 * (\text{Banda 5} - \text{Banda 4})}{\text{Banda 5} + 6 * \text{Banda 4} - 7.5 * \text{Banda 2} + 1} \quad \dots 1.4$$

Índice de Vegetación Ajustado con el Suelo (SAVI)

Este índice, ecuación 1.5, busca abordar algunas limitaciones del NDVI cuando es aplicado a zonas donde el suelo está expuesto. Este es calculado como una relación entre la banda roja y la banda del infrarrojo cercano con un ajuste por el brillo del suelo “L” igual a 0.5 que se ajusta a la mayoría de tipos de cobertura de suelo.

$$SAVI = \frac{NIR - R}{(NIR + R + L) * (1 + L)}$$

³ El canopeo es la capa superior de los árboles y el coeficiente “L” toma en cuenta el color de esta capa, la forma de las hojas y la geometría .

$$SAVI = \frac{Banda\ 5 - Banda\ 4}{(Banda\ 5 + Banda\ 4 + 0.5) * (1.5)} \quad \dots 1.5$$

Índice Modificado de Vegetación Ajustado con el Suelo (MSAVI)

La única diferencia con respecto a SAVI es la manera de calcular el coeficiente de corrección “L”, este factor es importante ya que describe la relación entre vegetación y suelo, y en el índice SAVI se asume como 0.5, en cambio en MSAVI “L” está en función de las bandas Roja e infrarrojo cercano. La fórmula se puede observar en la ecuación 1.6.

$$MSAVI = \frac{(2 * Banda\ 5 + 1) - \sqrt{(2 * Banda\ 4 + 1)^2 - 8 * (Banda\ 5 - Banda\ 4)}}{2} \quad \dots 1.6$$

Índice Normalizado de la Diferencia de Humedad (NDMI)

Este índice es un derivado de las bandas de infrarrojo cercano (NIR) y el infrarrojo medio (SWIR). La Banda SWIR es afectada por el contenido de agua de la vegetación y la estructura interna de la vegetación, mientras que la banda NIR es afectada por la estructura interna de las hojas y el contenido de materia seca de las mismas. Con la combinación de ambas bandas, como se muestra en la ecuación 1.7, se remueve la variación inducida por las hojas y mejora la precisión de la recuperación del contenido de agua de la vegetación.

$$NDMI = \frac{Banda\ 5 - Banda\ 6}{Banda\ 5 + Banda\ 6} \quad \dots 1.7$$

Índice Normalizado de Áreas Quemadas (NBR)

Este índice está pensado para resaltar áreas quemadas y estimar la severidad de incendios. Para el análisis de urbanización se usa el índice normalizado de la diferencia de áreas edificadas (NDBI) (Zha, Gao, & Ni, 2003), ambos NBR y NDBI tienen la misma fórmula, ecuación 1.8.

$$NBR = \frac{Banda\ 5 - Banda\ 7}{Banda\ 5 + Banda\ 7} \quad \dots 1.8$$

Índice Normalizado de Áreas Quemadas 2 (NBR2)

Es calculado de forma similar a NBR solo que sustituyendo la Banda NIR por SWIR1 para resaltar el contenido de agua de la vegetación.

$$NBR2 = \frac{\text{Banda 6} - \text{Banda 7}}{\text{Banda 6} + \text{Banda 7}} \quad \dots 1.9$$

Si bien la mayoría de los índices espectrales fueron pensados para resaltar parámetros diferentes a contaminantes atmosféricos, esto no los limita a que sean usados para intentar determinar otros parámetros, se espera que la inclusión de los índices espectrales en los modelos de Redes Neuronales ayude a mejorar dichos modelos, sino es el caso simplemente se eliminarán de los modelos.

1.3 Datos de contaminantes en la Ciudad de México

En la Ciudad de México la instancia encargada de recolectar esta información es la Secretaría del Medio Ambiente, en cumplimiento de las Normas Oficiales Mexicanas (NOM) en materia de salud ambiental que establecen los límites permisibles para contaminantes criterio. En el caso de PM 2.5 la norma vigente es NOM-025-SSA1-2014 que establece como límite de 24 horas $45 \mu\text{g}/\text{m}^3$ y como límite anual $12 \mu\text{g}/\text{m}^3$ esto es tomando en cuenta los valores promedios respectivos.

Además, existen normas técnicas que establecen los métodos de medición para distintos contaminantes, sin embargo, no se tienen este tipo de normas propias para la medición de concentraciones de PM por lo que se usan el método equivalente que recomienda United States Environmental Protection Agency (US EPA), una institución pionera en normas ambientales.

Para la medición en la Ciudad de México y área metropolitana se cuenta con el sistema de monitoreo atmosférico (SIMAT), este sistema cuenta con 4 diferentes sub sistemas:

-La red de meteorología y radiación solar (REDMET) lo integran 26 estaciones que determinan variables meteorológicas como son la velocidad y dirección del viento, temperatura, humedad y radiación solar.

-Para la medición de contaminantes se cuenta con 2 redes, la primera es la Red Automática de Monitoreo Atmosférico (RAMA) que cuenta con 32 estaciones para la medición de distintos contaminantes, sólo 23 estaciones cuentan con equipos para la medición de PM2.5, estas mediciones se realizan de manera automática cada hora durante todo el día los 365 días del año, aunque en ocasiones dejan de funcionar por cuestiones de mantenimiento y calibración.

La segunda es la Red Manual de Monitoreo Atmosférico (REDMA) la cual cuenta con 7 estaciones que miden la concentración de contaminantes cada 6 días, se encarga de medir concentraciones de PM10, PM2.5 y partículas suspendidas totales.

-Por último, cuenta con la red de Depósito Atmosférico que se encarga de determinar la variabilidad espacial y temporal de la deposición húmeda y seca en la Zona Metropolitana de la Ciudad de México.

Todos estos subsistemas están a cargo de la Secretaría del Medio Ambiente (SEDEMA) de la Ciudad de México. Como medidas de evaluación al sistema de monitoreo atmosférico se realiza auditorías externas en las que se verifica la precisión de los equipos de medición y el control que estos llevan, en 2018 se evaluaron 9 de las 34 estaciones automáticas, todas ellas cumpliendo con los parámetros de desviación estándar que se exigen en las normas.

En este trabajo se usarán las mediciones de PM2.5 de RAMA ya que sus mediciones cada hora resultan de mucha ayuda en comparación con el REDMA, además el paso de satélite es justo a las 11:00 horas tiempo local, y la medición también se realiza a esta hora lo que ayuda al modelo, ya que no se tendrán que interpolar para conocer los valores exactos.

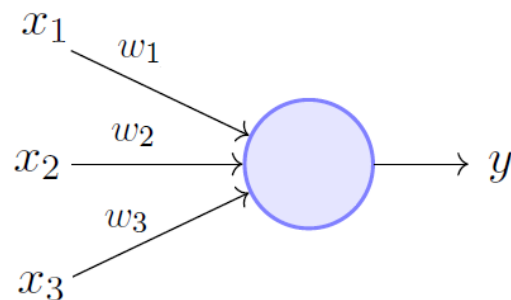
1.4 Redes Neuronales Artificiales

Desde hace ya cientos de años el hombre ha querido descifrar el funcionamiento del cerebro humano, ya que, aunque las computadoras se han hecho cada vez más rápidas y pueden resolver problemas con miles de operaciones en unos segundos, aún no pueden superar al hombre en tareas que a nosotros nos resultan sencillas, tales como el reconocimiento de imágenes y la clasificación de las mismas. Además, nos gustaría que las computadoras realizaran tareas más complejas como identificar células enfermas de las sanas o predecir fenómenos naturales.

Derivado de esta necesidad de computadoras que solucionen problemas más complejos surge la inteligencia artificial (IA), uno de los métodos de la IA son las Redes Neuronales Artificiales (RNA) inspirado en las células de nuestro propio cerebro, en específico el estudio de las neuronas, sus interconexiones, su participación como bloque fundamental del cerebro que es uno de los campos de investigación más dinámicos e importantes dentro de la biología.

Las redes neuronales artificiales desde su primer modelo en 1943 se han ido sofisticando con nuevos métodos década a década, además con computadoras que de igual forma se han ido haciendo cada vez más rápidas nos deja en una época con grandes posibilidades de aplicación e investigación en este campo del conocimiento, lo cual hasta hace unos años no sería posible.

Las RNA intentan modelar las capacidades del sistema nervioso al procesar la información, esto desde las señales de entrada de información, su procesamiento y la salida de la información. La unidad básica de estas redes es el perceptrón o neurona la cual recibe varias señales de entrada (datos) aplica una función y arroja un resultado como se muestra en la figura 1.3.



Perceptron Model (Minsky-Papert in 1969)

Figura 1.3 Diagrama de perceptrón simple

La idealización anterior se considera que la función es lineal ya que los parámetros se multiplican por coeficientes o “pesos sinópticos” y después se suman, siendo el resultado de esta suma el valor de salida. El siguiente elemento son las capas, estas capas se refieren a un conjunto de neuronas que reciben todas las mismas señales de entrada, pero cada una de estas tienen diferentes coeficientes que multiplican estas señales, lo que resulta en que cada neurona tenga diferentes señales de salida.

Si solo se tuvieran capas con perceptrones simples como el anterior, solo daría como resultados modelos de regresión lineales, lo cual no es muy útil ya que existen métodos más eficientes y más estudiados que llegarían a la solución óptima más rápido y sencillo, para añadir no linealidad a los modelos de redes neuronales se aplican funciones de activación.

1.4.1 Funciones de Activación

La función de activación tiene como principal propósito convertir el resultado de la suma del nodo en una señal de salida, aplicando alguna función no lineal, que será usada por la siguiente capa de la red, en la figura 1.4 se muestra la conceptualización de una neurona donde se aplica la función sigmoide como función de activación. Esta conversión es importante en las redes neuronales artificiales para agregar una respuesta compleja no-lineal entre las entradas y la señal de salida.

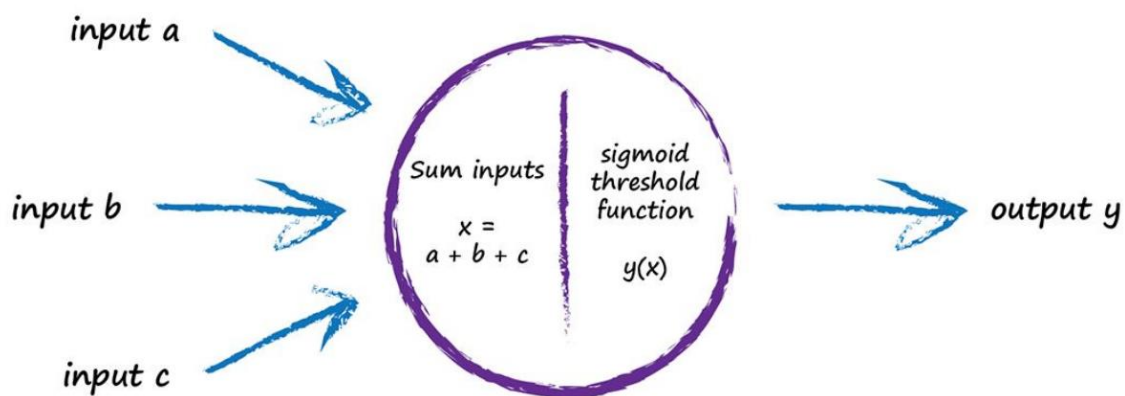


Figura 1.4 Diagrama de un perceptrón con función de activación sigmoide (Rashid, 2016)

Si no se aplicaran las funciones de activación se puede demostrar que el resultado de la red neuronal es simplemente una función polinomial lineal, este tipo de funciones tienen un poder de aprendizaje bastante limitado, ya que sería lo mismo que un modelo de regresión lineal.

Se requiere de la no-linealidad ya que no sólo es necesario aprender y aplicar modelos lineales, el mundo está lleno de problemas no lineales, es importante que las redes neuronales puedan aprender casi cualquier patrón en una variedad extensa de datos como son imágenes, videos, audios, discursos. Por eso se considera a las RNA como

aproximadores de funciones universales (Kurt Hornik, Maxwell Stinchcombe, & Halbert White, 1989).

Entre las funciones de activación más populares están:

- 1) **Función Sigmoide** es el nombre que se le da a la función $f(x) = \frac{1}{1+e^{-x}}$, fue una de las primeras funciones que se aplicaron a las redes neuronales artificiales ya que es fácil de entender y aplicar, pero presentan muchas desventajas por lo cual no son muy utilizadas excepto en textos introductorios de redes neuronales, figura 1.5:
 - Problema de desvanecimiento del gradiente, esto es que en valores extremos el gradiente es muy pequeño por lo que nunca convergen o tardan mucho en hacerlo
 - El cero no está centrado, esto hace que las capas subsiguientes se vuelvan a activar si la salida de la capa es cero.
 - "Mata neuronas", esta función hace que, si por alguna razón los alguno de los pesos sinópticos toma negativos extremos, esa neurona no vuelva a poder ser activada de nuevo en ninguna iteración posterior.

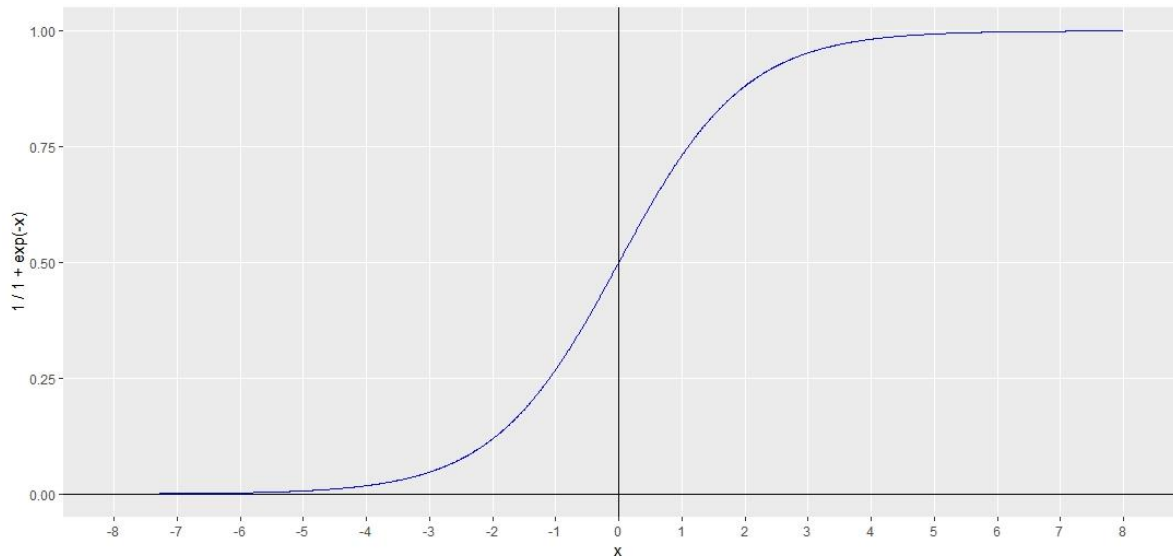


Figura 1.5 Gráfica de la función sigmoide. Elaboración propia.

- 2) **Tangente hiperbólico:** su fórmula matemática es $f(x) = \frac{e^{2x}-1}{e^{2x}+1}$ esta función presenta el cero centrado, figura 1.6, lo que resuelve un problema de la función

sigmoide, por ello en la práctica se prefiere el uso de esta función. Aun así, sufre el problema de desvanecimiento del gradiente.

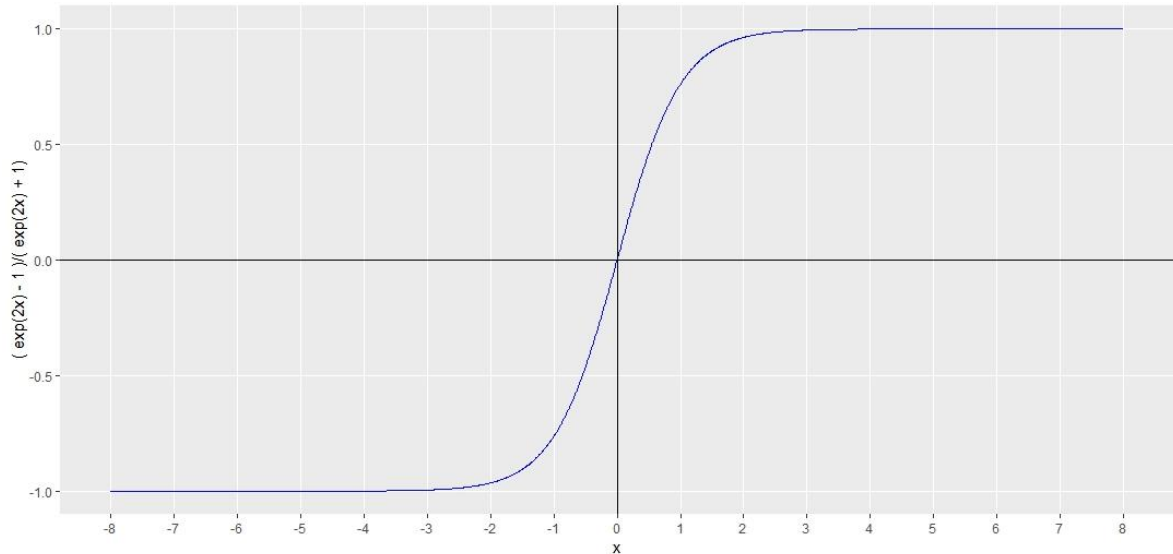


Figura 1.6 Gráfica de la función tangente hiperbólico. Elaboración propia.

3) Función lineal rectificadora (ReLU por sus siglas en inglés): Es la función más popular de los últimos años, ya que se ha demostrado que puede llegar a mejorar la convergencia hasta 6 veces respecto a la función Tanh. Se puede observar en la figura 1.7, su expresión matemática es:

$$R(x) \begin{cases} x & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Lo que la hace bastante simple y eficiente, evita todos los problemas de las dos funciones anteriores por lo que se usa prácticamente en todos los problemas en RNA, la única limitación es en usarla en las capas escondidas.

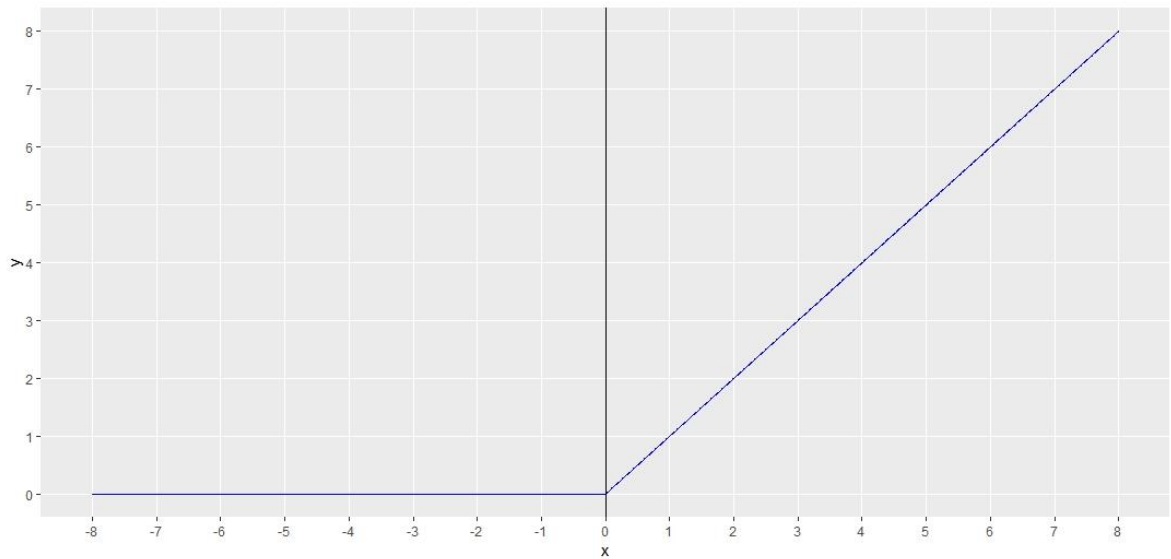


Figura 1.7 Gráfica de la función lineal rectificadora. Elaboración propia.

1.4.2 Red neuronal multicapa

Existen distintos tipos de arquitecturas de redes neuronales, éstas consisten en las diferentes formas en que están conectadas los perceptrones unos con otros, y la información de entrada que están recibiendo, ya que algunos modelos incluyen una variable de memoria que recuerda el resultado de las ultimas iteraciones de la red, en la figura 1.8 podemos ver algunas de estas arquitecturas.

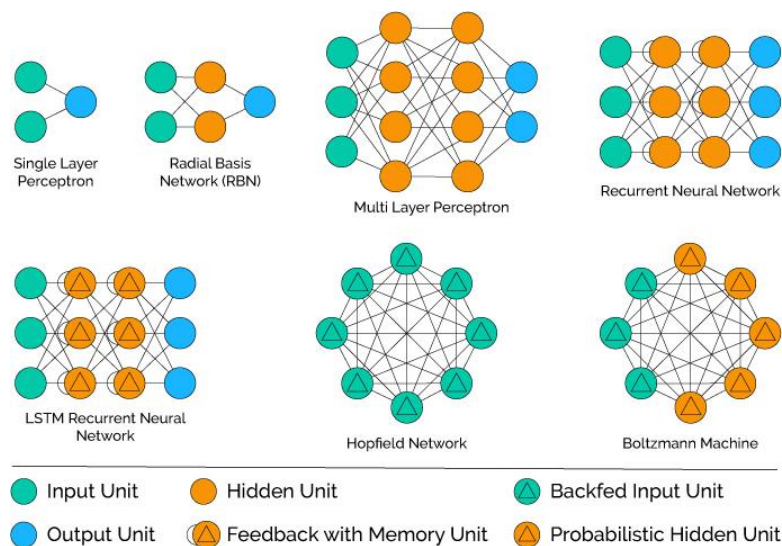


Figura 1.8 Diferentes arquitecturas de Redes Neuronales Artificiales.(Singh Gill, 2017)

La arquitectura adecuada para modelos de regresión lineal es el de múltiples capas de perceptrones cada una con diferente número de neuronas, todas las neuronas de una capa se interconectan con las neuronas de la capa anterior y siguiente, pero como sucede en la naturaleza algunas neuronas tienen una conexión más fuerte y otras más débiles dependiendo de los pesos sinápticos ($w_{x,y}$), la figura 1.9 ilustra mejor este proceso.

