



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

USO DE REDES NEURONALES ARTIFICIALES PARA LA OBTENCIÓN DE PERFILES DE  
CONCENTRACIONES DE OZONO EN LA ZONA METROPOLITANA DEL VALLE DE MÉXICO

TESIS  
QUE PARA OPTAR POR EL GRADO DE  
MAESTRO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

PRESENTA:  
YAZARET IRISNELIDA ARELLANO MONREAL

Directora de tesis:  
DRA. SUEMI RODRÍGUEZ ROMO  
FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN, UNAM

CIUDAD UNIVERSITARIA, CD. MX. ENERO, 2019



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



## AGRADECIMIENTOS

*Dedico el presente trabajo de tesis a las personas que más amo,  
Luis Yerald y Cristian, quienes con su cariño y apoyo,  
han hecho realidad uno de mis más grandes logros.*

*A mi madre, quien me ha brindado la confianza  
y una palabra de aliento.*

*A mi directora de tesis Dra. Suemi Rodríguez Romo,  
Quién ha puesto gran dedicación y esfuerzo,  
me ha guiado y ha compartido su conocimiento  
para que pudiera llevar a cabo el presente trabajo.*

*Al Consejo Nacional de Ciencia y Tecnología (CONACyT)  
por la beca que se me otorgó para la realización de mis  
estudios de maestría.*

*A la DGAPA-UNAM por el apoyo mediante  
el proyecto PAPIIT 101817.*

*Al Dr. Ricardo Paramont Hernández García que  
me ha brindado asesoría y ha compartido su conocimiento.*

*A mis compañeros y amigos que  
me han brindado su apoyo.*

*A los profesores que colaboraron con mi  
formación académica y profesional.*

# CONTENIDO

Glosario de las Estaciones de Monitoreo Atmosférico

Resumen

Capítulo 1 Introducción y objetivos .....	1
1.1. Introducción .....	1
1.1.1. Datos utilizados .....	1
1.1.2. Funcionamiento de las redes Multicapa de Perceptrones (MLP) -	2
- Entrenamiento de redes MLP .....	2
- Capacidad de generalización de las redes MLP .....	4
- Uso de las redes MLP .....	5
1.1.3. Utilización de las redes MLP para obtener perfiles de	
concentración de Ozono .....	5
1.2. Hipótesis .....	6
1.3. Objetivo General .....	6
1.4. Objetivos Particulares .....	7
Capítulo 2 Antecedentes .....	8
Capítulo 3 Marco Teórico .....	11
3.1. Metodología .....	11
3.2. Arquitectura de las redes MLP .....	11
- El perceptrón .....	13
3.3. Entrenamiento de las redes MLP .....	17
3.3.1. Propagación hacia adelante .....	18
3.3.2. Propagación hacia atrás .....	18
3.3.3. Uso del método momentum .....	21
Capítulo 4 Experimentación y Resultados .....	23
4.1. La base de datos original .....	23
4.2. Preparación de la base de datos .....	29

4.2.1. Distribución de los días utilizables a lo largo de 2017 -----	30
4.3. Seleccionando la topología de las redes MLP -----	32
4.3.1. Topologías propuestas para las redes MLP -----	32
4.3.2. Construcción de las tablas de entrenamiento y de prueba -----	33
- Selección de las estaciones de prueba y construcción de la tabla de prueba de un día -----	36
- Construcción de la tabla de entrenamiento del día -----	37
- Construcción de las tablas normalizadas de entrenamiento y de prueba de un día -----	37
- Cálculo del error relativo de ajuste de un perfil de concentraciones de ozono -----	38
4.3.3. Procedimiento para probar las doce topologías propuestas en esta tesis -----	39
4.3.4. Resultados de la aplicación del procedimiento para seleccionar la mejor topología -----	40
4.4. Obtención de perfiles diarios de concentración de ozono en la zona Metropolitana del Valle de México -----	42
4.4.1. Tabla de entrenamiento y de prueba normalizadas -----	43
4.4.2. Entrenamiento y prueba de las redes MLP -----	45
- Análisis de que tan correcto fue usar 50,000 épocas de entrenamiento -----	45
4.5. Resultados de los ajustes de los perfiles diarios de las concentraciones de ozono para las estaciones de prueba -----	48
4.5.1. Análisis de los Errores Relativos de Ajuste (ERA) mínimos, ERA máximos y la mediana de los ERA de las tres corridas -----	56
4.5.2. Caso especial estación Acolman -----	59
4.5.3. Análisis de los resultados de los ajustes de los perfiles diarios de las concentraciones de ozono para las estaciones de prueba con mayor ERA -----	63
4.5.4. Análisis de los resultados con los mejores y los peores ajustes de los perfiles diarios de las concentraciones de ozono -----	66

- Ejemplo del <i>mejor ajuste</i> de perfiles de concentración de ozono .....	67
- Ejemplo del <i>peor ajuste</i> de perfiles de concentración de ozono .....	74
4.5.5. Software desarrollado .....	81
Conclusiones .....	83
Bibliografía .....	86
Apéndice A .....	89
Apéndice B .....	96

## Glosario de las estaciones de monitoreo Atmosférico

<b>N°</b>	<b>Identificador</b>	<b>Nombre de estación de monitoreo</b>
1	ACO	Acolman
2	AJM	Ajusco Medio
3	AJU	Ajusco
4	ATI	Atizapán
5	BJU	Benito Juárez
6	CAM	Camarones
7	CCA	Centro de Ciencias de la Atmósfera
8	CHO	Chalco
9	CUA	Cuajimalpa
10	CUT	Cuautitlán
11	FAC	FES Acatlán
12	GAM	Gustavo A. Madero
13	HGM	Hospital General de México
14	INN	Instituto Nacional de Investigaciones Nucleares (ININ)
15	IZT	Iztacalco
16	LLA	Los Laureles
17	LPR	La Presa
18	MER	Merced
19	MGH	Miguel Hidalgo
20	MON	Montecillo
21	MPA	Milpa Alta
22	NEZ	Nezahualcóyotl
23	PED	Pedregal
24	SAG	San Agustín
25	SFE	Santa Fe
26	TAH	Tláhuac
27	TLA	Tlalnepantla
28	TLI	Tultitlán
29	UAX	UAM Xochimilco
30	UIZ	UAM Iztapalapa
31	VIF	Villa de las Flores
32	XAL	Xalostoc

# Resumen

Para registrar las concentraciones de contaminantes y de algunos parámetros relacionados con ellas, a lo largo de cada día del año en la Zona Metropolitana del Valle de México (ZMVM), la Secretaría de Medio Ambiente (SEDEMA), cuenta con el Sistema de Monitoreo Atmosférico (SIMAT). Entre los contaminantes cuyas concentraciones registra el SIMAT, se encuentra el ozono. La red del SIMAT encargada de hacerlo es la Red Automática de Monitoreo Ambiental (RAMA). La RAMA cuenta con varias estaciones distribuidas de manera *no uniforme* en la en la ZMVM.

El objetivo de este trabajo es estudiar con qué grado de error se puede obtener los perfiles diarios de concentraciones de ozono en puntos dentro de la ZMVM donde no hay estaciones de monitoreo, usando como datos los perfiles de concentraciones de ozono obtenidos por estaciones de la RAMA y las ubicaciones de esas estaciones, para entrenar redes multicapa de perceptrones (MLP) para que aprendan los perfiles de concentración de ozono en un día dado usando como datos de entrada las coordenadas (longitud y latitud).

Para comprobar el buen funcionamiento de las redes MLP entrenadas, algunas de las estaciones son usadas como estaciones de prueba, luego se compara el perfil obtenido para el punto donde se encuentran esas estaciones con el perfil registrado por ellas. A una medida de la diferencia entre los perfiles obtenidos por las redes entrenadas y los registrados prácticamente por las estaciones de prueba, se le llama aquí *error relativo de ajuste* (ERA). Las redes MLP fueron entrenadas con el método de propagación hacia atrás usando *momentum* para procurar evitar que el proceso de entrenamiento se entrapara en pequeños mínimos.

En esta investigación se trabajó con los datos de concentración de ozono de la RAMA para el año 2017. La mejor topología de las redes usadas tiene dos capas ocultas, en la primera de ellas, la capa 2, hay 8 perceptrones, en la segunda, la capa 3, hay 7 perceptrones (la capa de entrada tiene 2 distribuidores y la de salida 24

perceptrones, correspondientes a las coordenadas y el perfil diario de concentraciones de ozono -una medida por cada hora del día-, respectivamente).

Los ERAs obtenidos se clasificaron en tres categorías según su magnitud. La primera categoría corresponde a las estaciones ubicadas en la parte central de la Ciudad de México los ERAs promedio de estas estaban entre 15 y 20%; la segunda, a estaciones que se encuentran alrededor de las primeras, los ERAs de estas está entre el 20 y 25%; por último, la tercera categoría corresponde a estaciones en la periferia de la ZMVM, sus ERAs se encuentran entre el 25 y el 30%.

# Capítulo 1      Introducción y objetivos

## 1.1.      Introducción

El monitoreo de las concentraciones de contaminantes en el aire de las ciudades modernas, se ha convertido en una tarea de primordial importancia. Entre estos contaminantes el ozono es uno de los más importantes a medir (Bravo, Perrin, Sosa, & Torres, 1988). Desde 1986 en la Zona Metropolitana del Valle de México (ZMVM), se han instalado estaciones automáticas para monitorear las concentraciones de diferentes contaminantes en el aire, entre ellos el ozono (Bravo, Perrin, Sosa, & Torres, 1988). Estas estaciones, cuando operan sin contratiempos, registran durante todo el día, obteniendo una medición cada hora, las concentraciones de los contaminantes. Llamemos *perfil de concentraciones diario* o simplemente *perfil de concentraciones*, a la serie de tiempo constituida por los 24 datos diarios de algún contaminante.

El objetivo principal de este trabajo es estudiar la posibilidad de obtener, para un día dado, los perfiles de concentración de ozono en puntos donde no haya alguna estación, a partir de los perfiles obtenidos por las estaciones de monitoreo ese día. Para llevar a cabo esta tarea se usan redes neuronales artificiales del tipo multicapa de perceptrones por sus siglas en inglés MLP (*multilayer perceptron networks*).

### 1.1.1.      Datos utilizados

La Secretaría de Medio Ambiente (SEDEMA), de la Ciudad de México, publica en su portal en internet (SEDEMA, 2016), los datos del Sistema de Monitoreo Atmosférico (SIMAT); éste comprende cuatro redes de monitoreo: 1) la Red Automática de Monitoreo Atmosférico (RAMA), 2) la Red Manual de Monitoreo Atmosférico (REDMA), 3) la Red de Depósito Atmosférico (REDDA) y 4) la Red de Meteorología y Radiación Solar (REDMET).

En el mismo portal se muestran las ubicaciones de las diferentes estaciones de monitoreo en la ZMVM (SEDEMA, 2017). Dichas ubicaciones comprenden la latitud,

longitud y altura sobre el nivel del mar de cada una de las estaciones de monitoreo automáticas.

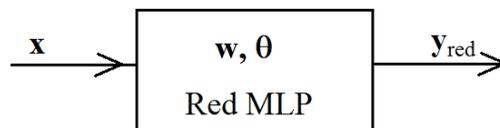
Los datos de estas cuatro redes, así como los de las ubicaciones de las estaciones, son de acceso público y se encuentran en el portal. Para este trabajo se usaron los datos de la red RAMA de 2017 (SEDEMA, 2017) y los de las coordenadas de ubicación de las estaciones.

### 1.1.2. Funcionamiento de las redes Multicapa de Perceptrones (MLP)

Una red MLP es una máquina de cómputo que se usa principalmente para dos tareas: clasificar y hacer regresiones. En este trabajo se utilizará para esta última. Para usar una red MLP, es necesario entrenarla antes. A continuación, se explicará en qué consiste este proceso.

#### Entrenamiento de las redes MLP

Funcionalmente, una red MLP que trabaja para hacer regresiones se puede esquematizar como se muestra en la figura I.1.



**Figura I.1** Representación funcional de una red MLP para hacer regresiones,  $\mathbf{x}$  representa el vector de entradas (variables independientes),  $\mathbf{y}_{\text{red}}$  el vector de salidas (variables dependientes),  $\mathbf{w}$  y  $\boldsymbol{\theta}$  son parámetros de la red.

El funcionamiento esquematizado en la figura **I.1** se puede representar en notación funcional con las ecuaciones **I.1**.

$$y_{1,red} = f_1(\mathbf{x} | \mathbf{w}, \boldsymbol{\theta}) \quad (\mathbf{I.1a})$$

$$y_{2,red} = f_2(\mathbf{x} | \mathbf{w}, \boldsymbol{\theta}) \quad (\mathbf{I.1b})$$

⋮

$$y_{n,red} = f_n(\mathbf{x} | \mathbf{w}, \boldsymbol{\theta}) \quad (\mathbf{I.1n})$$

Donde

- $\mathbf{x}$  es un vector que representa los datos de entrada a la red,
- $y_{i,red}$  son los valores que la red calcula para la  $\mathbf{x}$  alimentada,
- $\mathbf{w}, \boldsymbol{\theta}$  parámetros de la red.

Los valores de los vectores de parámetros  $\mathbf{w}$  y  $\boldsymbol{\theta}$  (denominados *pesos* y *umbrales*, respectivamente), se obtienen mediante un proceso de minimización de una función error (**I.2**).

$$E = \frac{1}{2} \sum_{i=1}^n (y_{i,red} - y_i)^2 \quad (\mathbf{I.2})$$

A ese proceso de minimización se le conoce como *entrenamiento de la red* o proceso de *aprendizaje de la red*. Para llevar a cabo el proceso de entrenamiento se debe crear una tabla llamada *tabla de entrenamiento* en la que cada uno de sus renglones o *ejemplos* está constituido por un vector de valores de la variable independiente,  $\mathbf{x}$ , y el correspondiente vector de valores de la variable dependiente,  $\mathbf{y}$  (ver tabla **I.1**).

Una forma de entrenar la red para calcular los valores de  $\mathbf{w}$  y  $\boldsymbol{\theta}$ , es la siguiente:

Para cada renglón o *ejemplo* de la tabla de entrenamiento:

- 1) se introduce a la red los valores de las variables independientes del renglón o *ejemplo*,

- 2) se hace que la red trabaje para calcular los valores de salida, que son los valores de las variables dependientes,
- 3) se calcula el valor de la función error (**I.2**),
- 4) se modifican los valores de los parámetros de la red,  $\mathbf{w}$  y  $\boldsymbol{\theta}$ , con el objeto de disminuir el valor de la función error.

**Tabla I.1** Ejemplo de tabla para entrenar una red MLP. La red hará regresión de  $m_N$  funciones que dependen de  $m_1$  variables independientes.

No. de ejemplo (o de renglón)	Valores de las Variables independientes	Valores de las Variables dependientes
1	$x_{1,1}, x_{2,1}, \dots, x_{m_1,1}$	$y_{1,1}, y_{2,1}, \dots, y_{m_N,1}$
2	$x_{1,2}, x_{2,2}, \dots, x_{m_1,2}$	$y_{1,2}, y_{2,2}, \dots, y_{m_N,2}$
$\vdots$	$\vdots$	$\vdots$
$r$	$x_{1,r}, x_{2,r}, \dots, x_{m_1,r}$	$y_{1,r}, y_{2,r}, \dots, y_{m_N,r}$
$\vdots$	$\vdots$	$\vdots$
$R$	$x_{1,R}, x_{2,R}, \dots, x_{m_1,R}$	$y_{1,R}, y_{2,R}, \dots, y_{m_N,R}$

### Capacidad de generalización de las redes MLP

Una de las características de las redes MLP es su capacidad de generalizar, es decir, su capacidad de calcular valores coherentes de las variables dependientes, cuando se les alimentan valores distintos de las variables independientes a aquellos con los que fueron entrenadas. Para comprobar la capacidad de generalización de una red, se elabora una *tabla de prueba* semejante a la tabla **I.1**, constituida por ejemplos con los que la red MLP no se ha entrenado. La tabla de prueba se usa como la tabla de entrenamiento, con la única diferencia de que no se aplica el paso (4), es decir, no se usa para modificar los valores de los parámetros de la red:

Para cada ejemplo de la tabla de prueba:

- 1) Se introduce a la red los valores de las variables independientes del ejemplo,

- 2) se hace que la red trabaje para calcular los valores de salida, que son los valores de las variables dependientes,
- 3) se calcula el valor de la función error **(I.2)** para reportarlo como error de prueba.

El proceso de entrenamiento consiste en ir modificando paulatinamente, mediante repetidas alimentaciones de la tabla entrenamiento, los parámetros  $\mathbf{w}$  y  $\boldsymbol{\theta}$  de la red, de manera tal que la función error **(I.2)** vaya disminuyendo tanto para la tabla de entrenamiento como para la tabla de prueba. El proceso se repite hasta que se obtiene un error lo suficientemente pequeño o hasta que el error promedio obtenido de la tabla de prueba, deja de disminuir y comienza a aumentar; en este último caso, son los valores de los parámetros  $\mathbf{w}$  y  $\boldsymbol{\theta}$  obtenidos en el mínimo valor de error, y son los que se le deben dejar a la red. De esta manera se mantiene la característica de generalización de la red, que es su principal virtud.

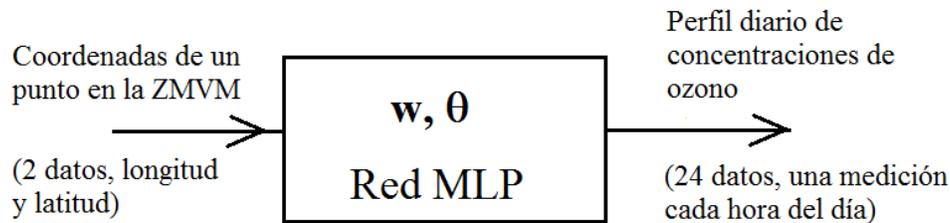
### **Uso de las redes MLP**

Una vez que una red MLP ha sido entrenada, se usa para obtener los valores de las variables dependientes correspondientes a valores de las variables independientes que no fueron usados para su entrenamiento, algunos autores le dan el nombre de etapa de *producción* de la red a esta etapa de la vida de la red (Freeman & Skapura, 1993).

#### 1.1.3. Utilización de redes MLP para obtener perfiles de concentración de ozono

Para este trabajo, se usaron los datos obtenidos de la SEDEMA, como se ha descrito en la sección 1.1.1 de esta introducción, se tomaron de ahí únicamente los datos de la concentración de ozono para el año de 2017, para elaborar las tablas de entrenamiento y de prueba para las redes MLP. Las variables independientes usadas fueron las coordenadas geográficas, longitud y latitud, en las que se ubican las diferentes estaciones automáticas de monitoreo ambiental en la ZMVM (referenciadas a

un punto en esa zona). Para este trabajo no se consideró el parámetro de la altura, porque se quería investigar si se podía obtener una buena aproximación de perfiles de concentración de ozono únicamente considerando las coordenadas (longitud y latitud). Las variables dependientes fueron las 24 mediciones que constituyen el perfil diario de concentración de ozono en el punto donde se ubica cada estación (ver figura I.2).



**Figura I.2** Representación funcional de las redes usadas en este trabajo. Para entrenamiento y prueba, las coordenadas del punto corresponden a las de una estación. Una vez entrenada la red se puede alimentar puntos que no correspondan necesariamente a estaciones de monitoreo.

## 1.2. Hipótesis

La hipótesis principal de la que se partió en la realización de este trabajo, es que es posible obtener los perfiles diarios de concentración de ozono en puntos donde no haya estaciones de monitoreo ambiental, en la zona metropolitana del valle de México, usando redes multicapa de perceptrones (redes MLP) utilizando como datos las coordenadas geográficas de las estaciones automáticas de monitoreo ambiental y los perfiles diarios de las concentraciones de ozono medidas por éstas, para entrenar dichas redes.

## 1.3. Objetivo General

Obtener los perfiles diarios de concentración de Ozono ( $O_3$ ) desconocidos en las diferentes coordenadas de la zona metropolitana del Valle de México, a partir de los datos conocidos usando redes multicapa de perceptrones entrenadas con los datos que proporciona la SEDEMA.

## 1.4. Objetivos Particulares

1.- Desarrollar un conjunto de programas para:

- a) Simular redes MLP y entrenarlas con el algoritmo de propagación hacia atrás, usando las bases de datos de concentración de Ozono del año 2017.
- b) Identificar y separar aquellos días en los que las estaciones de monitoreo obtuvieron el total de mediciones esperado.
- c) Seleccionar aleatoriamente un número fijo de estaciones para probar el entrenamiento de cada una de las redes.
- d) Seleccionar una muestra de datos para elegir la topología de red MLP más adecuada.

2.- Analizar los resultados del mejor y el peor ajuste del perfil de concentración de ozono.

## Capítulo 2                    Antecedentes

El uso de las redes neuronales artificiales (RNA), son herramientas que han ampliado el campo del cómputo en la investigación científica.

El uso de las RNA se ha motivado por los rápidos avances tecnológicos para tratar de modelar habilidades del cerebro humano. Según (Spellman, 1999) una red neuronal es un dispositivo de procesamiento implementado como un algoritmo de una red con elementos con procesamientos simples.

La mayoría de los sistemas del mundo real son no lineales, y los modelos lineales no capturan la esencia del fenómeno. Las RNA pueden manejar funciones no lineales, y pueden aprender de relaciones que no se conocen de otra manera, y son extremadamente eficientes computacionalmente una vez entrenadas (Spellman, 1999).

En 1995 se desarrollaron modelos de RNAs para realizar la predicción a corto plazo de la concentración de ozono en la Ciudad de México, con cinco estaciones de monitoreo (Ruiz, Mayora, Torres, & Ruiz, 1995). Algunas de estas estaciones de monitoreo se ocuparon para el trabajo actual pero con los datos del monitoreo del año 2017.

En 1998 se realizó un estudio en el que se muestra las primeras aplicaciones científicas de la red multicapa de perceptrones (MLP) en las ciencias atmosféricas, para obtener predicción, aproximación de funciones y clasificación de patrones (Gardner & Dorling, 1998).

Según (Spellman, 1999) y (Andrew, 1997) aseguran que con las RNAs se obtienen mejores resultados comparado con un modelo de regresión tradicional, y sobre todo cuando se utilizan una gran cantidad de datos, en este caso los datos relacionados con la concentración de ozono.

La relación entre el clima y el ozono es altamente compleja y no lineal; en el Reino Unido se obtuvieron estimaciones de la concentración de ozono utilizando una variable meteorológica, mediante la simulación de una red MLP (modelo basado en regresión) (Spellman, 1999).

Los desastres por smog se han presentado cada vez más frecuentemente y pueden tener graves consecuencias para el medio ambiente y para la salud. Además tomando en cuenta que actualmente lo que más se utiliza son las redes sociales, se realizó una investigación por (Chen, Chen, Wu, Hu, & Pan, 2017), en donde se propone un modelo de RNA con un enfoque analítico predictivo que utiliza los datos de los medios sociales y los sensores físicos para predecir cuál será el peligro para la salud relacionado con el smog al día siguiente.

Considerando los parámetros de concentración de siete contaminantes y cinco variables meteorológicas, según (Al, Abdul, & Bakheit, 2008) desarrolló varios modelos de RNAs para predecir la concentración de ozono de Kuwait.

Según (Wang & Lu, 2006) los datos de concentración de ozono son conocimiento, dicho conocimiento es utilizado en un modelo de red MLP, y nos dice que el rendimiento predictivo de la red MLP está determinado en gran medida por el algoritmo de entrenamiento que se elija.

En 2008 se realizó un estudio muy útil para apoyar decisiones en materia de contaminación de ozono en el área fronteriza de Mexicalli-Calexico, en el cual se desarrollaron 12 modelos de RNAs, en donde el objetivo de estos modelos era la predicción de concentraciones máximas de ozono para el día siguiente, considerando parámetros meteorológicos y los precursores de ozono (Salazar , Ordieres, Vergara, & Capuz, 2008).

Los episodios de ozono se han presentado en diferentes partes del mundo. Un episodio de ozono se le conoce cuando a una zona determinada llegan niveles muy altos de concentración de ozono que sobrepasan el criterio estipulado. Dentro de la comunidad Europea (Fontes, Silva, Silva, Barros, & Carvalho, 2014) desarrollaron una herramienta para identificar las fuentes de episodios de ozono y para este fin se utilizó un modelo de red MLP (se usó como un clasificador binario de la fuente de un episodio de ozono), se consideraron parámetros meteorológicos y los precursores de ozono.

El algoritmo de Propagación hacia atrás (BP) es uno de los métodos de entrenamiento más populares usado en redes MLP, una desventaja es que la convergencia es más lenta y los tiempos de entrenamiento son más largos. Según (Chien & Bin, 2002) el uso del método momentum ayuda a reducir el tiempo de entrenamiento y mejora la velocidad de convergencia. En Beijing se desarrolló un estudio para la predicción de la concentración de ozono utilizando una red neuronal con BP optimizada por un algoritmo genético (GA). Para este caso de estudio resultó suficiente usar solo parámetros meteorológicos para obtener buenos resultados (Feng, Zhang, Sun, & Zhang, 2011).

En el presente trabajo se realizó la estimación de perfiles de concentración de ozono desconocidos en las diferentes coordenadas de la zona metropolitana del Valle de México a partir de los datos conocidos, usando redes multicapa de perceptrones entrenadas con los datos del Año 2017 obtenidos de la SEDEMA.

## Capítulo 3 Marco Teórico

Para alcanzar los objetivos de este trabajo se propuso simular una red neuronal artificial multicapa de perceptrones (red MLP), para que, una vez entrenada, se pueda obtener el perfil de la concentración de ozono, para un día dado, en cualquier ubicación en la Zona Metropolitana del Valle de México (ZMVM).

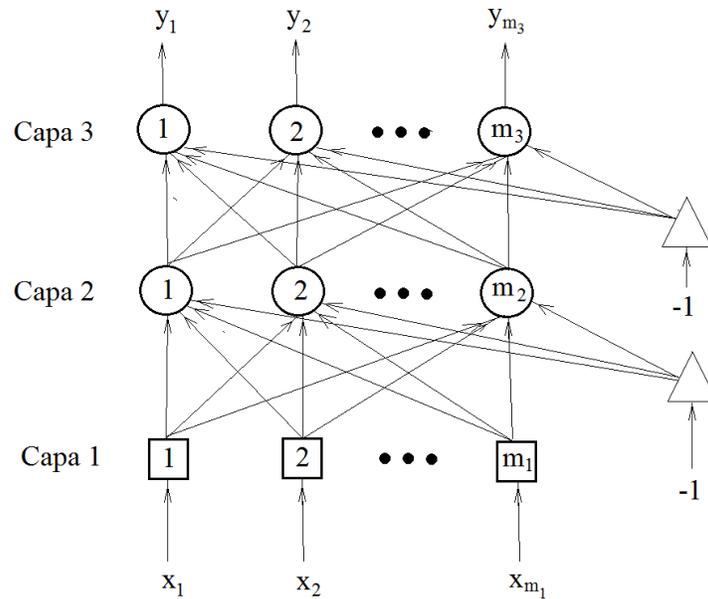
### 3.1. Metodología

El eje de operación de este trabajo está basado en la forma en la que trabajan las redes MLP. El programa que le sirve de núcleo es para entrenar y usar esas redes neuronales. Para describir las dos etapas principales de una red neuronal artificial, la etapa de entrenamiento y la de prueba, es necesario describir antes la arquitectura de las redes MLP.

### 3.2. Arquitectura de las redes MLP

Las redes MLP están constituidas por capas y éstas, a su vez por neuronas, como se muestra en la figura **3.1**.

Las neuronas de una capa están totalmente conectados con las neuronas de la siguiente y con los de ninguna otra capa más; excepto las neuronas de la capa de salida que no tienen una siguiente capa con la cual conectarse. Todas las neuronas de una capa se conectan únicamente con las neuronas de la siguiente capa.



**Figura 3.1.** Red multicapa de perceptrones. Se representa aquí una red de tres capas, es decir con una sola capa oculta. En esta figura las neuronas que trabajan como distribuidores de las entradas, los de la capa 1, se representan como cuadrados, las neuronas de polarización como triángulos y los perceptrones como círculos.

Las redes MLP se usan principalmente para dos tareas: para clasificar y para hacer regresión. Todas las redes que se usaron en este trabajo fueron para hacer regresión.

El número de neuronas de la capa de entrada, es igual al número de variables independientes, el número de neuronas de la de salida es igual al número de variables dependientes. Una red MLP puede tener una o más capas ocultas. El número de neuronas en las capas ocultas depende de la complejidad de las funciones de las que tiene que hacer regresión, entre más complejas sean, las capas ocultas deben tener más perceptrones.

Las neuronas son de tres tipos:

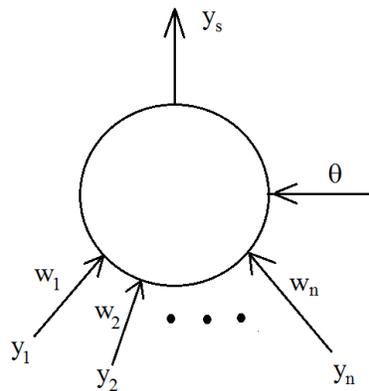
- 1) Distribuidores. Solo se encuentran en la capa de entrada o capa 1, su trabajo es simplemente distribuir la señal que pasa por ellos a todos las neuronas de la siguiente capa, la capa 2.

- 2) Perceptrones. Son los procesadores de la red, tienen varias entradas por las que reciben datos y de cada uno de ellos tiene una sola salida por la que entregan la información a las neuronas de la siguiente capa o al exterior de la red. Todas las capas de la red MLP están constituidas por ellos, excepto la capa de entrada, que está constituida por los distribuidores.
- 3) Neuronas de polarización. Estas neuronas tienen una sola entrada y varias salidas, cada una de ellas conectada a un perceptrón. Por su única entrada entra el valor constante -1, por sus salidas salen números de punto flotante que son los umbrales de los perceptrones.

### El perceptrón

Los principales componentes de una red neuronal artificial son los perceptrones, es un procesador que hace las tareas relativamente más sencillas, pero que, al interconectarse con otros, llega a realizar tareas más complejas. Como se muestra en la figura 3.2.

El perceptrón es un modelo de una neurona.