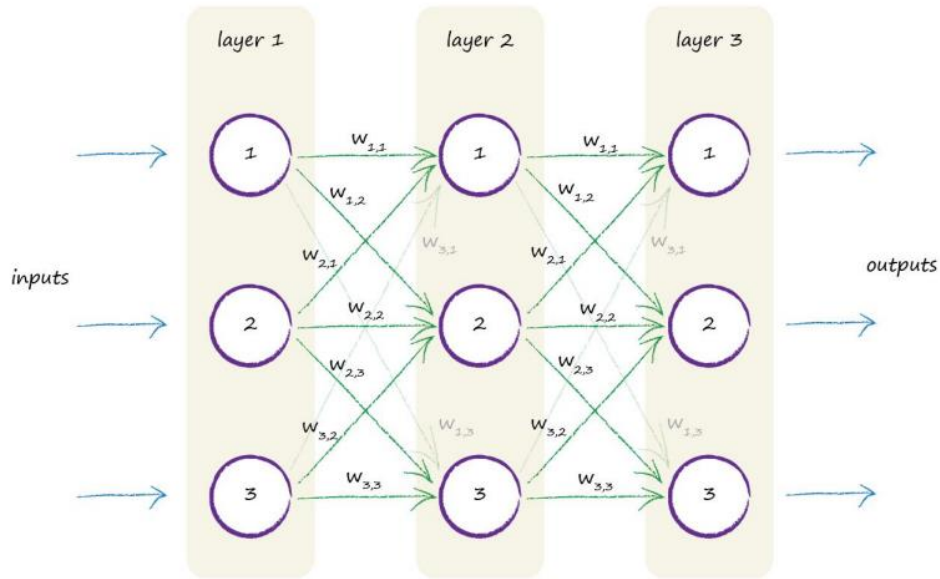


Figura 1.9 Arquitectura multicapa de Perceptrones. (Rashid, 2016)

Si en la red anterior las funciones son únicamente lineales, sin funciones de activación, se puede demostrar que para pasar de una capa a otra se puede representar por medio de matrices de la siguiente forma. Una demostración sencilla puede ser encontrada en la referencia ⁴ (Rashid, 2016).

$$\begin{bmatrix} w_{1,1} & w_{2,1} & w_{3,1} \\ w_{1,2} & w_{2,2} & w_{3,2} \\ w_{1,3} & w_{2,3} & w_{3,3} \end{bmatrix} \begin{pmatrix} entrada_1 \\ entrada_2 \\ entrada_3 \end{pmatrix} = \begin{pmatrix} salida_1 \\ salida_2 \\ salida_3 \end{pmatrix} = \begin{pmatrix} (w_{1,1} * e_1) + (w_{2,1} * e_2) + (w_{3,1} * e_3) \\ (w_{1,2} * e_1) + (w_{2,2} * e_2) + (w_{3,2} * e_3) \\ (w_{1,3} * e_1) + (w_{2,3} * e_2) + (w_{3,3} * e_3) \end{pmatrix}$$

Si se quisiera aplicar la función de activación se hace aplicando a cada elemento de salida. Para pasar a la siguiente capa, se usan las salidas de la capa 1 y se convierten en las entradas de la capa 2, se sigue con este mismo planteamiento hasta la última capa. Para empezar a usar un poco la terminología, la primera capa de la figura 1.9 se denomina la capa de entrada, la capa 2 se le denomina capa oculta y la última capa se le llama capa de salida, cada capa puede tener cualquier número de neuronas y se puede tener cualquier

⁴ Las variables e_1 , e_2 y e_3 corresponden a las variables de entrada_1, entrada_2 y entrada_3

número de capas ocultas, usualmente las primeras capas tienen mayor número de neuronas. El cómo elegir el número de capas y los elementos es aún un tema de estudio y generalmente se hacen recomendaciones empíricas.

Antes de explicar el último proceso de las redes neuronales, se necesita entender el objetivo de la red neuronal, este es poder predecir datos de salida únicamente conociendo los datos de entrada, pero para ello se necesitan grandes bases de datos, con datos de entrada y salida conocidos. Por ejemplo, se tienen mediciones de altura, peso, edad de muchas personas con su respectivo género y se quiere poder predecir el género solo con estos tres datos.

Hasta ahora se ha explicado el funcionamiento de una “red neuronal aleatoria” ya que al meter los datos de entrada nos arroja valores de salida, pero todos ellos aleatorios, y que dependen de las constantes de peso ($w_{x,y}$), pero el objetivo es que estas salidas coincidan con los datos. Esto se logra variando las w y para ello existen varios métodos, uno muy conocido es propagación hacia atrás, que consiste en que el error entre el valor arrojado por la red y el valor real, es dividido entre todos los pesos(w) de manera proporcional. Aunque la idea central de dividir el error entre todos los pesos es intuitiva, realmente no es la forma de actualizar los pesos, ya que las neuronas no se tratan de simples funciones lineales, es necesario aplicar la función de activación por lo que repartir el error no es tarea sencilla, basta observar la ecuación 1 que consiste en un nodo de salida de una red neuronal de 3 capas con 3 neuronas cada una y con función sigmoide como función de activación.

$$O_k = \frac{1}{1 + e^{-\sum_{j=1}^3 \left(w_{j,k} * \frac{1}{1 + e^{-\sum_{i=1}^3 (w_{i,j} * x_i)}} \right)}} \quad \dots 1.10$$

Donde:

O_k es la salida del nodo k

x_i es la entrada en el nodo i

$w_{i,j}$ son los pesos entre los nodos

Intentar adivinar los valores w exactos sería complicado incluso con una red neuronal pequeña, este acercamiento se le llama método de fuerza bruta y resultaría impráctico en redes mucho más grandes.

En vez de usar la fuerza bruta se suele utilizar un método llamado descenso por gradiente que consiste en modificar los valores de los pesos dependiendo de la pendiente de la función de costo (función de error), en la figura 1.10 se ilustra este método, obsérvese que lo que se intenta minimizar es el error.

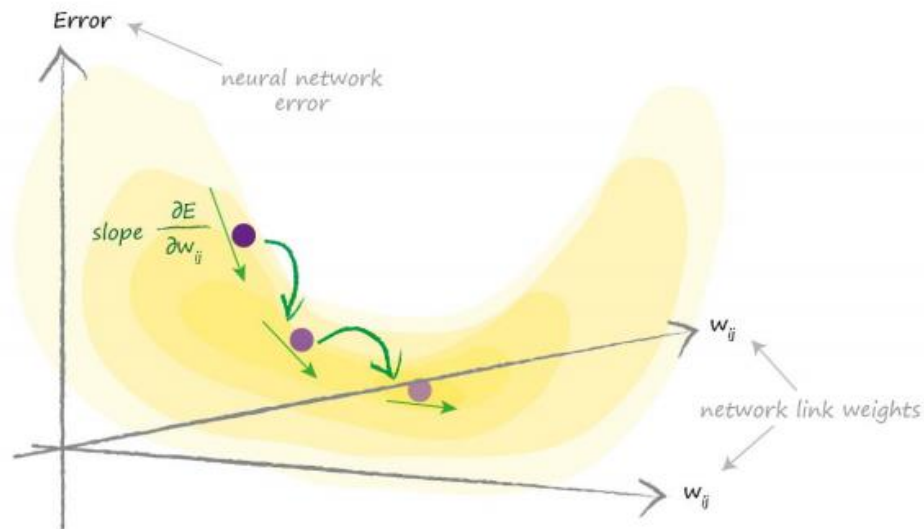


Figura 1.10 Esquema método del descenso por gradiente

1.4.3 Funciones de error

Antes de calcular $\frac{\delta E}{\delta w_{ij}}$ hay que hablar acerca de la función de error E . Se pueden aplicar cualquier tipo de función de error o de costo, esto depende del problema específico que se desee resolver.

Muchas veces usamos $e = (\text{valor medido} - \text{valor actual})$, pero E es la suma de todos los errores entonces es posible que al sumar todos los errores se cancelen y sumen 0, por lo cual no es una muy buena opción.

Otra función de error podría ser en valor absoluto $|\text{valor medido} - \text{valor actual}|$, pero este tiene la desventaja de no ser continuo cerca del mínimo, por lo que es difícil calcular este valor.

Otra función más popular es la de (valor medido – valor actual)² y esta es la función que se ocupará en este trabajo para los cálculos siguientes, sus ventajas son de que es continua en todo el rango, el gradiente es pequeño cerca del mínimo por lo que es difícil sobreestimar.

1.4.4 Propagación hacia atrás

Con las consideraciones previas, se está en condiciones de calcular $\frac{\delta E}{\delta w_{ij}}$, con la función de error = (Valor Medido – Valor actual)² y función de activación sigmoide.

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial}{\partial w_{jk}} (Vm_k - Va_k)^2 \quad \dots 1.111$$

Donde

Vm- Valor medido u objetivo

Va- Valor actual o valor de salida

Aplicando regla de la cadena

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial Va_k} * \frac{\partial Va_k}{\partial w_{jk}} \quad \dots 1.122$$

Resolviendo $\frac{\partial E}{\partial Va_k}$

$$\frac{\partial E}{\partial w_{jk}} = -2(Vm_k - Va_k) * \frac{\partial Va_k}{\partial w_{jk}} \quad \dots 1.13$$

Sustituyendo Va_k

$$\frac{\partial E}{\partial w_{jk}} = -2(Vm_k - Va_k) * \frac{\partial}{\partial w_{jk}} \text{sigmoid} \left(\sum_j w_{jk} Va_k \right) \quad \dots 1.14$$

El objetivo de este trabajo no es deducir la derivada parcial de la función sigmoide, se pueden consultar textos de cálculo para su deducción.

$$\frac{\partial}{\partial x} \text{sigmoid}(x) = \text{sigmoid}(x) * (1 - \text{sigmoid}(x)) \quad \dots 1.135$$

Continuando

$$\frac{\partial E}{\partial w_{jk}} = -2(Vm_k - Va_k) * \text{sigmoid}\left(\sum_j w_{jk} Va_k\right) * \left(1 - \text{sigmoid}\left(\sum_j w_{jk} Va_k\right)\right) \quad 1.146$$

$$* \frac{\partial}{\partial w_{jk}} \left(\sum_j w_{jk} Va_k\right)$$

$$\frac{\partial E}{\partial w_{jk}} = -2(Vm_k - Va_k) * \text{sigmoid}\left(\sum_j w_{jk} Va_k\right) * \left(1 - \text{sigmoid}\left(\sum_j w_{jk} Va_k\right)\right) * Va_k \quad \dots 1.157$$

La expresión anterior sirve para calcular el nuevo valor de w_{jk} para ello se utiliza la siguiente expresión, generalmente se usa el coeficiente α que se le conoce como factor de aprendizaje, este sirve para moderar la fuerza del cambio y asegurar que no exceder el valor buscado.

$$\text{nueva } w_{jk} = \text{vieja } w_{jk} - \alpha * \frac{\partial E}{\partial w_{jk}} \quad \dots 1.168$$

El aprendizaje de la red neuronal consiste en determinar valores de pesos sinápticos que permitan calcular los datos de salida con el menor error posible, esto se logra con el método de propagación hacia atrás, lo que conlleva a la realización de muchas iteraciones que dependen del tamaño de la base de datos de entrada, es prácticamente imposible llegar a una solución óptima por lo que existen criterios que detener las interacciones.

- Cuando se llega a un número de interacciones específico
- Cuando el error se considera suficientemente pequeño
- Cuando llega un punto que por más que se entrene los errores no se reducen

Recapitulando las redes neuronales artificiales es un método basado en la red neuronal humana, se compone por varias capas de neuronas, cada neurona aplica una ecuación lineal seguida por una función de activación, la función lineal depende de los pesos (w) y

generalmente esto son aleatorios al inicio del proceso. Para entrenar la red se le ingresa un conjunto de datos y los datos de salida son comparados con los valores reales a través de una función de error, para modificar los valores w se hace un proceso que se le denomina propagación hacia atrás en la cual se obtiene el gradiente de la función de error se multiplica por un coeficiente de aprendizaje y se le resta al valor anterior w .

Capítulo 2 Estado del arte

La contaminación atmosférica es un problema que se está estudiando mucho en los últimos años, tan sólo en los últimos 10 años se han publicado casi 1000 artículos relacionados con la contaminación atmosférica, se estudia desde su relación con variables atmosféricas, su composición química, su origen, la influencia de vegetación, estudios temporales, estudios espaciales, impacto sobre la salud humana, percepción remota para mediciones de contaminantes, etc.

Los artículos relacionados con el cálculo de PM_{2.5} por medio de percepción remota son aproximadamente 110 artículos en ninguno de ellos se ocupan los satélites Landsat, en cambio predomina el uso del satélite MODIS.

El artículo (Di et al., 2016) resulta bastante interesante ya que modela el mapa de concentraciones de PM_{2.5} con redes neuronales de todo EUA con ayuda del sensor MODIS, los datos de contaminantes fueron adquiridos de EPA Air Quality System, los datos atmosféricos los obtuvieron de North American Regional Reanalysis. Con ellos crearon una red neuronal, esta red se aplicó a mapas de Estados Unidos arrojando varios mapas de concentraciones de PM_{2.5} como el de la figura 2.1, que es similar al objetivo de este trabajo.

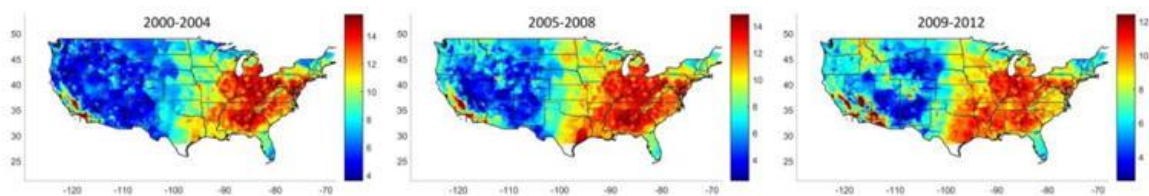


Figura 2.1 Red Neuronal entrenada para predecir Concentraciones de PM_{2.5} con una malla de celdas de 1kmx 1km en Estados Unidos de América.(Di et al., 2016)

Por otro lado, se han presentado muchos trabajos de calidad del aire para la Ciudad de México, aunque solo 3 de ellos usan Redes Neuronales para el cálculo de concentraciones

de contaminantes, esto es desde otra perspectiva que es aplicando a modelación de series de tiempo.

-En el trabajo de tesis (V. H. Torres, 2017) para obtener el título de Ingeniero Mecánico aplica las Redes Neuronales como modelo de predicción de series de tiempo, prediciendo puntos IMECA promedio diarios, figura 2.2. Uno de los puntos a destacar es que en este trabajo se usó algoritmos genéticos como forma de optimizar la red neuronal, en vez de los métodos tradicionales de Redes Neuronales.

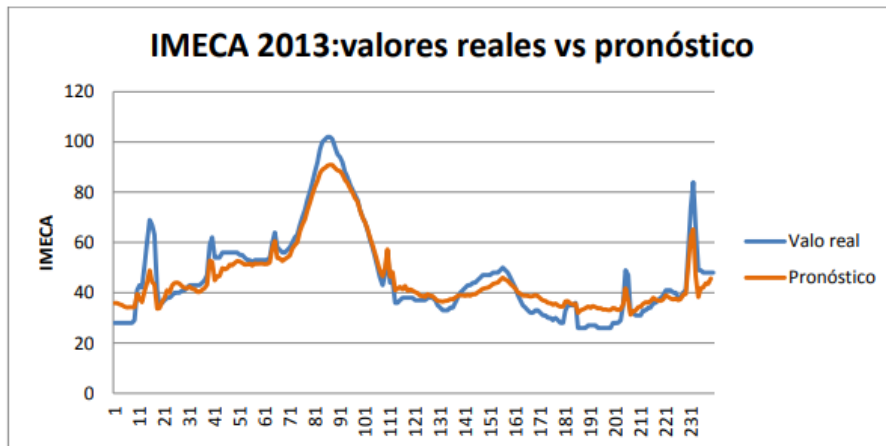
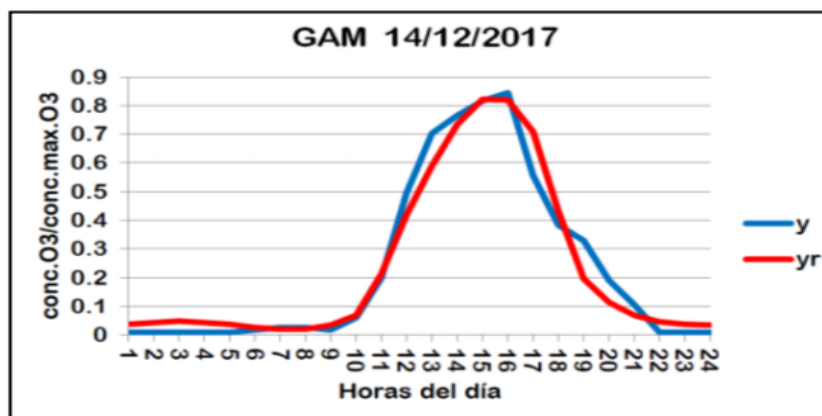


Figura 2.2 Gráfica IMECA mensual, real vs pronostico con RNA.(V. H. Torres, 2017)

- En la tesis de (Arellano Monreal, 2019) para obtener el título de Maestro en Ciencia e Ingeniería de la Computación propone de nuevo el uso de RNA como modelos de predicción de series de tiempo para concentraciones de ozono, a diferencia del trabajo de (Camacho González, 2018) la predicción que se hace es de tipo horario figura 2.3.



(c) Gustavo A. Madero

Figura 2.3 Gráfica de predicción concentraciones de ozono horario.(Arellano Monreal, 2019)

- En la tesis (Camacho Gonzáles, 2018) para obtener el título licenciado en ciencias de la Computación, realizo 22 redes neuronales con arquitecturas un poco diferentes, entre estos cambios están los datos de entrada, se modificó la función de costo, y se realizaron con y sin parámetros meteorológicos. Los resultados fueron series de tiempo de ozono por semana, algunos bastantes precisos como el de la figura 2.4. El único problema es que puede estar sobre ajustando para obtener esta precisión, en esta tesis no se hace mención a intentar aplicar este modelo a valores futuros, tampoco hace referencia a usar un conjunto de datos de evaluación.



Figura 2.4 Gráfica de predicción concentraciones de ozono promedio diario. (Camacho Gonzáles, 2018)

Aunque estas últimas tesis no siguen el método que se desea aplicar, es necesario analizar sus resultados, ya que usan los mismos datos de SEDEMA que se emplean en este trabajo, por ejemplo, de (Arellano Monreal, 2019) se pueden rescatar los resultados obtenidos, y haciendo especial hincapié en una de las conclusiones a las que llegó, mencionando que existe una correlación entre los errores y la altitud de las estaciones, por lo que se puede analizar la posibilidad de usar este parámetro en caso de tener bajas correlaciones.

Capítulo 3 Metodología

3.1 Descarga de imágenes satelitales

Para los modelos se necesitan datos de reflectancia TOA y reflectancia superficial SR, el Landsat cuenta precisamente con estos productos de nivel 2, esto significa que deben ser solicitados ya que requieren un procesamiento. Adicionalmente se ocupó banda 8 que no está en ninguno de los dos productos anteriores por lo que se descargó las imágenes de nivel 1 (DN).

Lo primero es saber las fechas que serán útiles, ya que muchas imágenes desafortunadamente presentan demasiada nubosidad y no serán utilizadas. Hasta el 31 de diciembre de 2018 el Landsat 8 ha capturado 123 escenas, 5 de ellas no están disponibles alguno de los productos, 74 no se pueden ocupar por la alta nubosidad por lo que nos deja con 44 escenas utilizables para nuestros propósitos.

La forma más eficiente que se encontró para descargar los tres productos que queríamos fue mediante la siguiente serie de pasos.

- 1) Crear una cuenta de United States Geological Survey (USGS)
- 2) Entrar en Earth Explorer <https://earthexplorer.usgs.gov/>
- 3) Ingresar las coordenadas cercanas al área de estudio, en la pestaña de conjunto de datos seleccionar “Landsat 8 OLI/TIRS C1 Level-1”, y en criterios adicionales seleccionar “solo imágenes de día”.
- 4) Excluir de la selección aquellas imágenes que no se vayan a ocupar. Exportar los resultados en formato valores separados por coma (.csv por sus siglas en inglés). Este archivo contendrá los metadatos de las escenas seleccionadas, lo que nos interesa es el Landsat Product Identifier que se almacena en un documento de texto y se nombra “espa_request.txt”
- 5) Se accede a la página <https://espa.cr.usgs.gov/> y se realiza una nueva orden, ahí nos pedirá el archivo de “espa_request.txt” y seleccionaremos los productos que necesitamos, “Input Products”, “Surface Reflectance”, “Top of Atmosphere Reflectance” y “Spectral Indices”. Incluso se podrían solicitar “Brightness Temperature” aunque en el proyecto se calculó con ayuda del software R.

- 6) Una vez hecha la orden se notificará por correo cuando las escenas estén disponibles para su descarga, se tiene un periodo de 1 semana para descargarlas después se eliminarán.

3.2 Datos de contaminantes

Se usarán solamente los datos de la red automática de monitoreo atmosférico (RAMA), los datos se encuentran de manera gratuita en la página <http://www.aire.cdmx.gob.mx>, se descargaron los datos desde el año 2013, año en que comenzó a operar el satélite Landsat 8, se optó por la base de datos en formato de valores separados por coma (.csv por sus siglas en inglés). La ubicación de las estaciones se muestra en la figura 3.1, se observa que existen estaciones en el Estado de México, aunque son administradas por la Ciudad de México, esto como forma de tener una mejor distribución de muestras, ya que gran parte de la contaminación no se genera dentro de la ciudad sino en la zona metropolitana.

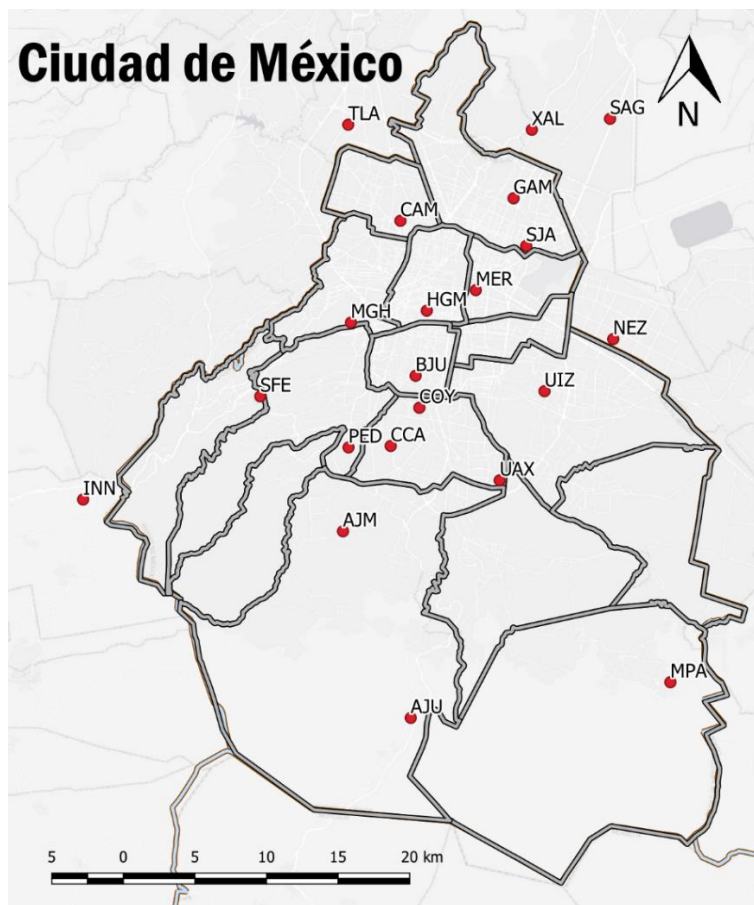


Figura 3.1 Mapa de ubicación de estaciones de medición PM2.5. Elaboración propia.

Se optó por usar el software R para el procesamiento de estos datos y realizar operaciones estadísticas con ellos, también para el manejo de las imágenes satelitales, por lo cual se volvió una herramienta indispensable en este trabajo.

El formato en que se presentan los datos es que cada fila representa un dato de una estación y de una fecha, lo que hace que la cantidad de filas sea tan grande que no se pueda abrir completamente en Excel. Por lo que el primer paso que se hizo es pasar estos datos a un formato más eficiente para ser manipulados con R. En la figura 3.2 se observa mejor los dos formatos, el inicial y el resultado después recomendarlos con R.

date	id_station	id_parameter	value	unit	Fecha	Estacion	CO	NO	NO2	NOX	O3	PM10	SO2	PM2.5	PMCO
01/01/2013 01:00	ACO	CO	1.5	15	01/01/2013 01:00	ACO	1.5	23	30	54	2	127	6	NA	NA
01/01/2013 01:00	ACO	NO	23	1	01/01/2013 01:00	ATI	NA	48	31	78	NA	NA	4	NA	NA
01/01/2013 01:00	ACO	NO2	30	1	01/01/2013 01:00	CAM	1.7	76	30	106	1	72	7	55	17
01/01/2013 01:00	ACO	NOX	54	1	01/01/2013 01:00	MON	NA	5	25	29	12	NA	NA	NA	NA
01/01/2013 01:00	ACO	O3	2	1	01/01/2013 01:00	CHO	NA	24	25	48	3	NA	5	NA	NA
01/01/2013 01:00	ACO	PM10	127	2	01/01/2013 01:00	COY	NA	49	29	78	3	NA	NA	75	NA

Figura 3.2 El lado izquierdo está el formato del archivo descargado y a la derecha con el formato después de recomendar datos. Elaboración propia.

El paso siguiente fue analizar los datos de PM2.5 con ayuda de un gráfico de caja, la figura 3.3 muestra el gráfico de caja de los promedios diarios, agrupados por mes. En la gráfica podemos observar que el límite diario ($45 \mu\text{g}/\text{m}^3$) es superado al menos unas 20 veces en los últimos 6 años, es preocupante que existan algunos días donde se presenta una concentración promedio de casi $90 \mu\text{g}/\text{m}^3$ que es el doble del límite diario de las normas.

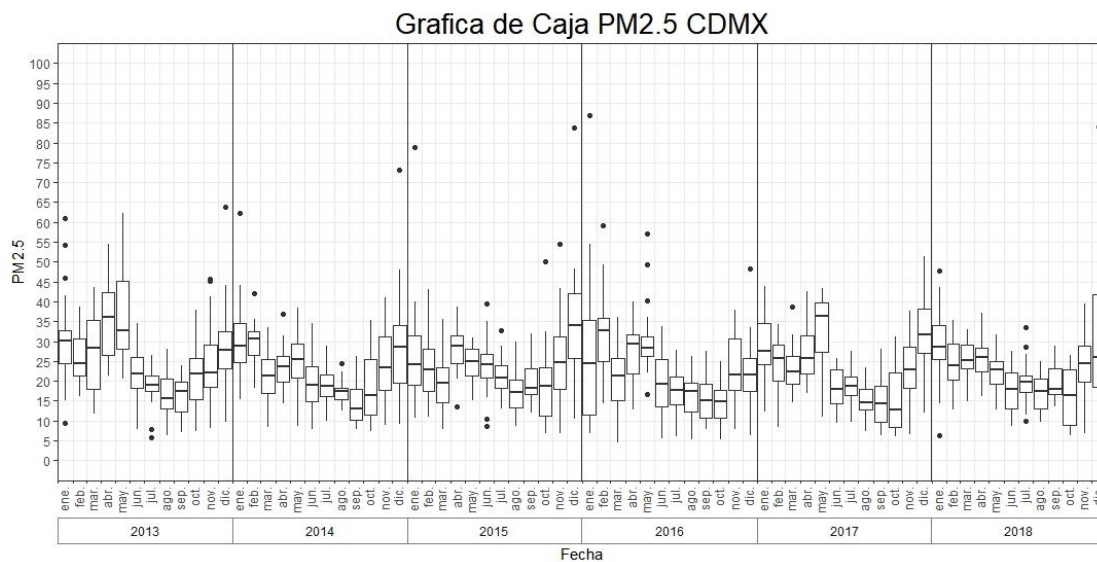


Figura 3.3 Gráfico de caja concentraciones promedio diarias PM2.5, 2013-2018. Elaboración propia.

La creación de un modelo que sólo necesitará como datos de entrada las imágenes satelitales, sería ideal, esto porque no se dependería de la disponibilidad de otra base de datos que no fuera la escena del satélite. Sin embargo, revisando la literatura la mayoría de modelos se apoyan en datos meteorológicos para la generación de modelos de Redes Neuronales Artificiales. Por ello se optó por descargar la base de datos del Programa de estaciones Meteorológicas del Bachillerato Universitario (PEMBU) desde <https://www.ruoa.unam.mx/pembu/#> , ya que a diferencia de los datos meteorológicos que proporciona SEDEMA los datos del PEMBU incluyen precipitación acumulada.

Las variables que incluye la base PEMBU son temperatura, humedad relativa, rapidez del viento, dirección del viento, rapidez de rachas, dirección de rachas, presión barométrica, precipitación, radiación solar, índice UV y dosis UV. Cuenta con 15 estaciones de monitoreo que se ubican en las 9 preparatorias de la UNAM, más los 5 Colegios de Ciencias y Humanidades y una más en el Centro de Ciencias de la Atmósfera de la UNAM, además las mediciones se realizan cada 30 minutos.

Aunque en este trabajo sólo se consideraron las variables de precipitación, humedad relativa y temperatura, esto para sencillez del modelo, en futuros trabajos se podrían considerar el uso de más variables.

Antes de poder ocupar esta base de datos hay que unirla ya que la forma en que está dividida por año y por estación lo que complica su manejo en R. Se realizó un script en R que une toda la información de las estaciones en un solo archivo. Posteriormente se realizaron las gráficas de la figura 3.4 y 3.5, para observar si las variables de precipitación y temperatura están correlacionadas con el PM.

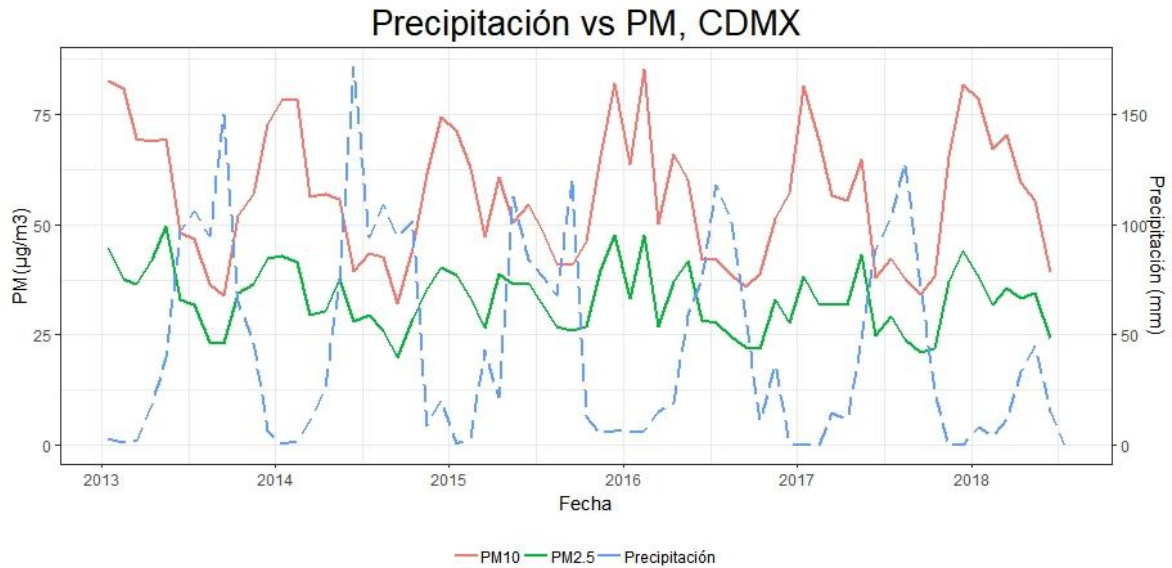


Figura 3.4 Gráfica de precipitación contra concentración de PM promedio. Elaboración propia.

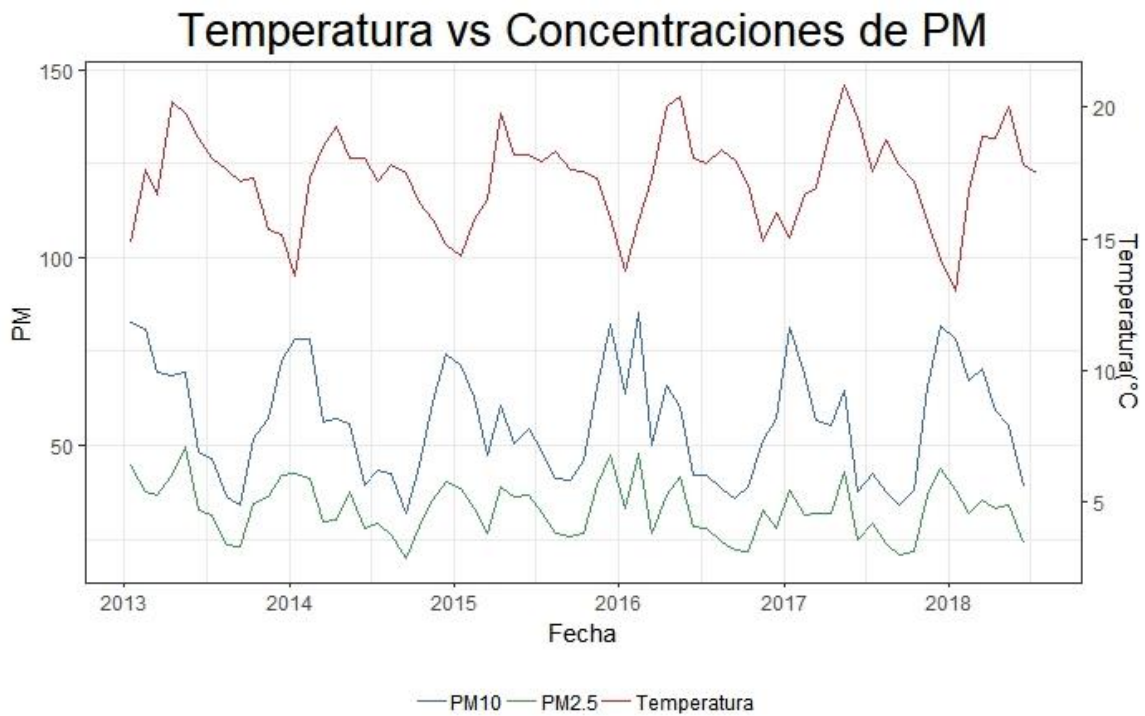


Figura 3.5 Gráfica de temperatura contra concentración de PM promedio. Elaboración propia.

Podemos concluir de la figura 3.4 que se tiene una relación entre la precipitación y las concentraciones de PM ya que cuando se presentan los máximos valores de precipitación la concentración promedio de PM disminuye, de igual forma se tiene una correlación entre temperatura y PM, cuando la temperatura disminuye las concentraciones de PM aumentan.

El problema que se tienen es que tanto precipitación y temperatura también estén correlacionadas, lo que generaría que habría que usar una u otra variables, se espera que el mismo modelo de Redes Neuronales seleccione la variable más útil.

3.3 Armado de la base de datos

Nos referimos con base de datos a una tabla que contienen toda la información que será ingresada a la Red Neuronal, en una columna estarán las mediciones de PM2.5 que es nuestra variable dependiente y en las demás columnas nuestras variables independientes o predictoras, una característica importante es que los datos no deben presentar valores nulos por lo que se tendrán filtrar los datos para limpiarlos.

3.3.1 Extracción de datos de imágenes satelitales

Una vez que se tienen las imágenes satelitales y los datos de contaminantes se prosiguió con la ubicación de las estaciones de monitoreo, en la tabla 3.1 se muestran las ubicaciones de las estaciones que miden PM2.5.

Tabla 3.1 Coordenadas de ubicación estaciones de medición PM2.5. Con datos de SEDEMA.

cve_estac	Nombre estación	longitud	latitud
AJU	Ajusco	-99.1626	19.15429
AJM	Ajusco Medio	-99.2077	19.27216
BJU	Benito Juárez	-99.1596	19.37046
CAM	Camarones	-99.1698	19.4684
CCA	Centro de Ciencias de la Atmósfera	-99.1761	19.32611
COY	Coyoacán	-99.1571	19.35026
GAM	Gustavo A. Madero	-99.0945	19.4827
HGM	Hospital General de México	-99.1522	19.41162
INN	Investigaciones Nucleares	-99.3805	19.29197
MER	Merced	-99.1196	19.42461
MGH	Miguel Hidalgo	-99.2027	19.40405
MPA	Milpa Alta	-98.9902	19.1769
NEZ	Nezahualcóyotl	-99.0282	19.39373
PED	Pedregal	-99.2041	19.32515

cve_estac	Nombre estación	longitud	latitud
SAG	San Agustín	-99.0303	19.53297
SFE	Santa fe	-99.2629	19.35736
SJA	San Juan Aragón	-99.0861	19.45259
TLA	Tlalnepantla	-99.2046	19.52908
UAX	UAM Xochimilco	-99.1036	19.30444
UIZ	UAM Iztapalapa	-99.0739	19.36079
XAL	Xalostoc	-99.0824	19.526

Estas estaciones se deben ubicar en el espacio en algún formato adecuado, por ejemplo, el formato shape (.shp). Para el procesamiento de datos espaciales, se usó en este trabajo el software libre QGIS, asegurándose en todo momento que el sistema coordinado de referencia fuera el mismo entre las imágenes vectoriales y las imágenes satelitales raster de Landsat 8.

Para extraer los valores de reflectancia de los píxeles donde se encuentran las estaciones, se plantearon dos formas:

La primera con la herramienta plugin Point Sampling Tool de QGIS, que tiene como desventaja principal el ser muy lento la adquisición de valores.

La otra es con el software libre R, con ayuda de la función `extract()` de la librería `raster`, con la desventaja de que es necesario escribir el código para extraer los valores de todas las fechas, pero con la gran ventaja de que una vez teniendo este script el proceso es mucho más rápido y práctico en el caso de agregar más imágenes o cambiar la ubicación de las estaciones.

Este último programa arroja una tabla en formato .csv como en la figura 3.6, donde se extrajeron las reflectancias atmosféricas, reflectancias superficiales, reflectancias TOA, índices espectrales y valores de otras imágenes incluidas en el producto descargado como son `Pixel_QA` y `Aerosol_QA`, que indican la presencia de errores en las mediciones o presencia de aerosoles de manera cualitativa.

Fecha	cve_estac	SR.B1	SR.B2	SR.B3	SR.B4	SR.B5	SR.B6	SR.B7	TOA.B1	TOA.B2	TOA.B3	TOA.B4	TOA.B5	TOA.B6	TOA.B7
20/05/2013	AJU	357	593	1075	1159	2297	2424	1871	1301	1230	1302	1286	2336	2351	1737
20/05/2013	AJM	1018	1093	1508	1858	2218	2399	2278	1652	1523	1620	1877	2242	2324	2113
20/05/2013	BJU	832	921	1098	1397	1771	2021	1700	1526	1395	1272	1460	1807	1959	1578

TOA.B7	ND.B8	TOA.B9	TOA.B10	TOA.B11	RA.B1	RA.B2	RA.B3	RA.B4	RA.B5	RA.B6	RA.B7	Pixel_QA	Aerosol_QA
1737	10934	13	304.1383667	301.1091003	944	637	227	127	39	-73	-134	322	160
2113	12476	10	308.8526306	305.0838928	634	430	112	19	24	-75	-165	322	66
1578	11095	9	307.8230286	304.1011353	694	474	174	63	36	-62	-122	322	96

Figura 3.6 Ejemplo formato de extracciones. Elaboración propia.

3.3.2 Adición de datos PM, humedad, temperatura y precipitación

El siguiente paso que se hizo fue incorporar al análisis los contaminantes, para ello hay que recordar que la imagen del satélite Landsat 8 es tomada a las 17:00 GTM que en hora local son las 11:00 am, por lo que es necesario filtrar los datos de SEDEMA para que coincidan con la hora y el día en que se tomaron las imágenes.

En cuanto a los datos de humedad y precipitación, como ya se había dicho, son obtenidos mediante la red PEMBU, por lo tanto, no coinciden con las ubicaciones de la RAMA por ello se propuso usar el promedio de humedad y temperatura a las 11:00 horas y añadir ese valor a todos los datos de ese día. En el caso de la precipitación se añadieron los valores acumulados de 1 día, 3 días, 5 días y 15 días antes de la toma de la imagen, esto con la intención intentar separar las imágenes en estiaje y lluvia.

3.4 Variables adicionales

Una vez ordenados los datos que se podrán ocupar para los modelos, es necesario todavía adecuar dos de las variables mencionadas anteriormente. Se tiene que convertir la variable “Aerosol QA” ya que ofrece información cualitativa de la cantidad de aerosoles presentes en ciertos pixeles o incluso la presencia de nubes, estos datos están en forma de número como se muestra en la tabla 3.2, por lo tanto, hay que crear “variables ficticias” (dummy variable) para que puedan ser ocupadas en el modelo. La variable ficticia ayuda a convertir variables categóricas en una variable de 0 y 1, cero en caso de que no presente la característica y uno en caso de que sí la presente, en este caso se crearon 4 variables ficticias “aerosol bajo”, “aerosol medio”, “aerosol alto” y “nubes”, por ejemplo, si la variable “Aerosol QA” en un pixel es igual a 224 se asigna un 1 a “aerosol alto” y 0 a las demás variables.

Tabla 3.2 Landsat 8 valores de pixel de la imagen SR_aerosol.(U.S. Geological Survey, 2018)

Attribute	Pixel Value
Fill	1
Valid Aerosol Retrieval (center pixel of 3x3 window)	2, 66, 130, 194
Water Pixel (or water pixel was used in the fill-the-window interpolation)	4, 68, 100, 132, 164, 196, 228
Cloud or Cirrus	8, 72, 136, 200
Cloud Shadow	16, 80, 144, 208
Non-center window pixel for which aerosol was interpolated from surrounding 3x3 center pixels	32, 96, 100, 160, 164, 224, 228
Low-level aerosol	66, 68, 72, 80, 96, 100
Medium-level aerosol	130, 132, 136, 144, 160, 164
High-level aerosol	194, 196, 200, 208, 224, 228

La otra variable ficticia que se creó es la variable “Estiaje”; la cual es una variable binaria que describe si se está en época de estiaje o lluvias. La importancia de saber si está en estiaje o lluvias es que cambia la reflectancia de algunas superficies en presencia o ausencia de agua. Por ejemplo, las plantas que tienen más agua en su estructura foliar, absorben más luz de la banda 4, también se presenta efectos de “lavado” de la atmósfera, además la humedad empieza a aumentar lo que genera diferentes efectos de absorción y dispersión de la luz en la atmósfera.

3.5 Base de datos adicionales

Se ocupó una base de datos ampliada, esto se consiguió incrementando en los alrededores de las estaciones de monitoreo ambiental puntos vecinos a los pixeles correspondientes a la estación. Con ellos y bajo la hipótesis de que en los alrededores cercanos de la estación se presentan prácticamente las mismas concentraciones de PM, se armó una base de datos similar a la anterior, pero 9 veces más grande, que se espera que ayude en el modelo de Redes Neuronales ya que estos modelos son mejores cuantos más datos se tengan.

3.6 Análisis de Variables

Los modelos de Redes Neuronales Artificiales se suelen considerar cajas negras ya que es difícil intentar predecir el efecto de variables individuales en el resultado del modelo, especialmente en el caso donde se presenten cientos de variables de entrada y además varias capas con cientos de perceptrones. Para el caso en estudio, se determinó crear una red neuronal lo más pequeña posible con las variables más significativas y con la menor cantidad de ellas, debido a que posteriormente se requiere aplicar el modelo a las imágenes

satelitales que contienen millones de píxeles, por tanto, requerirá tiempos de procesamiento muy grandes.

Se realizaron gráficas de PM vs variables predictoras con sus respectivas regresiones polinomiales, esto para clasificar las variables más prometedoras para predecir PM2.5 y distinguirlas. Este análisis se realiza previamente a la generación del modelo, sin descartar alguna variable al momento, posteriormente con estos datos y los resultados de varios modelos se podrán determinar las variables que no influyen en nuestra variable PM2.5.

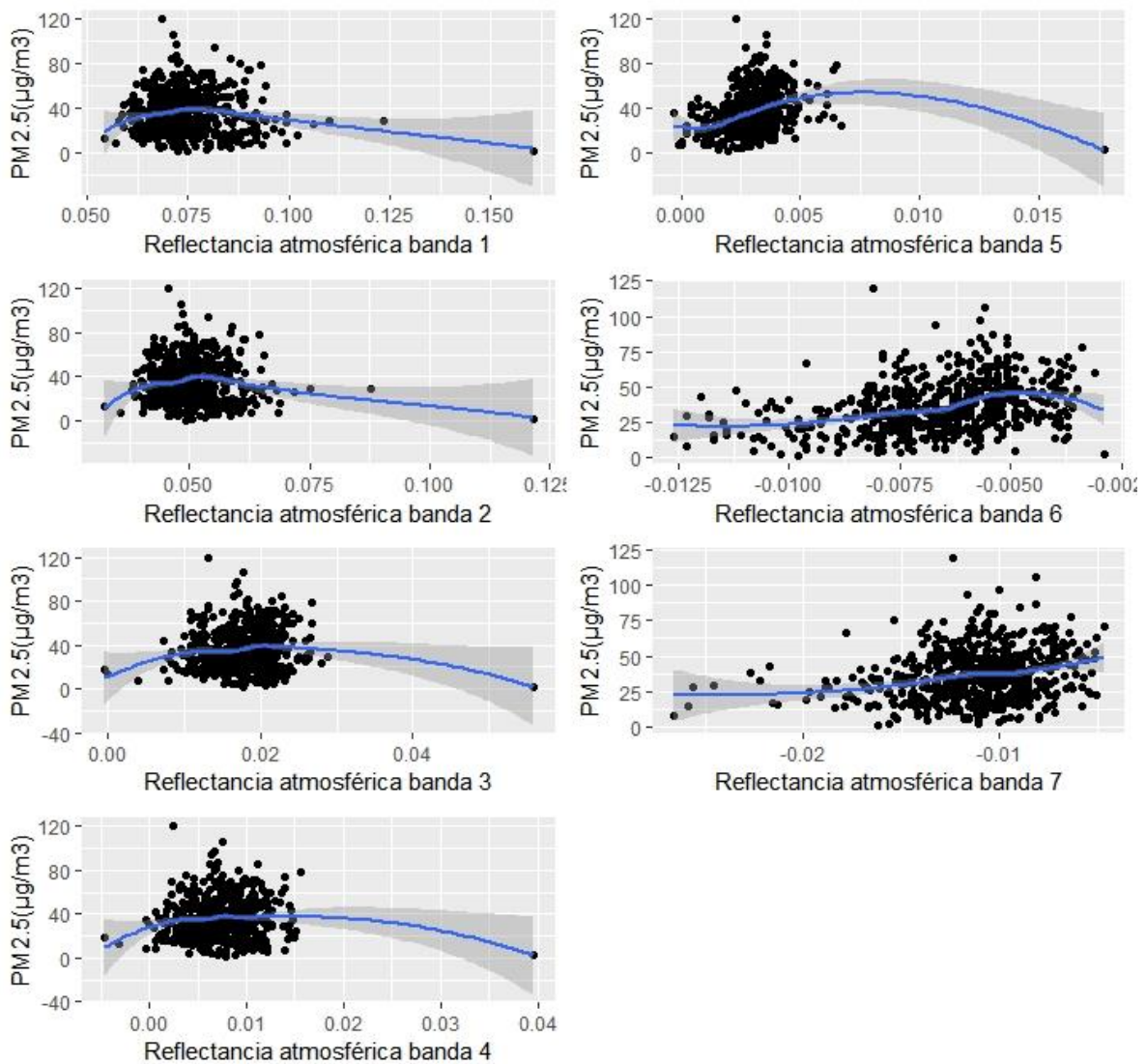


Figura 3.7 Gráficas de correlación entre reflectancias atmosféricas y PM2.5. Elaboración propia.