**Figura 3.2.** El perceptrón. Es el modelo funcional de una neurona.

La salida del perceptrón está determinada por una *función de activación*. Estas funciones tienen por lo general la forma de una “S” por eso se les llama *sigmoides*.

La función que se usó en este trabajo fue:

$$y_s = \frac{1}{e^{-\sigma} + 1} \quad (3.1)$$

Donde

$y_s$  es la salida del perceptrón,  
 $y_s \in (0, 1)$ , en (3.1).

$$\sigma = \sum_{i=1}^n w_i y_i - \theta \quad (3.2)$$

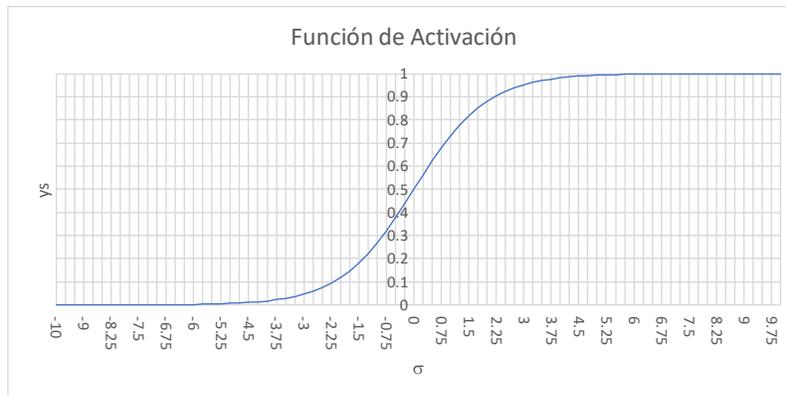
Donde

$w_i$  son los factores de eficiencia sináptica, o pesos de la red,  $w_i \in \mathbb{R}$

$y_i$  son las entradas al perceptrón,  $y_i \in (0, 1)$

$\theta$  es el umbral  $y_i \in \mathbb{R}$ .

En la figura 3.3 se muestra una gráfica de  $y_s$  contra  $\sigma$ .



**Figura 3.3** Función de activación de un perceptrón.

En la introducción se mencionó que:

- 1) antes de usar una red hay que entrenarla
- 2) para entrenarla hay que elaborar una tabla de entrenamiento y una tabla de prueba

- 3) para preservar la capacidad de generalización de una red hay que asignarle los valores de los pesos y los umbrales que hicieron que el error de prueba llegara al mínimo para usarla en la etapa de producción.

Analicemos más a detalle cada uno de estos puntos:

***1) Antes de usar una red hay que entrenarla***

Cuando se requiere que una red sea usada para hacer regresión, primero debe entrenarse para que aprenda las relaciones entre las variables dependientes e independientes. Para llevar a cabo esta tarea se aplican los pasos de propagación hacia adelante y propagación hacia atrás usando la tabla de entrenamiento, y propagación hacia adelante usando solo la tabla de prueba, para monitorear el avance del aprendizaje de las relaciones de las variables dependientes con las independientes y la capacidad de generalización de la red. El entrenamiento de la red se describe en la sección 3.3.

***2) Para entrenar una red MLP hay que elaborar una tabla de entrenamiento y una tabla de prueba***

De todos los datos de los que se dispone, ordenados en forma de ejemplos, esto es, valores de las variables independientes junto con los valores que les corresponden de las variables dependientes, se deben armar dos tablas: una tabla de entrenamiento y una tabla de prueba. Se recomienda que entre el 60 y el 90% de los ejemplos, se usen para construir la tabla de entrenamiento, y los restantes, es decir, entre el 10 el 40% para construir la tabla de prueba (Nunes, Hernane, Andrade, Bartocci, & Dos Reis, 2017).

***3) Para preservar la capacidad de generalización de una red hay que asignarle los valores de los pesos y los umbrales que hicieron que el error de prueba llegara al mínimo, para usarla en la etapa de producción.***

Durante el proceso de entrenamiento de una red, se van calculando dos errores: a) el error promedio por época de la tabla de entrenamiento y b) el error promedio por época de la tabla de prueba. A continuación, se define cada uno de ellos.

Error promedio por época de la tabla de entrenamiento:

$$\overline{err}_{\text{época entren}} = \frac{1}{R_e} \frac{1}{m_N} \frac{1}{2} \sum_{r_e}^{R_e} \sum_{i=1}^{m_N} (y_{i,r_e}^N - y_{i,r_e})^2 \quad (3.3)$$

Donde

$\overline{err}_{\text{época entren}}$  es el error promedio por época de la tabla de entrenamiento,

$y_{i,r_e}$  es el valor de la  $i$ -ésima variable dependiente del ejemplo o renglón  $r_e$  de la tabla de entrenamiento,

$y_{i,r_e}^N$  es el valor de la  $i$ -ésima variable dependiente obtenido por la red cuando se le alimenta el arreglo  $\mathbf{x}_{r_e}$  de variables independientes. El valor calculado para esta variable depende del valor de los pesos y de los umbrales que tiene la red en el momento que se le alimentó  $\mathbf{x}_{r_e}$ . En este trabajo, durante el entrenamiento, se aplicaba una corrección de los pesos y de los umbrales (propagación hacia atrás), inmediatamente después de calcular los valores de  $y_{i,r_e}^N$  para  $i = 1, \dots, m_n$ , (propagación hacia adelante),

$r_e$  subíndice que se refiere al  $r_e$ -ésimo ejemplo o renglón de la tabla de entrenamiento,

$R_e$  es el número de renglones de la tabla de entrenamiento,

$m_N$  es el número de neuronas en la capa de salida de la red,

Error promedio por época de la tabla de prueba:

$$\overline{err}_{\text{época prueba}} = \frac{1}{R_p} \frac{1}{m_N} \frac{1}{2} \sum_{r_p}^{R_p} \sum_{i=1}^{m_N} (y_{i,r_p}^N - y_{i,r_p})^2 \quad (3.4)$$

Donde

$\overline{err}_{\text{época prueba}}$  es el error promedio por época de la tabla de prueba,

$y_{i,r_p}$	es el valor de la $i$ -ésima variable dependiente del ejemplo o renglón $r_p$ de la tabla de prueba,
$y_{i,r_p}^N$	es el valor de la $i$ -ésima variable dependiente obtenido por la red cuando se le alimenta el arreglo $\mathbf{x}_{r_p}$ de variables independientes. A diferencia de lo que ocurre en la etapa de entrenamiento, los valores de los pesos y los umbrales de la red son los mismos para cada arreglo distinto $\mathbf{x}_{r_p}$ que se alimenta, ya que no se aplica ninguna corrección a los pesos y umbrales cuando se alimentan los ejemplos o renglones de la tabla de prueba,
$r_p$	subíndice que se refiere al $r_p$ -ésimo ejemplo o renglón de la tabla de entrenamiento,
$R_p$	es el número de renglones de la tabla de prueba,
$m_N$	es el número de neuronas en la capa de salida de la red,

Para entrenar una red se deben usar varias veces todos los ejemplos o renglones de la tabla de entrenamiento. Cada vez que se usan todos los ejemplos de la tabla de entrenamiento para entrenar la red se dice que se usó una *época*. Por extensión, cada vez que se usan todos los datos de la tabla de prueba, diremos aquí que se usó una *época de prueba*.

Cada vez que se aplica una época de entrenamiento se calcula  $\overline{err}_{época\ entren}$  e inmediatamente después se calcula  $\overline{err}_{época\ prueba}$ . Estos errores se van calculando a lo largo de las épocas. Una vez que se acaba la etapa de entrenamiento, los valores de los pesos y umbrales que se deben asignar a la red para ser usados durante su etapa de producción, deben ser los correspondientes al mínimo valor obtenido para el  $\overline{err}_{época\ prueba}$ . De esta forma se evita que la red se sobreentrene y pierda así su capacidad de generalización.

### 3.3. Entrenamiento de las redes MLP

Para entrenar una red hay que aplicar, de manera repetitiva, tres pasos:

Usando la tabla de entrenamiento (ver introducción y sección anterior):

- i. propagación hacia adelante y cálculo del error promedio de la época de entrenamiento.
- ii. propagación hacia atrás

Usando la tabla de prueba (ver introducción y sección anterior):

- iii. propagación hacia adelante y cálculo del error promedio de la época de prueba.

Mediante la propagación hacia adelante se obtiene, con los valores de sus parámetros  $\mathbf{w}$  y  $\boldsymbol{\theta}$  de ese momento, los valores de las variables dependientes correspondientes a los valores de las variables independientes que se le alimentaron. Una vez que se obtienen los valores de las variables dependientes calculados por la red, se puede calcular la diferencia entre los valores calculados por la red y los reportados en la tabla, con esa diferencia se puede calcular el error definido en la ecuación **(I.2)** (Gurney, 2003) y (Panda & Patra, 2013).

$$E = \frac{1}{2} \sum_{i=1}^n (y_{i,red} - y_i)^2 \quad \text{(I.2)}$$

#### 3.3.1. Propagación hacia adelante

Así se le llama al hecho de introducir los datos de un punto del dominio a la red neuronal, hacer que la red los procese y calcule las salidas.

### 3.3.2. Propagación hacia atrás

Este procedimiento consiste en ir minimizando la función error (**I.2**); el método de minimización usado es el llamado de *propagación hacia atrás* (*back propagation*). Este método se inicia suponiendo, por lo general de manera aleatoria, un punto inicial en el dominio de la función error. A continuación se calcula el vector gradiente en ese punto y se ubica otro punto sobre la dirección contraria a la señalada por ese vector, se repite el proceso y así se va avanzando en el dominio de la función error, siempre en dirección negativa del gradiente evaluado en el punto recién calculado, hacia *el* mínimo de la función; este algoritmo es la llamada *regla delta generalizada* (Haykin, 2009).

Aquí se ha supuesto que la función error tiene *un solo mínimo*, si no es así, si hay más de un mínimo, el método puede converger hacia un mínimo local. En este último caso ocurre que el error de entrenamiento sigue siendo alto<sup>1</sup> y el método ya no lo puede hacer disminuir porque el gradiente vale cero en el mínimo (o es muy cercano a cero si está en las vecindades de mínimo local). En otras palabras, la propagación hacia atrás no es un método de optimización global, y si la función no es unimodal, el mínimo al que converge depende, en gran medida, del punto inicial escogido donde comienza el proceso de optimización.

El procedimiento de propagación hacia atrás se aplica después de hacer la propagación hacia adelante, con este último procedimiento se calcula el error debido a la diferencia entre los valores obtenidos por la red y los valores esperados. La propagación hacia atrás usa esta información (el error que acaba de calcular la propagación hacia adelante), para actualizar los valores de todos los pesos y los umbrales de la red para tratar de disminuir el error. Lo hace en orden, capa por capa iniciando con la capa de salida y continuando con la capa que la antecede y así sucesivamente hasta terminar con la actualización de los pesos de la primera capa oculta.

---

<sup>1</sup> Aquí se está suponiendo que en las capas ocultas se ha fijado un número de **neuronas** lo suficientemente grande como para hacer que el error disminuya a mínimo global valores muy pequeños. Cuando el número de **neuronas** en esas capas no es lo suficiente grande, el de la función error puede tener un valor grande.

La actualización de los pesos y umbrales de una red MLP, usando el descenso por la dirección negativa del gradiente, se lleva a cabo por pasos, no de manera continua. Esto se puede expresar con las ecuaciones (3.5) y (3.6).

$$\mathbf{w}(t + 1) = \mathbf{w}(t) - \eta \text{grad}(E)_t \quad (3.5)$$

$$\boldsymbol{\theta}(t + 1) = \boldsymbol{\theta}(t) - \eta \text{grad}(E)_t \quad (3.6)$$

Las ecuaciones (3.5) y (3.6) describen el cambio de los vectores de los pesos y umbrales de la red del paso  $t$  al  $t+1$ , el cambio se hace siguiendo la dirección negativa del gradiente,  $\eta$  es el llamado factor de aprendizaje, que por lo general se fija en el intervalo (0, 1), (Nunes, Hernane, Andrade, Bartocci, & Dos Reis, 2017).

Las ecuaciones (3.5b) y (3.6b), son las versiones de las ecuaciones (3.5) y (3.6) por componentes de cada uno de los vectores  $\mathbf{w}$  y  $\boldsymbol{\theta}$ . En las ecuaciones que siguen se ha usado, con pequeñas modificaciones, la notación de Henseler (Henseler, 1995).

$$w_{i,j}^p(t + 1) = w_{i,j}^p(t) - \eta \left( \frac{\partial E}{\partial w_{i,j}^p} \right)_t \quad (3.5b)$$

$$\theta_i^p(t + 1) = \theta_i^p(t) - \eta \left( \frac{\partial E}{\partial \theta_i^p} \right)_t \quad (3.6b)$$

A los términos  $-\eta \frac{\partial E}{\partial w_{i,j}^p}$  y  $-\eta \frac{\partial E}{\partial \theta_i^p}$  en las ecuaciones (3.5b) y (3.6b) podríamos llamarlos  $\Delta w_{i,j}^p(t)$  y  $\Delta \theta_i^p(t)$  respectivamente, como se indica en las ecuaciones (3.7) y (3.8):

$$\Delta w_{i,j}^p(t) \equiv -\eta \left( \frac{\partial E}{\partial w_{i,j}^p} \right)_t \quad (3.7)$$

$$\Delta \theta_i^p(t) \equiv -\eta \left( \frac{\partial E}{\partial \theta_i^p} \right)_t \quad (3.8)$$

Entonces las actualizaciones de los pesos y umbrales se pueden expresar con las ecuaciones (3.9) y (3.10).

$$w_{i,j}^p(t+1) = w_{i,j}^p(t) + \Delta w_{i,j}^p(t) \quad (3.9)$$

$$\theta_i^p(t+1) = \theta_i^p(t) + \Delta \theta_i^p(t) \quad (3.10)$$

Donde

- $i$             neurona al que llega el arco
- $j$             neurona de la que sale el arco
- $p$             capa

Según Henseler (Henseler, 1995), los términos  $\Delta w_{i,j}^p(t)$  y  $\Delta \theta_i^p(t)$  se calculan según las ecuaciones (3.11), (3.12) y (3.13)

$$\Delta w_{i,j}^p(t) \equiv -\eta (\delta_i^p y_j^{p-1})_t \quad (3.11)$$

$$\Delta \theta_i^p(t) \equiv +\eta (\delta_i^p)_t \quad (3.12)$$

Donde

$$\delta_i^p = \begin{cases} y_i^p (1 - y_i^p) (y_i^p - y_i) & \text{para } p = N \\ y_i^p (1 - y_i^p) \sum_{j=1}^{m_{p+1}} w_{ji}^{p+1} \delta_j^{p+1} & \text{para } 1 < p < N \end{cases} \quad (3.13)$$

### 3.3.3. Uso del método del *momentum*

Una buena forma de evitar que el proceso de descenso en la dirección negativa del gradiente caiga en mínimos locales pequeños, es usar el método llamado del *momentum*. (Pedrycz, 1998). Este método consiste en tomar en cuenta no solo el último

valor del gradiente, sino *teóricamente* todos los valores del gradiente de cada uno de los puntos por los que el método de minimización ha pasado, para calcular el siguiente valor de los pesos y de los umbrales. Esto se logra calculando los nuevos valores de los pesos y de los umbrales usando las ecuaciones (3.14) y (3.15):

$$w_{i,j}^p(t+1) = w_{i,j}^p(t) + (\Delta w_{i,j}^p)_t^* \quad (3.14)$$

$$\theta_i^p(t+1) = \theta_i^p(t) + (\Delta \theta_i^p)_t^* \quad (3.15)$$

Donde

$$(\Delta w_{i,j}^p)_t^* = -\eta(\delta_i^p y_j^{p-1})_t + \alpha(\Delta w_{i,j}^p)_{t-1}^* \quad (3.16)$$

$$(\Delta w_{i,j}^p)_0^* = 0 \quad (3.17)$$

$$(\Delta \theta_i^p)_t^* = +\eta(\delta_i^p)_t + \alpha(\Delta \theta_i^p)_{t-1}^* \quad (3.18)$$

$$(\Delta \theta_i^p)_0^* = 0 \quad (3.19)$$

$\alpha$  es un parámetro en el intervalo [0, 1).

# Capítulo 4 Experimentación y Resultados

## 4.1. La base de datos original

La información sobre las concentraciones de ozono reportadas por la SEDEMA obtenidas por su red RAMA, vienen en forma de tablas. En la figura 4.1 se muestra una parte de la tabla del año 2017, que es el año cuyos datos se usaron para este trabajo.

	A	B	C	D	E	F	
1	FECHA	HORA	ACO	AJM	AJU	ATI	BJ
2	01/01/2017	1	-99	30	-99	8	
3	01/01/2017	2	-99	29	-99	5	
4	01/01/2017	3	-99	30	-99	2	
5	01/01/2017	4	4	31	6	2	
6	01/01/2017	5	5	30	5	2	
7	01/01/2017	6	5	31	5	6	
8	01/01/2017	7	5	31	5	4	
9	01/01/2017	8	7	35	5	4	
10	01/01/2017	9	8	31	7	6	
11	01/01/2017	10	18	33	25	20	
12	01/01/2017	11	37	50	43	40	
13	01/01/2017	12	54	65	50	50	
14	01/01/2017	13	67	75	59	58	
15	01/01/2017	14	74	82	57	69	
16	01/01/2017	15	76	82	54	66	
17	01/01/2017	16	73	71	58	69	
18	01/01/2017	17	71	62	57	69	
19	01/01/2017	18	47	54	50	58	
20	01/01/2017	19	41	50	42	39	
21	01/01/2017	20	40	48	13	30	
22	01/01/2017	21	35	45	10	14	
23	01/01/2017	22	22	44	11	5	
24	01/01/2017	23	15	42	15	2	
25	01/01/2017	24	3	39	13	2	
26	02/01/2017	1	4	35	15	1	
27	02/01/2017	2	4	35	12	1	
28	02/01/2017	3	3	41	10	5	
29	02/01/2017	4	3	40	9	4	
30	02/01/2017	5	2	33	12	7	
31	02/01/2017	6	3	33	11	2	
32	02/01/2017	7	5	29	5	1	

**Figura 4.1** Parte de la tabla del año de 2017 de la red RAMA de la SEDEMA con los datos de las concentraciones del ozono reportadas en ppb (partes por billón).

En las columnas A y B se muestra la fecha y la hora, respectivamente, de la medición tomada, se puede observar que se reporta una medición por hora. Y a partir de la columna C en adelante se muestra las mediciones de cada estación de monitoreo,

donde cada columna representa una estación de monitoreo. Para identificar cada una de las estaciones de monitoreo se utiliza solo tres siglas. Por ejemplo en la primera fila de la columna C se muestra el identificador de la primera estación automática de monitoreo que es ACO, que corresponde a la estación de monitoreo llamada Acolman. Entonces en la primera fila de las siguientes columnas se reportan, por orden alfabético, los identificadores de las demás estaciones: AJM (Ajusco Medio), AJU (Ajusco), ATI (Atizapán), en la figura solo se alcanza a ver parte del identificador de la siguiente estación de monitoreo: BJ... En la tabla 4.1 se muestra los identificadores y el nombre de la estación de monitoreo correspondiente.

En la columna D (figura 4.1), iniciando inmediatamente debajo del identificador de la estación AJM (Ajusco Medio), se muestran, en las siguientes 24 filas, los datos de las concentraciones de ozono del primer día del mes de enero. En la columna C, en las tres filas inmediatas abajo del identificador de la estación ACO (Acolman), se muestra en cada una de esas entradas el número -99; esto indica que la estación no reportó la medición de la concentración de ozono en la hora correspondiente a cada una de esas entradas de la tabla, en este caso la una, las dos y las tres de la madrugada del primero de enero de 2017. Entonces a los valores -99 y 0 se les reconoce como valores *nulos*, para este trabajo.

La tabla 4.1 contiene las mediciones para los 365 días del año de 2017, esto es  $24 \times 365 = 8,760$  mediciones por cada una de las estaciones. En la tabla 4.1 se muestran las estaciones automáticas de monitoreo ambiental que estuvieron operando en el año de 2017, se puede observar que operaron 34 estaciones. En el momento en el que se tomaron los datos de la ubicación de las estaciones, no estaban reportados los de dos de ellas: Coyoacán, COY, y San Juan de Aragón, SJA, por lo que no se tomaron en cuenta para este trabajo, o sea que se trabajó con los datos de las 32 estaciones restantes.

Para este trabajo, se transformaron las coordenadas de las estaciones, se pasaron de coordenadas esféricas a coordenadas en el plano cartesiano (ver apéndice A). La nueva representación de las coordenadas se calculó con respecto a la longitud de la estación más al oeste, INN (Instituto Nacional de Investigaciones Nucleares), y la latitud de la estación más al sur, AJU (Ajusco). De acuerdo a lo anterior se obtuvieron

las coordenadas de longitud = -99.38052 y latitud = 19.154674, este punto se tomó como origen del plano cartesiano. No se tomó en cuenta la altitud para este trabajo, porque se quería investigar si se podía obtener una buena aproximación de los perfiles de concentración de ozono únicamente considerando las coordenadas latitud y longitud. La nueva representación de las coordenadas se muestra en la tabla **4.2**. En esta tabla se omitieron las estaciones COY y SJA por la razón ya expuesta en esta sección.

**Tabla 4.1** Estaciones automáticas de monitoreo ambiental de la red RAMA de la SEDEMA que operaron en 2017 y su ubicación.

<b>Identificador</b>	<b>Nombre</b>	<b>Latitud (grados)</b>	<b>Longitud (grados)</b> -con respecto al meridiano de Greenwich-	<b>Altitud (msnm)</b>
ACO	Acolman	19.635501	-98.912003	2198
AJM	Ajusco Medio	19.2721	-99.207658	2619
AJU	Ajusco	19.154674	-99.162459	2953
ATI	Atizapán	19.576963	-99.254133	2341
BJU	Benito Juárez	19.371612	-99.158969	2250
CAM	Camarones	19.468404	-99.169794	2233
CCA	Centro de Ciencias de la Atmósfera	19.3262	-99.1761	2280
CHO	Chalco	19.266948	-98.886088	2253
COY	Coyoacán	ND	ND	ND
CUA	Cuajimalpa	19.365313	-99.291705	2704
CUT	Cuautitlán	19.722186	-99.198602	2263
FAC	FES Acatlán	19.482473	-99.243524	2299
GAM	Gustavo A. Madero	19.4827	-99.094517	2227
HGM	Hospital General de México	19.411617	-99.152207	2234
INN	Instituto Nacional de Investigaciones Nucleares	19.291968	-99.38052	3082
IZT	Iztacalco	19.384413	-99.117641	2238
LLA	Los Laureles	19.578792	-99.039644	2230
LPR	La Presa	19.534727	-99.11772	2302
MER	Merced	19.42461	-99.119594	2245
MGH	Miguel Hidalgo	19.40405	-99.202603	2366
MON	Montecillo	19.460415	-98.902853	2252
MPA	Milpa Alta	19.1769	-98.990189	2594
NEZ	Nezahualcóyotl	19.393734	-99.028212	2235
PED	Pedregal	19.325146	-99.204136	2326
SAG	San Agustín	19.532968	-99.030324	2241
SFE	Santa Fe	19.357357	-99.262865	2599
SJA	San Juan de Aragón	ND	ND	ND
TAH	Tláhuac	19.246459	-99.010564	2297
TLA	Tlalnepantla	19.529077	-99.204597	2311
TLI	Tultitlán	19.602542	-99.177173	2313
UAX	UAM Xochimilco	19.304441	-99.103629	2246
UIZ	UAM Iztapalapa	19.360794	-99.07388	2221
VIF	Villa de las Flores	19.658223	-99.09659	2242
XAL	Xalostoc	19.525995	-99.0824	2160

ND = no disponible cuando se realizó este trabajo

msnm = metros sobre el nivel del mar

**Tabla 4.2** Estaciones automáticas de monitoreo ambiental cuyos datos se usaron para este trabajo, con sus coordenadas expresadas en el plano cartesiano, en el punto: longitud = -99.38052, latitud = 19.154674, se ubica el origen del plano (**ver apéndice A**).

<b>Identificador</b>	<b>Nombre</b>	$x_1$ (km) longitud	$x_2$ (km) latitud
ACO	Acolman	52.0967	53.4655
AJM	Ajusco Medio	19.2214	13.0572
AJU	Ajusco	24.2473	0.0000
ATI	Atizapán	14.0536	46.9564
BJU	Benito Juárez	24.6353	24.1224
CAM	Camarones	23.4317	34.8852
CCA	Centro de Ciencias de la Atmósfera	22.7305	19.0728
CHO	Chalco	54.9783	12.4843
CUA	Cuajimalpa	9.8758	23.4220
CUT	Cuautitlán	20.2284	63.1045
FAC	FES Acatlán	15.2333	36.4496
GAM	Gustavo A. Madero	31.8021	36.4748
HGM	Hospital General de México	25.3872	28.5708
INN	Instituto Nacional de Investigaciones Nucleares	0.0000	15.2664
IZT	Iztacalco	29.2308	25.5458
LLA	Los Laureles	37.9037	47.1598
LPR	La Presa	29.2220	42.2600
MER	Merced	29.0136	30.0155
MGH	Miguel Hidalgo	19.7835	27.7293
MON	Montecillo	53.1141	33.9968
MPA	Milpa Alta	43.4028	2.4714
NEZ	Nezahualcóyotl	39.1749	26.5823
PED	Pedregal	19.6130	18.9556
SAG	San Agustín	38.9400	42.0644
SFE	Santa Fe	13.0826	22.5373
TAH	Tláhuac	41.1372	10.2060
TLA	Tlalnepantla	19.5617	41.6317
TLI	Tultitlán	22.6112	49.8006
UAX	UAM Xochimilco	30.7889	16.6533
UIZ	UAM Iztapalapa	34.0968	22.9195
VIF	Villa de las Flores	31.5716	55.9921
XAL	Xalostoc	33.1494	41.2890

**Tabla 4.3** Número de *días válidos* para cada estación. Se llama aquí *día válido* a aquel en el que la estación registró un perfil de concentraciones completo (24 medidas de concentración de ozono por día, una cada hora)

Identificador	Nombre	Número de días válidos
ACO	Acolman	193
AJM	Ajusco Medio	237
AJU	Ajusco	191
ATI	Atizapán	175
BJU	Benito Juárez	219
CAM	Camarones	181
CCA	Centro de Ciencias de la Atmósfera	201
CHO	Chalco	201
CUA	Cuajimalpa	245
CUT	Cuautitlán	221
FAC	FES Acatlán	152
GAM	Gustavo A. Madero	225
HGM	Hospital General de México	228
INN	Instituto Nacional de Investigaciones Nucleares	187
IZT	Iztacalco	203
LLA	Los Laureles	196
LPR	La Presa	197
MER	Merced	222
MGH	Miguel Hidalgo	183
MON	Montecillo	161
MPA	Milpa Alta	207
NEZ	Nezahualcóyotl	154
PED	Pedregal	209
SAG	San Agustín	194
SFE	Santa Fe	197
TAH	Tláhuac	187
TLA	Tlalnepantla	203
TLI	Tultitlán	176
UAX	UAM Xochimilco	213
UIZ	UAM Iztapalapa	194
VIF	Villa de las Flores	175
XAL	Xalostoc	238

## 4.2. Preparación de la base de datos

Se armó una nueva base de datos a partir de la base original. Para cada estación, se tomaron los días que no tenían ni un solo valor nulo (esto es, ni una hora en la que la estación se quedó sin registrar su medición de la concentración de ozono). En la tabla **4.3** se muestra para cada estación el número días con el perfil de concentraciones de ozono completo; es decir, con sus 24 mediciones de concentración (una por hora). En este trabajo a ese día se le denomina *día válido*.

De manera similar, llámese *estación válida* para un día determinado, a aquella estación que obtuvo un perfil completo de concentración de ozono en ese día.

Hubo un total de 279 días con 4 o más *estaciones válidas*, esto es el 76.44% de los 365 días del año 2017. En la tabla **4.4** se muestran el número de días relacionado con el número de *estaciones válidas*.

Usando esta tabla se puede observar que hubo 275 días en los que operaron al menos 15 *estaciones válidas*, de un total de 279 días, es decir el 98.57% de los días. El restante 1.43%, es decir 4 días, operaron 4, 11, 13 y 14 *estaciones válidas* en cada uno de ellos.

El promedio ponderado de estaciones válidas por día fue de 22.81 estaciones.

**Tabla 4.4** Relación del número de días en los que operaron los diferentes números de *estaciones válidas*.

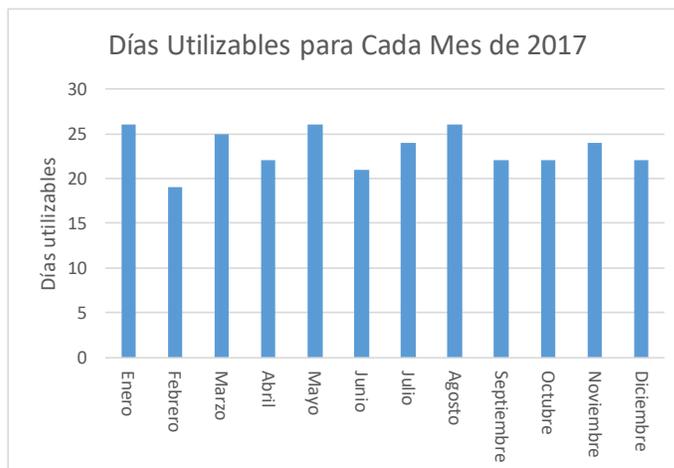
Número de <i>estaciones válidas</i>	Número de días con el número de <i>estaciones válidas</i> indicado en la columna 1	Porcentaje de días en los que operó el número de <i>estaciones válidas</i> de la columna 1
25	38	13.62
24	29	10.39
23	28	10.04
21	27	9.68
22	25	8.96
26	22	7.89
27	19	6.81
20	17	6.09
19	15	5.38
18	13	4.66
29	13	4.66
16	9	3.23
17	8	2.87
28	8	2.87
15	2	0.72
30	2	0.72
4	1	0.36
11	1	0.36
13	1	0.36
14	1	0.36
Total	279	100

#### 4.2.1. Distribución de los *días utilizables* a lo largo de 2017

Se llama en este trabajo *día utilizable* a aquel día con al menos 4 *estaciones válidas* (solo hubo un día con 4 *estaciones válidas*, además hubo tres días que tuvieron 11, 13 y 14 *estaciones válidas* respectivamente, y por otro lado los días restantes tuvieron 15 o más *estaciones válidas* -ver tabla 4.4). En la tabla 4.5 se muestra como quedaron repartidos estos 279 días a lo largo de los meses del año 2017.