Se observa en la figura 3.7 que las variables de reflectancia atmosférica producen regresiones polinomiales casi rectas y horizontales además de mucha dispersión respecto a la línea de regresión, esto habla de que existe correlación entre estas variables y PM2.5 aunque sea poca. Una cosa a destacar es los valores de la reflectancia atmosférica en las bandas 6 y 7 dan negativos, que son una consecuencia del algoritmo aplicado por él Landsat 8, en trabajos posteriores se podría indagar más al respecto de este comportamiento.

En la figura 3.8 se hizo un análisis similar, ahora comparando índices espectrales contra PM2.5, inmediatamente podemos concluir que NDVI, EVI, SAVI y MSAVI son muy similares por lo que únicamente se necesitará ocupar una de estas variables en el modelo, sólo faltaría determinar cuál es la que produce mejores resultados. Podemos resaltar que NBR2 parece ser la más prometedora al tener mejor dispersión de puntos en su rango de valores.

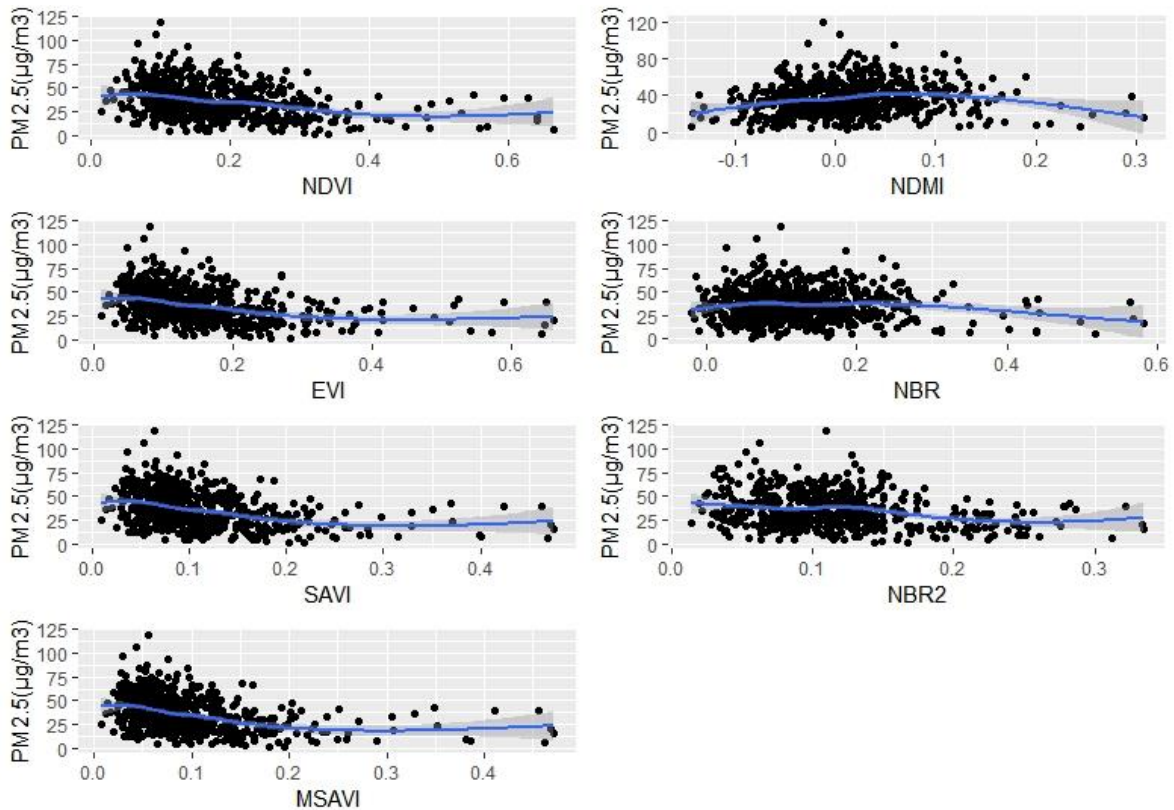


Figura 3.8 Gráficas de correlación entre índices espectrales y PM2.5. Elaboración propia.

Las gráficas anteriores no son precisamente para encontrar correlaciones, sino tendencias, la figura 3.9 se realiza un matriz con el propósito de conocer la correlación entre variables predictoras, estos gráficos muestran la dispersión entre dos variables predictoras, además

del lado izquierdo se muestra con un número el índice de correlación de Pearson que tienen estas variables entre sí.

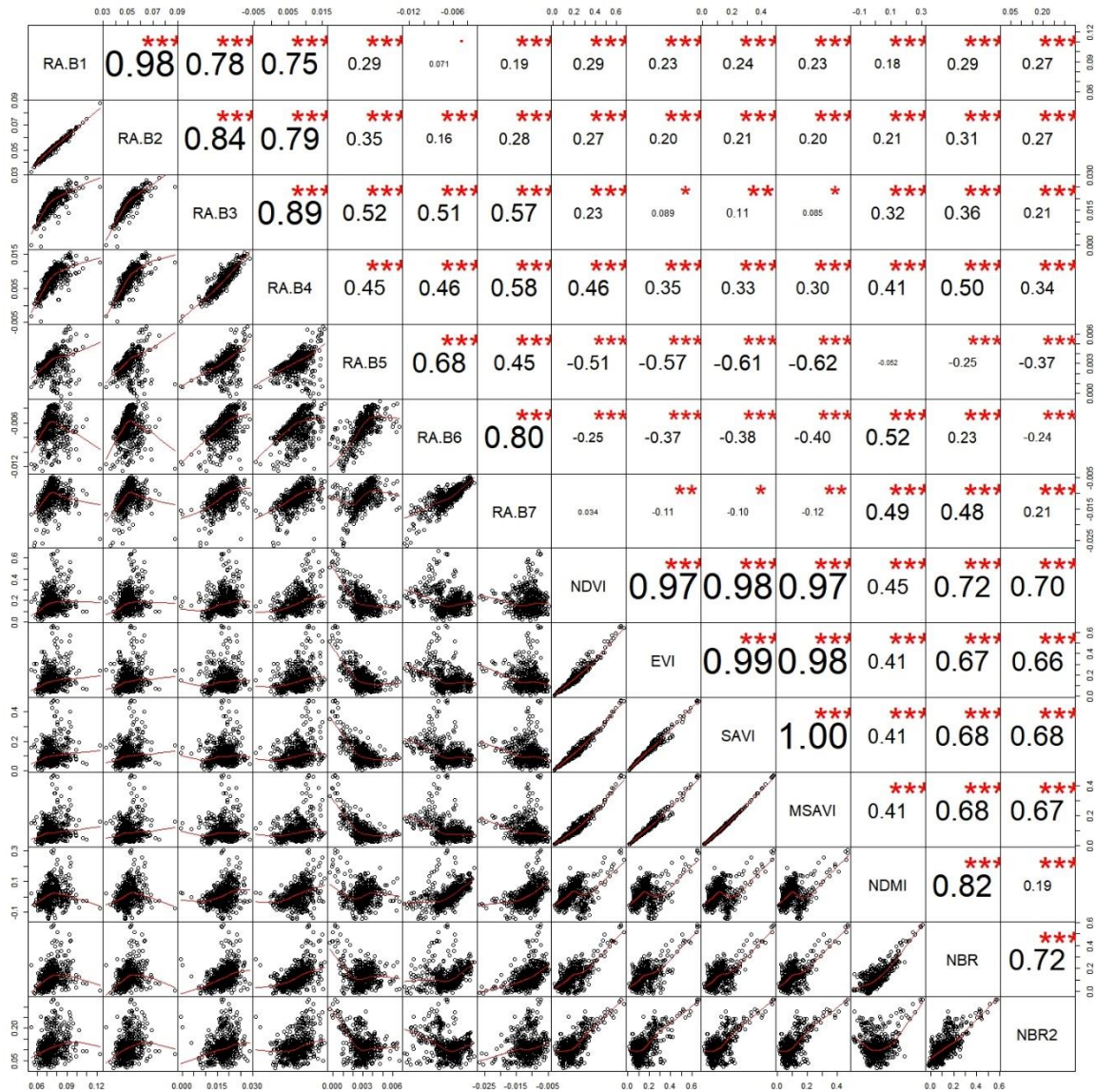


Figura 3.9 Matriz de correlaciones entre variables independientes. Elaboración propia.

Observando la figura anterior verificamos que existe mucha correlación entre NDVI, EVI, SAVI y MSAVI, además observamos una correlación muy grande entre la banda 1 y la banda 2. Esto no descarta a alguna de estas dos variables, ya que las pequeñas diferencias entre las mismas pueden aportar información importante a los modelos.

Todas las relaciones anteriores se basan en modelos lineales, en cambio en redes neuronales se espera que se formen relaciones no lineales más complejas, por lo que,

aunque en regresiones lineales las variables no presenten mucha correlación no significa que en el modelo no vayan a ser ocupadas. Como ya se había escrito, descartar variables solo será posible después de crear modelos, analizarlos y compararlos con la información de este capítulo, para así tomar una buena decisión.

3.7 Creación de la Red Neuronal

3.7.1 Tensorflow y Keras

En palabras de los propios creadores:

TensorFlow™ es una biblioteca de software libre que se utiliza para realizar cálculos numéricos mediante diagramas de flujo de datos. Los nodos de los diagramas representan operaciones matemáticas y las aristas reflejan las matrices de datos multidimensionales (tensores) comunicadas entre ellas. Gracias a la flexibilidad de la arquitectura, solo necesitas una API para desplegar el sistema informático de una o varias CPU o GPU en un escritorio, servidor o dispositivo móvil. En su origen, TensorFlow fue fruto del trabajo de investigadores e ingenieros de Google Brain Team que formaban parte de la organización de investigación del aprendizaje automático de Google. Su objetivo era realizar investigaciones en el campo del aprendizaje automático y las redes neuronales profundas. A pesar de que este era su propósito inicial, se trata de un sistema lo bastante general como para poder aplicarse en muchos otros campos. (Martín Abadi, Ashish Agarwal, Paul Barham et al., 2015)

Existe un dicho de que no hay que reinventar la rueda, y en el caso de Redes Neuronales Artificiales Tensorflow es una de estas ruedas, ya que es una librería disponible para R, que aprovecha de manera eficiente el CPU y GPU de la computadora en que se esté trabajando, además se ajusta perfectamente para realizar casi cualquier tipo de arquitectura de redes neuronales.

La desventaja principal de Tensorflow es la manera de armar la red neuronal, ya que todas sus variables se alojan en tensores, los cuales no se utilizan de manera generalizada. Para solucionar este pequeño problema se optó por usar Keras, que es una interfaz de programación pensada para trabajar con Tensorflow, pero con un conjunto de instrucciones más accesibles para el usuario, además es muy fácil crear y modificar redes neuronales, y como trabaja con Tensorflow es bastante rápido. En la figura 3.10 se muestran las líneas de código en R que se necesitan para la creación de un modelo Redes Neuronales, lo cual demuestra lo accesible que es este programa.

```

model <- keras_model_sequential()
model %>%
  layer_dense(units = 10, activation = 'relu', input_shape = c(n)) %>%
  layer_dropout(rate = 0.1) %>%
  layer_dense(units = 10, activation = 'relu') %>%
  layer_dropout(rate = 0.1) %>%
  layer_dense(units = 1, activation = 'tanh')

model %>% compile(
  loss = 'logcosh',
  optimizer = optimizer_rmsprop(),
  metrics = r_square
)

history <- model %>% fit(
  x_entrenamiento, y_entrenamiento,
  epochs = 100, batch_size = 50,
  validation_split = 0.1 ,
  verbose=0,
  view_metrics = F
)

```

Figura 3.10 Ejemplo de código de Redes Neuronales en R usando Karas. Elaboración propia.

3.7.2 Configuración

La configuración se refiere a la elección de los valores que tendrán los parámetros, esto es el número de capas escondidas, el número de perceptrones que tendrá cada una de ellas, las funciones de activación, las capas de entrada, las capas de salida, la función de costo, tamaño del conjunto de validación, etc.

Realmente no existe una configuración estándar de las Redes Neuronales Artificiales, ya que depende de algunos factores inherentes al problema que se quiere resolver y al conjunto de datos mismo, aunque se tienen una serie de lineamientos para alcanzar un modelo adecuado.

Algunos de estos lineamientos son los siguientes, la capa de entrada será solo una y el número de nodos dependerán de los datos de entrada, la capa de salida será solo una y en el caso de modelos de regresión solo consistirá en un nodo, en cuanto a las capas escondidas “Una sola capa es más que suficiente para la mayoría de los problemas” (Doug, 2018) y para el número de nodos o perceptrones se suele usar la siguiente fórmula:

$$N_h = \frac{N_s}{(\alpha * (N_i + N_0))} \quad \dots 3.1$$

N_h =Número de nodos de la capa oculta

N_i = Número de nodos de entrada (variables independientes)

N_o = Número de nodos de salida (variable dependiente)

N_s = Número de muestras en el conjunto de datos de entrenamiento

α = Factor arbitrario usualmente entre 2-10

Se iniciará con un modelo de una sola capa oculta con 100 nodos siguiendo la fórmula anterior, la función de activación será ReLU, la capa de salida constará de solo un nodo con función de activación tanh. La función de costo del modelo será “logcosh” ya que es una de las funciones más recomendadas para modelos de regresión (Grover, 2018), se optimizará con el método “rmsProp” que es un método donde el descenso por gradiente se hace realizando promedios de los lotes anteriores, en vez de un descenso constante lo que brinda flexibilidad al modelo, especialmente cuando hay gradientes pequeños este método puede seguir aprendiendo (Lyon, 2017). En cuanto al número de iteraciones del modelo se proponen 1000 y el tamaño de lote (batch size) será igual a 600.

En la figura 3.11 se muestran los resultados de esta configuración inicial, la gráfica de dispersión entre valores predichos y valores reales ayuda de manera visual a revisar la calidad del modelo, ya que si los puntos se aproximan a la recta de identidad mejor será el modelo. El histograma ayuda a identificar si el modelo está estimando de mayores o menores concentraciones, además, aunque no se trata de regresión lineal, lo ideal sería que la media de los residuos sea 0. Con esta misma figura se puede determinar si el modelo será capaz de hacer buenas predicciones, esto gracias a que al principio se dividieron los datos en 2 grupos, uno para entrenar a la red y el otro para probarla, si se aplica el modelo a la prueba se puede graficar tanto dispersión como el histograma de residuos, si ambos tienen la misma forma que con el conjunto de datos de entrenamiento se puede decir que el modelo tiene una predicción adecuada.

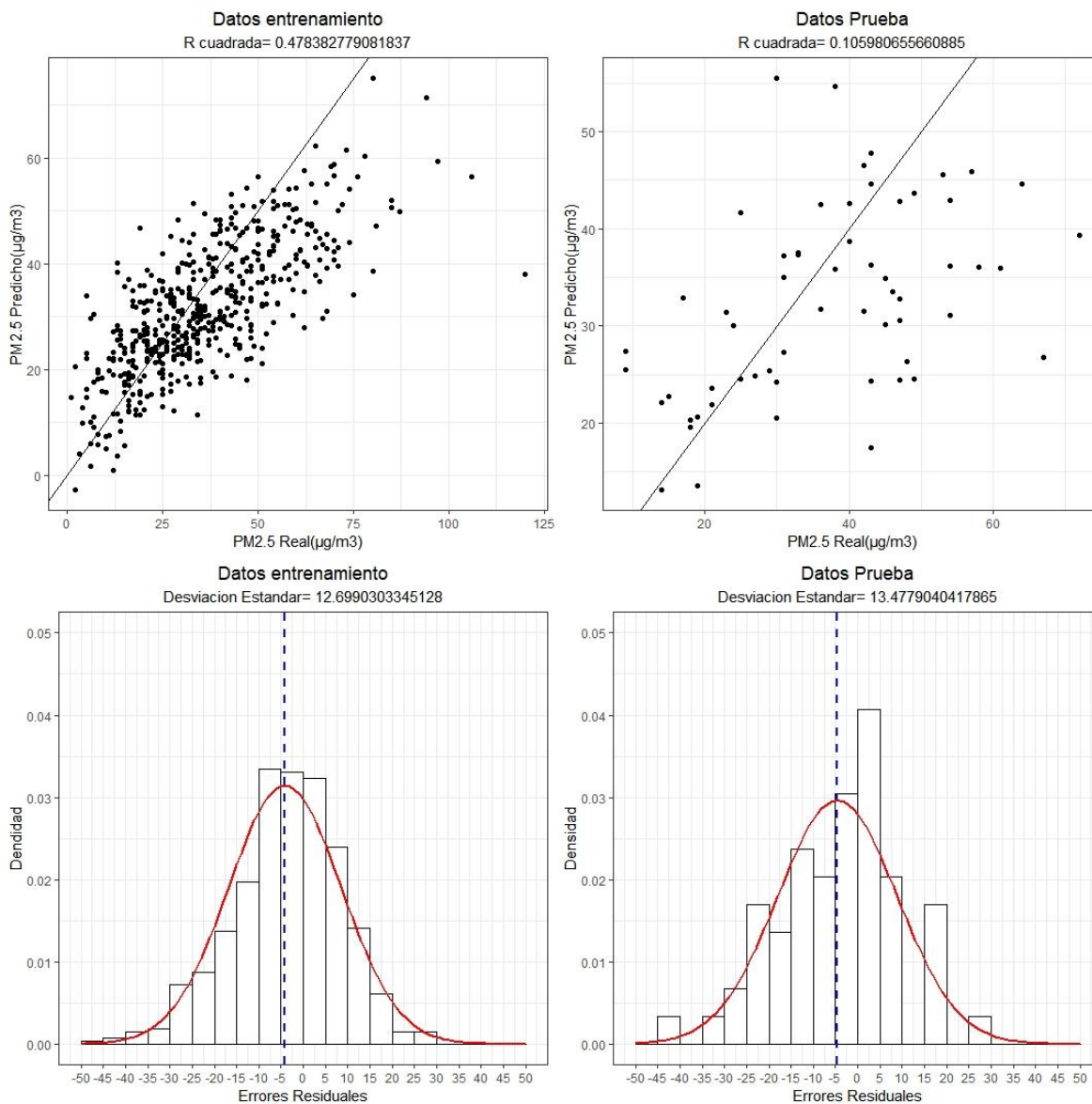


Figura 3.11 Gráficas de dispersión valores predichos y valores reales, conjunto de datos solo estaciones número de datos=834. Elaboración propia.

En esta primera red neuronal se observa que la gráfica de dispersión de los datos de prueba no tiene mucha correlación, esto indica que el modelo no es capaz de predecir fuera de los datos de entrenamiento del modelo, la desviación estándar tampoco coincide con el entrenamiento.

El primer objetivo será lograr la mejor correlación en los datos de entrenamiento pero que se mantenga en los datos de prueba, también que la media de los errores residuales se acerque al 0, ya que para esta etapa se encuentra en en -5. Lo primero que se puede hacer

aumentar el número de iteraciones que realiza la red, si eso no funciona se tendrá que aumentar en número de nodos.

Existe ocasiones en que la red neuronal presenta problemas como el desvanecimiento del gradiente, por lo que la red neuronal ya no puede converger y se “atora” en un mínimo local, en la figura 3.12 se observa un problema de estos donde por más iteraciones que se hagan siempre predice valores por debajo de los reales. Este problema empezó a suceder cuando se pusieron más de 200 nodos en la primera capa, por lo que se propondrá añadir una segunda capa oculta.

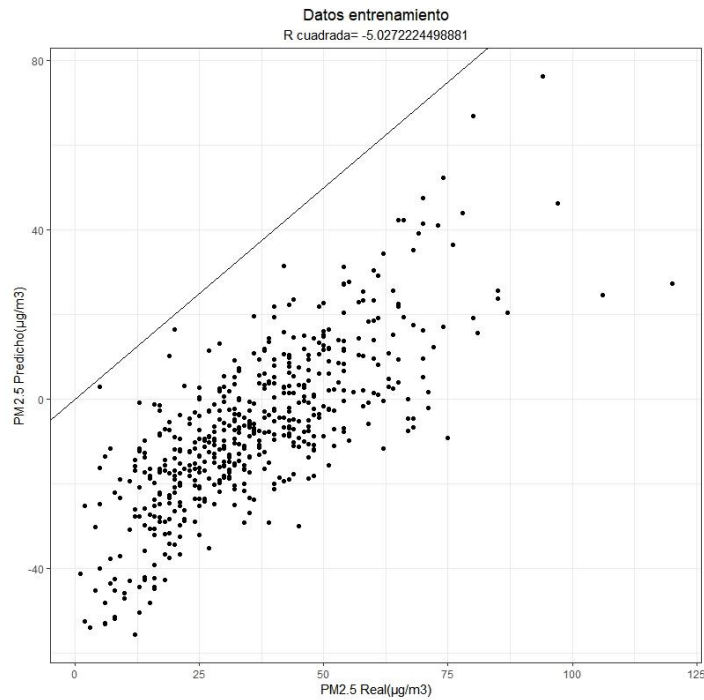


Figura 3.12 Gráfica de dispersión modelo de RNA fallido. Elaboración propia

Después de varias etapas de crear distintos tipos de modelos donde se cambiaron número de nodos, número de capas ocultas, número de iteraciones y cambio de funciones de activación se llegó a la conclusión de que no se puede mejorar el modelo con los datos de entrenamiento y que a su vez también mejore la predicción con los datos de prueba, esto muy probablemente se debe a que los datos que tenemos no son suficientes para que la red aprenda todos los patrones que relacionan nuestras variables predictoras y PM2.5. En la figura 3.13 se observa un modelo de 2 capas ocultas con 40 nodos cada uno de ellos y tan solo después de 2000 iteraciones logra dar una muy buena regresión al conjunto de

datos de entrenamiento, sin embargo, al aplicar este modelo a los datos de prueba no se observa ninguna relación entre predicción y valores reales, la desviación estándar aumenta el doble, lo que nos indica que no es un muy buen modelo predictor, se hizo un sobreajuste, esto pasa con todos los modelos: cuando se maximiza el R^2 de los datos de entrenamiento la correlación en los datos de prueba disminuye.

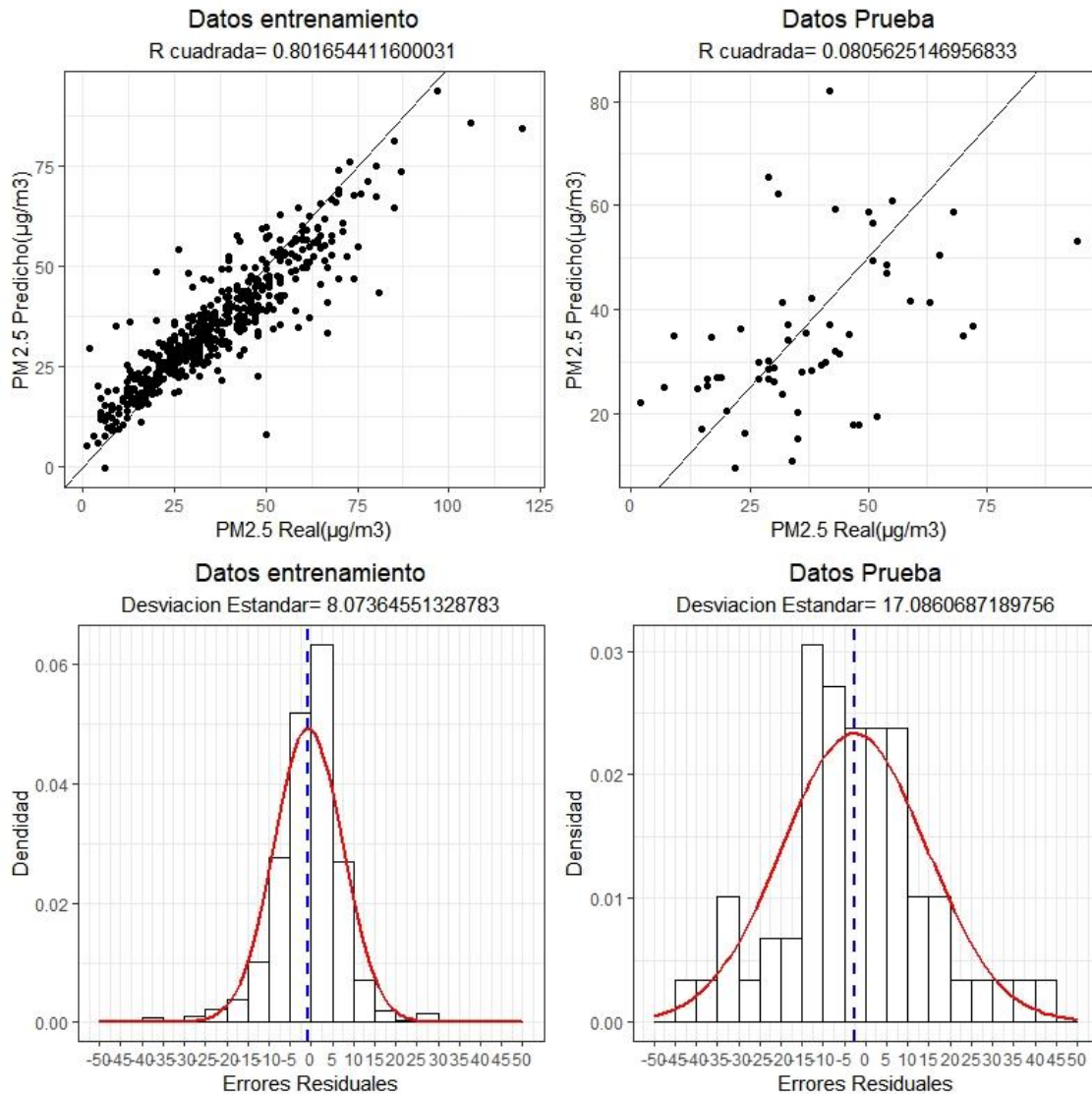


Figura 3.13 Gráficas de dispersión valores predichos y valores reales, modelo 2 número de datos=834. Elaboración propia

Se optó por utilizar la base de datos donde se añadieron puntos de extracción en los pixeles vecinos de la estación de monitoreo ambiental, esta base de datos es 9 veces más grande por lo que se esperaría una correlación menor y una desviación estándar mayor. En la figura

3.14 se creó un modelo con los mismos parámetros que la anterior, 2 capas ocultas con 40 nodos cada una de ellas, el resultado fue un modelo con una correlación de 0.63 y una desviación estándar de 11.1, la cual no es excelente, pero es mejor de lo que se esperaba, además tiene una característica muy importante y es que estos parámetros se casi se mantienen en los datos de prueba.

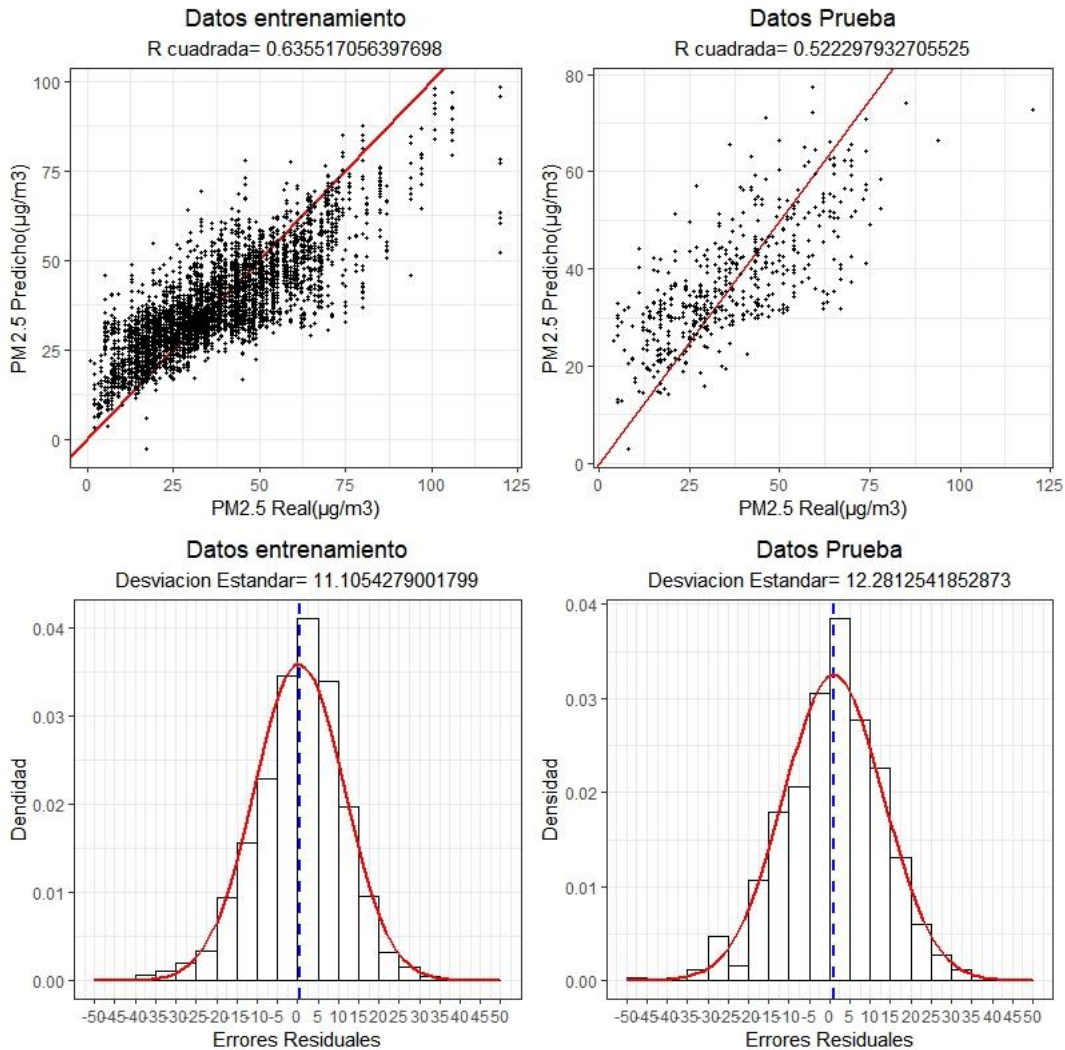


Figura 3.14 Gráficas de dispersión valores predichos y valores reales, modelo 3, conjunto de datos con pixeles vecinos, número de datos=7781.

Realizando más modelos se encontró que para este problema es mejor tener varias capas ocultas, ya que una capa de 10000 nodos y que contiene 270,001 parámetros que optimizar tarda en realizarse unos 2 min, y arroja valores de R^2 0.52 y desviación estándar de 12.56. En cambio, una red con 2 capas ocultas y 100 nodos cada una de ellas, tiene 22,901

parámetros y solo tarda 55 segundos en calcularse además mejora bastante el coeficiente de R^2 y desviación estándar, 0.748 y 8.8 respectivamente. Por lo que se usará esta última configuración como base para ir descartando variables.

3.8 Selección de variables predictoras

Tener muchas variables en nuestro modelo de Redes Neuronales impacta negativamente, ya que en el aprendizaje el mismo descarta las variables, el problema radica en que se desea aplicar este modelo a imágenes satelitales donde cada variable del modelo es igual a una imagen la cual contienen millones de píxeles, por lo que el resultado de aplicar una sola imagen más, añade miles de millones de operaciones que tiene que realizar la computadora, por ello estamos buscando minimizar el número de variables que ocupará el modelo.

Las primeras variables que se descartaron fueron las de precipitación, humedad y temperatura, ya que al ser los promedios de toda la escena no aportan información precisa de los puntos de muestreo lo que afectó al resultado de los modelos, haciendo predicciones muy homogéneas en toda la escena.

Para descartar más variables se ideó un método similar a los que se usan en regresiones lineales, que es ir descartando variable a variable, calcular el modelo y ver el parámetro de correlación de Pearson y la desviación estándar, al dividir estos dos factores se obtiene un parámetro que describe si se mejoró el modelo, entre menor sea el valor significa que el modelo mejoró al remover esa variable, por lo que se removerá la variable que tenga el menor de este parámetro. En la figura 3.15 se observa una gráfica de barras en la que se indica el valor de este parámetro (R^2 /Desviación estándar), la barra color rojo indica el caso en que ninguna variable⁵ se removió, los que están a la izquierda de esta barra significa que los modelos mejoraron y del lado derecho los modelos que empeoraron. Como comentario adicional, debido a que los modelos no están siendo optimizados al máximo y los valores iniciales de los parámetros de los modelos son generados aleatoriamente, los resultados pueden variar incluso si se generan varias veces con los mismos parámetros de entrada, por ello se realizaron 3 iteraciones para calcular R^2 y la desviación estándar por

⁵ “Ninguna” significa que en esa iteración no se removió variable, pero anteriormente ya se habían removido algunas variables

cada variable, posteriormente se obtiene un promedio, un ejemplo de estas tablas se puede ver en la figura A1 del anexo.

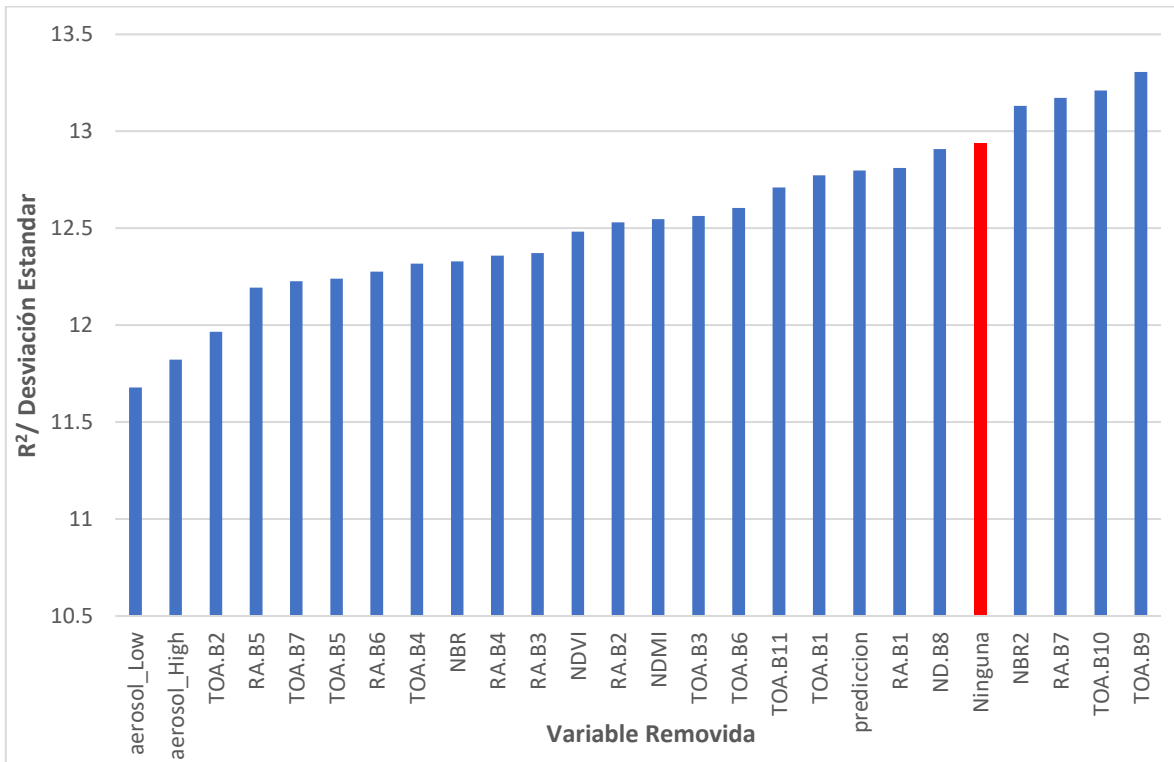


Figura 3.15 Gráfico de barras R²/ desviación estándar de modelos con remoción de una variable. Elaboración propia.

Se empezará quitando la variable aerosol_Low y se continuará con este procedimiento removiendo una variable por iteración hasta que la barra roja llegue del lado izquierdo, esto significará que si removemos alguna variable más el modelo empeorará. Si el lector se pregunta por qué eliminar una a una las variables y no todas de golpe, es debido a que existe correlación entre las variables y quitar varias de estas puede hacer que los modelos empeoren mucho, en cambio quitar una sola de estas variables correlacionadas no modificará tanto el modelo.

El algoritmo que se creó para ir eliminando variables se detuvo una vez una vez eliminadas 15 variables de 27 que teníamos originalmente, en la figura 3.16 se observan las variables resultantes, eliminar cualquiera de estas variables significaría empeorar el modelo. Este fue un paso bastante tardado, unas 16 horas, debido a las capacidades de procesamiento del equipo y que este método no es muy eficiente. No existe mucha información respecto a la optimización de la arquitectura de Redes Neuronales, es por ello que se propuso este método, similar al método “Eliminación progresiva” de modelos lineales.

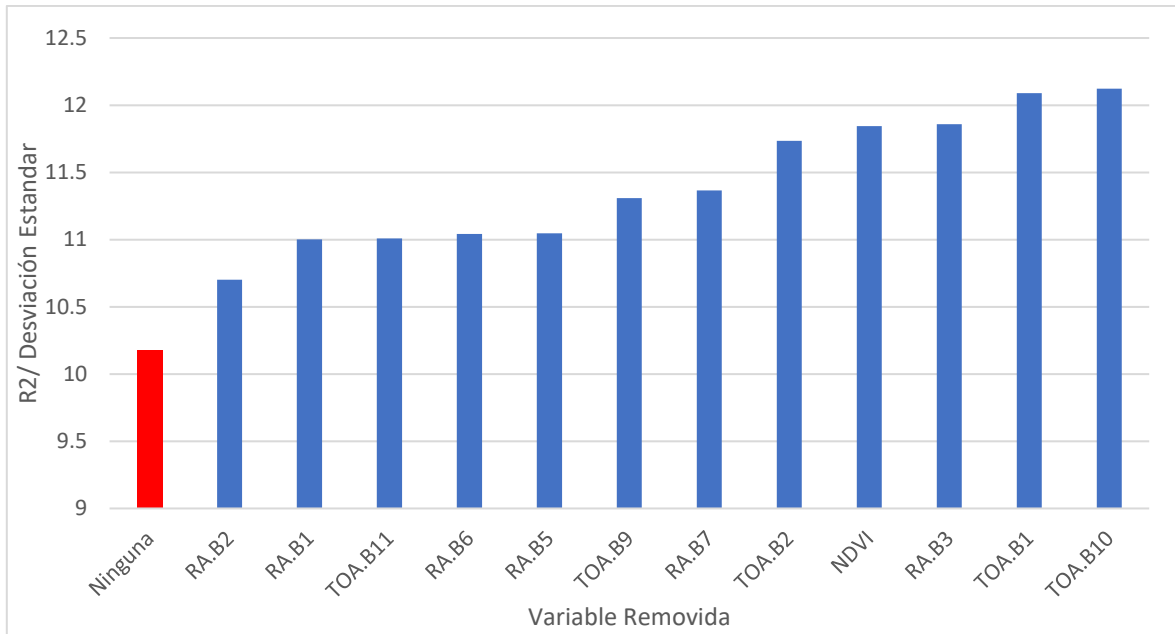


Figura 3.16 Gráfico de barras R^2 / desviación estándar de modelos una vez removidas 15 variables.
Elaboración propia.

Se debe mencionar que se encontró un pequeño error en el algoritmo, este fue que cada que se hacía una iteración para remover una variable y ver sus efectos, se seleccionaba una muestra distinta para los datos de entrenamiento de la red, esto implica que no estaban en igualdad de condiciones los distintos modelos para poder compararse y eliminar variables. Aunque esto parece un error grave en realidad no lo es, ya que la toma de muestras fue aleatoria y el número de elementos bastante grande respecto a la población (90%), esto hace que las muestras en teoría sean bastante parecidas y las mejoras de los modelos generados por el conjunto de datos sean mínimas. Aun así, se propone realizar otro algoritmo que funcione similar al método “Introducción progresiva” de regresiones lineales, al ir agregando una variable de las que eliminamos y verificar si mejora el modelo significativamente, esta vez usando el mismo conjunto de datos de entrenamiento para cada iteración.

Este procedimiento fue más rápido y solo añadió una variable TOA.B3, ya que añadir alguna otra variable no mejoraría significativamente los modelos.

3.9 Aplicación de modelo

Posteriormente se procede a la creación de un modelo bien entrenado, esto se logra aumentando el número de iteraciones hasta que la función de costo llegue a su mínimo. La arquitectura del modelo que se aplicó es la siguiente.

- Capa de entrada: de 13 variables RA.B1, RA.B2, RA.B3, RA.B5, RA.B6, RA.B7, TOA.B1, TOA.B2, TOA.B3, TOA.B9, TOA.B10, TOA.B11 y NDVI.
- Capas ocultas: 2, la primera con 100 nodos y la segunda con 60 nodos, ambas activadas con la función ReLU, además tienen un factor 0.1 para resetear nodos aleatorios en la fase de entrenamiento, esto ayuda a prevenir un sobreajuste.
- Capa de salida: consiste en un solo nodo activado con la función tangh.
- Se realizaron 10,000 iteraciones en el entrenamiento, en lotes (batch)⁶ de 600 elementos.

Un esquema de la arquitectura de la red se puede observar en la figura 3.17, donde los colores azules indican unos pesos positivos y los rojos negativos, la intensidad del color indica la magnitud de estos. Adicionalmente se proporcionan un ejemplo de la matriz de pesos en el anexo tabla 1, es impráctico mostrar la otra matriz ya que cuenta con 101 renglones y 60 columnas.

⁶ Los lotes o batch en inglés sirven para dividir el conjunto de datos y aplicar el modelo de regresión solo a este lote, se utilizan para mejorar la velocidad de entrenamiento y ocupar menos memoria (Ittxer, 2018)

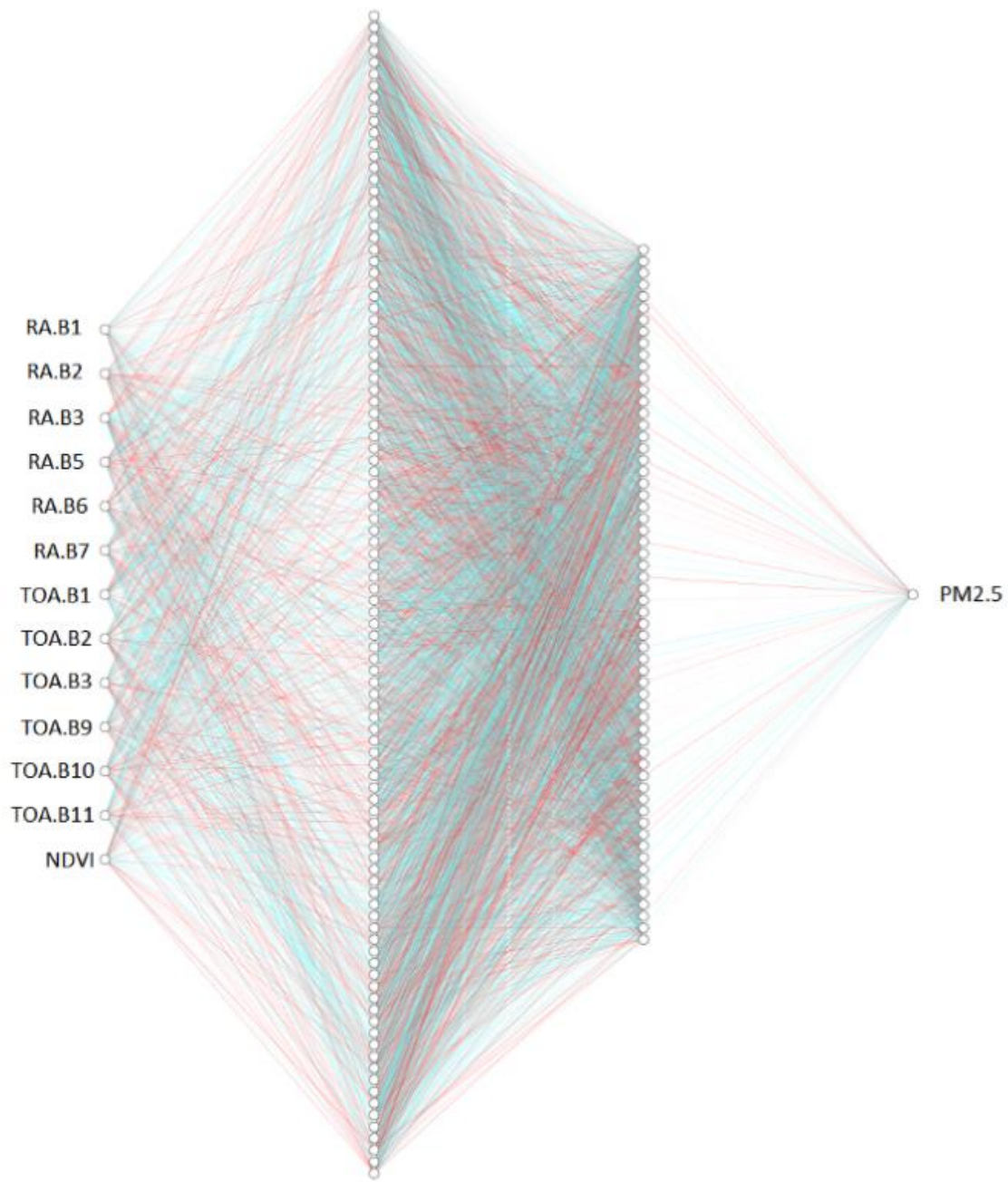


Figura 3.17 Arquitectura del modelo de Red Neuronal Artificial. Elaboración propia.