**Tabla 4.5** Distribución de los *días utilizables* a lo largo del año de 2017. Se llama en este trabajo *día utilizable* a aquel día con al menos 4 *estaciones válidas*.

Mes	Número de días utilizables
Enero	26
Febrero	19
Marzo	25
Abril	22
Mayo	26
Junio	21
Julio	24
Agosto	26
Septiembre	22
Octubre	22
Noviembre	24
Diciembre	22
Total	279



**Gráfica 4.1** Distribución de días útiles en los diferentes meses de 2017.

### Conclusión de esta sección

Se puede concluir que:

1) A pesar de que se eliminaron todos los días que tenían algún valor nulo, quedó una buena cantidad de días para trabajar, 279 días, que son el 76.44% de los 365 días del año. Esto significa que se trabajó con una buena base de datos, con datos suficientes y completos.

2) Aunque el límite inferior de *estaciones válidas* fijado para seleccionar un *día utilizable* fue más bien bajo, 4 estaciones, solo uno de los *días utilizables* resultó tener 4 *estaciones válidas*, los tres inmediatos siguientes por su cantidad de estaciones, tuvieron 11, 13 y 14 *estaciones válidas* cada uno. Esto es, los días con menor número de

*estaciones válidas* tuvieron el 12.5%, 37.35%, 40.63% y 43.75% de las 32 estaciones que monitorearon la concentración de ozono a lo largo del año 2017 en la ZMVM. El promedio diario de *estaciones válidas* para los *días utilizables* fue de 22.81 estaciones, el 71.28% de esas 32 estaciones.

3) Los 279 días utilizables estuvieron bien repartidos en los 12 meses del año. El mes que tuvo menor número de días utilizables fue febrero con 19 días, los que mayor número de días utilizables fueron enero, mayo y agosto cada uno con 26; en parte se puede explicar la diferencia porque febrero tiene 28 días y los otros tres meses tienen 31 días cada uno. El promedio de días utilizables por mes fue de 23.25 días con una mediana de 23 días y una desviación estándar de 2.165 días.

### 4.3. Seleccionando la topología de las redes MLP

Una vez armada la base de datos con la que se trabajaría, se procedió a seleccionar la topología de entre un grupo de topologías que se propuso. Se usó una red MLP por cada día del año, se decidió usar la misma topología que se seleccionara para cada uno de esos días.

#### 4.3.1. Topologías propuestas para las redes MLP

El número de neuronas en la capa de entrada queda fijado por el número de variables independientes, en este caso son 2: las coordenadas en el plano cartesiano,  $x_1$  y  $x_2$ , es decir el punto de ubicación en la ZMVM. El número de neuronas de la capa de salida queda fijado por el número de variables dependientes, en este caso es el perfil de concentraciones diario de ozono, que consta de 24 mediciones, por eso el número de neuronas de la capa de salida es de 24. Se decidió que el número máximo de capas ocultas sea de 3, el motivo es porque considerar más de 3 capas ocultas posiblemente ya no se perciba un cambio considerable, además que se invierte más tiempo en cómputo.

El número de neuronas que queda por ser fijado es el de las capas ocultas. Se propusieron 12 topologías, mismas que se muestran en la tabla 4.6.

**Tabla 4.6** Doce topologías propuestas para las redes MLP

No. de topología Propuesta:	No. de capas	No. de neuronas en la Capa de Entada	No. de neuronas en la Capa de Salida	No. de neuronas en la Capa Oculta 1	No. de neuronas en la Capa Oculta 2	No. de neuronas en la Capa Oculta 3
1	3	2	24	5	0	0
2	3	2	24	10	0	0
3	3	2	24	15	0	0
4	3	2	24	20	0	0
5	4	2	24	3	2	0
6	4	2	24	5	5	0
7	4	2	24	8	7	0
8	4	2	24	10	10	0
9	5	2	24	2	2	1
10	5	2	24	4	3	3
11	5	2	24	5	5	5
12	5	2	24	7	7	6

Para proponer esta topología se pensó en cuatro números distintos de neuronas para las capas ocultas: 5, 10, 15 y 20 neuronas; luego se pensó en escoger combinaciones para ordenarlos en una, dos y tres capas ocultas, procurando que la diferencia entre el número de neuronas de cada capa oculta fuera mínima.

Una de las fases importantes para probar las topologías propuestas es la construcción de tablas de entrenamiento y de prueba para entrenar las redes con cada una de las topologías propuestas. A continuación, se describe cómo se elaboraron estas tablas.

#### 4.3.2. Construcción de las tablas de entrenamiento y de prueba

##### Número de estaciones de prueba

Se recomienda que entre el 10 y 40 % de los ejemplos se usen para construir la tabla de prueba según (Nunes, Hernane, Andrade, Bartocci, & Dos Reis, 2017). Sin

embargo considerando que hay solo un día que tiene 4 estaciones utilizables, para este caso se consideró adecuado que se tomará el 50% para estaciones de prueba y 50% para estaciones de entrenamiento, es decir, para este caso 2 estaciones de prueba y 2 estaciones de entrenamiento. Entonces para este trabajo se decidió establecer una regla para elegir el número de estaciones de prueba, la cual se describe a continuación.

Para cada día, se usará la siguiente regla para fijar el número de estaciones de prueba:

$$n_{ep} = \begin{cases} 5 & \text{para } n_{ev} > 10 \\ \left\lfloor \frac{n_{ev}}{2} \right\rfloor & \text{para } 4 \leq n_{ev} \leq 10 \end{cases} \quad (4.1)$$

Donde

$n_{ep}$  es el número de estaciones de prueba para un día dado,  
 $n_{ev}$  es el número de estaciones válidas para ese día.

El número de estaciones de entrenamiento es el número restante de las estaciones válidas para ese día:

$$n_{ee} = n_{ev} - n_{ep} \quad (4.2)$$

Donde

$n_{ee}$  es el número de estaciones de entrenamiento para el día en cuestión.

Para calcular el número de estaciones de prueba y de entrenamiento se usaron las ecuaciones (4.1) y (4.2). Los porcentajes de los números de estaciones de prueba y de entrenamiento para cada uno de los números de estaciones válidas por día, se muestran en la tabla 4.7.

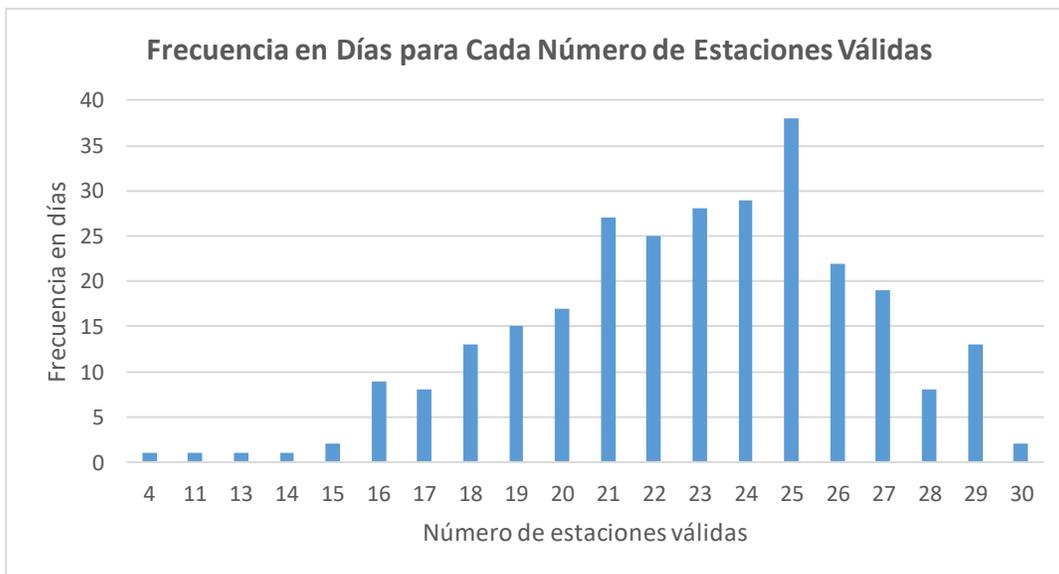
**Tabla 4.7** Porcentajes de estaciones de prueba y de entrenamiento para los días utilizables con diferentes números de estaciones válidas.

Número de estaciones válidas	Frecuencia en días, para cada número de estaciones válidas	Número de estaciones de prueba	Número de estaciones de entrenamiento	Porcentaje de estaciones de prueba	Porcentaje de estaciones de entrenamiento
25	38	5	20	20.00	80.00
24	29	5	19	20.83	79.17
23	28	5	18	21.74	78.26
21	27	5	16	23.81	76.19
22	25	5	17	22.73	77.27
26	22	5	21	19.23	80.77
27	19	5	22	18.52	81.48
20	17	5	15	25.00	75.00
19	15	5	14	26.32	73.68
18	13	5	13	27.78	72.22
29	13	5	24	17.24	82.76
16	9	5	11	31.25	68.75
17	8	5	12	29.41	70.59
28	8	5	23	17.86	82.14
15	2	5	10	33.33	66.67
30	2	5	25	16.67	83.33
4	1	2	2	50.00	50.00
11	1	5	6	45.45	54.55
13	1	5	8	38.46	61.54
14	1	5	9	35.71	64.29

Total 279

En la gráfica 4.2 se muestra la relación entre las columnas 1 y 2 de la tabla 4.7.

Como se puede observar en la tabla 4.7, solo hay dos casos en los que el número de estaciones de prueba está por debajo del 60%. Eso ocurrió en 2 de los 279 días utilizables: el caso del día con 4 estaciones válidas y el del día con 11 estaciones válidas. En el primero el porcentaje de estaciones de prueba es del 50% y en el segundo es del 54.55%. Por tanto, se puede afirmar que se siguió la recomendación de fijar entre un 60 y 90% de datos para entrenamiento y entre un 10 y 40% de datos para prueba en la gran mayoría de los casos (277 de 279, esto es el 99.28%).



**Gráfica 4.2** Frecuencia en días para cada número de estaciones válidas. Por ejemplo, la frecuencia para 25 estaciones válidas es de 38, esto significa que fueron 38 días utilizables que trabajaron con 25 estaciones válidas.

### **Selección de las estaciones de prueba y construcción de la tabla de prueba de un día**

Para un día dado, se escogieron al azar  $n_{ep}$  (ver ec. 4.1) estaciones de prueba, de entre todas las estaciones válidas para ese día.

Una vez escogidas las estaciones de prueba para un día dado, se procedió a construir la tabla de prueba del día. Por cada estación seleccionada, se armó un ejemplo o renglón de la tabla. La primera parte del renglón, la de las variables independientes, se integró con las coordenadas,  $x_1$  y  $x_2$  de la estación; la segunda parte del renglón, la de las variables dependientes,  $y_i$ ,  $i = 1, \dots, 24$ , con las 24 concentraciones de ozono que constituían el perfil de concentraciones del día escogido. Con todas las estaciones válidas seleccionadas al azar como estaciones de prueba para un día, se procedió a la construcción de la tabla de prueba de ese día.

### **Construcción de la tabla de entrenamiento del día**

Con el resto de las estaciones válidas del día, esto es, con las estaciones válidas que no fueron escogidas como estaciones de prueba, se armó la tabla de entrenamiento para ese día. El procedimiento para armar cada ejemplo o renglón de la tabla de entrenamiento fue el mismo que se usó para armar un ejemplo o renglón de la tabla de prueba.

### **Construcción de las tablas normalizadas de entrenamiento y de prueba de un día**

Cabe mencionar que para construir las tablas normalizadas de entrenamiento y de prueba, es necesario escalar o normalizar los valores de las variables. Para este trabajo se utilizará el término “normalizar”.

Como ya se ha mencionado en este trabajo, la función de activación para calcular la salida de cada uno de los perceptrones, toma valores en el intervalo  $[0, 1]$ . Por esta razón es necesario normalizar (escalar) los valores de las variables dependientes de las tablas de entrenamiento y de prueba, para que tomen valores en ese intervalo. También es conveniente normalizar los valores de las variables independientes. Para llevar a cabo esta tarea, se buscan los valores mayores en las estaciones válidas del día de: 1) la componente  $x_1$  de la ubicación, 2) la componente  $x_2$  de la ubicación y 3) el mayor valor de todas las concentraciones de ozono, obteniéndose así los valores de las variables  $x_{1,max}$ ,  $x_{2,max}$  y  $y_{max}$ . A continuación, se dividen los valores de las tablas de entrenamiento y de prueba entre los valores máximos correspondientes a cada variable.

Es con las tablas normalizadas de entrenamiento y de prueba para un día, se llevan a cabo el entrenamiento de la red MLP para ese día.

### **Cálculo del error relativo de ajuste de un perfil de concentraciones de ozono**

Para este trabajo, se consideró conveniente usar un error relativo ya que se puede expresar en porcentaje (%), además tomando en cuenta que las mediciones de los datos de concentración de ozono mostraban variaciones dependiendo del día, estación de monitoreo o debido a la estación del año. Por ejemplo hay estaciones de monitoreo que normalmente tienen mediciones muy altas de concentración de ozono, y hay otras estaciones de monitoreo que normalmente tienen mediciones bajas de concentración de ozono. Debido a estas variaciones en la medición es adecuado usar un error relativo.

Para calcular que tan bueno ha sido el ajuste de un perfil de concentraciones de ozono, se compara el perfil obtenido con la red con el perfil que se tiene como dato mediante la ecuación 4.3.

$$err\_rel = \frac{|\mathbf{y}^N - \mathbf{y}|}{0.5(|\mathbf{y}^N| + |\mathbf{y}|)} \quad \text{para } |\mathbf{y}^N| + |\mathbf{y}| > 0 \quad (4.3)$$

Donde

$err\_rel$  es el error relativo de ajuste del perfil de una estación,

$|\mathbf{a}|$  es la norma del vector  $\mathbf{a}$  de  $n$  componentes:

$$|\mathbf{a}| = \sqrt{a_1 + \dots + a_n} \quad (4.4)$$

$\mathbf{y}^N$  es el perfil de concentraciones obtenido por la red MLP,

$\mathbf{y}$  es el perfil normalizado de concentraciones, se obtiene de la tabla normalizada correspondiente.

### 4.3.3. Procedimiento para probar las doce topologías propuestas en esta tesis

Este procedimiento se divide en dos etapas:

1) cálculo de los errores relativos promedio de ajuste de todas las topologías para todos los días escogidos y todas las repeticiones fijadas,

2) cálculo de los promedios y desviaciones estándares de los errores relativos promedio calculados en la etapa (1), para seleccionar, con base en los promedios obtenidos y las desviaciones estándares la mejor topología.

**Procedimiento 4.1** Procedimiento para probar las doce topologías propuestas en este trabajo.

#### **Inicio**

De entre los 279 días utilizables escoger al azar una muestra de 15 días.

**para**  $i = 1$  **hasta** 15 /\*15 días de prueba\*/

Construir las tablas normalizadas de entrenamiento y de prueba para el día  $i$ -ésimo como se ha descrito en las secciones precedentes inmediatas

**para**  $j = 1$  **hasta** 12 /\*12 topologías para probarse\*/

**para**  $k = 1$  **hasta** 5 /\*5 repeticiones de la prueba\*/

Entrenar una red MLP con la  $j$ -ésima topología usando las tablas de entrenamiento y prueba normalizadas

Calcular el error de ajuste relativo promedio de las estaciones de prueba para el  $i$ -ésimo día con la  $j$ -ésima topología en  $k$ -ésima repetición:  $\overline{err\_rel}_{i,j,k}$

Calcular el promedio de los errores relativos promedio para cada topología y las desviaciones estándares de los errores relativos promedio de cada topología:

$\overline{err\_rel}_j, desv\_est\_erp_j$  para  $j=1, \dots, 12$

#### **Fin**

El número de redes que se entrenaron fue entonces de  $12 \times 15 \times 5 = 900$  redes, cada una por 50,000 épocas.

#### 4.3.4. Resultados de la aplicación del procedimiento para seleccionar la mejor topología

Se aplicó el procedimiento que se acaba de describir en el rubro anterior. Para el entrenamiento se usaron 50,000 épocas. El valor del factor de aprendizaje usado fue de:  $\eta = 0.005$ . Los valores iniciales de los pesos se obtuvieron al azar. Se almacenaron 11 listas de valores de los pesos obtenidos a lo largo del entrenamiento: para  $t = 0, 5,000, 10,000, \dots, 50,000$  épocas, de entre estos valores de las listas de los pesos, se escogió la lista de pesos que daba el menor valor del error (ec. **I.2**) de prueba para obtener el perfil de concentraciones de ozono,  $\mathbf{y}^N$ , y calcular con él el error relativo promedio de ajuste:

$$\overline{err\_rel}_{tipo\_est} = \frac{\sum_{i=1}^{n_{ep}} err\_rel_i}{n_{ep}} \quad (4.5)$$

Donde

$tipo\_est$	subíndice que indica el tipo de estación del que se calcula el promedio, $tipo\_est$ puede ser de entrenamiento o de prueba: $tipo\_est \in \{\text{entrenamiento, prueba}\}$ ,
$\overline{err\_rel}_{tipo\_est}$	es el promedio de los errores relativos de ajuste de las estaciones $tipo\_est$ para el día en cuestión,
$err\_rel_i$	es el error relativo de ajuste de la $i$ -ésima estación $tipo\_est$ (ver ec. <b>4.3</b> ).

También se calculó la desviación estándar de los errores relativos de ajuste:

$$desv\_est\_er_{tipo\_est} = \sqrt{\frac{\sum_{i=1}^{n_{e,tipo\_est}} (err\_rel_i - \overline{err\_rel})^2}{n_{e,tipo\_est}}} \quad (4.6)$$

Una vez que se calculan los errores relativos de ajuste promedio (ERA promedio) y la desviación estándar de los ERAs, se pueden calcular los promedios de

los ERAs promedio y los promedios de las desviaciones estándares de los ERAs, como se muestra en las ecuaciones (4.7) y (4.8).

$$\overline{\overline{err\_rel}}_{tipo\_est} = \frac{\sum_{i=1}^{n_{días}} \overline{err\_rel}_{tipo\_est,i}}{n_{días}} \quad (4.7)$$

Donde

$n_{días}$  es el número de días en los cuales se calcularon los promedios de los errores relativos de ajuste,

$\overline{err\_rel}_{tipo\_est,i}$  error relativo promedio de ajuste de las estaciones  $tipo\_est$  el  $i$ ésimo día,

$\overline{\overline{err\_rel}}_{tipo\_est}$  es el promedio de los errores relativos promedio de ajuste de los  $n_{días}$ , para las estaciones de  $tipo\_est$ .

$$\overline{\overline{desv\_est\_er}}_{tipo\_est} = \frac{\sum_{i=1}^{n_{días}} \overline{desv\_est\_er}_{tipo\_est,i}}{n_{días}} \quad (4.8)$$

Donde

$\overline{desv\_est\_er}_{tipo\_est,i}$  es la desviación estándar de los errores relativos de ajuste del  $i$ -ésimo día,

$\overline{\overline{desv\_est\_er}}_{tipo\_est}$  es el promedio de las desviaciones estándares de los errores relativos de los  $n_{días}$ .

Los resultados de las pruebas de las topologías se muestran en la tabla 4.8. En ella se puede observar que el valor del promedio del ERA promedio menor, es de 21.1660, correspondiente a la topología 7. A esa topología le corresponde el segundo valor más pequeño de la desviación estándar de los promedios de los ERAs, 7.2881.

**Tabla 4.8** Resultados de entrenar 12 topologías distintas (ver tabla 4.6), en 15 días, cada uno de ellos con las estaciones de prueba elegidas al azar y repitiendo 5 veces el entrenamiento (recuérdese que cada vez que se inicia una red los pesos iniciales se escogen al azar). Obsérvese que el mínimo  $\overline{\overline{err\_rel}}_{prueba}$  es el de la topología número 7, a esa misma topología le corresponde la segunda menor desviación estándar promedio.

No. de Topología (ver tabla 4.6)	$\overline{\overline{err\_rel}}_{prueba}$	$\overline{desv\_est\_er}_{prueba}$
1	21.4251	7.7634
2	21.7604	7.8546
3	21.7040	8.0123
4	21.8115	8.0469
5	22.0056	8.2275
6	21.3725	7.3842
<b>7</b>	<b>21.1660</b>	<b>7.2881</b>
8	21.3309	7.1653
9	24.1231	8.1546
10	22.7987	7.9513
11	22.4867	8.1477
12	22.3751	7.9294

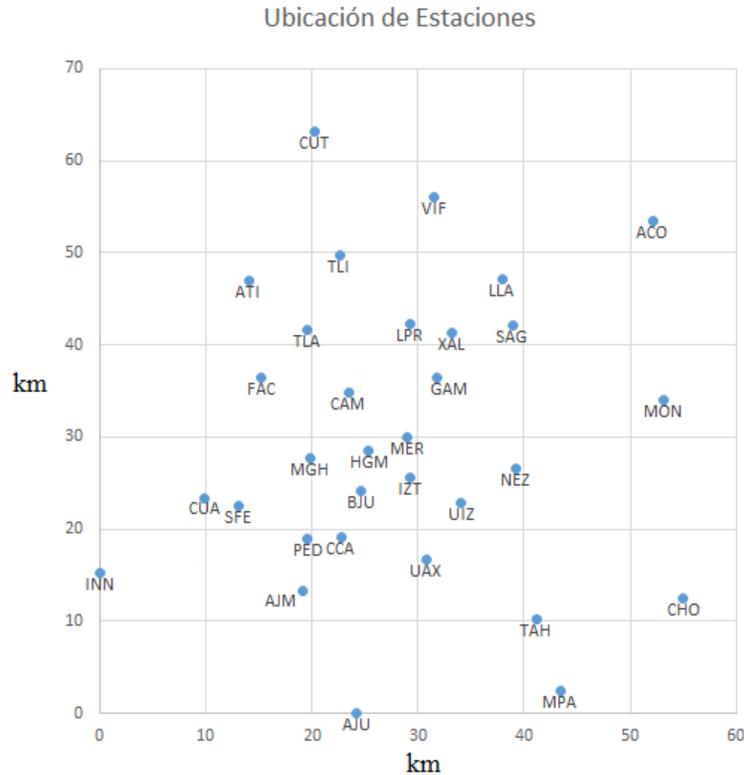
En la tabla 4.6, se puede ver que esa topología corresponde a una red de 4 capas con 2 neuronas en la capa de entrada, 8 en la primera capa oculta, 7 en la segunda capa oculta y 24 en la capa de salida. Esta fue la topología seleccionada.

#### 4.4. Obtención de perfiles diarios de concentración de ozono en la zona Metropolitana del Valle de México

##### Ubicación de las estaciones

Para cada uno los 279 días utilizables de 2017, se entrenó una red MLP. En la figura 4.2 se muestra la posición de todas las estaciones automáticas de monitoreo

ambiental que operaron en 2017, excepto la de Coyoacán (COY) y la de San Juan de Aragón (SJA) porque no están reportadas sus coordenadas en el portal de la SEDEMA.



**Figura 4.2** Ubicación de las estaciones automáticas de monitoreo (en coordenadas cartesianas -ver **apéndice A**), de la ZMVM que operaron en 2017 y cuya ubicación está reportada por la SEDEMA. En la tabla **4.2** se reportan los nombres de las estaciones.

En la figura **4.2** se muestran los identificadores de las estaciones; sus nombres y coordenadas se muestran en la tabla **4.2**.

#### 4.4.1. Tablas de entrenamiento y de prueba normalizadas

Para cada día utilizable se escogieron al azar 5 estaciones de prueba<sup>2</sup>, con esas estaciones se construyó una tabla de prueba para cada día. Las tablas de entrenamiento,

<sup>2</sup> Excepto para un solo día, el único que tuvo solo 4 estaciones válidas (ver tabla **4.7**), para ese día se seleccionaron al azar dos estaciones de prueba, las otras 2 se usaron para entrenamiento.

también una para cada día, se construyeron con los datos de las demás estaciones válidas del día en cuestión, que no fueron escogidas como estaciones de prueba.

Las tablas de entrenamiento y de prueba tienen la forma que se muestra en la tabla 4.9. En esa tabla, en cada ejemplo o renglón, en las dos primeras columnas van los valores de las variables independientes, estas son las coordenadas en el plano cartesiano de las estaciones; en las siguientes 24 columnas van los valores de las concentraciones de ozono a lo largo del día (una medida por cada hora del día), a las que hemos llamado *perfil diario de concentraciones de ozono*. Todos los valores de las tablas están normalizados, de tal manera que están en el intervalo  $[0, 1]$ , como ya se ha mencionado en este trabajo. La forma en la que se normalizó fue la siguiente: se obtenía el valor máximo para cada una de las variables tomando en cuenta las dos tablas, la de entrenamiento y la de prueba. Por ejemplo, para la coordenada  $x_1$ :

$$x_{1,max} = \max\{\mathbf{x}_{1,entren} \cup \mathbf{x}_{1,prueba}\} \quad (4.9)$$

Donde

$x_{1,max}$	valor máximo de los valores de la variable $x_1$ en las tablas de entrenamiento y de prueba,
$\mathbf{x}_{1,entren}$	vector que contiene todos los valores de la variable $x_1$ en la tabla de entrenamiento,
$\mathbf{x}_{1,prueba}$	vector que contiene todos los valores de la variable $x_1$ en la tabla de prueba.

Para obtener la tabla con los valores normalizados, se dividió cada valor de  $x_1$  de la tabla de entrenamiento y de prueba, entre  $x_{1,max}$ . De igual manera se procedió para todas las demás variables de las tablas para obtener los valores normalizados.

**Tabla 4.9** Forma de las tablas de prueba y de entrenamiento para cada día utilizable de 2017. En las primeras dos columnas van las coordenadas de la estación, en las siguientes 24, los perfiles diarios de las concentraciones de ozono.

$x_{1,1}$	$x_{2,1}$	$y_{1,1}$	$\dots$	$y_{24,1}$
$x_{1,2}$	$x_{2,2}$	$y_{1,2}$	$\dots$	$y_{24,2}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$x_{1,r}$	$x_{2,r}$	$y_{1,r}$	$\dots$	$y_{24,r}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$x_{1,R}$	$x_{2,R}$	$y_{1,R}$	$\dots$	$y_{24,R}$

#### 4.4.2. Entrenamiento y prueba de las redes MLP

Por cada uno de los 279 días utilizables de 2017, se entrenó una red MLP con la topología obtenida en la sección 4.3.4 y se armaron las tablas de entrenamiento y prueba como se describió en la sección anterior. Se usaron 50,000 épocas para entrenar cada red MLP, es decir, se alimentó ese número de veces la tabla de entrenamiento; por cada renglón alimentado de esa tabla se aplicó la propagación hacia atrás para tratar de disminuir el error de entrenamiento.

Se hicieron tres repeticiones de las corridas para los 279 días utilizables, en cada repetición se usaron estaciones de prueba al azar, de manera tal que, para un día dado en el calendario, el conjunto de estaciones de prueba era distinto de una repetición a otra, ya que era muy poco probable que se escogieran las mismas estaciones de prueba. A las *repeticiones* se les llamará de aquí en adelante *corridas*.

##### **Análisis de que tan correcto fue usar 50,000 épocas de entrenamiento**

Los pesos iniciales se fijaban al azar. Para cada época, se calculaba el error de entrenamiento y el de prueba. Se almacenaban 11 valores de los pesos de la red para los

siguientes números de épocas de entrenamiento:  $t = 0, 5000, 10000, 15000, 20000, 25000, 30000, 35000, 40000, 45000$  y  $50000$  épocas. Para cada red se escogía la lista de valores de pesos que hacía que el valor del error de prueba fuera el menor registrado. En la tabla **4.10** se relaciona el número de épocas de entrenamiento y el número de redes cuyo valor de prueba mínimo se obtuvo con ese número de épocas para cada una de las tres corridas.

**Tabla 4.10** Número de días en el que se obtuvo el error mínimo de prueba durante el entrenamiento para diferentes números de épocas

No. de épocas de entrenamiento	Corrida 1 Número de días	Corrida 2 Número de días	Corrida 3 Número de días	Número Promedio de días
5000	50	48	44	47.33
10000	39	36	42	39.00
15000	37	39	34	36.67
20000	35	33	29	32.33
25000	17	24	33	24.67
30000	19	17	18	18.00
35000	18	16	14	16.00
40000	14	19	11	14.67
45000	15	9	10	11.33
50000	35	38	44	39.00
Total:	279	279	279	279

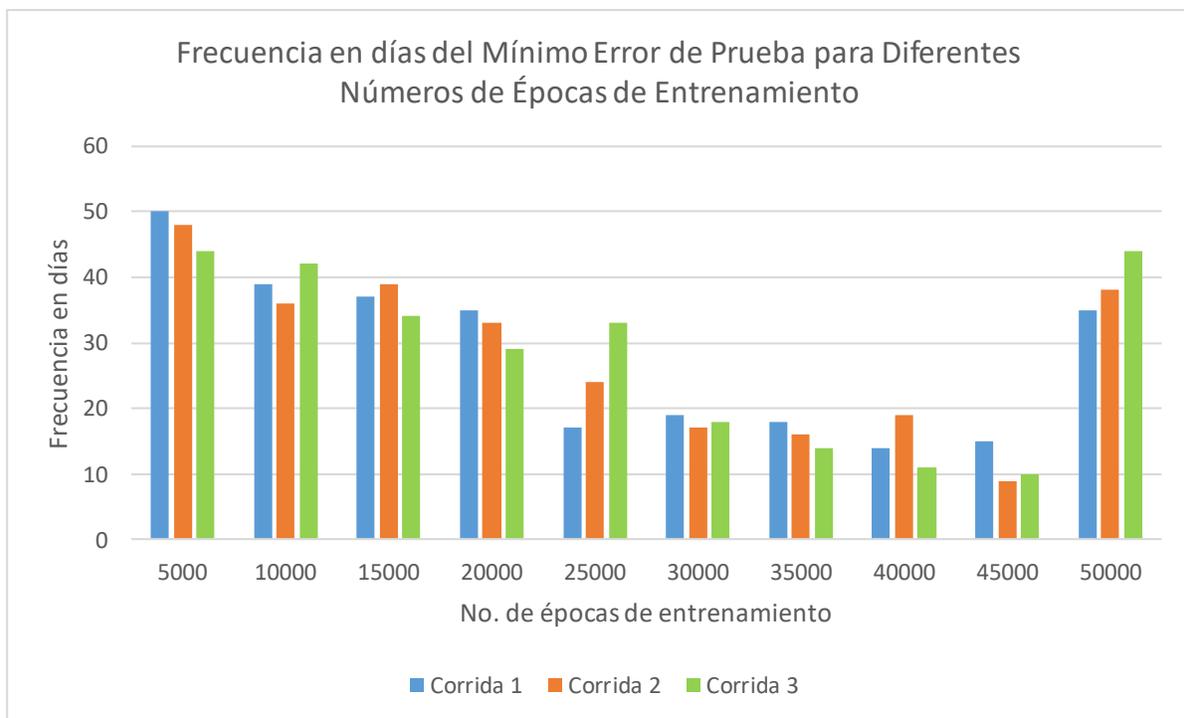
En la gráfica **4.3** se representan los datos de la tabla **4.10**

Para números de época de  $5,000$  a  $45,000$ , se puede asegurar que se llegó a un error mínimo de prueba porque, para los datos registrados, *antes y después* de valores en ese intervalo, el error era mayor; pero para valores de  $50,000$  épocas de entrenamiento, al decir *mínimo error de prueba*, se hace referencia a que es el menor error de los valores que se tienen registrados *antes* de él, *pero no después de él*, ya que no se continuó con el entrenamiento, y, por tanto, se carece de datos del error de prueba después de las  $50,000$  épocas de entrenamiento para poder afirmar que era un mínimo.