La desviación estándar del error y el coeficiente de correlación R^2 resultaron ser bastante buenos en nuestros datos de entrenamiento como se observa en la figura 3.18, además se obtuvo la raíz del error cuadrático medio que fue 7.16, en los datos de prueba bajan estos

tres parámetros, pero manteniendo la tendencia de los otros datos, por lo que se considera apto para aplicar a las escenas.

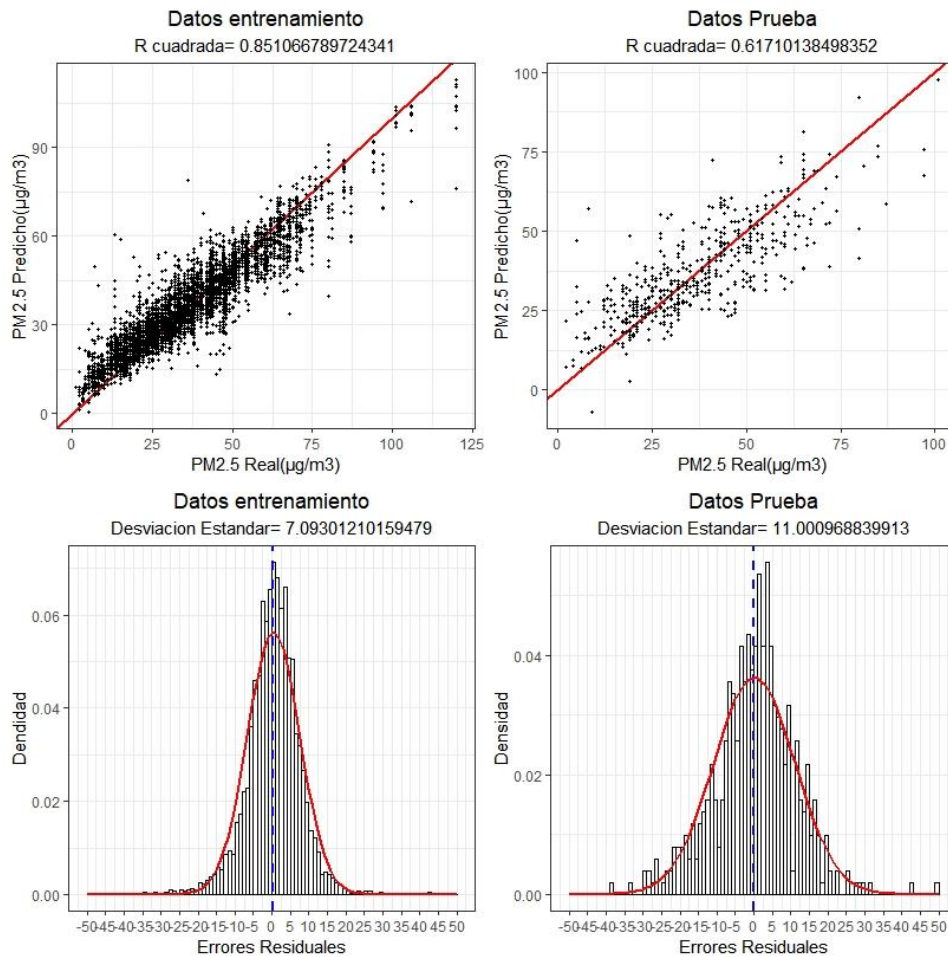


Figura 3.18 Gráficas de dispersión del primer modelo aplicado. Elaboración propia.

El aplicar un modelo generado con keras es un paso sencillo pero tardado, lo primero que se realizó fue pasar cada banda que se ocupa de forma ráster a matriz con una función incluida en la librería ráster de R, después estas matrices se pasaron a forma vectorial para luego unir las en una tabla donde cada columna representa una banda y las filas el valor de los píxeles, esta forma es idéntica a las tablas que se ocupan para ingresar los datos al modelo, posteriormente se aplica el modelo a esta tabla por lo que nos dará por resultado un vector con los valores predichos de PM2.5. A partir de aquí se aplican los pasos inversos que son, convertir el vector en matriz para posteriormente georreferenciarlo y convertirlo en un ráster.

Capítulo 4 Análisis de resultados

El resultado obtenido al aplicar este primer modelo se observa en la figura 4.1, éste se aplicó a la escena que presentaba mayores concentraciones. Lo primero a destacar que se observa con una inspección visual es que este modelo es diferente a los modelos de interpolación ya que estos suponen zonas completas con altas concentraciones y zonas con bajas concentraciones, en nuestro modelo esto no sucede, existen concentraciones altas, bajas y medias en una misma zona, no existe homogeneidad perfecta.

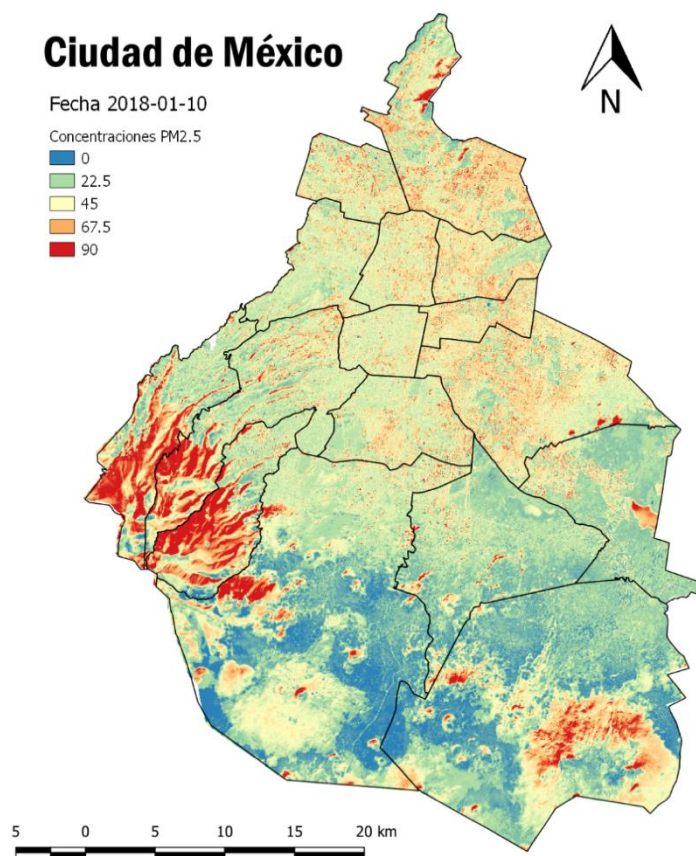


Figura 4.1 Mapa de concentraciones aplicando el 1er modelo de Redes Neuronales Generado. Elaboración propia.

Otras características a destacar de la figura 4.1 son las estimaciones de concentraciones altas de la zona suroeste, ya que estas corresponden a las sombras de la zona montañosa del Ajusco, añadido a esto se puede mencionar el cambio demasiado brusco entre zonas urbanizadas y vegetación, ya que en esta última zona el modelo predice bajas concentraciones, pero el cambio tan brusco en estas zonas de transición no parece

correcto, encima de cuerpos de agua también existe incertidumbre ya que predice concentraciones altas aunque es posible que el modelo confunda con concentraciones altas de humedad. Estos detalles implican que el modelo aún no es adecuado en ciertas zonas, principalmente en aquellas donde no se han tomado mediciones nunca por lo que el modelo está extrapolando información, la forma más adecuada sería ir y realizar mediciones en campo en estos sitios, para este trabajo se optó por realizar más modelos tomando en cuenta otras variables, el siguiente modelo se realizó eliminando la variable NDVI con la intención de reducir la influencia de vegetación en el modelo. En la figura 4.2 se muestra el modelo donde se eliminó la variable NDVI, notamos que el modelo prácticamente es el mismo en la zona urbana, pero en la zona con vegetación cambia, prediciendo concentraciones más bajas, y en zonas montañosas predice concentraciones más altas.

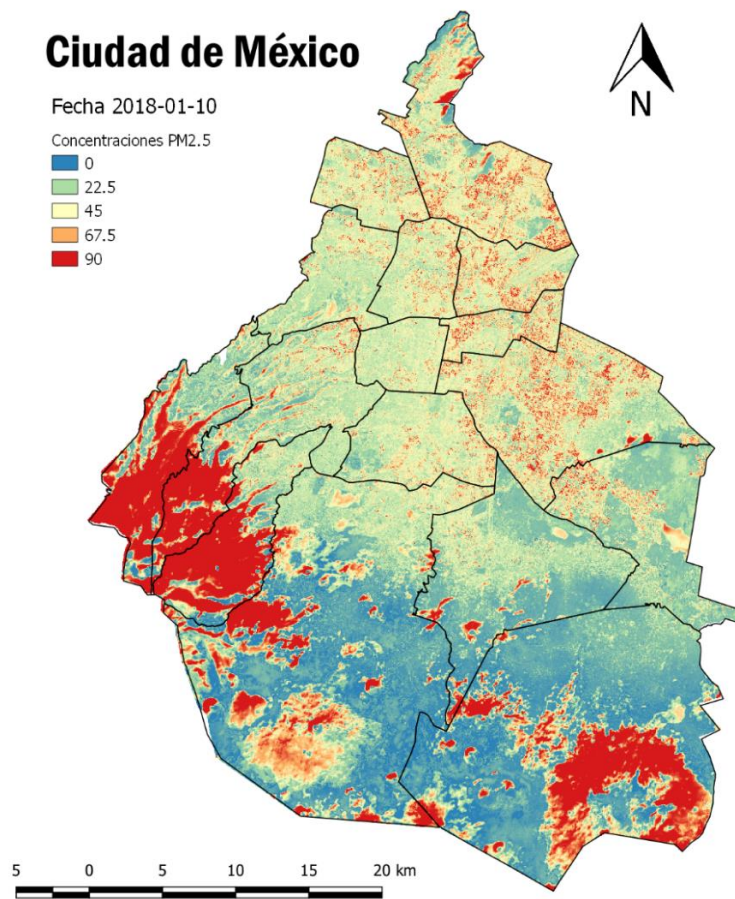


Figura 4.2 Mapa de concentraciones PM2.5 modelo 2(sin NDVI). Elaboración propia.

Se crearon otros modelos después de seguir observando zonas con altas concentraciones en zonas montañosas, logrando en uno de estos modelos que disminuyeran dichas concentraciones, pero esto sería amañar resultados haciendo que estos se adaptaran a lo

que nosotros creemos que es lo correcto, por eso se decidió únicamente reportar estos dos modelos ya que fueron producto de todo un procedimiento en el que se fueron descartando variables por una razón matemática y no una razón de empirismo puro.

Analizando todos los valores resultantes de aplicar el modelo 1 a la CDMX nos resulta un gráfico de barras como en la figura 4.3. Lo primero es que el modelo presenta pequeños errores ya que aún existen valores menores de 0, otra característica es que se casi se forma una distribución normal.

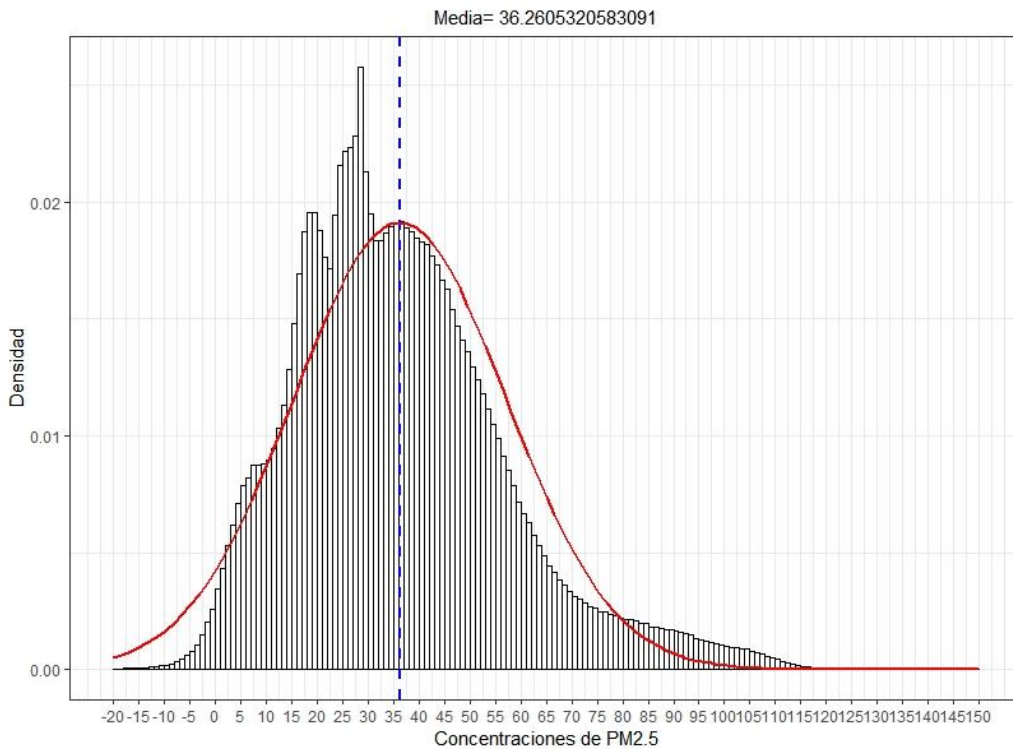


Figura 4.3 Distribución de valores predichos de PM2.5 tomando en cuenta toda la CDMX. Elaboración propia.

Otra característica que se puede observar del modelo es que el modelo predice que dentro los parques de la ciudad se presentan concentraciones un poco más bajas que en sus alrededores, no se sabe si esto es una situación real o un error del modelo debido a la presencia de vegetación, en la figura 4.4 se muestra los alrededores del bosque de Chapultepec.

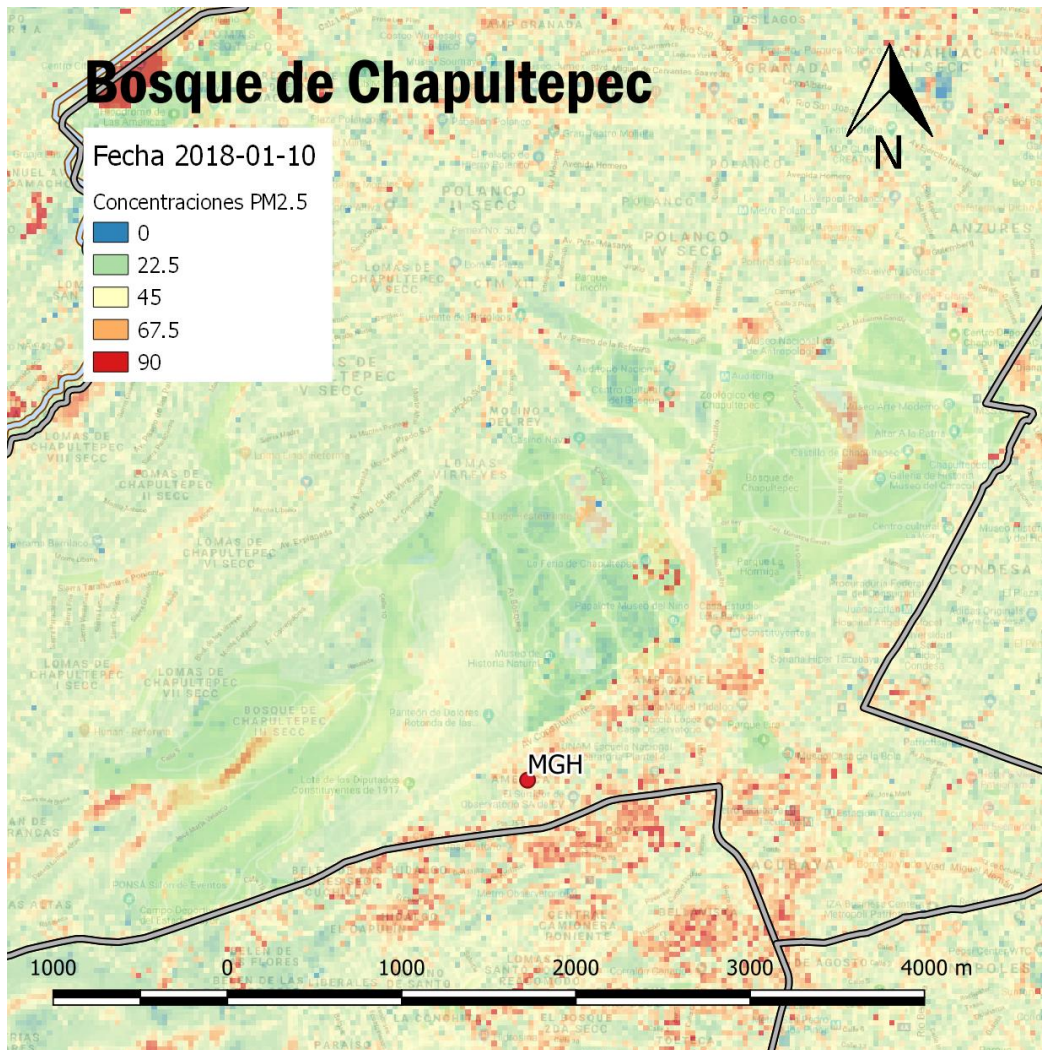


Figura 4.4 Concentraciones de PM2.5 en alrededores Bosque de Chapultepec. Elaboración propia.

Este modelo fue creado con datos donde las estaciones se encuentran en un ambiente de ciudad, donde hay mucha presencia de calles, edificios, vehículos, por ello al sur de la ciudad donde empieza a existir zonas con mayor vegetación se tienen la incertidumbre si el modelo está calculando de manera adecuada en estas zonas, ya que hay un cambio abrupto de las concentraciones predichas, en presencia de vegetación.

Aunque ya no entra dentro de la zona de estudio se observó en el estado de México una zona industrial, figura 4.5, en ella existe un conjunto de naves industriales y sólo sobre algunas de estas naves se mostraron altas concentraciones de PM2.5, lo que es un indicativo que estos modelos pueden servir como una forma de localizar puntos de emisión muy específicos esto gracias a la alta resolución espacial del Landsat 8.

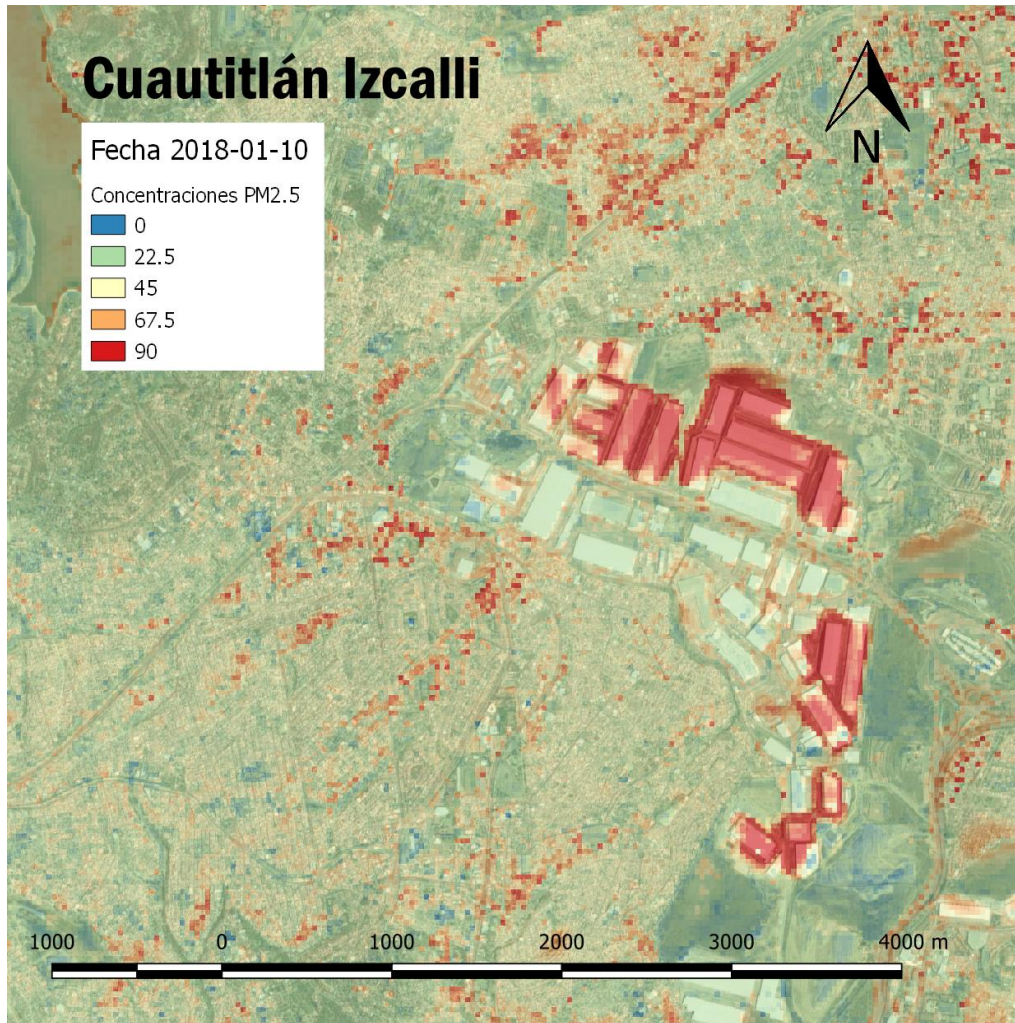


Figura 4.5 Concentraciones de PM2.5 sobre naves industriales, Hacienda del parque, Cuautitlán Izcalli, 2018-01-10. Elaboración propia.

4.1 Análisis Temporal

Resulta impráctico mostrar todos los mapas de concentraciones de las 44 fechas analizadas en este trabajo por lo que solo se analizara las de 2015 ya es el año con intervalos más constantes en la toma de las imágenes satelitales.

Lo primero que se realizó es el análisis de las mediciones realizadas de los días y hora en que fueron tomadas las imágenes, una forma fácil de visualizar esta información es por medio del diagrama de caja mostrada en la figura 4.6, en esta gráfica podemos observar que la mediana es casi $40 \mu\text{g}/\text{m}^3$ las dos primeras fechas posteriormente el 24 de abril de 2015 baja un poco hasta los $35 \mu\text{g}/\text{m}^3$, los dos días con mayores concentraciones son en junio y diciembre.