Cabe la posibilidad de que el error podría haber seguido bajando si se hubiera continuado con el entrenamiento. Esto sugiere que se debió usar un número mayor de

épocas de entrenamiento para decidir en cuál de ellas se alcanzaba un mínimo. Para responder a este cuestionamiento, se analiza que porcentaje de los casos cayó en esta situación: en promedio fueron 39 (ver tabla 4.10), esto da un porcentaje de 13.98%, o sea que en un 86.02% se puede afirmar que se llegó a un mínimo. Se ha considerado en este trabajo que este último es un porcentaje aceptable.

El subir el número de épocas de entrenamiento, por ejemplo, a unas 60,000, hubiera aumentado el tiempo de cómputo en un 20% y no hubiera garantizado que no se obtuvieran algunos casos en los que se tuviera la misma situación descrita. Fue por esta razón, por un compromiso entre la precisión y el tiempo disponible para hacer las corridas, que se decidió mantener como máximo 50,000 épocas de entrenamiento.



**Gráfica 4.3** Frecuencia en días del mínimo error de prueba para diferentes números de épocas de entrenamiento y para cada una de las tres corridas. Por ejemplo, para la corrida 1, la frecuencia para un número de 5000 épocas de entrenamiento es de 50, es decir, hubo 50 días, de los 279 utilizables, en los que el mínimo del error de prueba se obtuvo a las 5000 épocas de entrenamiento.

## 4.5. Resultados de los ajustes de los perfiles diarios de las concentraciones de ozono para las estaciones de prueba

Los resultados de las tres corridas se agruparon para cada una de las estaciones de prueba. A continuación, se muestran las tablas concentradoras de los resultados para cada corrida.

**Tabla 4.11. Corrida 1.** Tabla concentradora de los datos de las estaciones de prueba.

Identificador de la estación	Frecuencia	Error relativo promedio	Desviación estándar de los errores relativos promedio	Error relativo mínimo	Error relativo máximo	Mediana del error relativo promedio
SAG	38	14.12	4.16	4.67	23.74	12.54
IZT	58	14.63	4.59	5.88	26.47	13.82
HGM	47	14.66	7.00	6.04	39.43	12.65
VIF	35	15.31	4.78	7.27	27.07	14.84
LLA	33	15.37	4.37	8.03	26.31	15.05
TLI	42	15.40	5.89	6.34	42.34	14.18
BJU	41	15.68	5.05	5.64	30.14	14.64
CAM	41	16.08	5.69	5.96	37.36	15.83
UAX	49	16.70	5.91	7.98	38.01	14.38
UIZ	46	16.76	6.46	8.13	32.87	14.63
CCA	53	16.89	6.33	6.27	40.29	16.84
XAL	53	16.99	6.57	6.78	34.80	15.22
PED	40	18.35	8.13	8.79	53.76	15.99
MGH	49	18.56	6.14	8.82	40.40	16.88
TLA	45	18.86	5.98	7.71	33.24	18.33
TAH	35	19.08	5.49	8.59	32.91	17.37
GAM	52	19.81	5.25	8.18	36.20	19.84
FAC	43	20.02	6.58	10.56	41.92	19.17
LPR	40	21.69	7.89	7.88	41.26	19.92
MER	56	22.06	6.75	11.33	41.76	21.66
CUT	43	22.15	5.63	12.59	40.91	20.91
ATI	40	22.36	8.41	7.86	41.56	20.23
CHO	43	22.86	6.82	11.09	37.29	22.99
MON	29	23.25	10.50	9.66	60.67	20.78
ACO	41	24.14	17.95	12.10	102.03	21.28
AJM	54	24.86	9.41	10.62	58.73	23.16
NEZ	35	24.91	7.62	14.47	44.85	22.92
AJU	34	25.05	8.53	11.81	55.73	22.92
MPA	51	27.44	9.33	11.03	47.56	27.43
SFE	42	29.32	10.81	15.29	65.13	25.68
CUA	42	30.18	12.22	9.80	68.85	27.50
INN	42	30.57	11.83	11.11	56.21	27.74

**Tabla 4.12. Corrida 2.** Tabla concentradora de los datos de las estaciones de prueba.

Identificador de la estación	Frecuencia	Error relativo promedio	Desviación estándar de los errores relativos promedio	Error relativo mínimo	Error relativo máximo	Mediana del error relativo promedio
IZT	40	14.66	5.22	7.01	29.07	13.69
UAX	47	15.19	6.87	5.15	41.47	14.04
HGM	43	15.28	5.03	6.84	29.34	14.65
SAG	36	15.82	5.54	8.25	40.99	15.06
XAL	55	15.93	5.17	5.95	33.03	14.40
LLA	45	16.06	5.95	7.16	30.84	14.63
CAM	49	16.22	4.77	9.38	28.64	14.98
BJU	53	16.32	5.92	8.09	34.25	14.83
CCA	44	16.61	5.56	7.18	30.32	16.69
MGH	29	16.65	5.24	7.46	29.08	15.38
TLA	40	16.90	5.15	7.54	31.63	16.14
UIZ	38	17.49	6.71	7.26	33.14	18.12
TAH	49	17.56	6.60	8.75	38.11	15.90
TLI	32	17.73	6.87	8.13	33.32	16.38
PED	38	17.88	5.68	7.15	33.25	17.71
VIF	36	18.17	5.38	9.40	28.12	17.03
GAM	42	20.43	7.31	12.53	50.20	18.78
FAC	39	20.84	5.55	11.44	38.09	21.52
LPR	52	21.27	7.91	10.31	45.30	19.48
MER	56	21.67	7.49	9.54	51.88	20.72
NEZ	22	23.02	6.34	15.40	40.30	21.80
CUT	48	23.11	7.94	12.30	53.27	21.12
ATI	35	24.42	9.97	9.21	45.13	25.92
ACO	49	24.80	15.21	10.58	103.00	22.15
CHO	43	24.97	8.53	13.04	53.21	23.20
MON	35	25.37	10.34	8.71	48.92	23.72
AJM	59	26.25	9.56	6.60	48.59	26.98
SFE	40	26.87	7.94	14.03	50.38	25.77
MPA	40	29.32	12.28	10.32	71.20	28.12
AJU	41	29.77	16.24	11.55	76.02	22.28
CUA	68	29.84	13.06	9.93	70.03	26.77
INN	49	31.01	12.80	13.68	71.09	28.06

**Tabla 4.13. Corrida 3.** Tabla concentradora de los datos de las estaciones de prueba.

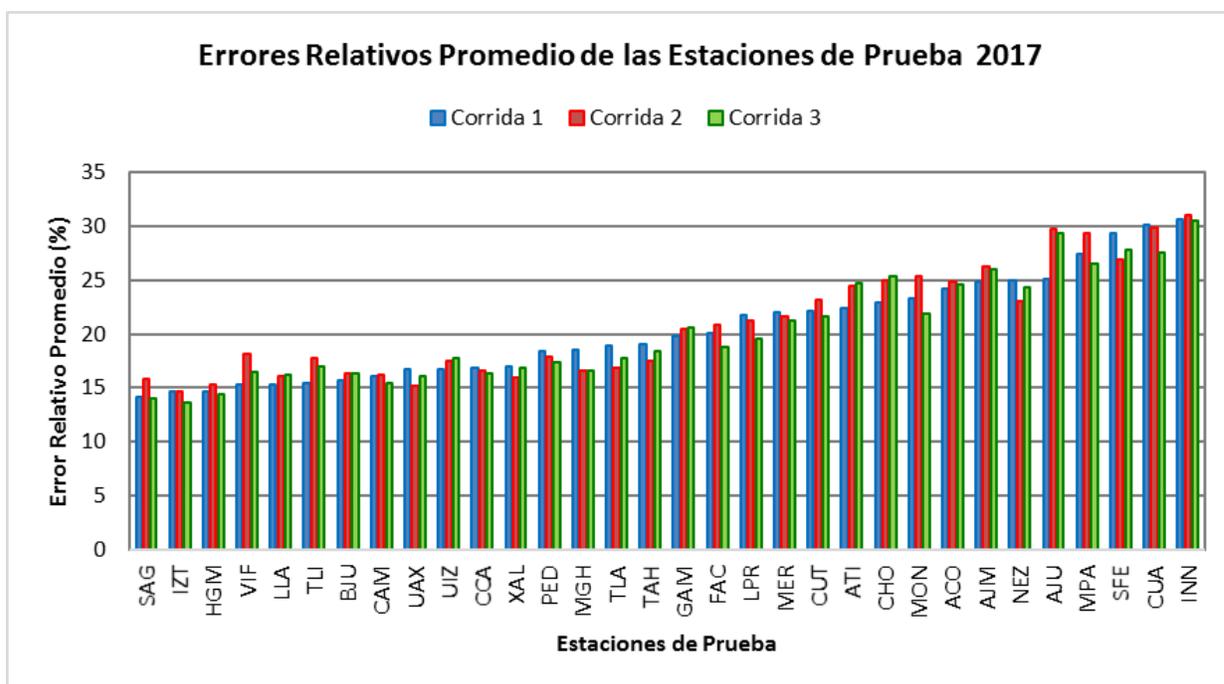
Identificador de la estación	Frecuencia	Error relativo promedio	Desviación estándar de los errores relativos promedio	Error relativo mínimo	Error relativo máximo	Mediana del error relativo promedio
IZT	52	13.68	4.40	7.04	28.43	13.60
SAG	41	13.98	3.20	7.85	19.68	14.45
HGM	58	14.46	5.35	7.25	34.19	13.36
CAM	40	15.50	5.34	8.56	30.88	14.70
UAX	63	16.03	5.12	7.98	31.84	15.11
LLA	32	16.25	4.93	9.52	25.71	16.24
BJU	45	16.39	5.49	7.39	31.91	15.96
CCA	47	16.41	5.89	7.63	32.34	14.69
VIF	32	16.44	6.15	9.09	40.18	15.57
MGH	46	16.60	5.36	9.63	35.12	15.59
XAL	44	16.86	5.39	8.77	32.74	16.62
TLI	40	17.04	5.11	6.91	31.17	17.16
PED	44	17.34	7.21	7.03	37.16	14.87
TLA	42	17.77	6.69	6.55	40.95	16.58
UIZ	43	17.81	6.50	8.96	37.49	17.10
TAH	42	18.35	7.07	8.63	44.41	16.69
FAC	30	18.79	7.42	7.83	40.70	16.95
LPR	39	19.52	6.37	9.79	39.28	18.87
GAM	45	20.63	6.09	10.24	43.83	20.80
MER	43	21.20	7.79	10.18	45.95	19.67
CUT	45	21.61	6.74	10.09	43.28	21.25
MON	34	21.93	6.70	10.04	38.49	21.59
NEZ	32	24.28	7.32	12.42	39.75	22.07
ACO	48	24.59	16.06	9.57	100.15	21.38
ATI	31	24.67	9.13	9.88	43.22	24.10
CHO	45	25.39	9.08	12.29	46.58	24.48
AJM	53	26.01	9.36	7.56	55.87	24.88
MPA	53	26.50	9.03	9.43	54.74	26.41
CUA	47	27.54	13.13	8.39	60.38	25.43
SFE	49	27.78	11.14	13.18	51.95	25.06
AJU	40	29.30	12.12	14.91	70.15	24.83
INN	47	30.47	11.23	14.74	74.23	28.50

En las tablas 4.11, 4.12 y 4.13, los datos de las estaciones están ordenados según su Error Relativo de Ajuste (ERA) promedio. La estación que resultó con el menor ERA promedio va primero. Un ejemplo de cómo se deben interpretar estas tablas es el

siguiente: en la tabla 4.11 se observa que la estación IZT fue seleccionada 40 veces en la segunda corrida; es decir, fue seleccionada 40 días como estación de prueba junto con otras estaciones. De esas 40 veces, su ERA mínimo fue de 7.01, su ERA máximo de 29.7, su ERA promedio de 14.66, la desviación estándar de sus ERA fue de 5.22 y la mediana de sus ERA 13.69.

Todos estos datos, junto con los demás datos de la tabla, permiten darse una idea de cómo resultó el ajuste de la estación: tuvo un ERA mínimo bueno, el tercero mejor de la segunda corrida; sus ERA promedio y máximo fueron los mejores de la segunda corrida; la desviación estándar de sus ERA fue la tercera mejor de esa corrida y a partir de la mediana de sus ERA y las ERA mínima y máxima, se puede ver que para la mitad de las veces que fue seleccionada, esto es 20 veces, sus ERA fueron iguales o mayores a 7.01 y menores a 13.69 y la otra mitad fueron mayores a 13,69 y menores o iguales a 29.07 en la corrida 2.

En la gráfica 4.4 se grafican los ERA promedio para las tres corridas.



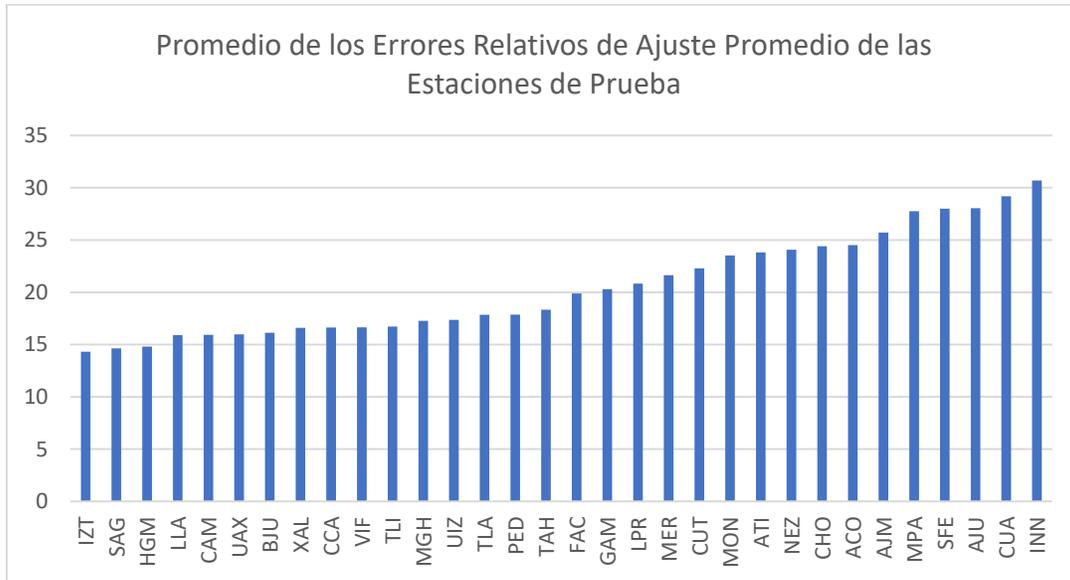
**Gráfica 4.4** Errores relativos de ajuste promedio de cada una de las estaciones de prueba para las tres corridas.

En la tabla 4.14 se obtienen los promedios de los ERA promedio,  $\overline{\overline{\text{ERA}}}$ , para cada una de las estaciones de prueba.

**Tabla 4.14** Promedio de las tres corridas de los errores relativos de ajuste promedio,  $\overline{\overline{\text{ERA}}}$ .

Identificador de la estación	Corrida 1	Corrida 2	Corrida 3	Promedio
SAG	14.12	15.82	13.98	14.64
IZT	14.63	14.66	13.68	14.32
HGM	14.66	15.28	14.46	14.80
VIF	15.31	18.17	16.44	16.64
LLA	15.37	16.06	16.25	15.89
TLI	15.40	17.73	17.04	16.72
BJU	15.68	16.32	16.39	16.13
CAM	16.08	16.22	15.50	15.93
UAX	16.70	15.19	16.03	15.98
UIZ	16.76	17.49	17.81	17.35
CCA	16.89	16.61	16.41	16.64
XAL	16.99	15.93	16.86	16.59
PED	18.35	17.88	17.34	17.86
MGH	18.56	16.65	16.60	17.27
TLA	18.86	16.90	17.77	17.84
TAH	19.08	17.56	18.35	18.33
GAM	19.81	20.43	20.63	20.29
FAC	20.02	20.84	18.79	19.89
LPR	21.69	21.27	19.52	20.83
MER	22.06	21.67	21.20	21.64
CUT	22.15	23.11	21.61	22.29
ATI	22.36	24.42	24.67	23.82
CHO	22.86	24.97	25.39	24.41
MON	23.25	25.37	21.93	23.52
ACO	24.14	24.80	24.59	24.51
AJM	24.86	26.25	26.01	25.71
NEZ	24.91	23.02	24.28	24.07
AJU	25.05	29.77	29.30	28.04
MPA	27.44	29.32	26.50	27.75
SFE	29.32	26.87	27.78	27.99
CUA	30.18	29.84	27.54	29.18
INN	30.57	31.01	30.47	30.68

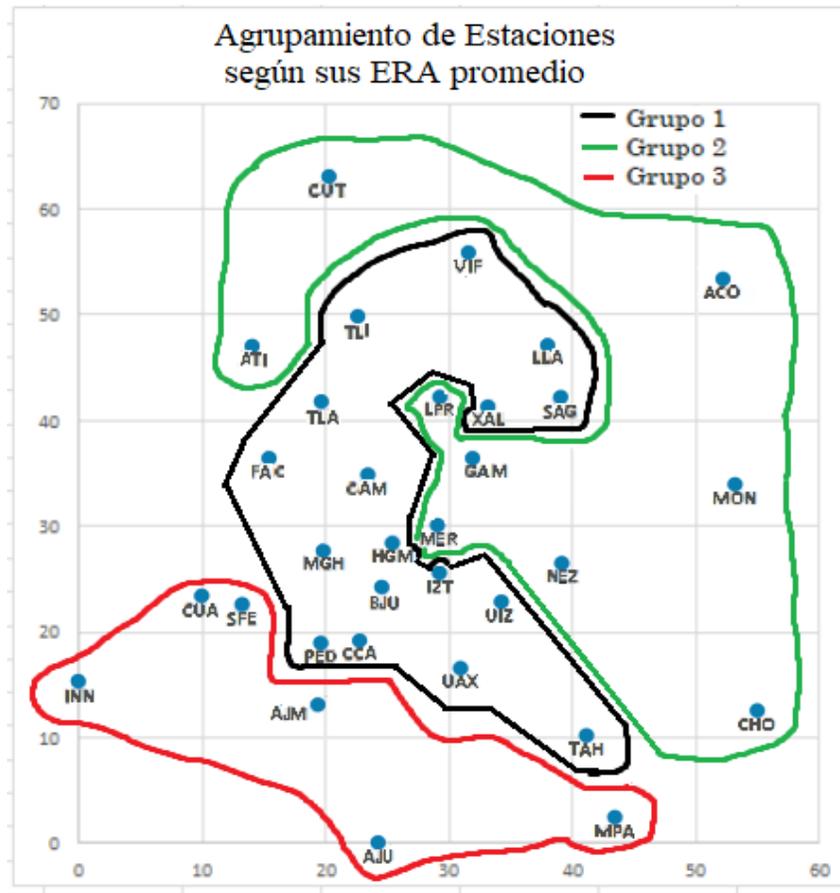
En la gráfica 4.5 se grafican los  $\overline{\overline{\text{ERA}}}$  promediados de las tres corridas (columna 5 de la tabla 4.14).



**Gráfica 4.5.** Promedio de las tres corridas de los errores relativos de ajuste promedio de cada una de las estaciones de prueba, ordenados del menor al mayor.

Dividamos las estaciones de monitoreo en tres grupos según sus errores relativos de ajuste, ERA, cuando fueron escogidas como estaciones de prueba. En la tabla 4.15 se muestra una división en tres grupos.

En la figura 4.3 se muestran las estaciones en el plano cartesiano agrupadas según su  $\overline{\overline{\text{ERA}}}$ , de la tabla 4.15.



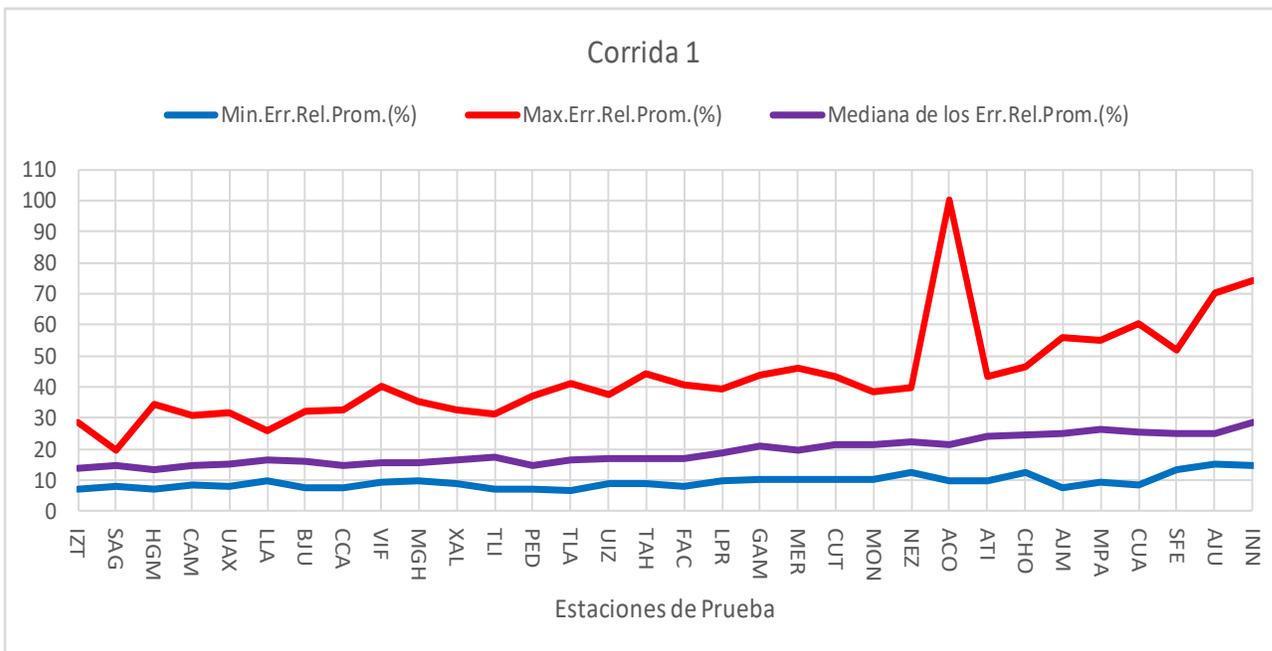
**Figura 4.3** Estaciones agrupadas según sus  $\overline{ERA}$ ; en el Grupo 1 (negro) están las estaciones con  $14.64\% \leq \overline{ERA} < 20\%$ , en el Grupo 2 (verde) están las estaciones con  $20\% < \overline{ERA} < 25\%$ , y en el Grupo 3 (rojo) están las estaciones con  $25\% < \overline{ERA} < 31$

**Tabla 4.15** Estaciones de monitoreo agrupadas según su  $\overline{\overline{\text{ERA}}}$ .

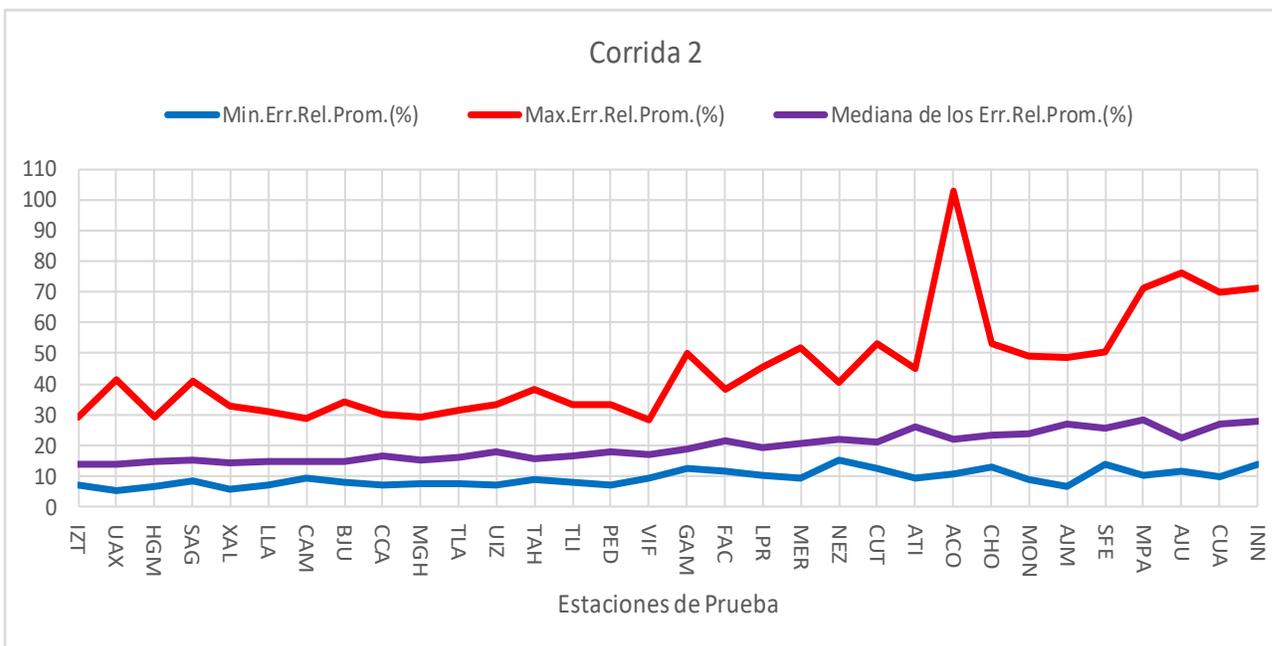
Grupo	Condición	Estaciones (ordenadas según sus, $\overline{\overline{\text{ERA}}}$ la de menor $\overline{\overline{\text{ERA}}}$ va primero)	Observaciones
1	Estaciones con: $14.64\% \leq \overline{\overline{\text{ERA}}} < 20\%$	IZT, SAG, HGM, LLA, CAM, UAX, BJU, XAL,CCA,VIF, TLI, MGH, UIZ, TLA, PED, TAH, FAC	Estaciones con menor error relativo promedio
2	Estaciones con: $20\% < \overline{\overline{\text{ERA}}} < 25\%$	GAM, LPR, MER, CUT, MON, ATI, NEZ, CHO, ACO	Caso Especial: ACO ACO tiene un error muy grande sólo en diciembre
3	Estaciones con: $25\% < \overline{\overline{\text{ERA}}} < 31\%$	AJM, MPA, SFE, AJU, CUA, INN	Tiene mayor $\overline{\overline{\text{ERA}}}$ y existe la posibilidad que tenga relación con la <b>Altura</b> de dichas estaciones, la Altura no fue considerada en este trabajo.

#### 4.5.1. Análisis de los Errores Relativos de Ajuste (ERA) mínimos, ERA máximos y la mediana de los ERA de las tres corridas

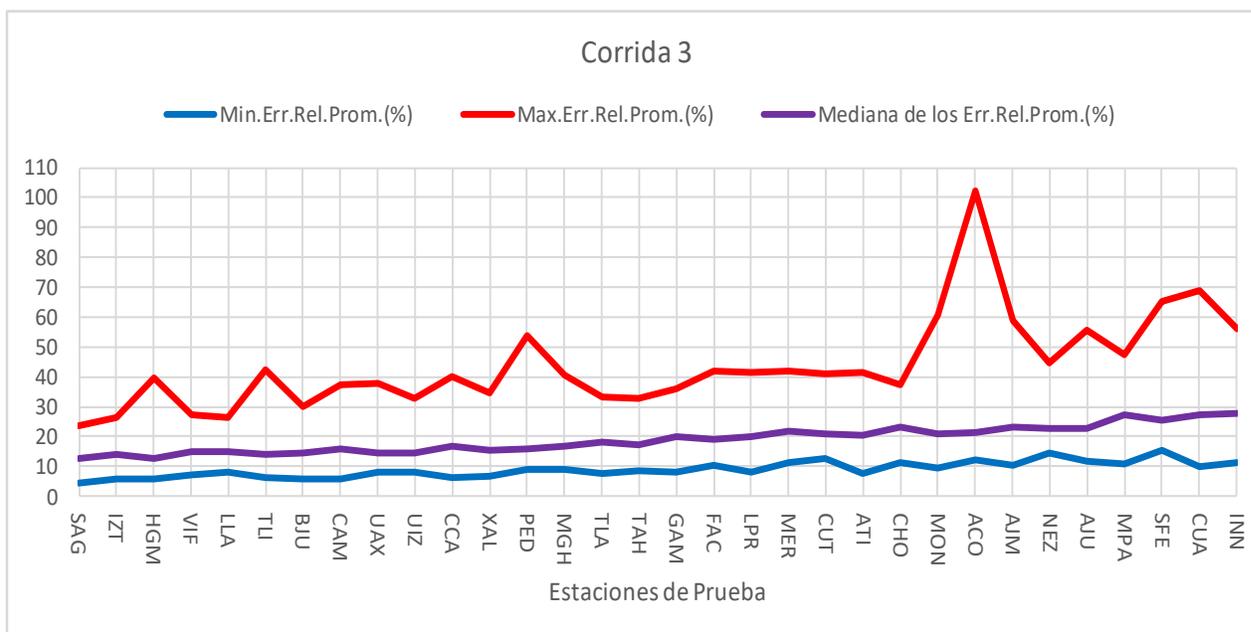
En las gráficas **4.6**, **4.7** y **4.8**, se muestran los valores de los ERA mínimos, máximos y de la mediana de los era para cada una de las estaciones.



**Gráfica 4.6.** Corrida 1. Valores de los ERA mínimos, máximos y la mediana de los ERA de las estaciones de monitoreo para el año 2017.



**Gráfica 4.7.** Corrida 2. Valores de los ERA mínimos, máximos y la mediana de los ERA de las estaciones de monitoreo para el año 2017.



**Gráfica 4.8.** Corrida 3. Valores de los ERA mínimos, máximos y la mediana de los ERA de las estaciones de monitoreo para el año 2017.

En estas gráficas podemos observar lo siguiente:

- 1) Los ERA mínimos de todas las estaciones en todas las corridas son menores al 20%.
- 2) Los ERA máximos sobrepasan el 20% para todas las estaciones con excepción de un solo caso: SAG, San Agustín, en la corrida 1, en ese caso el ERA es de 19.68%, muy cercana a 20%.
- 3) Las medianas de los ERA para todas las estaciones en todas las corridas, inicia con valores superiores a 10 y menores a 15%.
- 4) Las medianas de los ERA para todas las estaciones en todas las corridas, es siempre menor a 30%.
- 5) Para la mayoría de las estaciones el ERA máximo fue mayor a 30%.
- 6) Los ERA máximos presentan picos mucho mayores que la mediana.
- 7) En las tres corridas, para los ERA máximos, se destaca un pico que sobresale sobre todos los demás, ese pico corresponde a la estación ACO, Acolman.

Por lo visto en los anteriores puntos se puede inferir lo siguiente:

1) Como consecuencia del punto (1) se puede afirmar que por lo menos hubo un caso (esto es, un día) para cada una de las estaciones, tratadas como estación de prueba, en las que el ERA fue menor de 20%. Esto se cumplió para las tres corridas.

2) Como consecuencia del punto (4), para cada estación, al menos la mitad de todas las mediciones del ERA fueron menores de 30%. Esto es cierto para las tres corridas.

3) Como consecuencia del punto (5) se puede afirmar que hubo muy pocas estaciones, tratadas como estaciones de prueba, en las que *todos* sus ERA registrados fueron menores de 30% para las tres corridas.

4) Se presentó algún factor o combinación de factores que hicieron que la estación ACO, Acolman, presentara valores máximos del ERA mayores que los de las demás estaciones.

En la siguiente subsección se hablará más de este tema.

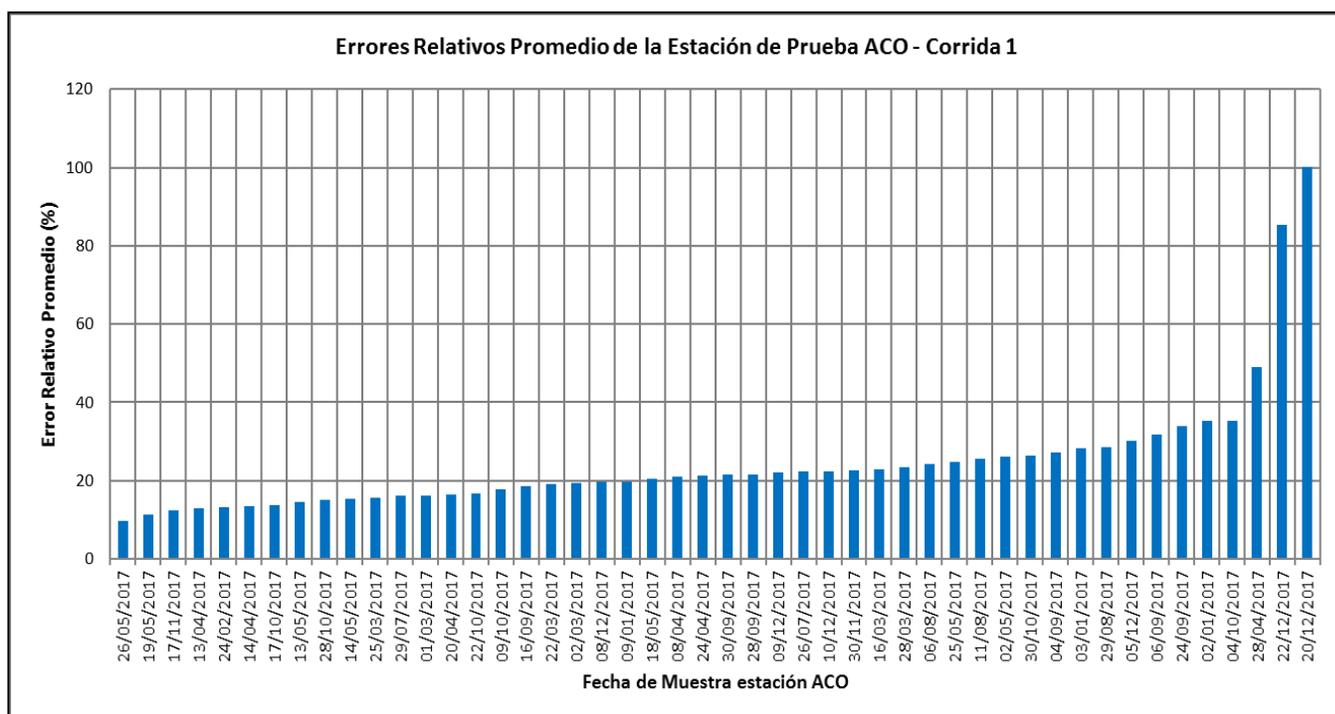
#### 4.5.2. Caso especial estación Acolman

En las gráficas **4.6**, **4.7** y **4.8**, en las gráficas de los ERA máximos, salta a la vista un valor mayor a todos los demás: el de la estación ACO, Acolman. Veamos todos sus valores del ERA, cuando fue escogida como estación de prueba, para las tres corridas; se encuentran en la tabla **4.16**, en las gráficas **4.9**, **4.10** y **4.11** se muestra la gráfica de esos datos.

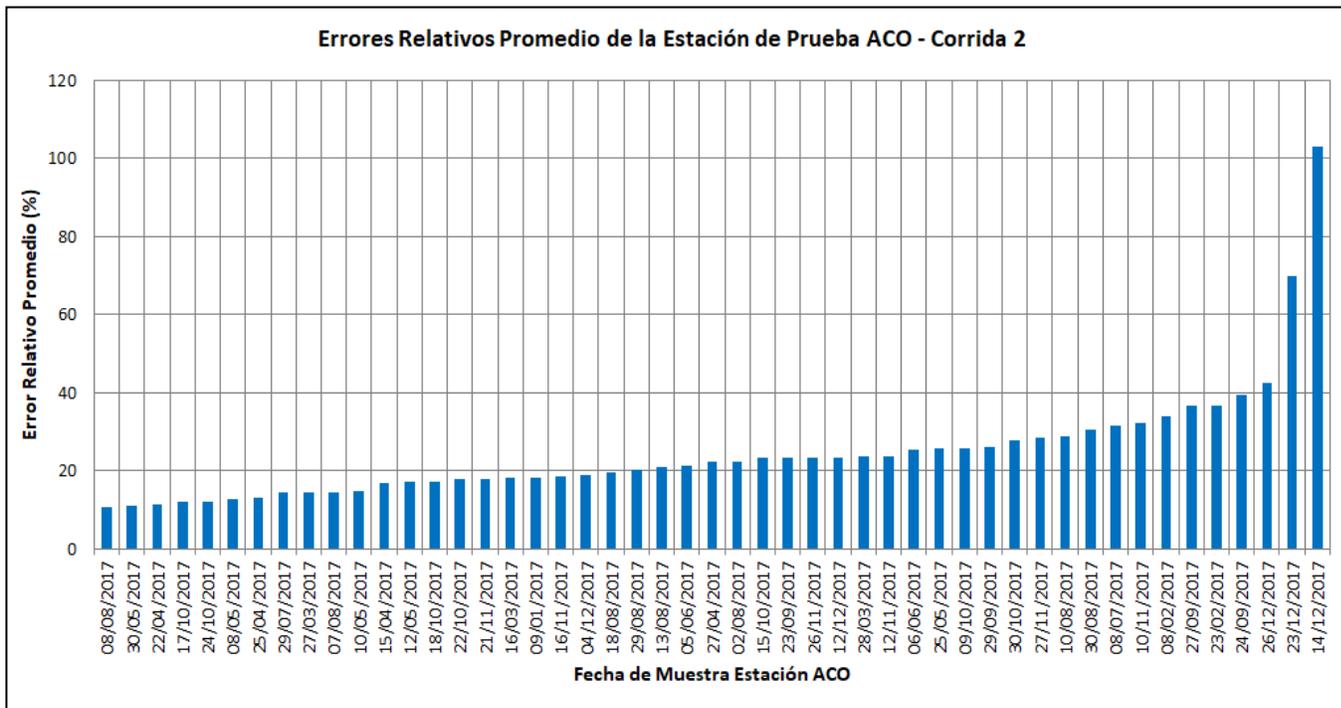
**Tabla 4.16** Errores relativos de ajuste de la estación ACO, Acolman, todas las veces que fue escogida como estación de prueba.

Errores Relativos de Ajuste de la Estación ACO, Acolman					
Corrida 1		Corrida 2		Corrida 3	
Fecha	ERA (%)	Fecha	ERA (%)	Fecha	ERA (%)
26/05/2017	9.57	08/08/2017	10.58	19/05/2017	12.10
19/05/2017	11.42	30/05/2017	11.16	24/02/2017	12.65
17/11/2017	12.35	22/04/2017	11.51	28/10/2017	13.01
13/04/2017	12.98	17/10/2017	12.16	11/11/2017	13.25
24/02/2017	13.12	24/10/2017	12.17	07/08/2017	13.36
14/04/2017	13.37	08/05/2017	12.87	24/05/2017	14.32
17/10/2017	13.80	25/04/2017	13.23	30/04/2017	14.82
13/05/2017	14.68	29/07/2017	14.41	22/05/2017	15.46
28/10/2017	15.16	27/03/2017	14.49	15/04/2017	15.57
14/05/2017	15.38	07/08/2017	14.58	18/05/2017	16.25
25/03/2017	15.70	10/05/2017	14.79	03/12/2017	16.32
29/07/2017	16.07	15/04/2017	16.87	24/11/2017	16.51
01/03/2017	16.14	12/05/2017	17.27	12/10/2017	16.87
20/04/2017	16.35	18/10/2017	17.29	24/08/2017	17.12
22/10/2017	16.68	22/10/2017	17.80	04/12/2017	17.47
09/10/2017	17.79	21/11/2017	17.91	14/03/2017	17.77
16/09/2017	18.52	16/03/2017	18.11	09/01/2017	18.37
22/03/2017	19.16	09/01/2017	18.16	22/08/2017	18.41
02/03/2017	19.26	16/11/2017	18.64	29/07/2017	20.02
08/12/2017	19.56	04/12/2017	19.06	16/09/2017	20.35
09/01/2017	19.62	18/08/2017	19.52	10/03/2017	21.28
18/05/2017	20.42	29/08/2017	20.27	18/09/2017	21.54
08/04/2017	20.95	13/08/2017	20.88	24/04/2017	21.58
24/04/2017	21.25	05/06/2017	21.39	02/05/2017	21.90
30/09/2017	21.50	27/04/2017	22.15	05/03/2017	22.10
28/09/2017	21.60	02/08/2017	22.48	27/07/2017	22.17
09/12/2017	22.18	15/10/2017	23.22	25/05/2017	22.23
26/07/2017	22.44	23/09/2017	23.29	08/01/2017	22.29
10/12/2017	22.46	26/11/2017	23.46	14/07/2017	22.85
30/11/2017	22.69	12/12/2017	23.51	29/09/2017	22.93
16/03/2017	22.84	28/03/2017	23.62	23/02/2017	23.32
28/03/2017	23.38	12/11/2017	23.80	20/11/2017	23.87
06/08/2017	24.32	06/06/2017	25.26	12/11/2017	25.30
25/05/2017	24.73	25/05/2017	25.60	02/01/2017	25.97
11/08/2017	25.70	09/10/2017	25.82	03/01/2017	26.72
02/05/2017	26.07	29/09/2017	26.09	01/06/2017	27.34
30/10/2017	26.51	30/10/2017	27.68	06/01/2017	28.08

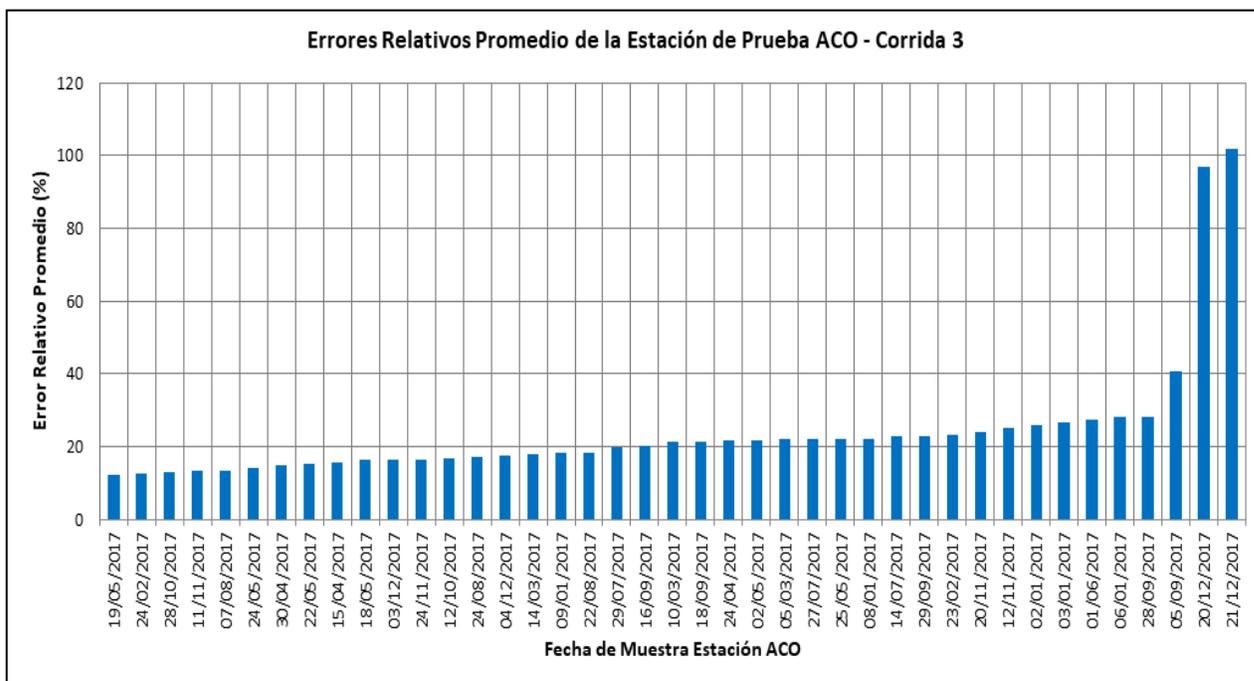
04/09/2017	27.30	27/11/2017	28.51	28/09/2017	28.27
03/01/2017	28.15	10/08/2017	28.97	05/09/2017	40.82
29/08/2017	28.60	30/08/2017	30.35	20/12/2017	97.12
05/12/2017	30.23	08/07/2017	31.65	21/12/2017	102.03
06/09/2017	31.80	10/11/2017	32.12		
24/09/2017	33.85	08/02/2017	34.10		
02/01/2017	35.18	27/09/2017	36.61		
04/10/2017	35.27	23/02/2017	36.74		
28/04/2017	48.92	24/09/2017	39.52		
22/12/2017	85.29	26/12/2017	42.64		
20/12/2017	100.15	23/12/2017	69.96		
		14/12/2017	103.00		



**Gráfica 4.9. Corrida 1.** Valores de los ERA mínimos, máximos y la mediana de los ERA de las estaciones de monitoreo para el año 2017.



**Gráfica 4.10. Corrida 2.** Valores de los ERA mínimos, máximos y la mediana de los ERA de las estaciones de monitoreo para el año 2017.



**Gráfica 4.11. Corrida 3.** Valores de los ERA mínimos, máximos y la mediana de los ERA de las estaciones de monitoreo para el año 2017.

En las gráficas **4.9**, **4.10** y **4.11**, observando más a detalle se puede identificar cuáles son las fechas que tuvieron los ERA máximos de la estación ACO. En la tabla **4.17** se muestra las fechas de los días que tuvieron los ERA máximos de cada corrida de estación ACO. Se tomaron dos días con ERA máximo de cada corrida ya que son en los que se observa un mayor incremento del ERA. Y se puede apreciar que los valores de ERA máximos corresponden al mes de diciembre para las 3 corridas; esto quiere decir que es en el mes de diciembre cuando incrementa el ERA de la estación ACO.

**Tabla 4.17** Errores relativos de ajuste máximos de la estación ACO, Acolman, (ACO como estación de Prueba)

Errores Relativos de Ajuste de la Estación ACO, Acolman							
Corrida 1			Corrida 2			Corrida 3	
Fecha	ERA (%)		Fecha	ERA (%)		Fecha	ERA (%)
22/12/2017	85.29		23/12/2017	69.96		20/12/2017	97.12
20/12/2017	100.15		14/12/2017	103.00		21/12/2017	102.03

En el Apéndice **B** se muestra los resultados de los ejemplos de los ERA máximos de la estación ACO, es decir un ejemplo del peor ajuste de la estación ACO para cada corrida, con la finalidad de analizar el caso de la estación ACO. También se muestra un ejemplo con el ERA mínimo de la estación ACO, es decir de las tres corridas se seleccionó un ejemplo con el *mejor ajuste* de la estación ACO.

#### 4.5.3. Análisis de los resultados de los ajustes de los perfiles diarios de las concentraciones de ozono para las estaciones de prueba con mayor ERA

Como se mencionó en la sección **4.5**, las estaciones de monitoreo se dividieron en tres grupos según sus ERAs, como se muestra en la tabla **4.15**. Retomando esta división en grupos, falta analizar el tercer grupo.

Este tercer grupo está formado por las estaciones de prueba que obtuvieron un mayor  $\overline{\overline{ERA}}$ , que en este caso está entre 25% y 31%, y está formado por seis estaciones: AJM, MPA, SFE, AJU, CUA y INN. En la tabla **4.18** se obtienen los promedios de los ERA promedio,  $\overline{\overline{ERA}}$ , para las estaciones de prueba que forman parte del grupo 3.

En la gráfica **4.5** también se puede observar el promedio de las tres corridas de los  $\overline{\overline{ERA}}$  de cada una de las estaciones de prueba. Ahora nos vamos a enfocar sólo en las estaciones con mayor  $\overline{\overline{ERA}}$  (considerando  $\overline{\overline{ERA}} > 25\%$ ). En este caso son seis estaciones que conforman el grupo con mayor  $\overline{\overline{ERA}}$ , las cuales son AJM (Ajusco Medio), MPA(Milpa Alta), SFE(Santa Fe), AJU(Ajusco), CUA(Cuajimalpa), y INN(Instituto Nacional de Investigaciones Nucleares).

**Tabla 4.18** Promedio de las tres corridas de los errores relativos de ajuste promedio,  $\overline{\overline{ERA}}$  de las estaciones que forman parte del grupo 3.

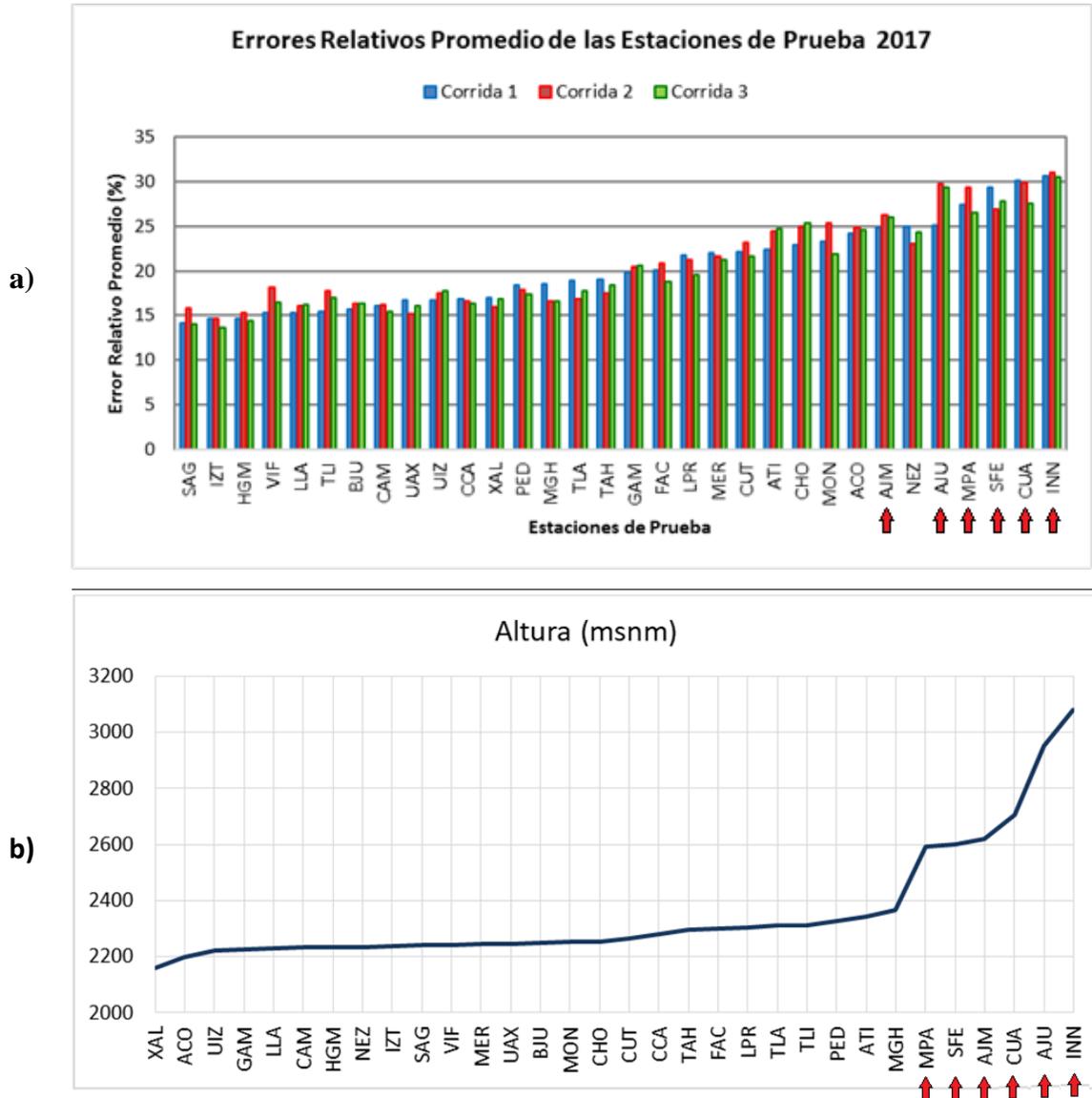
Identificador de la estación	Corrida 1	Corrida 2	Corrida 3	Promedio
AJM	24.86	26.25	26.01	25.71
MPA	27.44	29.32	26.50	27.75
SFE	29.32	26.87	27.78	27.99
AJU	25.05	29.77	29.30	28.04
CUA	30.18	29.84	27.54	29.18
INN	30.57	31.01	30.47	30.68

En la figura **4.4**, se muestra la relación de las estaciones que pertenecen al grupo 3, con la altura. Esto es porque las estaciones del grupo 3 tienen mayor  $\overline{\overline{ERA}}$  y existe la posibilidad que tenga relación con la *Altura* de dichas estaciones. Las estaciones de monitoreo que pertenecen al grupo 3, son las estaciones de prueba que tienen mayor altura como se muestra en la figura **4.4b**.

En la figura **4.4a** de cada corrida se muestra las estaciones de prueba que tuvieron mayores  $\overline{\overline{ERA}}$ . En la figura **4.4a** y **4.4b** se observa que aparentemente existe una relación de las estaciones de prueba que tienen los máximos  $\overline{\overline{ERAs}}$  (del grupo 3) con

respecto a la altura, ya que coincide que las estaciones con mayor altura pertenecen al grupo 3.

Para este trabajo no se consideró como parámetro de estudio la altura.



**Figura 4.4.** Relación de estaciones del grupo 3 con mayor  $\overline{\overline{ERA}}$ , con la altura.

a) Muestra las estaciones de prueba de las 3 corridas, que tuvieron mayores  $\overline{\overline{ERAs}}$  señaladas con una flecha roja (estaciones del grupo 3); b) Muestra la Altura de las estaciones de monitoreo, y se encuentran señaladas con una flecha roja las estaciones con mayor altura.

#### 4.5.4. Análisis de los resultados con los mejores y los peores ajustes de los perfiles diarios de las concentraciones de ozono

En esta sección se presentan los resultados de los mejores y los peores ajustes de los perfiles diarios de concentración de ozono que se obtuvieron de las tres corridas. Para esto se decidió tomar solo tres ejemplos con los mejores ajustes de las tres corridas, en la tabla **4.19** se muestran los tres *mejores ajustes* de los perfiles diarios de la concentración de ozono, se considera que los *mejores ajustes* son los que tienen el valor mínimo del porcentaje de ERA promedio, de las tres corridas.

El día con el mejor ajuste fue el *19 de Junio de 2017* que se obtuvo de la corrida 2, y cuyo ERA promedio es de **7.87%**.

El segundo día con mejor ajuste fue el *19 de Junio de 2017* que se obtuvo de la corrida 1, y cuyo ERA promedio es de **9.30%**. El tercer día con mejor ajuste fue el *20 de Junio de 2017* que se obtuvo de la corrida 2, y cuyo ERA promedio es de **9.94%**.

Se puede checar los detalles de los mejores ajustes en la tabla **4.19**, donde se muestran el porcentaje del ERA promedio, el número de estaciones totales, las estaciones de prueba, la fecha y la corrida, de cada uno de los ejemplos. Así mismo en la tabla **4.20** se puede checar los detalles de los peores ajustes, de cada uno de los ejemplos.

**Tabla 4.19** Porcentajes del Error Relativo de Ajuste Promedio que se obtuvieron de los mejores ajustes de los perfiles diarios de la concentración de ozono, de las tres corridas.

MEJORES AJUSTES					
#	Error Relativo Promedio (%)	Núm. Est. Totales	Estaciones de Prueba	Fecha	Corrida
1	7.87	27	AJM, ATI, IZT, TLI, UAX	19 de Junio de 2017	2
2	9.30	27	FAC, MER, MON, TAH, TLA	19 de Junio de 2017	1
3	9.94	28	CAM, CUA, LLA, SAG, UAX	20 de Junio de 2017	2

Y en la tabla 4.20 se muestran los tres *peores ajustes* de los perfiles diarios de la concentración de ozono. Para este caso se considera que los *peores ajustes* son los que tienen el valor máximo del porcentaje de ERA promedio, de las tres corridas.

El día con el peor ajuste fue el *14 de Diciembre de 2017* que se obtuvo de la corrida 2, y cuyo ERA promedio es de **44.79%**.

El segundo día con peor ajuste fue el *21 de Diciembre de 2017* que se obtuvo de la corrida 1, y cuyo ERA promedio es de **43.43%**. El tercer día con peor ajuste fue el *4 de Diciembre de 2017* que se obtuvo de la corrida 2, y cuyo ERA promedio es de **40.73%**.

**Tabla 4.20** Porcentajes del Error Relativo de Ajuste Promedio que se obtuvieron de los peores ajustes de los perfiles diarios de la concentración de ozono, de las tres corridas.

PEORES AJUSTES					
#	Error Relativo Promedio (%)	Núm. Est. Totales	Estaciones de Prueba	Fecha	Corrida
1	44.79	21	ACO, AJU, GAM, LPR, LLA	14 de Diciembre de 2017	2
2	43.43	23	ACO, FAC, HGM, INN, MER	21 de Diciembre de 2017	3
3	40.73	23	ACO, CUA, MER, MPA, SAG	04 de Diciembre de 2017	2

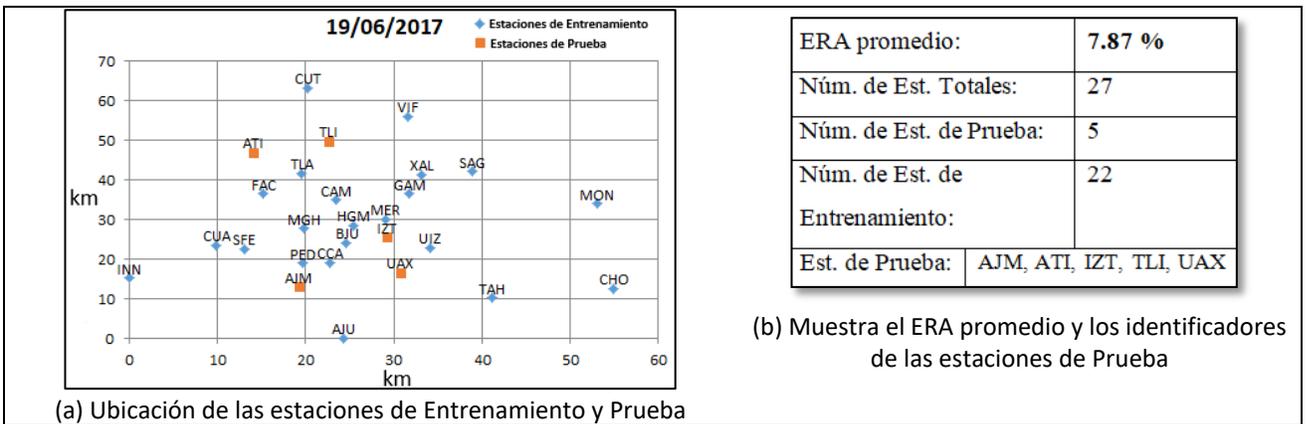
En las siguientes secciones se mostrará los resultados obtenidos del mejor y el peor ajuste de perfil de concentración de ozono.