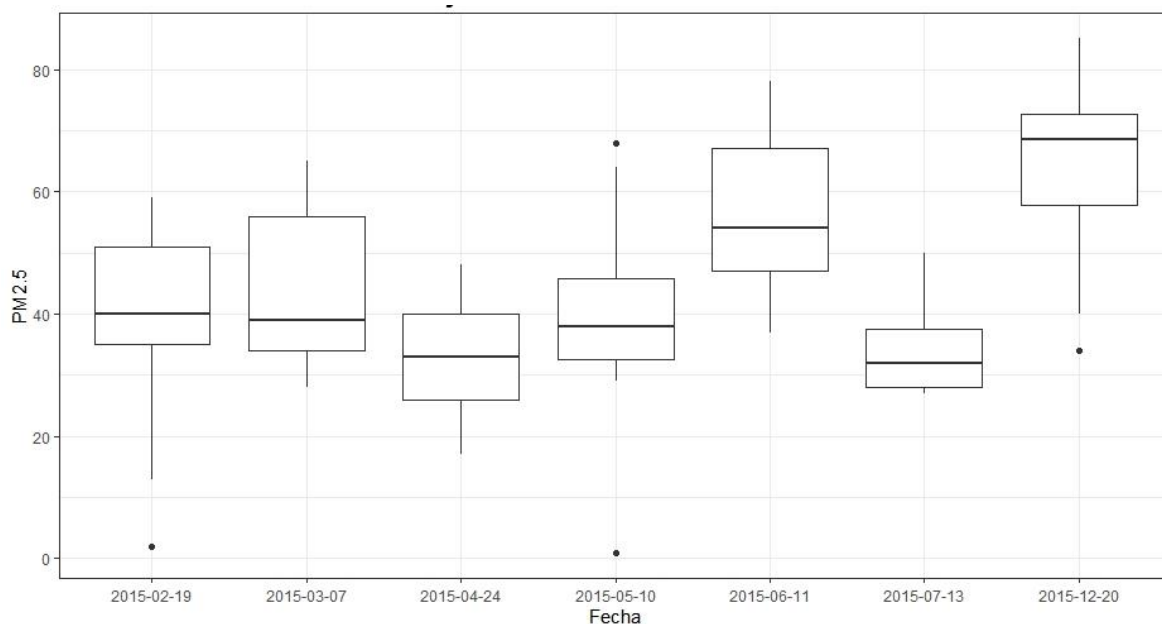


Figura 4.6 Gráfica de caja concentraciones PM2.5. Elaboración propia.

Los datos anteriores son comparados visualmente con las figuras 4.7 a la 4.13, por ejemplo, el primer día la mayor parte de la ciudad esta en color amarillo que corresponde a una concentración de $45 \mu\text{g}/\text{m}^3$ además se presentan prácticamente las mismas concentraciones en la segunda fecha. En la tercera fecha, figura 4.9, se muestra una tendencia hacia el color verde que significa que las concentraciones de PM2.5 bajan, al igual que en la gráfica de caja. En el cuarto día, figura 4.10, se presenta un balance entre el color amarillo y verde lo que significaría que las concentraciones aumentaron, pero no tanto como en los primeros días. Ya para la quinta fecha que corresponde al 11-06-2015, figura 4.11, se tienen una imagen con predominancia naranja y amarilla lo cual se traduce en concentraciones entre $45 - 67.5 \mu\text{g}/\text{m}^3$, que corresponden a lo observado en la figura 4.6. Para la figura 4.12 las concentraciones vuelven a bajar, al observarse que predomina el verde y el amarillo. Por último, en la fecha de 20-12-2015, figura 4.13, se observa que la mayor parte del centro de la ciudad se encuentra en naranja incluso algunos puntos en rojo que indica concentraciones entre $67.5 - 90 \mu\text{g}/\text{m}^3$, y esto coincide con las mediciones ya que este día fue el mas contaminando de todos.

Ciudad de México

Fecha 2015-02-19

Concentraciones PM2.5

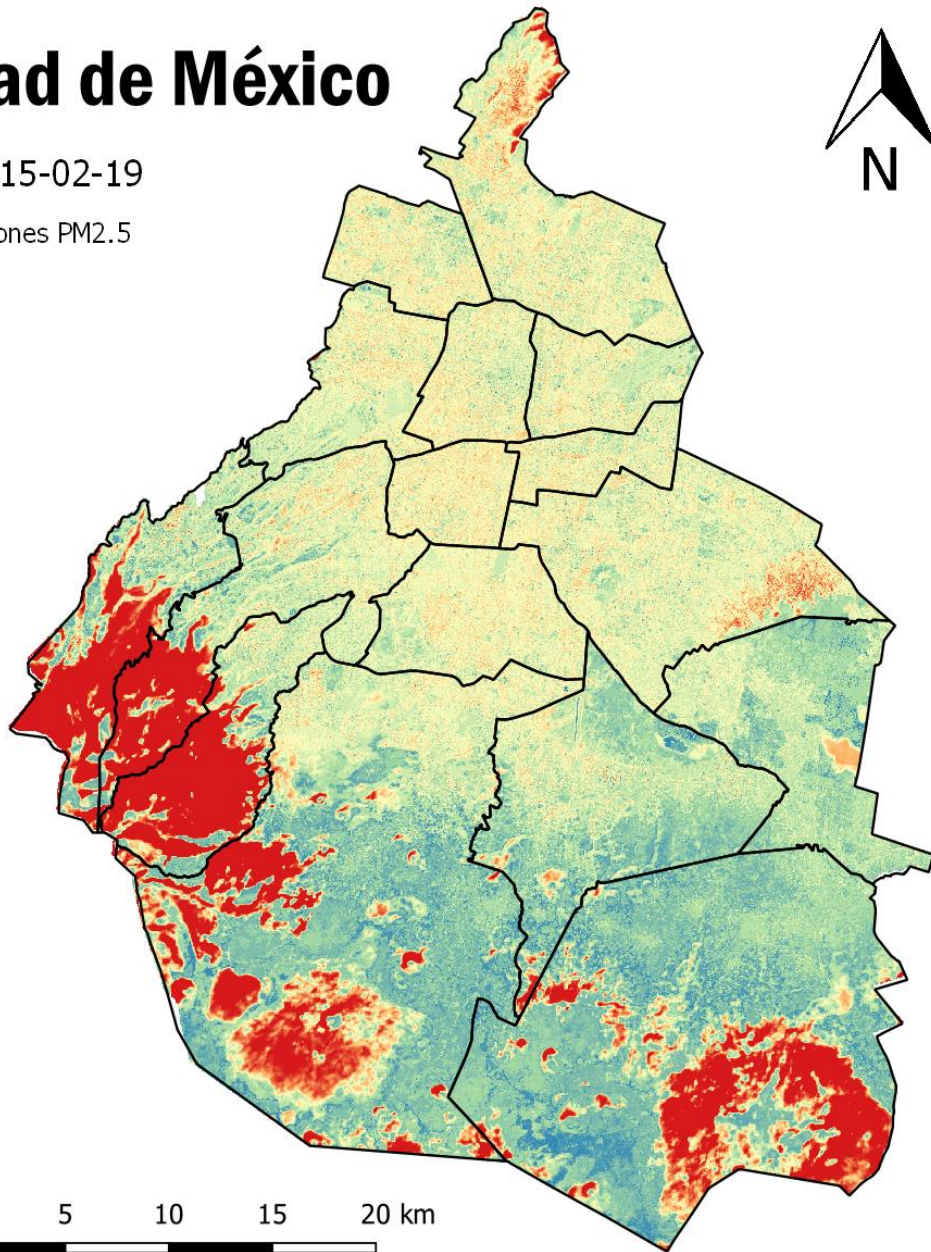


Figura 4.7 Mapa concentraciones PM2.5, 2015-02-19. Elaboración propia.

Ciudad de México

Fecha 2015-03-07

Concentraciones PM2.5

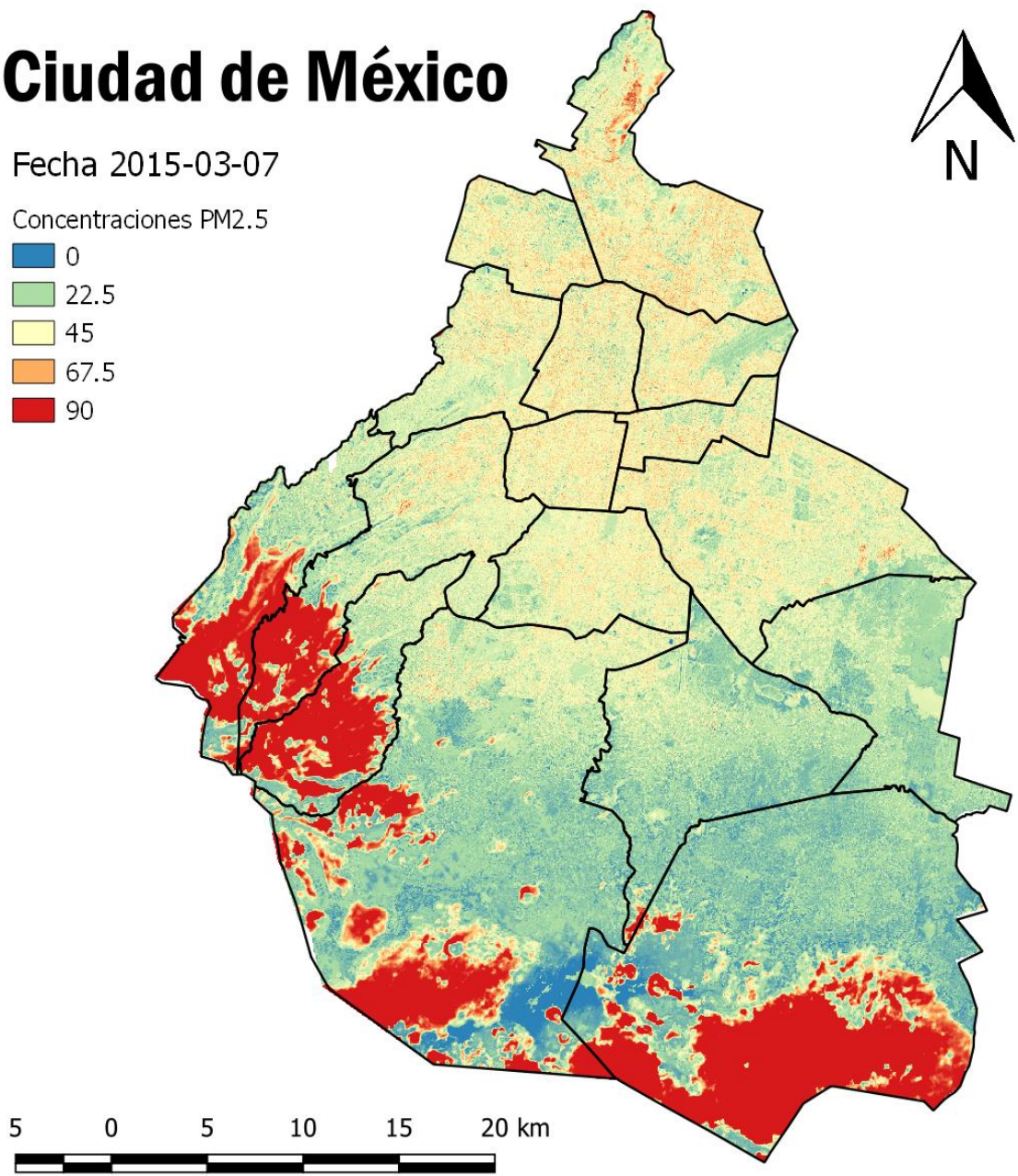


Figura 4.8 Mapa concentraciones PM2.5, 2015-03-07. Elaboración propia.

Ciudad de México

Fecha 2015-04-24

Concentraciones PM2.5

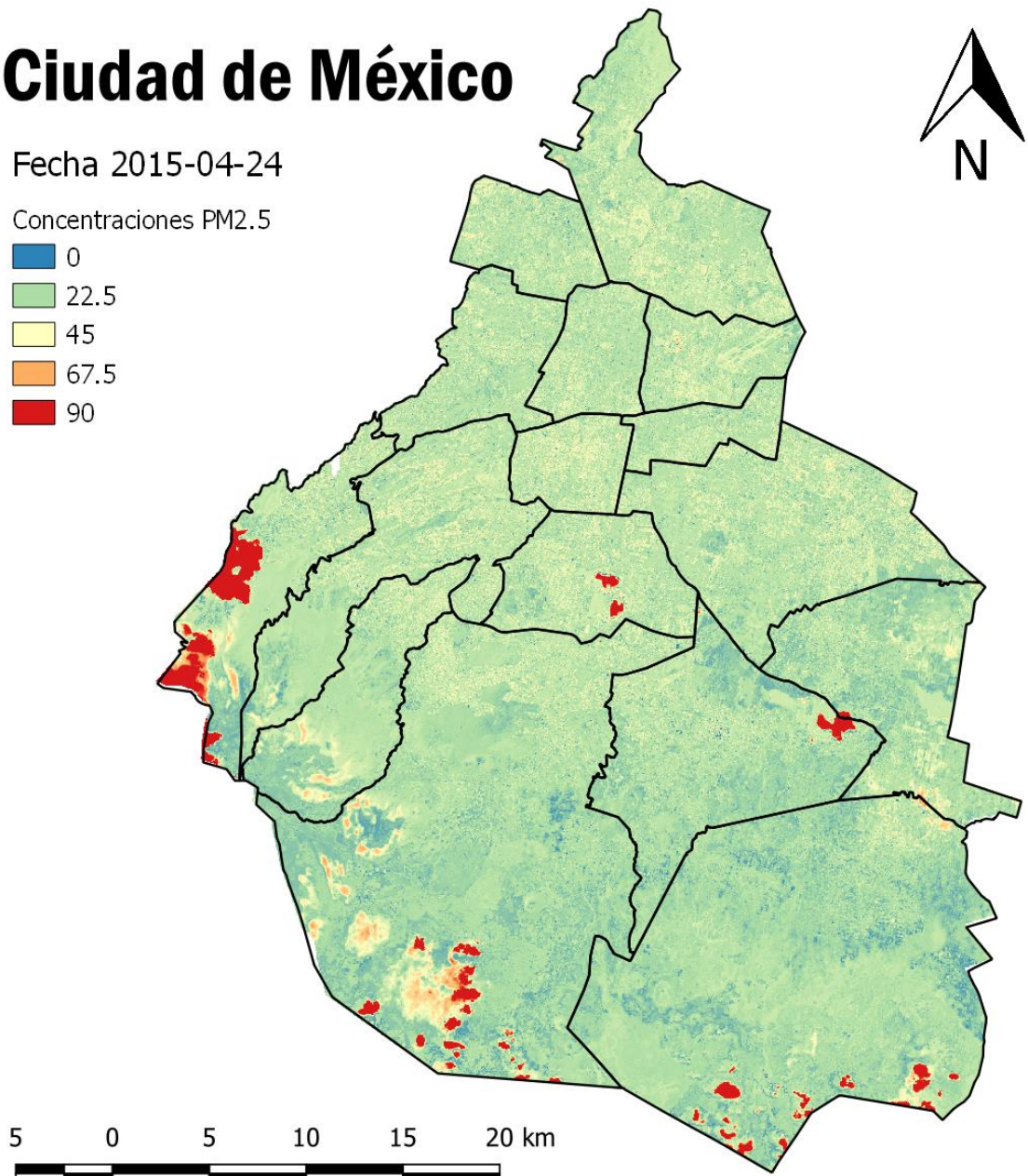


Figura 4.9 Mapa concentraciones PM2.5, 2015-04-24. Elaboración propia.

Ciudad de México

Fecha 2015-05-10

Concentraciones PM2.5

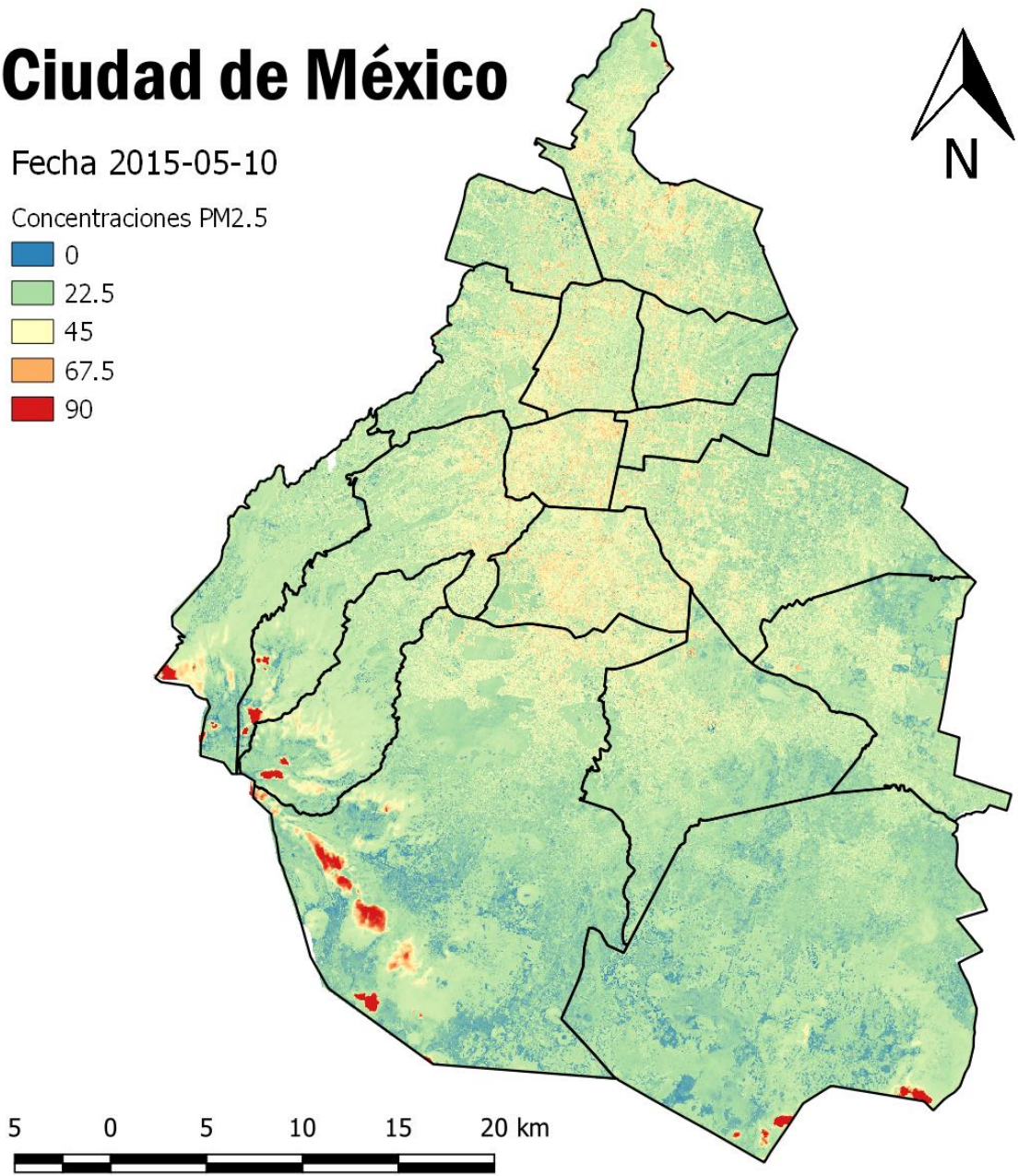


Figura 4.10 Mapa concentraciones PM2.5, 2015-05-10. Elaboración propia.

Ciudad de México

Fecha 2015-06-11

Concentraciones PM2.5

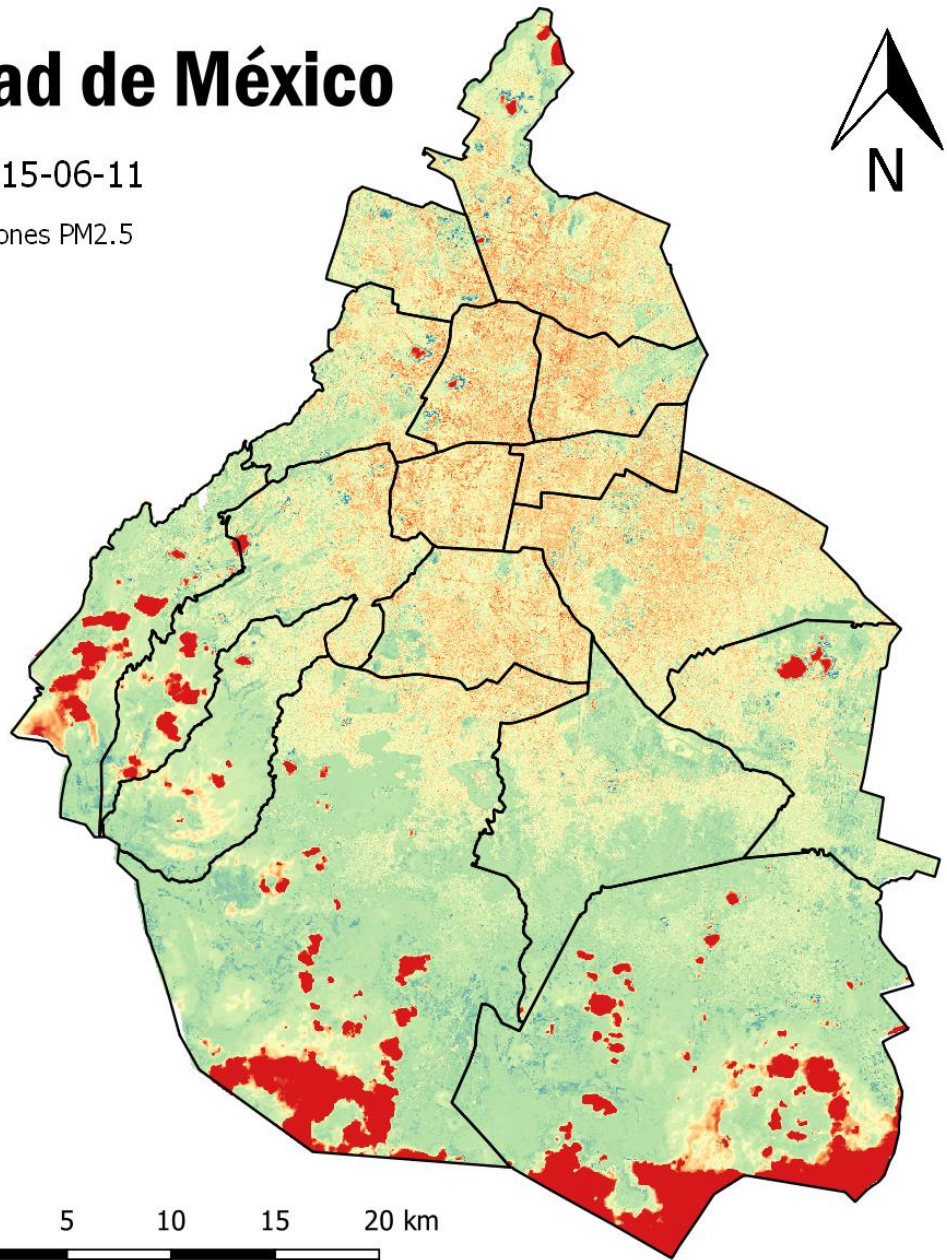
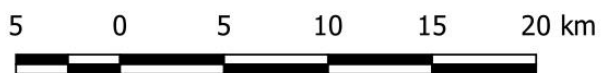


Figura 4.11 Mapa concentraciones PM2.5, 2015-06-11. Elaboración propia.