### **Ejemplo del *mejor ajuste* de perfiles de concentración de ozono**

En la sección anterior se mostró algunos detalles acerca de los tres *mejores ajustes* de perfiles de concentración de ozono, sin embargo en la sección actual se eligió sólo el mejor de los tres, es decir se eligió el día con el *mejor ajuste* de los perfiles de concentración de ozono, donde se mostrarán los resultados de los ajustes obtenidos de cada una de las estaciones de entrenamiento y de prueba.

El día con el *mejor ajuste* fue el 19 de Junio de 2017 que se obtuvo de la corrida 2, y cuyo ERA promedio es de 7.87%, como se observa en la tabla 4.19.

Y en la figura 4.5.I. se muestra el ejemplo del *mejor ajuste* correspondiente al 19 de Junio de 2017; en la figura 4.5.Ia se muestra la ubicación de las estaciones de Entrenamiento y las de Prueba; en la figura 4.5.Ib muestra el ERA promedio de 7.87%, e indica 22 estaciones de entrenamiento, 5 estaciones de prueba, y un total de 27 estaciones utilizables. Además muestra los identificadores de las estaciones de prueba AJM (Ajusco Medio), ATI (Atizapán), IZT (Iztacalco), TLI (Tultitlán) y UAX (UAM Xochimilco); dichas estaciones fueron elegidas aleatoriamente.

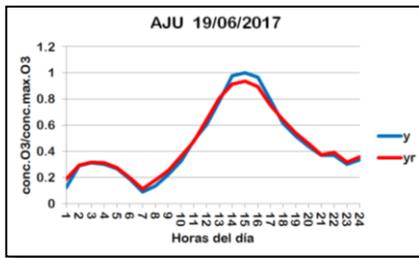


**Figura 4.5.I.** Ejemplo del *mejor ajuste* de los perfiles de concentración de ozono, correspondiente al 19 de Junio de 2017. (a) Ubicación de las estaciones de Entrenamiento y Prueba; (b) Muestra el ERA promedio, el número de estaciones de Entrenamiento y Prueba, y los identificadores de las estaciones de Prueba.

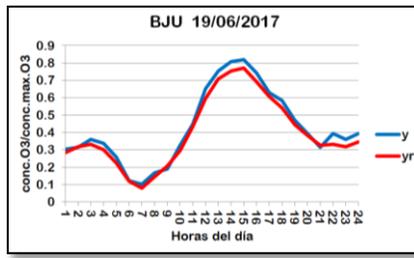
En la figura 4.5.II se muestra los perfiles diarios de concentración de ozono para cada estación obtenidos con la red MLP entrenada; en los incisos (a), (b),..., (v) se muestra los ajustes de cada una de las estaciones de entrenamiento correspondientes al ejemplo del *mejor ajuste* de los perfiles de concentración de ozono del 19 de Junio de 2017. En general, se observa que se obtuvieron buenos ajustes de los perfiles de concentración de ozono de las estaciones de entrenamiento.

En la gráfica **4.12**, se muestra porcentajes del ERA de cada una de las estaciones de entrenamiento, correspondiente al ejemplo del *mejor ajuste* de los perfiles de concentración de ozono del 19 de Junio de 2017, y se observa que la estación SFE (Santa Fe) es la que obtuvo el ERA máximo de **20.54%**, que es un poco más elevado con respecto a las demás estaciones de entrenamiento, y la estación INN (Instituto Nacional de Investigaciones Nucleares) obtuvo el ERA mínimo de **4.88%**. Sin embargo, el que se considera es el ERA promedio de **8.84%** de las estaciones de entrenamiento, esto es, que se tuvo buen ajuste en general.

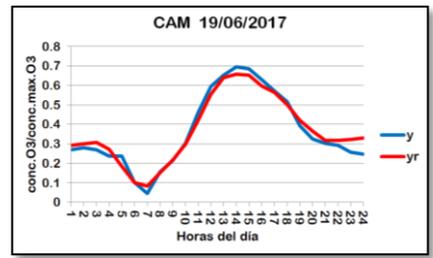
Cabe mencionar que para este trabajo se le da mayor importancia a los resultados de los ajustes de las estaciones de prueba que a los ajustes de las estaciones de entrenamiento, porque el objetivo es tener una buena generalización para las estaciones de prueba. Sin embargo también se reportan los resultados de los ajustes de las estaciones de entrenamiento.



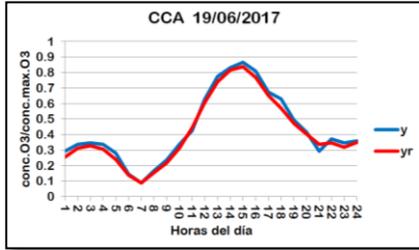
(a) Ajusco



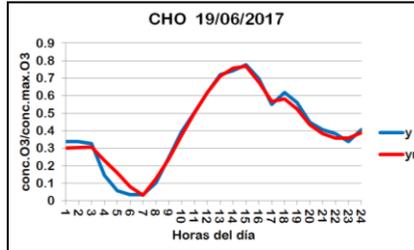
(b) Benito Juárez



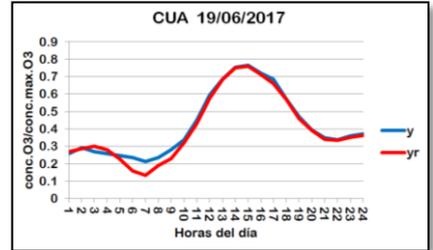
(c) Camarones



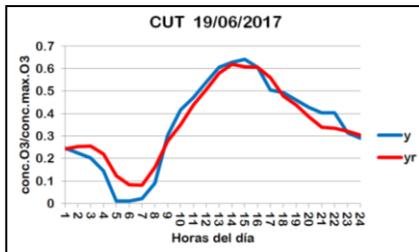
(d) Centro de Ciencias de la Atmosfera



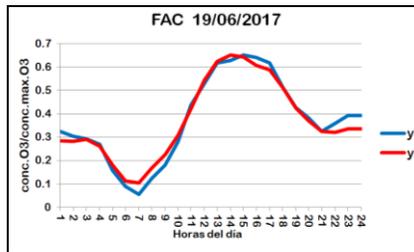
(e) Chalco



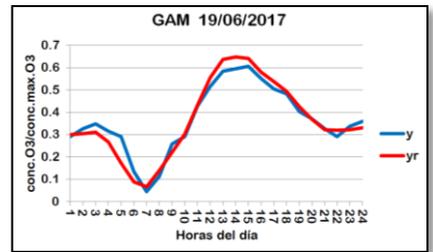
(f) Cuajimalpa



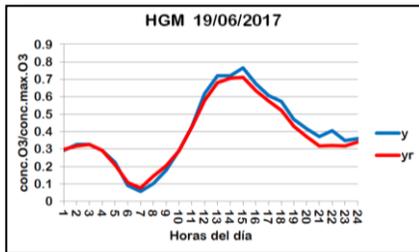
(g) Cuautitlán



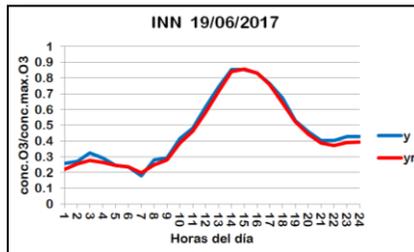
(h) FES Acatlán



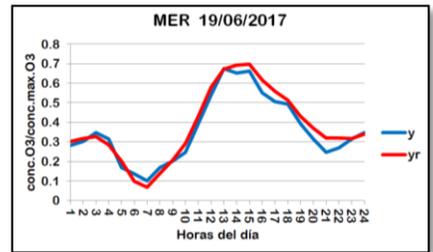
(i) Gustavo A. Madero



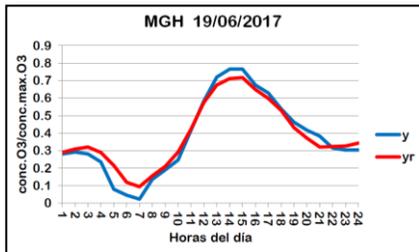
(j) Hospital General de México



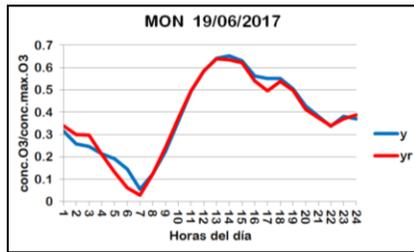
(k) Instituto Nacional de Investigaciones Nucleares (ININ)



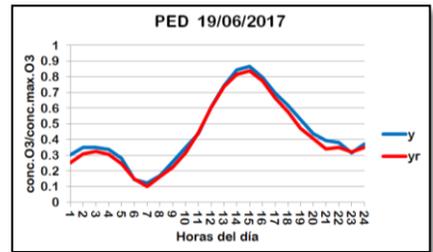
(l) Merced



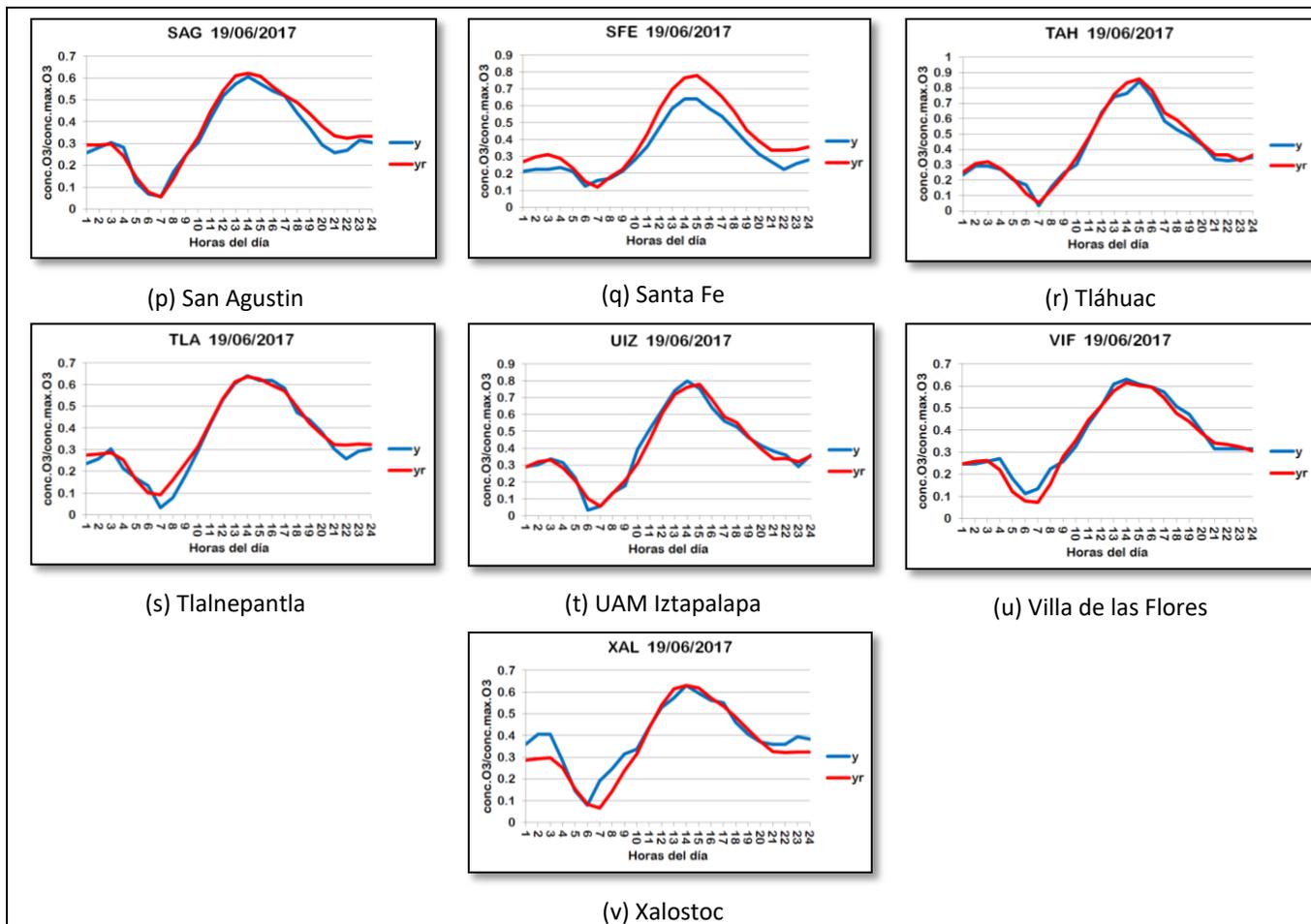
(m) Miguel Hidalgo



(n) Montecillo



(o) Pedregal



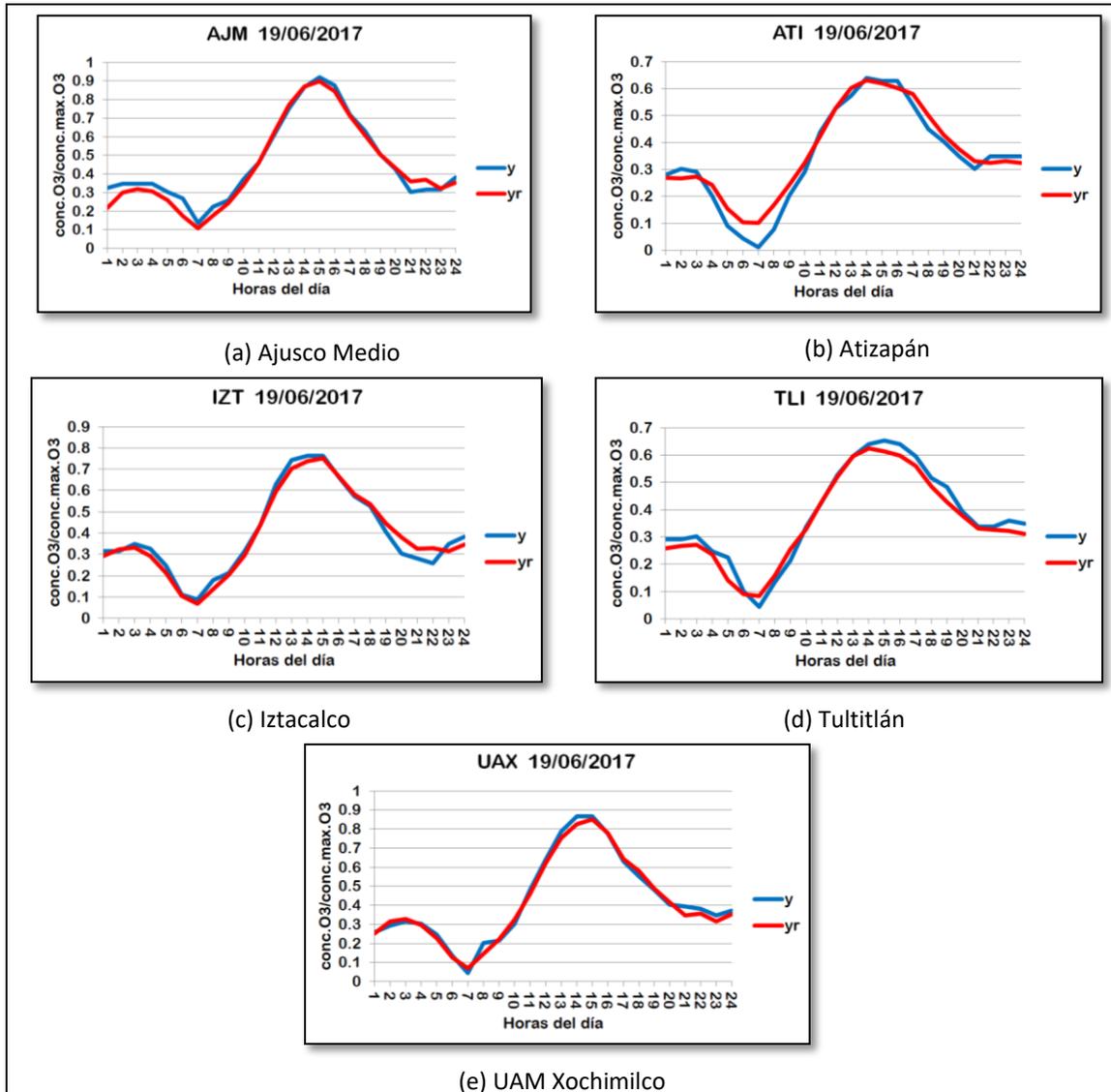
**Figura 4.5.II.** Muestra los Perfiles diarios de concentración de ozono para cada estación obtenidos con la red MLP entrenada. En los incisos (a), (b), ..., (v) se muestra los ajustes de cada una de las estaciones de entrenamiento, correspondientes al ejemplo del *mejor ajuste* de los perfiles de concentración de ozono del 19 de Junio de 2017.



**Grafica 4.12.** Porcentajes de ERA de las estaciones de entrenamiento, correspondiente al ejemplo del *mejor ajuste* de los perfiles de concentración de ozono.

A continuación se mostrarán los resultados de los ajustes de los perfiles de concentración de ozono de las estaciones de prueba correspondientes al *mejor ajuste*; éstos son de suma importancia ya que son los que se toman en cuenta para este trabajo.

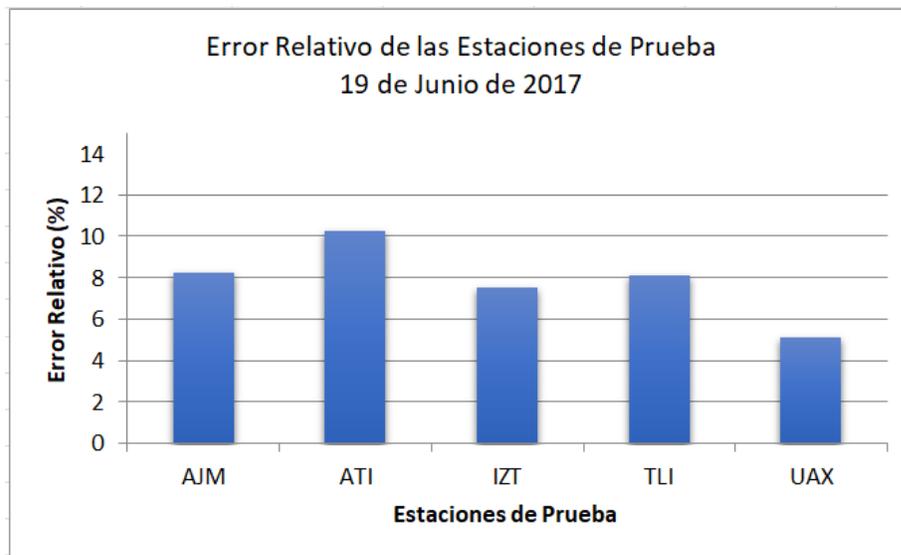
En la figura **4.5.III** se muestra los perfiles diarios de concentración de ozono para cada estación obtenidos con la red MLP entrenada; en los incisos (a), (b),..., (e) se muestra los ajustes de cada una de las estaciones de prueba correspondientes al ejemplo del *mejor ajuste* de los perfiles de concentración de ozono del 19 de Junio de 2017. En general, se observa que se obtuvieron buenos ajustes de los perfiles de concentración de ozono de las estaciones de prueba: AJM (Ajusco Medio), ATI (Atizapan), IZT (Iztacalco), TLI (Tultitlán) y UAX (UAM Xochimilco).



**Figura 4.5.III.** Muestra los Perfiles diarios de concentración de ozono para cada estación obtenidos con la red MLP entrenada. En los incisos (a), (b), (c), (d) y (e) se muestra los ajustes de cada una de las estaciones de Prueba, correspondientes al ejemplo del *mejor ajuste* de los perfiles de concentración de ozono del 19 de Junio de 2017.

En la gráfica **4.13**, se muestra los porcentajes del ERA de cada una de las estaciones de prueba, correspondiente al ejemplo del *mejor ajuste* de los perfiles de concentración de ozono, y se observa que la estación ATI (Atizapan) es la que obtuvo el ERA máximo de **10.27%**, y la estación UAX (UAM Xochimilco) obtuvo el ERA

minimo de **5.15%**. Entonces para las estaciones de prueba el resultado que se considera es el ERA promedio que es de **7.87%** . Este resultado es el mejor ajuste que se obtuvo de las tres corridas y corresponde al 19 de Junio de 2017.



**Gráfica 4.13.** Porcentajes de ERA de las estaciones de prueba, correspondiente al ejemplo del mejor ajuste de los perfiles de concentración de ozono.

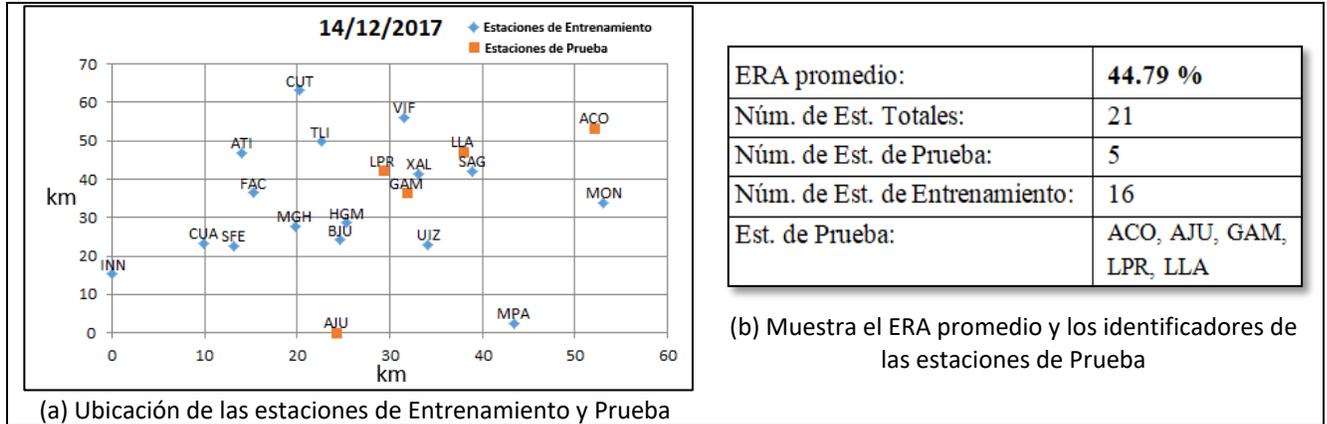
### **Ejemplo del *peor ajuste* de perfiles de concentración de ozono**

En la sección **4.5.4**, se mostró algunos detalles acerca de los peores ajustes de perfiles de concentración de ozono. Sin embargo en la sección actual, se eligió sólo un ejemplo, que en este caso sólo el peor ajuste para poder apreciar más a detalle los resultados.

El día con el peor ajuste fue el *14 de Diciembre de 2017* que se obtuvo de la corrida 2, y cuyo ERA promedio es de **44.79%**, como se observa en la tabla **4.20**.

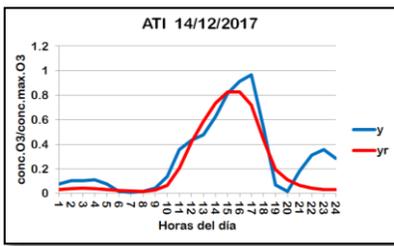
En la figura **4.6.I**. se muestra el ejemplo del *peor ajuste* correspondiente al 14 de Diciembre de 2017; en la figura **4.6.Ia** se muestra la ubicación de las estaciones de Entrenamiento y las de Prueba; en la figura **4.6.Ib** muestra el ERA promedio de **44.79%**, e indica 16 estaciones de entrenamiento, 5 estaciones de prueba, y un total de 21 *estaciones utilizables*. Además muestra los identificadores de las estaciones de

prueba ACO (Acolman), AJU (Ajusco), GAM (Gustavo A. Madero), LPR (La Presa) y LLA (Los Laureles), dichas estaciones fueron elegidas aleatoriamente.

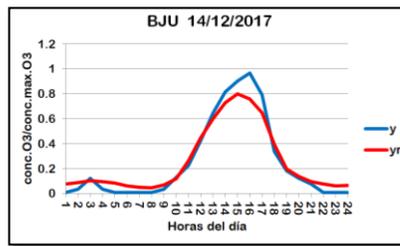


**Figura 4.6.I.** Ejemplo del *peor ajuste* de los perfiles de concentración de ozono, correspondiente al 14 de Diciembre de 2017. (a) Ubicación de las estaciones de Entrenamiento y Prueba; (b) Muestra el ERA promedio, el número de estaciones de Entrenamiento y Prueba, y los identificadores de las estaciones de Prueba.

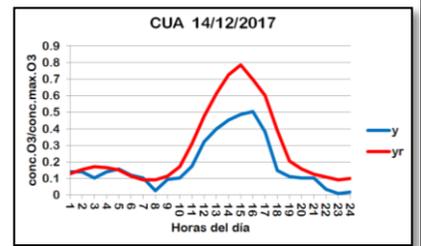
En la figura **4.6.II** se muestra los perfiles diarios de concentración de ozono para cada estación obtenidos con la red MLP entrenada; en los incisos (a), (b),..., (p) se muestra los ajustes de cada una de las estaciones de entrenamiento correspondientes al ejemplo del *peor ajuste* de los perfiles de concentración de ozono del 14 de Diciembre de 2017.



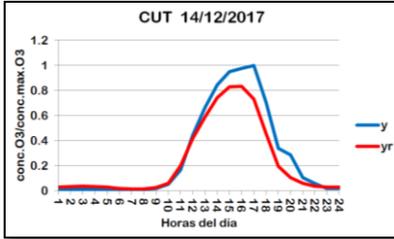
(a) Atizapán



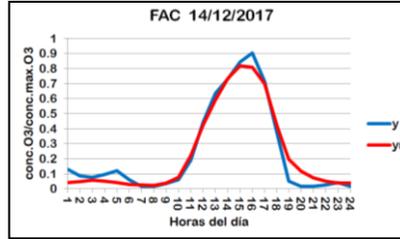
(b) Benito Juárez



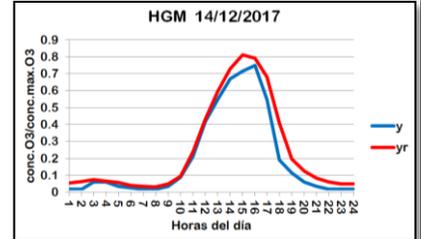
(c) Cuajimalpa



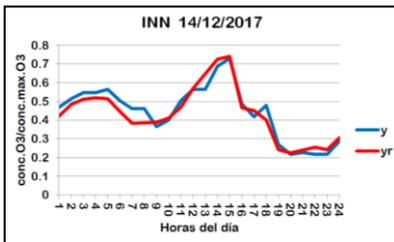
(d) Cuautitlán



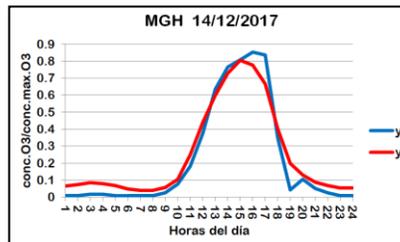
(e) FES Acatlán



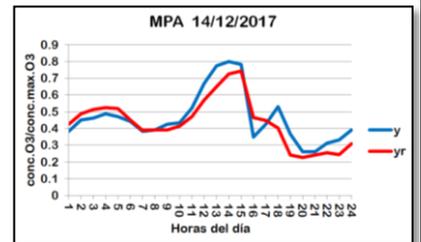
(f) Hospital General de México



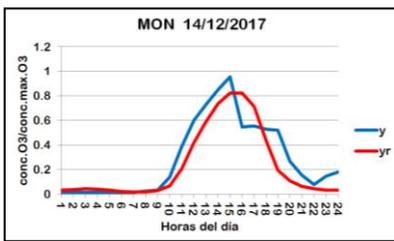
(g) Instituto Nacional de Investigaciones Nucleares (ININ)



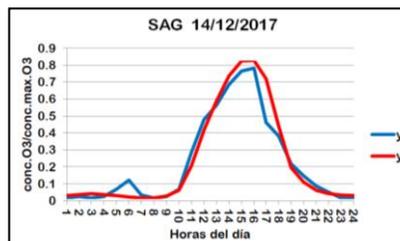
(h) Miguel Hidalgo



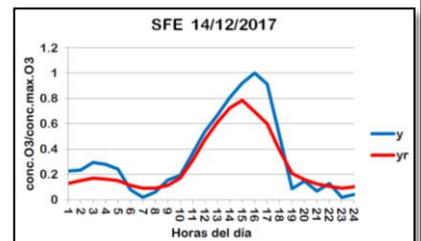
(i) Milpa Alta



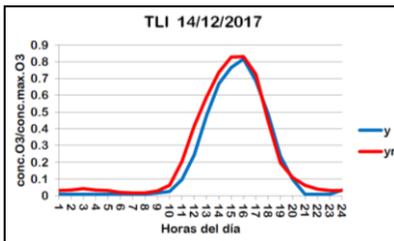
(j) Montecillo



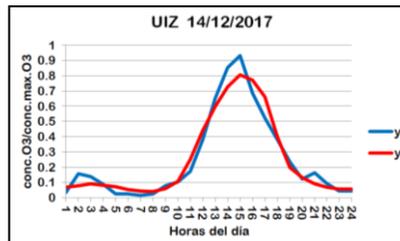
(k) San Agustín



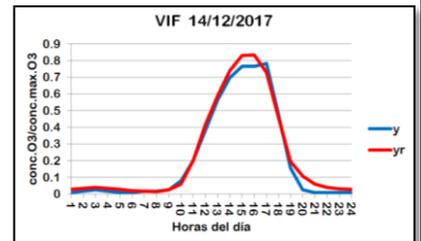
(l) Santa Fe



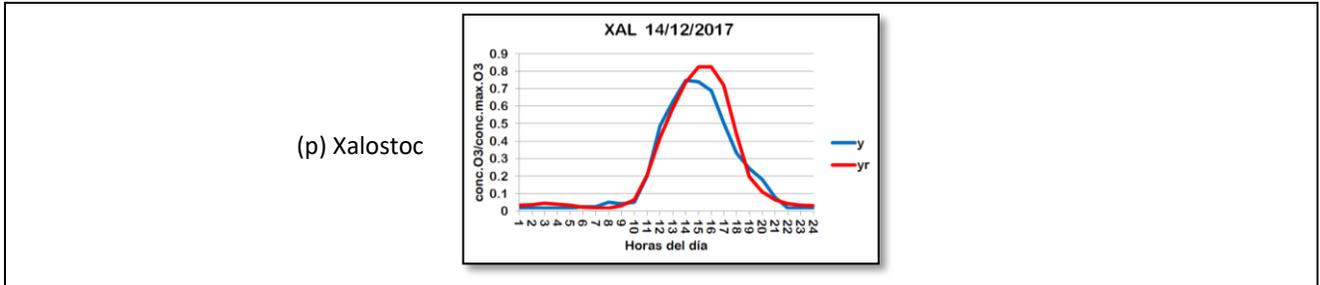
(m) Tultitlán



(n) UAM Iztapalapa



(o) Villa de las Flores



**Figura 4.6.II.** Muestra los Perfiles diarios de concentración de ozono para cada estación

obtenidos con la red MLP entrenada. En los incisos (a), (b), ..., (p) se muestra los ajustes de cada una de las estaciones de entrenamiento, correspondientes al ejemplo del *peor ajuste* de los perfiles de concentración de ozono del 14 de Diciembre de 2017.

En la gráfica **4.14** se muestra porcentajes del ERA de cada una de las estaciones de entrenamiento, correspondiente al ejemplo del *peor ajuste* de los perfiles de concentración de ozono del 14 de Diciembre de 2017, y se observa que la estación CUA (Cuajimalpa) es la que obtuvo el ERA máximo de **45.24%**, que es un poco más elevado con respecto a las demás estaciones de entrenamiento, y la estación INN (Instituto Nacional de Investigaciones Nucleares) obtuvo el ERA mínimo de **9.24%**. Sin embargo, el que se considera es el ERA promedio de **21.32%** de las estaciones de entrenamiento, esto es, que se tuvo ajuste regular en general.



**Grafica 4.14.** Porcentajes de ERA de las estaciones de entrenamiento, correspondiente al ejemplo del peor ajuste de los perfiles de concentración de ozono.

A continuación se mostrarán los resultados de los ajustes de los perfiles de concentración de ozono de las estaciones de prueba correspondientes al *peor ajuste*. Estos son de suma importancia ya que son los que se toman en cuenta para este trabajo.

En la figura **4.6.III** se muestra los perfiles diarios de concentración de ozono para cada estación obtenidos con la red MLP entrenada; en los incisos (a), (b),..., (e) se muestra los ajustes de cada una de las estaciones de prueba correspondientes al ejemplo del *peor ajuste* de los perfiles de concentración de ozono del 14 de Diciembre de 2017.

Como se muestra en la figura **4.6.III** los resultados obtenidos de las estaciones de prueba para el ejemplo del *peor ajuste* de los perfiles de concentración de ozono, son cinco estaciones: ACO (Acolman), AJU (Ajusco), GAM (Gustavo A. Madero), LPR (La Presa) y LLA (Los Laureles). Y se puede observar que las estaciones GAM, LPR y LLA (figuras **4.6.IIIc**, **4.6.III d** y **4.6.IIIe** respectivamente) obtuvieron buenos ajustes de los perfiles de concentración de ozono. Por otro lado en las estaciones ACO y AJU (figuras **4.6.IIIa** y **4.6.IIIb** respectivamente) se puede observar que no obtuvieron buenos ajustes de los perfiles de concentración de ozono.

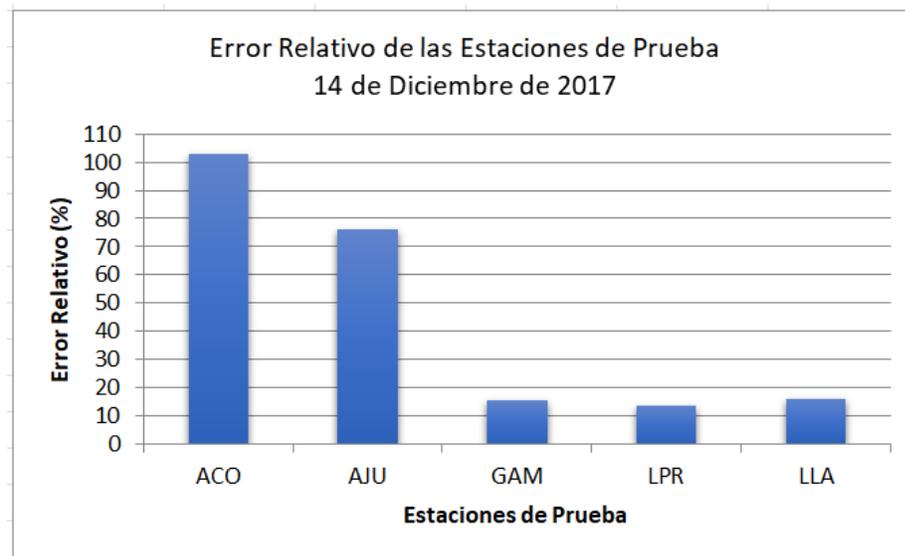
En la gráfica **4.15** se muestra los porcentajes del ERA de cada una de las estaciones de prueba, correspondiente al ejemplo del *peor ajuste* de los perfiles de concentración de ozono, y se observa que la estación ACO (Acolman) es la que obtuvo el ERA máximo de **103.00%**, y la estación LPR (La Presa) obtuvo el ERA mínimo de **13.45%**. Entonces para las estaciones de prueba el resultado que se considera es el ERA promedio que es de **44.79%**. Este resultado es el *peor ajuste* que se obtuvo de las tres corridas y corresponde al 14 de Diciembre de 2017.

Como se muestra en la grafica **4.15**, analizando más a detalle los ERA de las cinco estaciones de prueba son solo dos estaciones que realmente tienen un ERA muy grande en comparación de las otras estaciones. La estación ACO (Acolman) con un ERA de **103.00%** y la estación AJU (Ajusco) con un ERA de **76.02%** son las estaciones de prueba que presentan el ERA muy grande. Anteriormente en la sección **4.5.2** se ha mencionado que la estación ACO es un caso especial que se analizó individualmente, debido a que precisamente en el mes de diciembre presenta errores extremadamente grandes. Lo que ocurre en esa temporada es que la medición del perfil de concentración

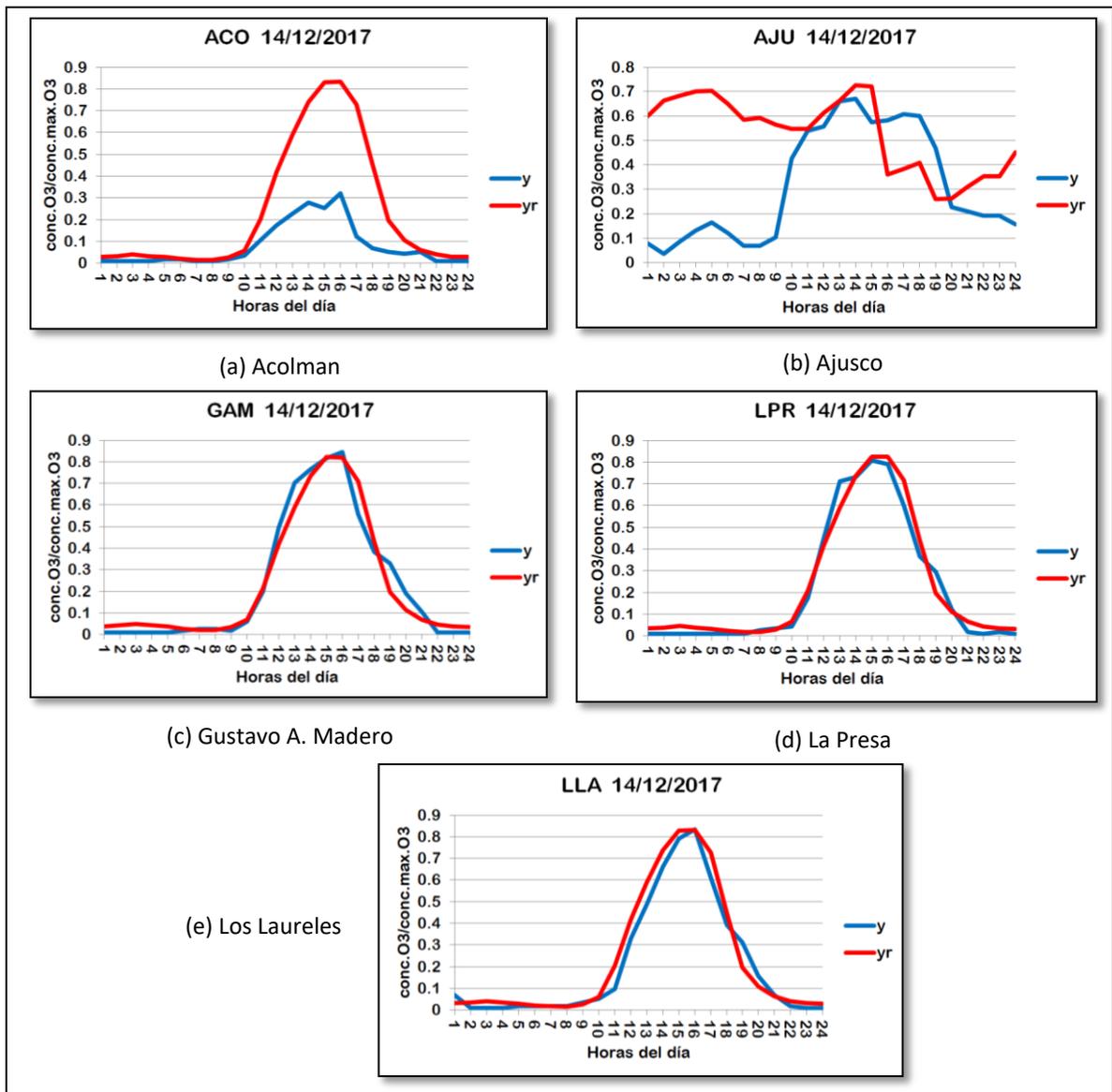
de ozono de la estación ACO disminuye y es mayor la medición del perfil de concentración de ozono obtenido por la red, por esta razón incrementa el error. Por otro lado con lo que respecta a la estación AJU, existe sospecha de una posible relación del incremento del error debido a la *Altura*, como se menciona en la sección **4.5.3**.

Por otro lado para las tres estaciones de prueba restantes GAM, LPR y LLA tienen un ERA menor que **16%** cada una, por lo cual no representan un error significativo.

En conclusión, para las estaciones de prueba donde se obtuvo un ERA promedio de **44.79%**, resulta ser el peor error de ajuste de los perfiles de concentración de ozono que se obtuvo de las tres corridas, y corresponde al 14 de Diciembre de 2017. Sin duda el error es grande debido a que son dos estaciones de prueba las que influyen en el incremento del error.



**Grafica 4.15.** Porcentajes de ERA de las estaciones de prueba, correspondiente al ejemplo del peor ajuste de los perfiles de concentración de ozono.



**Figura 4.6.III.** Muestra los Perfiles diarios de concentración de ozono para cada estación obtenidos con la red MLP entrenada. En los incisos (a), (b), (c), (d) y (e) se muestra los ajustes de cada una de las estaciones de Prueba, correspondientes al ejemplo del *peor ajuste* de los perfiles de concentración de ozono del 14 de Diciembre de 2017.

#### 4.5.5. Software desarrollado

Se implementaron cada uno de los programas específicamente para este trabajo, y los programas se fueron desarrollaron conforme se fueron requiriendo.

Una vez que se descargó la base de datos de las mediciones de la concentración de ozono (SEDEMA, 2018), fue necesario crear una nueva base de datos donde dichos datos no tuvieran valores nulos, por eso se creó el programa “separador de días sin nulos” y en la tabla 4.21 se muestra una breve descripción.

Se propuso un conjunto de 12 topologías, y se desarrollaron los programas necesarios para probarlas y seleccionar así la más adecuada. Además se usó la información sobre las concentraciones de ozono del año 2017 para entrenar y probar las redes propuestas. Se usó el programa “Simulador de RNAs multicapa de perceptrones de validación” y en la tabla 4.21 se muestra una breve descripción.

Una vez que se obtuvo la topología más adecuada, se utilizan los programas creados para simular las redes neuronales artificiales MLP aplicado a los datos de todo el año 2017.

En la tabla **4.21** se muestra una lista de los principales programas desarrollados y la descripción de sus funciones.

**Tabla 4.21.** Programas desarrollados y la descripción de las tareas que realizan.

PROGRAMA	DESCRIPCIÓN
Separador de días sin nulos	Este programa se encarga de seleccionar de la base de datos original, para cada una de las estaciones de monitoreo, los días en los que no se presentó ningún dato nulo*, creando así una base que solo contiene días sin nulos.
Seleccionador de estaciones totalmente activas en un día dado	Dado un día del año, este programa selecciona las estaciones en las que no se presentó ni un dato nulo, es decir, selecciona solo las <i>estaciones válidas</i> .
Seleccionador al azar de estaciones de prueba	Dado un día, selecciona aleatoriamente, un número especificado de estaciones de la base constituida por días sin nulos.
Simulador de RNAs multicapa de perceptrones	Para simular redes neuronales artificiales multicapa de perceptrones,
	Se crearon todos los programas necesarios para procesar la información que se alimenta a esas redes,
	Se crearon todos los programas para procesar la información obtenida con las redes.
Simulador de RNAs multicapa de perceptrones de validación.	Para simular redes neuronales artificiales multicapa de perceptrones, específicamente para probar cada una de las topologías propuestas.

\* Se llama aquí *dato nulo* o simplemente *nulo*, cuando el dato de concentración de ozono de alguna estación, en alguna hora de algún día, no se reportó (tal vez, la estación no funcionó por alguna causa en esa hora).

## Conclusiones

El uso de las redes neuronales artificiales (RNA), son herramientas que han ampliado el campo del cómputo en la investigación científica. Sin duda esta fue la principal motivación para este trabajo.

El objetivo principal fue la obtención de los perfiles diarios de concentración de Ozono desconocidos en las diferentes coordenadas de la zona metropolitana del Valle de México (ZMVM), a partir de los datos conocidos, usando redes multicapa de perceptrones entrenadas con los datos que proporciona la SEDEMA. Para esto, se usaron como parámetros las coordenadas geográficas (longitud y latitud) y los perfiles de concentración de ozono de las estaciones de monitoreo. Para cumplir lo anterior se usaron redes MLP entrenadas con el algoritmo de propagación hacia atrás con momentum y factor de aprendizaje constante. Para este trabajo no se consideró el parámetro de la altura, porque se quería investigar si se podía obtener una buena aproximación de perfiles de concentración de ozono únicamente considerando las coordenadas (longitud y latitud).

Se utilizaron los datos de concentración de ozono registrados el año 2017. A pesar de que se eliminaron los datos de los días que tenían algún valor nulo, quedó una buena cantidad de datos para trabajar, los datos de 279 días utilizables, que son el 76.44% de los 365 días del año que tuvieron perfiles completos de concentración de ozono en la ZMVM. Los 279 días utilizables estuvieron bien repartidos en los 12 meses del año. El promedio de días utilizables por mes fue de 23.25 días. Con un total de 32 estaciones de monitoreo, el promedio diario de estaciones válidas para los días utilizables fue de 22.81 estaciones, que corresponde al 71.28% de esas 32 estaciones.

Se realizaron tres repeticiones o corridas, cuyos resultados se concluye lo siguiente:

❖ Los ERA mínimos de todas las estaciones en todas las corridas son menores al 20%, por lo tanto se puede afirmar que por lo menos hubo un caso (esto es, un día) para cada una de las estaciones, tratadas como estación de prueba, en las que el ERA fue menor de 20%. Esto se cumplió para las tres corridas.

❖ Los ERA máximos sobrepasan el 20% para todas las estaciones con excepción de un solo caso: SAG, San Agustín, en la corrida 1. En ese caso el ERA es de 19.68%, muy cercana a 20%.

❖ Las medianas de los ERA para todas las estaciones en todas las corridas, inicia con valores superiores a 10 y menores a 15%.

❖ Las medianas de los ERA para todas las estaciones en todas las corridas, es siempre menor a 30%, por lo tanto para cada estación, al menos la mitad de todas las mediciones del ERA fueron menores de 30%. Esto es cierto para las tres corridas.

❖ Para la mayoría de las estaciones el ERA máximo fue mayor a 30%, por lo tanto se puede afirmar que hubo muy pocas estaciones, tratadas como estaciones de prueba, en las que todos sus ERA registrados fueron menores de 30% para las tres corridas.

En las tres corridas, para los ERA máximos, se destaca un pico que sobresale sobre todos los demás, ese pico corresponde a la estación ACO, Acolman. Se le llamó “Caso especial ACO (Acolman)”.

Según los resultados obtenidos de las tres corridas, las estaciones de monitoreo se agruparon en tres grupos con respecto a los resultados de sus  $\overline{\overline{\text{ERA}}}$ :

1.- Al primer grupo pertenecen las estaciones de prueba que obtuvieron un  $\overline{\overline{\text{ERA}}}$  entre 14.64% y 20%. Grupo formado por 17 estaciones de monitoreo. A este grupo pertenecen la mayoría de las estaciones con un  $\overline{\overline{\text{ERA}}}$  menor del 20%. Esto quiere decir que sí se cumple el objetivo principal de este trabajo, ya que las estaciones de este grupo obtuvieron los perfiles diarios de la concentración de ozono (antes desconocidos) con un muy buen ajuste, a partir de los datos conocidos.

2.- Al segundo grupo pertenecen las estaciones de prueba que obtuvieron un  $\overline{\overline{\text{ERA}}}$  entre 20% y 25%. Grupo formado por 9 estaciones de monitoreo. En este grupo tienen un  $\overline{\overline{\text{ERA}}}$  mayor del 20% y menor del 25%, se puede decir que estas estaciones obtuvieron los perfiles diarios de la concentración de ozono (antes desconocidos) con

un buen ajuste, a partir de los datos conocidos. Dentro de este grupo destaca el caso especial ACO (Acolman). La estación ACO sólo en el mes de diciembre presenta un ERA exagerado. Sin embargo el resto del año tiene ERA considerablemente menor, que sino considerara el mes de diciembre la estación ACO pertenecería al primer grupo.

En el caso de la estación ACO, la razón del incremento del ERA es porque en el mes de diciembre se detectan mediciones bajas de concentraciones de ozono de la estación ACO, y la estimación de la red sale por encima de la realidad, es decir la red hace la estimación alta de la concentración de ozono. Otro factor que también puede influir es la falta de datos periféricos de la estación ACO.

3.- Al tercer grupo pertenecen las estaciones de prueba que obtuvieron un  $\overline{\overline{\text{ERA}}}$  entre 25% y 31%, y está formado por seis estaciones: AJM, MPA, SFE, AJU, CUA y INN. En este grupo tienen un  $\overline{\overline{\text{ERA}}}$  mayor del 25%, se puede decir que las estas estaciones obtuvieron los perfiles diarios de la concentración de ozono (antes desconocidos) con un ajuste regular, a partir de los datos conocidos. Se observó que son seis estaciones de este grupo y tiene los  $\overline{\overline{\text{ERA}}}$  máximos. Sin embargo existe una coincidencia con dichas estaciones, ya que son las mismas seis estaciones que tienen mayor altura. Por esta razón existe la posibilidad de que el parámetro Altura tenga relación con el incremento del  $\overline{\overline{\text{ERA}}}$ . También puede influir la falta de datos periféricos.

Se desarrollaron los programas específicamente para este trabajo, y los programas se desarrollaron conforme se fueron requiriendo, checar sección **4.5.5**.

Como trabajo futuro se pueden incluir los parámetros meteorológicos como: velocidad del viento, dirección del viento, temperatura del aire, humedad y radiación solar; y parámetro de altitud (altura).

## Bibliografía

- Al, S., Abdul, S., & Bakheit, C. (2008). Combining principal component regression and Artificial Neural Networks for more accurate predictions of ground-level Ozone. *Elsevier Science. Environmental Modelling & Software*, 23 , 396-403.
- Andrew, C. (1997). Comparing Neural Networks and Regression Models for Ozone Forecasting. *Journal of Air & Waste Management Association*, 47:6, 653-663, DOI:10.1080/10473289.1997.10463925.
- Bravo, H., Perrin, F., Sosa, R., & Torres, R. (1988). Importancia de la contaminación atmosférica por ozono en la zona metropolitana de la Ciudad de México. *Revista Ciecias UNAM, Centro de Ciencias de la Atmosfera.*, 36-39. Importancia de la contaminación atmosférica por ozono en la zona metropolitana de la ciudad de México. Obtenido de: <http://www.revistaciencias.unam.mx/images/stories/Articles/12/CNS01205.pdf>
- Chen, J., Chen, H., Wu, Z., Hu, D., & Pan, J. (2017). Forecasting smog-related health hazard based on social media and physical sensor. *Elsevier Science. Information Systems*, 64, 281-291.
- Chien, Y., & Bin, L. (2002). A Backpropagation Algorithm with Adaptive Learning Rate and Momentum coefficient. *IEEE*, 1218-1223.
- Feng, Y., Zhang, W., Sun, D., & Zhang, L. (2011). Ozone concentration forecast method based on Genetic Algorithm optimized Backpropagation Neural Network and Support Vector Machine data classification. *Elsevier Science. Atmospheric Environment*, 45, 1979-1985.
- Fontes, T., Silva, L., Silva, M., Barros, N., & Carvalho, A. (2014). Can Artificial Neural Networks be used to predict the origin of Ozone episodes? *Elsevier Science. Science of the Total Environment*, 488-489, 197-207.

- Freeman, J., & Skapura, D. (1993). *Redes Neuronales: Algoritmos, aplicaciones y técnicas de programación*. Wilmington, Delaware, E.U.A.: Addison-Wesley Iberoamericana.
- Gardner, M., & Dorling, S. (1998). Artificial Neural Networks (the multilayer perceptron) - a review of applications in the Atmospheric Sciences. *Elsevier Science. Atmospheric Environment*, 32 (14/15), 2627-2636.
- Gurney, K. (2003). *An Introduction to Neural Networks*. London: CRC Press.
- Haykin, S. (2009). *Neural Networks and Learning Machines*. Upper Saddle River, New Jersey. (3a Ed.): Prentice Hall.
- Henseler, J. (1995). Back Propagation. En P. Braspenning, F. Thuijsman, & A. Weijters(Eds.), *Artificial Neural Networks: An Introduction to ANN Theory and Practice* (págs. 35- 66). Berlin Heidelberg New York: Springer.
- NASA. (21 de Abril de 2018). *Earth Fact Sheet*. Obtenido de <https://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html>
- Nunes, I., Hernane, D., Andrade, R., Bartocci, L., & Dos Reis, S. (2017). *Artificial Neural Networks: A practical course*. Switzerland: Springer.
- Panda, M., & Patra, M. (2013). *Soft Computing: Concepts and Techniques*. New Delhi, Boston: University science Press.
- Pedrycz, W. (1998). *Computational Intelligence: An Introduction*. Winnipeg, Canada: CRC Press.
- Ruiz, J., Mayora, O., Torres, J., & Ruiz, L. (1995). Short-term Ozone forecasting by Artificial Neural Networks. *Elsevier Science. Advances in Engineering Software*, 23, 143-149.
- Salazar , E., Ordieres, J., Vergara, E., & Capuz, S. (2008). Development and comparative analysis of tropospheric Ozone prediction models using linear and Artificial intelligence - based models in Mexicali, Baja California (México) and Calexico, California (US). *Elsevier Science. Environment Modelling & Software*, 23, 1056-1069.

SEDEMA. (20 de Diciembre de 2016). *Calidad del aire*. Obtenido de Página Oficial:  
<https://www.sedema.cdmx.gob.mx/>

SEDEMA. (16 de Octubre de 2017). *Calidad del aire*. Ubicación de las estaciones de monitoreo. Obtenido de:  
<http://www.aire.cdmx.gob.mx/default.php?opc=%27ZaBhnml=&dc=%27ZA==>

SEDEMA. (23 de Enero de 2018). *Calidad del aire*. Bases de Datos de concentración de ozono de Red RAMA. Obtenido de:  
<http://www.aire.cdmx.gob.mx/default.php?opc=%27aKBh%27>

Spellman, G. (1999). An Application of Artificial Neural Networks to the prediction of surface ozono concentrations in the United Kingdom. *Elsevier Science. Applied Geography, 19*, 123-136.

Wang, D., & Lu, W.-Z. (2006). Ground - level Ozone prediction using Multilayer Perceptron trained with an innovative hybrid approach. *Elsevier Science. Ecological Modelling, 1 (98)*, 332-340.

# Apéndice A

## Las coordenadas usadas en este trabajo para ubicar puntos en la Zona Metropolitana del Valle de México

*Las coordenadas de las estaciones de monitoreo tomadas del portal de la SEDEMA, estaban medidas en grados.*

En el portal de la SEDEMA, se muestra las coordenadas y la altura sobre el nivel del mar de cada una de las estaciones de monitoreo (SEDEMA, 2017). Las coordenadas vienen dadas en grados, tanto para la longitud como para la latitud. Para medir la longitud, se toma como referencia el meridiano de Greenwich; este meridiano está en el grado 0, al ángulo que gira en la dirección este (es decir, gira en sentido contrario a las manecillas del reloj) hasta llegar al meridiano opuesto ( $180^\circ$ ), se le asigna el signo positivo; el ángulo que gira en la dirección oeste (es decir, que gira en el sentido de las manecillas del reloj) hasta llegar al meridiano opuesto, se le asigna el signo negativo. Para medir la latitud, se toma como grado 0 el ecuador, para el hemisferio norte, al ángulo que gira desde el ecuador hasta el polo norte (de  $0$  a  $90^\circ$ ), se le asigna el signo positivo.

### El punto de referencia fijado en este trabajo

En este trabajo se tomó como punto de referencia la longitud de la estación de monitoreo más hacia el oeste (INN, Investigaciones Nucleares, longitud:  $-99.38052^\circ$ , latitud:  $19.291968^\circ$ ) y la latitud de la estación de monitoreo más hacia el sur (AJU, Ajusco, longitud:  $-99.162459^\circ$ , latitud:  $19.154674^\circ$ ), es decir, el punto: longitud:  $-99.38052^\circ$ , latitud:  $19.154674^\circ$  se tomó como referencia, asignándole las coordenadas (0, 0).

### Representando las coordenadas de las estaciones de monitoreo en un plano cartesiano.

Por lo general, la gran mayoría de las personas, están más acostumbradas a medir las distancias de un punto a otro de la superficie terrestre, en kilómetros u otra

unidad de distancia, en lugar de hacerlo en grados, por eso se decidió en este trabajo ubicar las estaciones en un plano cartesiano cuyos ejes están acotados en kilómetros. Por otro lado, el haber escogido el origen en las coordenadas mencionadas en el párrafo anterior, hace que todas las coordenadas de las estaciones de monitoreo queden ubicadas en el cuadrante positivo de los ejes cartesianos (así, todas las coordenadas de las estaciones quedan representadas con números positivos). Para trabajar con distancias en lugar de trabajar con grados, los ángulos reportados de las coordenadas (longitud y latitud), se convirtieron en arcos sobre la superficie terrestre (suponiendo que la tierra es esférica y tiene un radio de 6,371 km (NASA, 2018)). Los arcos de las longitudes medidos desde la longitud del origen fijado hasta las longitudes de las ubicaciones de las estaciones, se tomaron como las abscisas de las estaciones en el plano cartesiano; los arcos de las latitudes medidos desde la latitud del origen fijado hasta las latitudes de las estaciones se tomaron como las ordenadas de las estaciones en dicho plano.

#### **Cálculo del error en el que se incurrió por haber representado puntos sobre la superficie de una esfera en un plano.**

Para estimar el error resultante de haber representado las ubicaciones de las estaciones de monitoreo expresadas en coordenadas esféricas en un plano cartesiano, se calculará el error máximo en el que se incurrió por tal hecho.

De las estaciones de monitoreo cuyos datos se usaron para este trabajo, las más alejadas entre sí son la estación Cuautitlán (CUT) y la estación Milpa Alta (MPA). En la tabla **A.1** se muestran sus ubicaciones junto con la del punto de origen fijado en este trabajo.

Se calculará el error en el que se incurre usando las coordenadas de estas dos estaciones, ya que ese es el error máximo. En la tabla **A.2** se muestran las coordenadas de esas dos estaciones tomando como longitud de referencia la del punto de origen. En la tabla **A.2b** se escribe la misma información que la reportada en la tabla **A.2**, pero con los ángulos medidos en radianes.

**Tabla A.1.** Coordenadas de las estaciones más alejadas entre sí, cuyos datos se utilizaron en este trabajo, y del punto que se eligió como origen también para este trabajo. Nótese que las longitudes están medidas con respecto del meridiano de Greenwich.

	CUT	MPA	Punto del origen fijado en este trabajo
Longitud (grados) -con respecto al meridiano de Greenwich-	-99.198602	-98.990189	-99.38052
Latitud (grados)	19.722186	19.1769	19.154674

**Tabla A.2.** Coordenadas de las estaciones más alejadas entre sí, y del punto que se eligió como origen; las longitudes y las latitudes están medidas con respecto al punto de origen.

	CUT	MPA	Punto del origen fijado en este trabajo
Longitud (grados) -con respecto al punto de origen fijado-	0.181918	0.390331	0
Latitud (grados)	0.567512	0.022226	0

**Tabla A.2b.** Coordenadas de las estaciones más alejadas entre sí, y del punto que se eligió como origen; las longitudes y las latitudes están medidas con respecto al punto de origen. Obsérvese que aquí los ángulos están medidos en radianes.