Ciudad de México

Fecha 2015-07-13

Concentraciones PM2.5

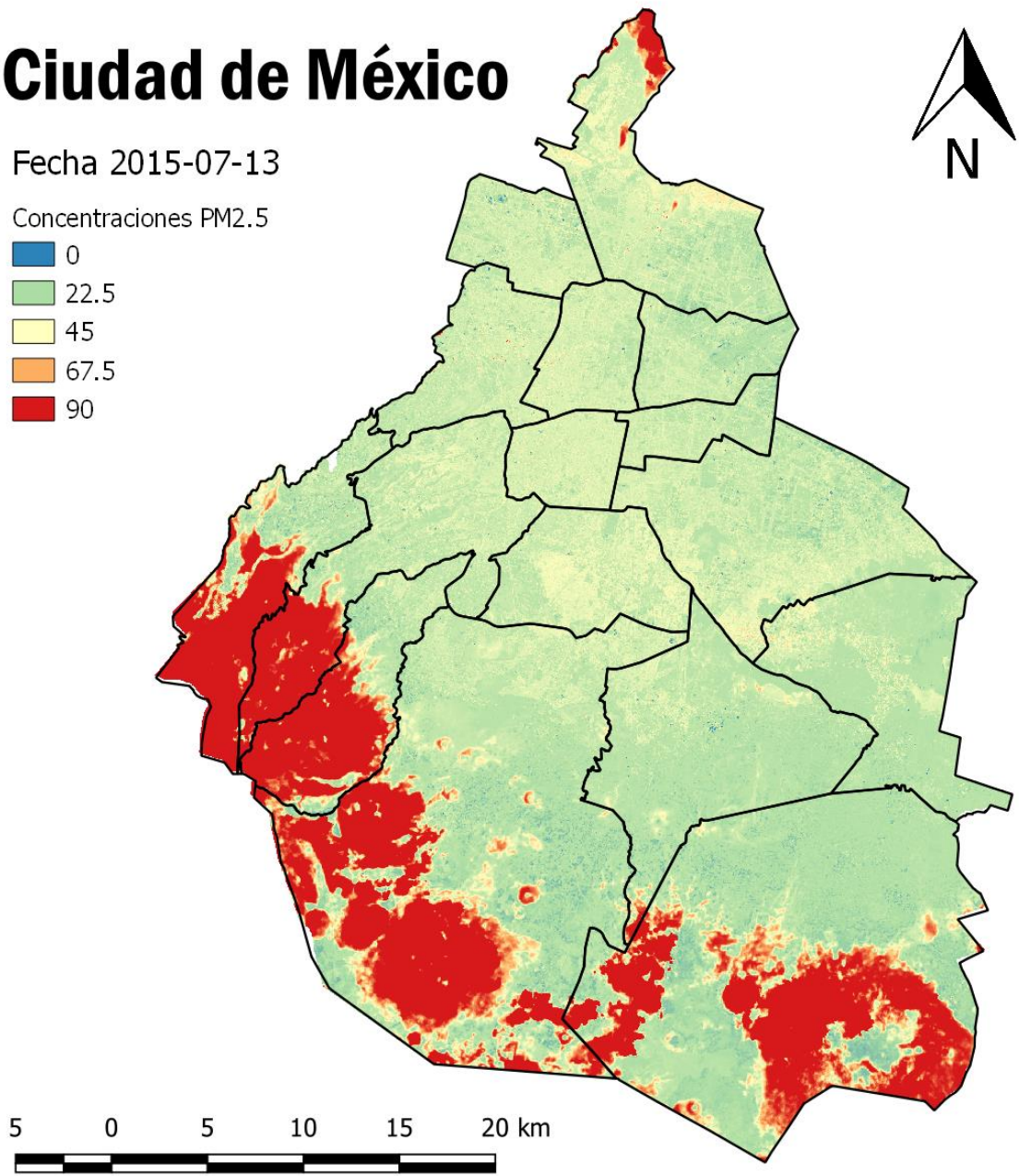


Figura 4.12 Mapa concentraciones PM2.5, 2015-07-13. Elaboración propia.

Ciudad de México

Fecha 2015-12-20

Concentraciones PM2.5

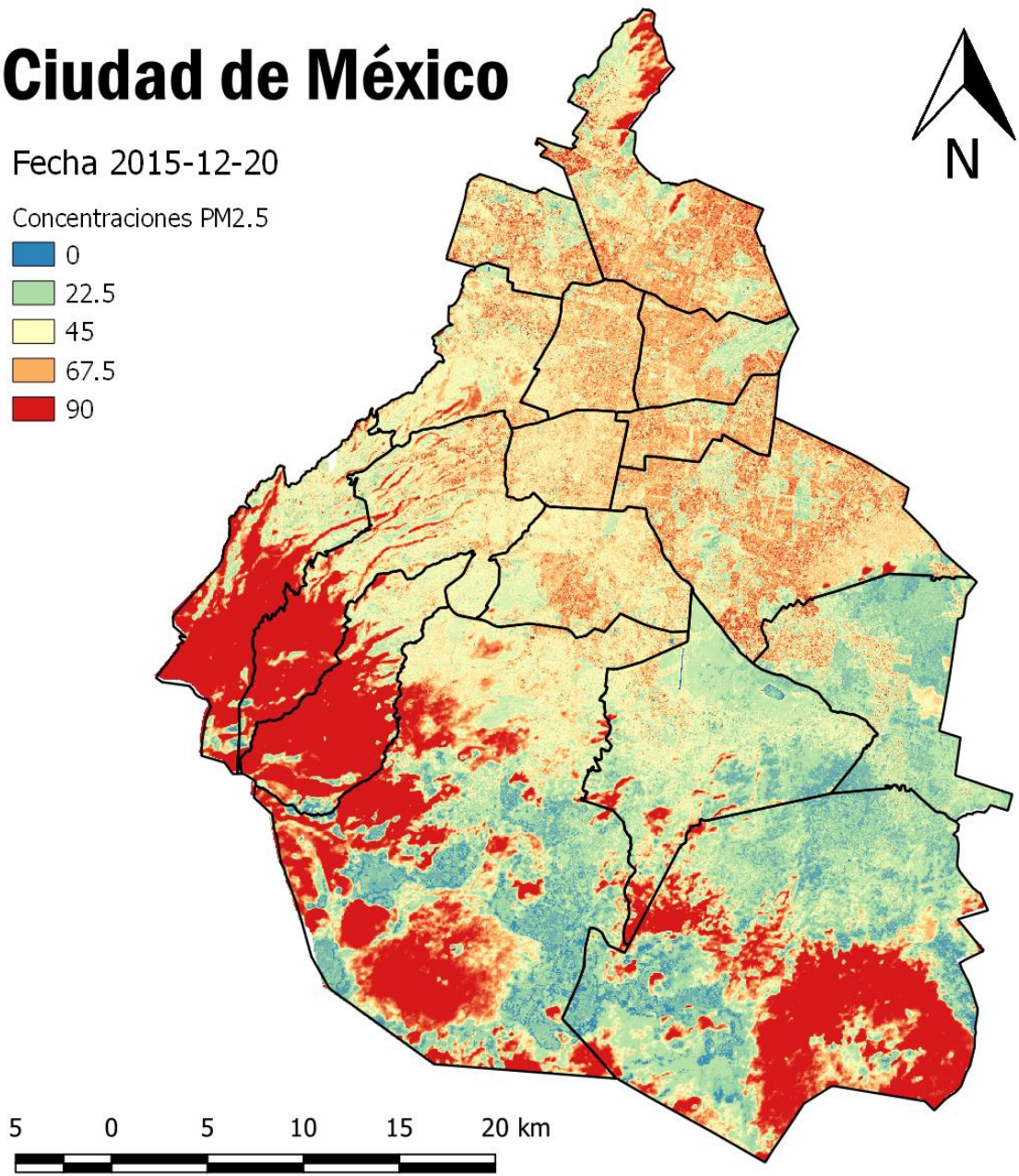


Figura 4.13 Mapa concentraciones PM2.5, 2015-12-20. Elaboración propia.

4.2 Comprobación del modelo

Aunque para realizar una verdadera comprobación del modelo es necesario ir a realizar mediciones a campo en distintas partes de la ciudad, esto es complicado que se requiere muchas mediciones en el momento exacto que pasa el satélite, por ello se aplicó el modelo a una fecha donde sus ninguno de sus datos fue parte de la generación del modelo.

Se usó la imagen del 29 enero de 2019, la cual no había sido usada para la generación de modelos. Se aplicó el modelo como se muestra en la figura 4.6, obteniendo un mapeo con concentraciones de alrededor de $40 \mu\text{g}/\text{m}^3$.

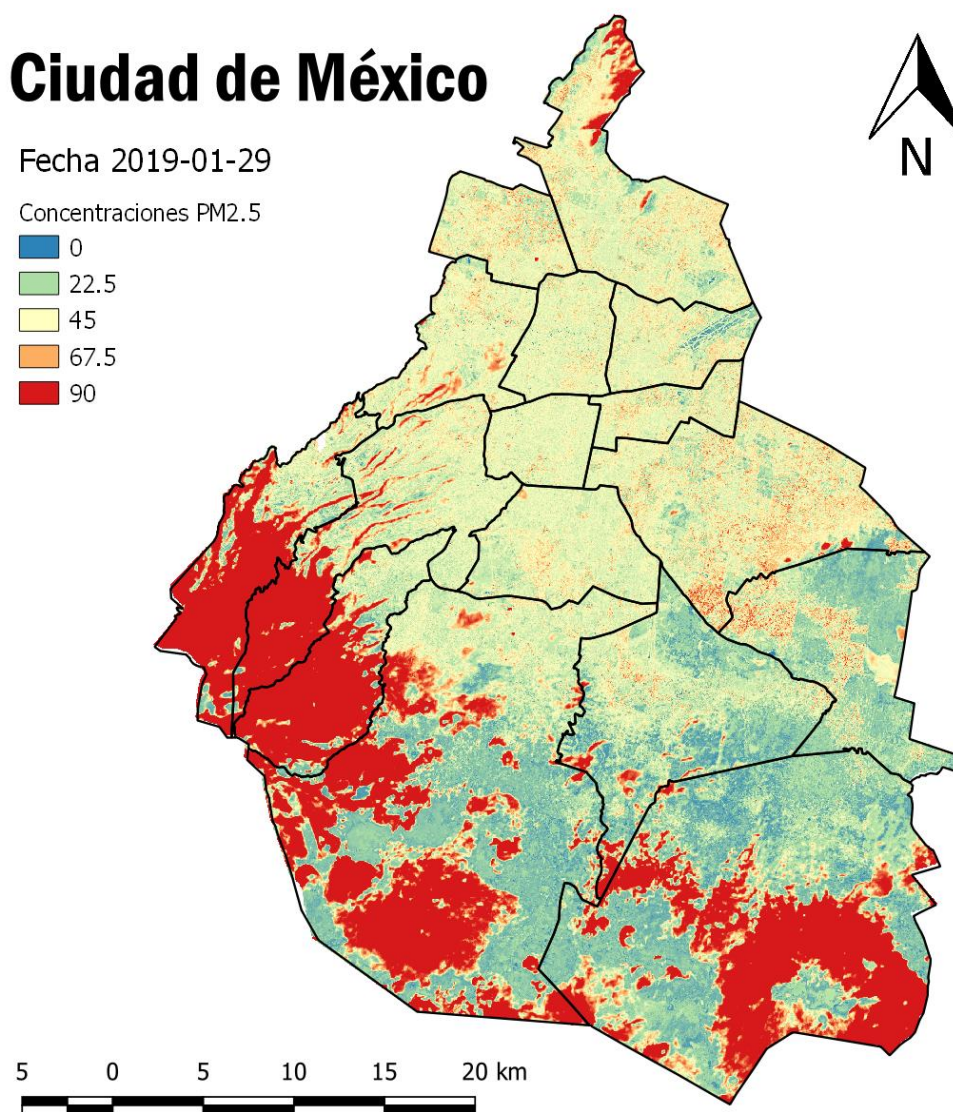


Figura 4.14 Mapa concentraciones PM2.5 2019-01-29. Elaboración propia.

Con este mapa y los datos de contaminantes medidos por RAMA, que en esa fecha solo se tienen datos de 13 estaciones, se obtuvieron los siguientes resultados R^2 de 0.53 y un error cuadrático medio de 12.04.

Conforme estén disponibles nuevas imágenes se podrían ir usando como forma de comprobación del modelo para tener una mayor certidumbre del mismo, aunque como ya se ha mencionado no se puede sustituir a la medición en campo.

Capítulo 5 Conclusiones y recomendaciones

Se logró la obtención de 45 imágenes satelitales Landsat 8, con las cuales se obtuvo la información de reflectancia tanto superficial y reflectancia en el tope de la atmósfera, con esta información y los datos de contaminantes de la Ciudad de México se obtuvo un modelo de regresión con el uso de redes neuronales para estimar las concentraciones de PM_{2.5} en toda la Ciudad de México, lo cual demuestra que las Redes Neuronales en conjunto con imágenes satelitales son una poderosa herramienta en la ingeniería para determinar la distribución de contaminantes PM_{2.5}, por lo tanto el objetivo principal de este trabajo se cumple.

El modelo generado está limitado a zonas urbanas, debido a la naturaleza de las mediciones, por lo que se recomendaría que se realizaran mediciones cerca de cuerpos de agua, zonas montañosas, zonas con alta vegetación mediciones directas sobre vías primarias, para poder realizar modelos más generales.

Uno de los principales desafíos que se enfrenta el modelo son la escasez de datos, ya que los modelos de redes neuronales necesitan muchos datos para entrenarse, en este trabajo se mitigó este problema con el uso de los píxeles adyacentes a la medición. Para estudios futuros convendría añadir los datos de la red de monitoreo atmosférico de Toluca, ya que se encuentra dentro de la imagen del Landsat 8.

Esta metodología puede ser aplicada para la investigación de otros contaminantes atmosféricos dentro de la CDMX ya que se cuenta con mediciones con las mismas características. Además, el uso de estos modelos puede ser generalizado a cualquier tipo de ciudad o poblado que cuente con las mediciones suficientes.

A través de la generación de este modelo, aplicado a toda la escena, resultaron mapas de distribución de concentraciones de PM_{2.5} en un punto en el tiempo. Los mapas que se

generaron pueden ser un punto de partida para aplicar modelos de transporte de contaminantes y así obtener mapas precisos de distribución espacial en intervalos de tiempo posteriores a la toma de la imagen satelital. También éstos pueden ser usados en la detección de puntos de emisión o puntos donde se acumulan contaminantes, y así ayudar en la toma de decisiones sobremedidas de mitigación de riesgos por exposición a concentraciones altas de contaminantes.

Anexo

Variable_removida	R ² -1	Desviación Estandar 1	R ² -2	Desviación Estandar 2	R ² -3	Desviación Estandar 3	R ₂ prom	Desviación Estandar prom	R2/SD	%
aerosol_Low	0.7499	8.84445928	0.7461885	8.841644226	0.756945	8.627152456	0.7510113	8.77108532	11.6790331	0.902707
aerosol_High	0.731851	9.097639153	0.7645554	8.510872381	0.738645	8.816356652	0.7450171	8.808289396	11.8229362	0.91383
TOA.B2	0.7474154	8.857111627	0.7420194	8.816749469	0.736886	8.966950461	0.7421068	8.880270519	11.9662977	0.924911
RA.B5	0.7411707	8.839309491	0.7556752	8.653864552	0.709907	9.413610442	0.7355843	8.968928161	12.1929299	0.942428
TOA.B7	0.7602674	8.585030043	0.7270864	9.206525672	0.725771	9.268934775	0.7377082	9.020163497	12.2272785	0.945083
TOA.B5	0.7332437	9.101493038	0.7228797	9.039373126	0.746702	8.820585798	0.7342751	8.987150654	12.2394873	0.946026
RA.B6	0.7279771	9.115800997	0.7409156	8.885776128	0.727069	8.955971236	0.7319871	8.985849454	12.2759668	0.948846
TOA.B4	0.7219957	9.23405802	0.7363589	9.068586667	0.745255	8.840107802	0.7345365	9.047584163	12.3174065	0.952049
NBR	0.7418919	8.945406656	0.7296536	9.151435174	0.729839	9.043883274	0.733795	9.046908368	12.3289324	0.95294
RA.B4	0.7491792	8.800387263	0.7422999	8.85560872	0.7073	9.516260532	0.7329265	9.057418838	12.357882	0.955177
RA.B3	0.7312329	9.08649676	0.7332319	9.056918408	0.733512	9.051203605	0.732659	9.064872924	12.3725677	0.956312
NDVI	0.7221829	9.233830664	0.732285	9.066915191	0.723692	8.88884927	0.7260535	9.063198375	12.4828252	0.964835
RA.B2	0.7377006	8.868635273	0.6896998	9.281049431	0.736604	8.965986809	0.7213347	9.038557171	12.5303235	0.968506
NDMI	0.7256986	9.115376622	0.7081367	8.982624487	0.733794	9.099502095	0.722543	9.065834401	12.547121	0.969804
TOA.B3	0.7249242	9.210586547	0.7210941	9.243286098	0.73696	8.97185972	0.7276595	9.141910788	12.5634462	0.971066
TOA.B6	0.7148128	9.360742205	0.7102416	9.2515749	0.750787	8.81454819	0.7252805	9.142288432	12.6051753	0.974291
TOA.B11	0.715533	9.183929394	0.7319989	9.14442099	0.72468	9.280804198	0.7240707	9.203051528	12.7101563	0.982406
TOA.B1	0.730672	9.132330485	0.7211047	9.332106305	0.721803	9.298538576	0.7245267	9.254325122	12.7729255	0.987257
prediccion	0.7410971	8.774178315	0.7020521	9.548732683	0.715013	9.295120837	0.7193873	9.206010612	12.7970149	0.989119
RA.B1	0.7167009	9.287519253	0.7051545	9.587022173	0.736818	8.779573949	0.7195577	9.218038458	12.8107012	0.990177
ND.B8	0.7104202	9.19524135	0.7121123	9.121274542	0.714852	9.274174172	0.7124613	9.196896688	12.9086254	0.997746
Ninguna	0.7193069	9.30624	0.7193069	9.30624	0.719307	9.30624	0.7193069	9.30624	12.9377878	1
NBR2	0.7164285	9.318970304	0.7084757	9.524470599	0.719404	9.312443109	0.7147693	9.385294671	13.1305224	1.014897
RA.B7	0.7173349	9.376775724	0.6991592	9.540150219	0.723006	9.265640032	0.7131665	9.394188658	13.1725034	1.018142
TOA.B10	0.6985697	9.630467717	0.6975831	9.669759098	0.741644	8.940812498	0.7125991	9.413679771	13.2103456	1.021067
TOA.B9	0.67664	9.78407515	0.7162285	9.113781291	0.723545	9.262529599	0.7054711	9.386795346	13.3057128	1.028438

Figura A1. Cálculo de R2 y desviación estándar al remover variables primera iteración. Elaboración propia.

Bibliografía

- Anenberg, S. C., Horowitz, L. W., Tong, D. Q., & West, J. J. (2010). An estimate of the global burden of anthropogenic ozone and fine particulate matter on premature human mortality using atmospheric modeling. *Environmental Health Perspectives*, 118(9), 1189–1195. <https://doi.org/10.1289/ehp.0901220>
- Arellano Monreal, Y. (2019). *USO DE REDES NEURONALES ARTIFICIALES PARA LA OBTENCIÓN DE PERFILES DE CONCENTRACIONES DE OZONO EN LA ZONA METROPOLITANA DEL VALLE DE MÉXICO*. Universidad Nacional Autónoma de México.
- Camacho Gonzáles, P. (2018). *Sistema de Redes Neuronales para la predicción de ozono en la CDMX y el área metropolitana*. Universidad Nacional Autónoma de México.
- Canada Centre for Remote Sensing. (2016). *Fundamentals of Remote Sensing*, 258. [https://doi.org/10.1016/0301-4207\(76\)90065-9](https://doi.org/10.1016/0301-4207(76)90065-9)
- Department of the Interior U.S. (2018). *Landsat 8 Data Users Handbook*, 8(October), 2–4.
- Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., & Schwartz, J. (2016). Assessing PM 2.5 Exposures with High Spatiotemporal Resolution across the Continental United States Graphical abstract HHS Public Access. *Environ Sci Technol*, 50(9), 4712–4721. <https://doi.org/10.1021/acs.est.5b06121>
- Doug. (2018). How to choose the number of hidden layers and nodes in a feedforward neural network? Retrieved from <https://stats.stackexchange.com/questions/181/how-to-choose-the-number-of-hidden-layers-and-nodes-in-a-feedforward-neural-netw>
- Grover, P. (2018). 5 Regression Loss Functions All Machine Learners Should Know. Retrieved from <https://heartbeat.fritz.ai/5-regression-loss-functions-all-machine-learners-should-know-4fb140e9d4b0>
- Itidxer. (2018). What is batch size in neural network? Retrieved from <https://stats.stackexchange.com/q/153535>
- Jones, H. G., & Vaughan, R. (2010). *Remote Sensing of Vegetation* (First). Oxford University.

- Kraus, R., Szalay, C., Meyer, C., Kilian, O., & Schnettler, R. (2007). Die Distale Radiusfraktur - Eine Torwartverletzung bei Kindern und Jugendlichen. *Sportverletzung-Sportschaden*, 21(4), 177–179. <https://doi.org/10.1055/s-2007-963706>
- Kurt Hornik, Maxwell Stinchcombe, & Halbert White. (1989). Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, 2, 359–366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
- Liang, S., Fang, H., & Chen, M. (2001). Atmospheric correction of Landsat ETM+ land surface imagery-Part I: Methods. *IEEE Transactions on Geoscience and Remote Sensing*, 39(11), 2490–2498. <https://doi.org/10.1109/36.964986>
- Lyon, R. F. (2017). Neural Networks for Machine Learning. *Human and Machine Hearing*, 419–440. <https://doi.org/10.1017/9781139051699.031>
- Martín Abadi, Ashish Agarwal, Paul Barham, E. B., Zhifeng Chen, Craig Citro, Greg S. Corrado, A. D., Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, I. G., Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Y. J., Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, M. S., Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, J. S., ... Yuan Yu, and X. Z. (2015). TensorFlow. Retrieved from <https://www.tensorflow.org/>
- NOM-251-SSA1-2009. (2009). NORMA Oficial Mexicana NOM-251-SSA1-2009. *Control*, 52(27), 52–70.
- Rashid, T. (2016). *Make Your Own Neural Network* (1st ed.). CreateSpace Independent Publishing Platform.
- Servicio Geológico Minero Argentino. (2018). Sensores Remotos. Retrieved from http://www.segemar.gov.ar/igrm/sensores_remotos/
- Singh Gill, N. (2017). What are some neural network architectures? Retrieved from <https://www.quora.com/What-are-some-neural-network-architectures>
- Torres, N., & Vivianco, V. (2018). *Comparación en la estimación de material particulado PM10 usando imágenes satelitales Landsat 7, Landsat 8 y MODIS en Quito*. Universidad Politécnica Salesiana. Retrieved from <https://core.ac.uk/download/pdf/84704028.pdf>
- Torres, V. H. (2017). Estudio de la contaminación ambiental de la Ciudad de México

mediante los modelos redes neuronales artificiales y algoritmos genéticos. Retrieved from <http://tesis.ipn.mx/handle/123456789/20461>

U.S. Geological Survey. (2018). LANDSAT 8 SURFACE REFLECTANCE CODE (LaSRC) Version 4.3 PRODUCT: Product Guid, (December), 40. <https://doi.org/10.1016/J.RSE.2016.04.008>

United States Geological Survey. (2018a). How does Landsat 8's 12-bit data improve data products? Retrieved from <https://landsat.usgs.gov/how-does-landsat-8s-12-bit-data-improve-data-products>

United States Geological Survey. (2018b). Landsat Satellite Missions. Retrieved from https://www.usgs.gov/land-resources/nli/landsat/landsat-satellite-missions?qt-science_support_page_related_con=2#qt-science_support_page_related_con

Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24(3), 583–594. <https://doi.org/10.1080/01431160304987>