	CUT	MPA	Punto del origen fijado en este trabajo
Longitud (radianes) -con respecto al punto de origen fijado-	0.00317507	0.00681256	0
Latitud (radianes)	0.00990495	0.00038792	0

En la Tabla **A.3** se muestran las coordenadas de las estaciones más alejadas entre sí, y del punto que se eligió como origen; las longitudes y las latitudes están medidas con respecto al punto de origen. Para obtener las coordenadas que se muestran en la tabla **A.3**, se usaron las ecuaciones **(A.1)** y **(A.2)**.

$$x_1 = r \lambda_o \quad (\text{A.1})$$

$$x_2 = r \varphi_o \quad (\text{A.2})$$

Donde

$\lambda_o$  es la longitud con respecto al punto fijado en este trabajo (ver tabla **A.2b**),

$\varphi_o$  es la latitud con respecto al punto fijado en este trabajo (ver tabla **A.2b**),

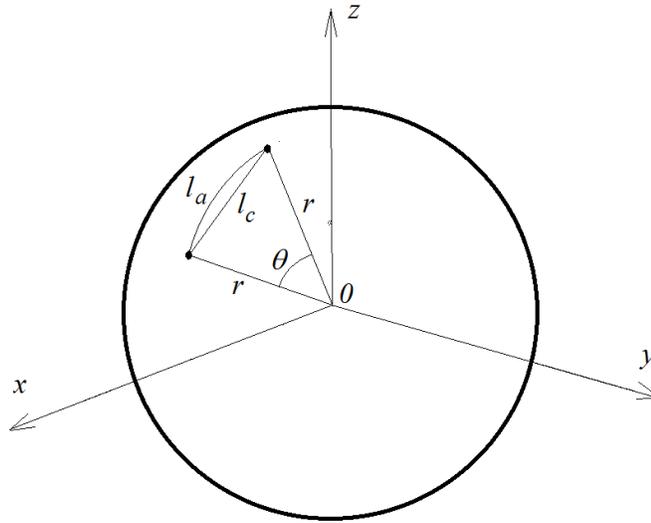
$r$  es el radio terrestre promedio, 6,371 km.

Nótese que al calcular así cada una de las coordenadas, se está suponiendo que la longitud del arco en la esfera es muy similar a la longitud de la línea que es su proyección en el plano (la cuerda).

**Tabla A.3** Coordenadas de los dos puntos más alejados entre sí, usados en este trabajo, representadas en un plano cartesiano.

	CUT	MPA	Punto del origen fijado en este trabajo
Abscisa, $x_1$ (km) -con respecto al punto de origen fijado-	20.2283587	43.402827	0
Ordenada, $x_2$ (km) -con respecto al punto de origen fijado-	63.1044552	2.4714184	0

Comparemos la longitud de una cuerda que une dos puntos ubicados en la superficie de una esfera, con la longitud del arco que los une pasando en esa superficie (ver figura **A.1**).



**Figura A.1.** Se muestra un arco que une dos puntos en la superficie de una esfera, su longitud es  $l_a$ , también se muestra una cuerda que pasa a través de la esfera y une esos mismos dos puntos, su longitud es  $l_c$ ;  $r$  es el radio terrestre.

Se transformarán las coordenadas esféricas de cada uno de los puntos en coordenadas cartesianas, para ello se usarán las ecuaciones (A.3), (A.4) y (A.5).

$$x = r \cos(\lambda) \cos(\varphi) \quad (\text{A.3})$$

$$y = r \cos(\lambda) \text{sen}(\varphi) \quad (\text{A.4})$$

$$z = r \text{sen}(\lambda) \quad (\text{A.5})$$

Donde

$\lambda$  es la longitud del punto en cuestión con respecto al meridiano de Greenwich medida en radianes,

$\varphi$  es la latitud del punto en cuestión medida en radianes,

$r$  es el radio terrestre promedio, 6,371 km,

$x, y, z$  son las coordenadas cartesianas del punto en cuestión.

Aplicando las ecuaciones (A.3), (A.4) y (A.5) a las coordenadas esféricas de las estaciones CUT y MPA, se obtienen sus coordenadas cartesianas (ver tabla A.4).

**Tabla A.4.** Coordenadas cartesianas de las estaciones más alejadas entre sí.

	Coordenadas cartesianas (x, y, z)
CUT	(-958.70729, -5920.1528, 2149.95631)
MPA	(-940.32137, -5943.5422, 2092.78351)

La distancia euclídea entre esos dos puntos en la esfera es la longitud del segmento de recta que los une, es decir la longitud de la cuerda  $l_c$  (ver figura A.1). El valor de  $l_c$  se puede calcular con la ecuación (A.4).

$$l_c = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (\text{A.6})$$

Donde los subíndices 2 y 1 se refieren a las coordenadas de los dos puntos.

Por otro lado, para calcular la longitud del arco  $l_a$ , se puede proceder de la siguiente manera: calcular el ángulo  $\theta$  en radianes y multiplicarlo por el radio terrestre  $r$ .

$$\theta = 2 \operatorname{sen}^{-1} \left( \frac{l_c}{2r} \right) \quad (\text{A.7})$$

$$l_a = r\theta \quad (\text{A.8})$$

Defínase el error relativo con la ecuación (A.7):

$$\operatorname{err}_{rel} = \frac{l_a - l_c}{0.5(l_a + l_c)} \quad (\text{A.9})$$

Aplicando la ecuación (A.6) y las ecuaciones (A.7) y (A.8) se obtienen los valores de  $l_c$  y  $l_a$ ; luego se puede aplicar la ecuación (A.9) para obtener el error relativo. En la tabla A.5 se muestran los resultados.

**Tabla A.5.** Longitudes del arco, de la cuerda y del error relativo. El error relativo porcentual,  $\% err_{rel}$ , es el error relativo multiplicado por cien.

$l_c =$	64.4502557
$l_a =$	64.45053052
$err_{rel} =$	$4.2641 \times 10^{-6}$
$\% err_{rel} =$	$4.2641 \times 10^{-4}$

Entonces, el máximo error relativo porcentual en el que se incurre por suponer que la longitud del arco que une dos puntos es igual a la de la cuerda que une esos dos mismos puntos, es de  $4.2641 \times 10^{-4}$ , es decir, un poco más de cuatro diezmilésimas de un punto porcentual, el cual es un *error despreciable*.

**Nota:** en todos los cálculos para obtener las tablas de este apéndice, se usaron 17 cifras decimales, en las tablas solo se muestra una parte de ellas.

## Apéndice B

### Ejemplos de los ERA máximos y mínimos de la estación

#### Acolman

En esta sección se muestra tres ejemplos de los ERA máximos de la estación ACO, es decir un ejemplo del peor ajuste de la estación ACO para cada corrida, con la finalidad de analizar el caso de la estación ACO. También se muestra un ejemplo con el ERA mínimo de la estación ACO, es decir de las tres corridas se seleccionó un ejemplo con el *mejor ajuste* de la estación ACO.

De la tabla 4.17, se selecciona un ejemplo de cada corrida que tenga el máximo ERA de la estación ACO (donde ACO es estación de prueba en todos los casos).

**Tabla 4.17** Errores relativos de ajuste máximos de la estación ACO, Acolman, (ACO como estación de Prueba)

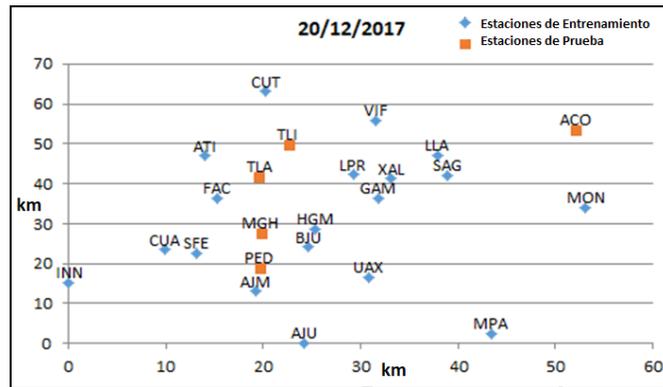
Errores Relativos de Ajuste de la Estación ACO, Acolman							
Corrida 1			Corrida 2			Corrida 3	
Fecha	ERA (%)		Fecha	ERA (%)		Fecha	ERA (%)
22/12/2017	85.29		23/12/2017	69.96		20/12/2017	97.12
20/12/2017	100.15		14/12/2017	103.00		21/12/2017	102.03

Para la corrida 1, en la figura B.1 se muestra el ejemplo del *peor ajuste* de la estación ACO, dicho ejemplo corresponde al 20 de Diciembre de 2017; en la figura B.1a se muestra la ubicación de las estaciones de Entrenamiento y las de Prueba; en la figura B.1b muestra el ERA de la estación ACO de **100.15%**, e indica 19 estaciones de entrenamiento, 5 estaciones de prueba, y un total de 24 *estaciones utilizables*, además muestra el identificador de las estaciones de entrenamiento (LLA, MON, SAG y VIF) considerando solo estaciones vecinas más cercanas a la estación ACO.

En la figura B.1c se muestra la gráfica de los resultados de ajuste de la estación ACO, en la cual se puede apreciar que tiene un error de ajuste bastante grande. Por

último, en las siguientes subfiguras se puede apreciar los resultados de ajuste de las estaciones de entrenamiento vecinas cercanas a la estación ACO, en la figura **B.1d** estación LLA (Los Laureles), en la figura **B.1e** estación MON (Montecillo), en la figura **B.1f** estación SAG (San Agustín), y en la figura **B.1g** estación VIF (Villa de las Flores).

(a) Ubicación de las estaciones de Entrenamiento y Prueba

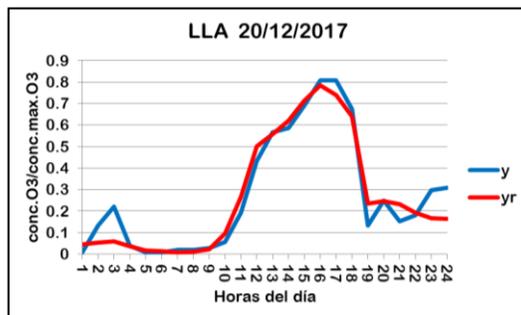


ERA de Est. ACO:	<b>100.15 %</b>
Núm. de Est. Totales :	24
Núm. de Est. de Prueba :	5
Núm. de Est. de Entrenamiento :	19
Est. vecinas más cercanas:	LLA, MON, SAG, VIF,

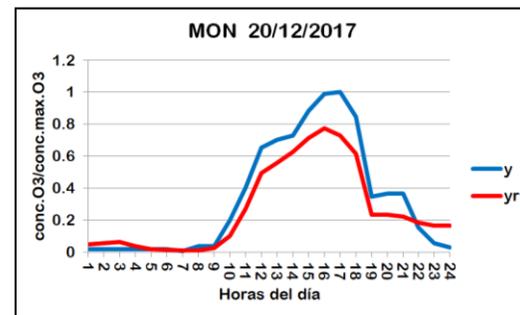
(b) Muestra el ERA de ACO



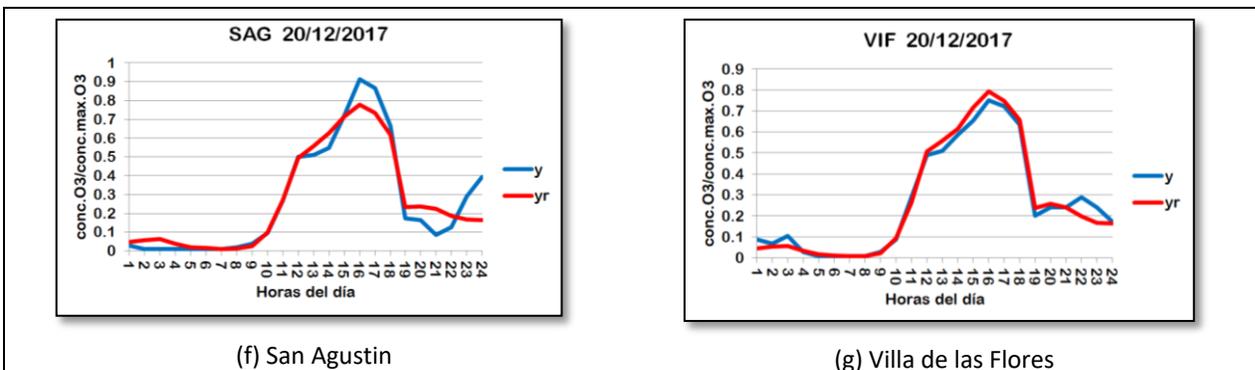
(c) Acolman



(d) Los Laureles



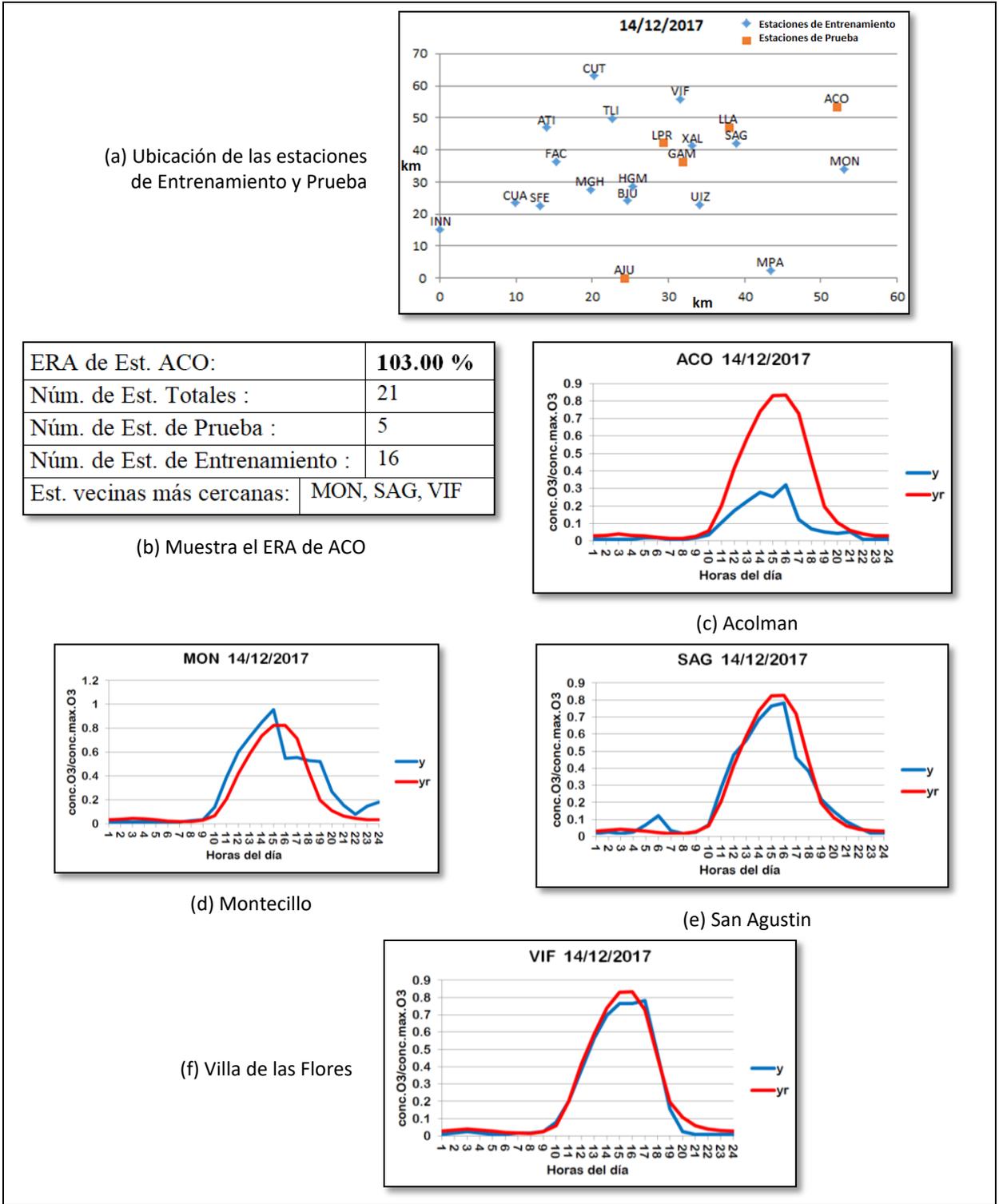
(e) Montecillo



**Figura B.1.** Ejemplo del *peor ajuste* de la estación ACO obtenido de la corrida 1, correspondiente al 20 de Diciembre de 2017. (a) Ubicación de las estaciones de Entrenamiento y Prueba; (b) Muestra el ERA de la estación ACO y otros datos; (c) Resultado del Ajuste de la estación de prueba ACO; En (d), (e), (f) y (g) se muestran los resultados de los ajustes de las estaciones de entrenamiento, solo estaciones cercanas a ACO.

Para la corrida 2, en la figura **B.2** se muestra el ejemplo del *peor ajuste* de la estación ACO, dicho ejemplo corresponde al 14 de Diciembre de 2017; en la figura **B.2a** se muestra la ubicación de las estaciones de Entrenamiento y las de Prueba; en la figura **B.2b** muestra el ERA de la estación ACO de **103.00%**, e indica 16 estaciones de entrenamiento, 5 estaciones de prueba, y un total de 21 *estaciones utilizables*. Además muestra el identificador de las estaciones de entrenamiento (MON, SAG y VIF) considerando solo estaciones vecinas más cercanas a la estación ACO.

En la figura **B.2c** se muestra la gráfica de los resultados de ajuste de la estación ACO, en la cual se puede apreciar que tiene un error de ajuste bastante grande. Por último, en las siguientes subfiguras se puede apreciar los resultados de ajuste de las estaciones de entrenamiento vecinas cercanas a la estación ACO, en la figura **B.2d** estación MON (Montecillo), en la figura **B.2e** estación SAG (San Agustín), y en la figura **B.2f** estación VIF (Villa de las Flores).

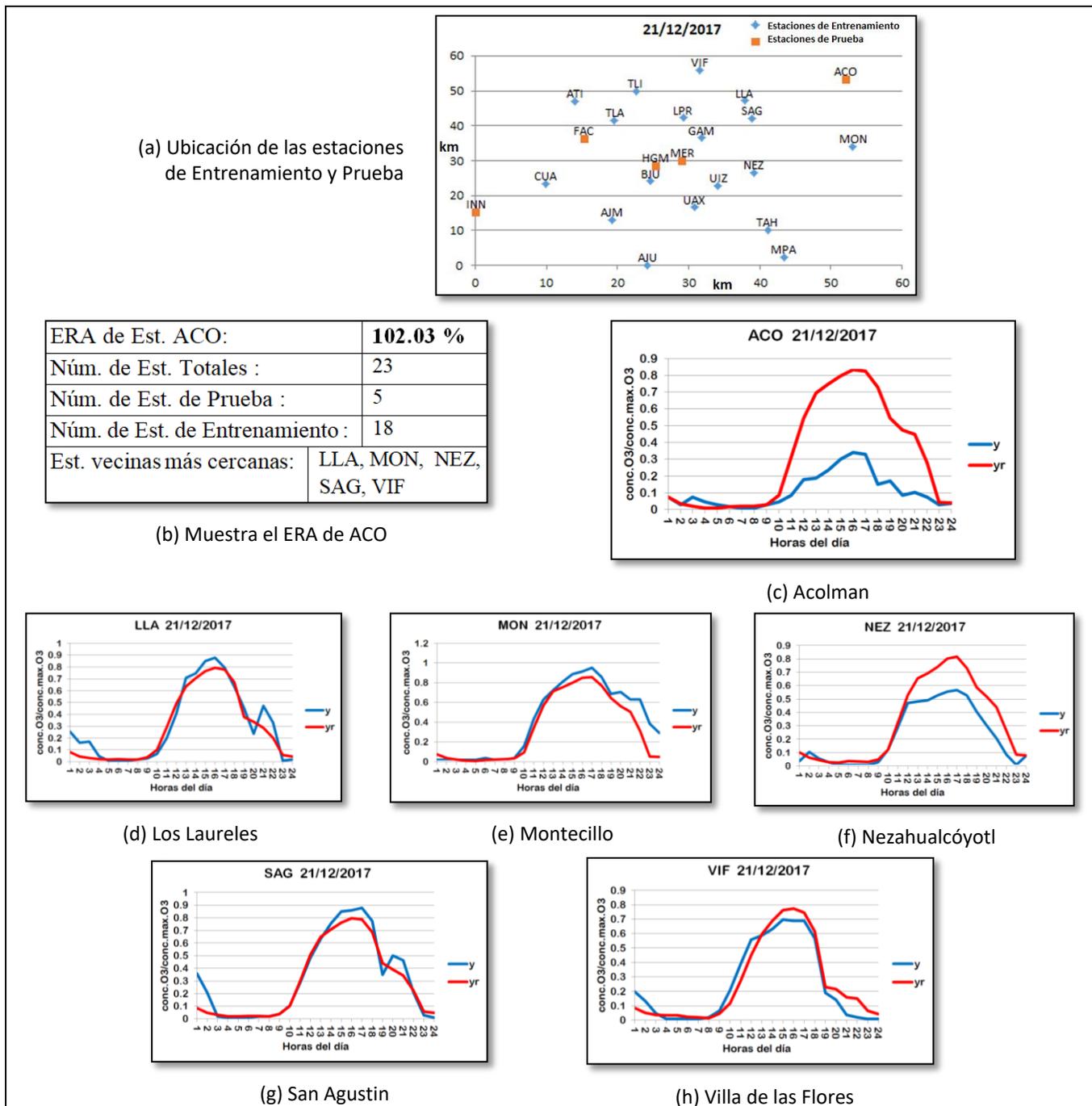


**Figura B.2.** Ejemplo del *peor ajuste* de la estación ACO obtenido de la corrida 2, correspondiente al 14 de Diciembre de 2017. (a) Ubicación de las estaciones de Entrenamiento y Prueba; (b) Muestra el ERA de la estación ACO y otros datos; (c) Resultado del Ajuste de la estación de prueba ACO; En (d), (e) y (f) se

muestran los resultados de los ajustes de las estaciones de entrenamiento, solo estaciones cercanas a ACO.

Y para la corrida 3, en la figura **B.3** se muestra el ejemplo del *peor ajuste* de la estación ACO, dicho ejemplo corresponde al 21 de Diciembre de 2017; en la figura **B.3a** se muestra la ubicación de las estaciones de Entrenamiento y las de Prueba; en la figura **B.3b** muestra el ERA de la estación ACO de **102.03%**, e indica 18 estaciones de entrenamiento, 5 estaciones de prueba, y un total de *23 estaciones utilizables*. Además muestra el identificador de las estaciones de entrenamiento (LLA, MON, NEZ, SAG y VIF) considerando solo estaciones vecinas más cercanas a la estación ACO.

En la figura **B.3c** se muestra la gráfica de los resultados de ajuste de la estación ACO, en la cual se puede apreciar que tiene un error de ajuste bastante grande. Por último, en las siguientes subfiguras se puede apreciar los resultados de ajuste de las estaciones de entrenamiento vecinas cercanas a la estación ACO, en la figura **B.3d** estación LLA (Los Laureles), en la figura **B.3e** estación MON (Montecillo), en la figura **B.3f** estación NEZ (Nezahualcóyotl), en la figura **B.3g** estación SAG (San Agustín), y en la figura **B.3h** estación VIF (Villa de las Flores).



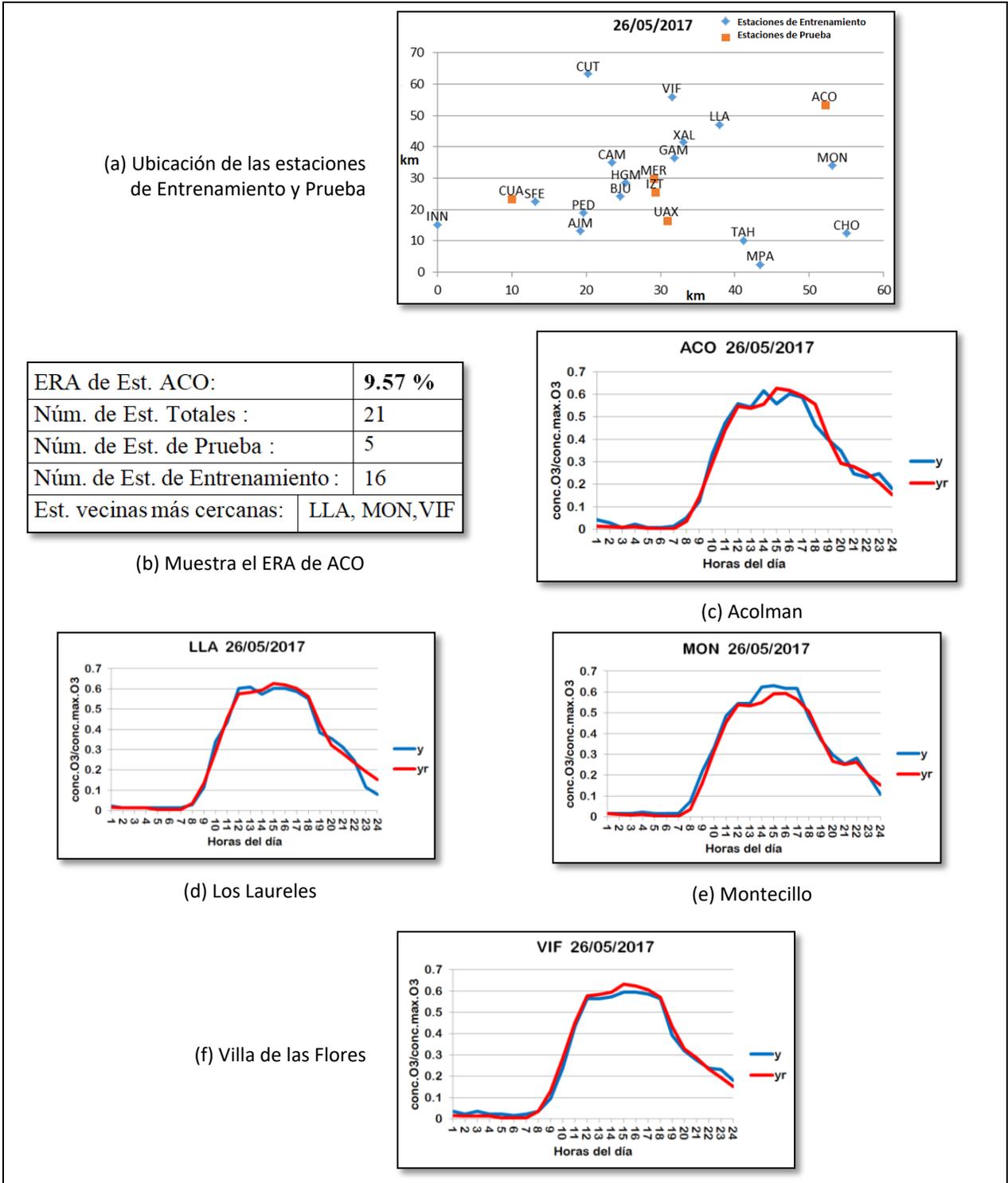
**Figura B.3.** Ejemplo del *peor ajuste* de la estación ACO obtenido de la corrida 3, correspondiente al 21 de Diciembre de 2017. (a) Ubicación de las estaciones de Entrenamiento y Prueba; (b) Muestra el ERA de la estación ACO y otros datos; (c) Resultado del Ajuste de la estación de prueba ACO; En (d), (e), (f), (g) y (h) se muestran los resultados de los ajustes de las estaciones de entrenamiento, solo estaciones cercanas a ACO.

Como se mencionó anteriormente, en los ejemplos de los *peores ajustes* de la estación ACO (correspondientes a las figuras **B.1**, **B.2** y **B.3**), se puede observar que en cada una de las corridas los ERAs máximos de la estación ACO son errores altos, pero la razón de este error alto no es debido a que incremente la concentración de ozono en esos días, sino que el error alto es debido a que disminuye la concentración de ozono en esos días (esto ocurre sólo en el mes de diciembre). Es decir, que el *perfil obtenido de la red* resultó con niveles altos de concentración de ozono, y que al compararlo con el *perfil real* que es el que obtuvo niveles muy bajos de concentración de ozono, se observa la gran diferencia entre los niveles de concentración lo que provoca que se tenga un mayor error.

Sin embargo se desconoce cuál es el factor o combinación de factores que hicieron que la estación ACO tuviera disminución de la concentración de ozono en las mediciones obtenidas en el mes de diciembre.

Hasta ahora se observaron los ejemplos con los máximos ERAs de la estación ACO, los peores ajustes, pero también es necesario analizar el día con el mínimo ERA de la estación ACO, es decir, el mejor ajuste. Para este caso, de las tres corridas se seleccionó sólo un ejemplo correspondiente al día con el mínimo ERA de la estación ACO. Por lo tanto en la figura **B.4** se muestra el ejemplo del *mejor ajuste* de la estación ACO, dicho ejemplo corresponde al 26 de Mayo de 2017; en la figura **B.4a** se muestra la ubicación de las estaciones de Entrenamiento y las de Prueba; en la figura **B.4b** muestra el ERA de la estación ACO de **9.57%**, e indica 16 estaciones de entrenamiento, 5 estaciones de prueba, y un total de 21 *estaciones utilizables*. Además muestra el identificador de las estaciones de entrenamiento (LLA, MON y VIF) considerando solo estaciones vecinas más cercanas a la estación ACO.

En la figura **B.4c** se muestra la gráfica de los resultados de ajuste de la estación ACO, en la cual se puede apreciar que tiene un error de ajuste muy bueno. Por último, en las siguientes subfiguras se puede apreciar los resultados de ajuste de las estaciones de entrenamiento vecinas cercanas a la estación ACO, en la figura **B.4d** estación LLA (Los Laureles), en la figura **B.4e** estación MON (Montecillo), y en la figura **B.4f** estación VIF (Villa de las Flores).



**Figura B.4.** Ejemplo del *mejor ajuste* de la estación ACO, correspondiente al 26 de Mayo de 2017. (a) Ubicación de las estaciones de Entrenamiento y Prueba; (b) Muestra el ERA de la estación ACO y otros datos; (c) Resultado del Ajuste de la estación de prueba ACO; En (d), (e) y (f) se muestran los resultados de los ajustes de las estaciones de entrenamiento, solo estaciones cercanas a ACO.

Según el ejemplo anterior del *mejor ajuste* de la estación ACO (figura B.4), es la que tuvo el mínimo ERA de las tres corridas. Con esto se puede observar que las estaciones vecinas cercanas a la estación ACO tienen más o menos la misma forma y perfil de concentración de ozono, es decir que tienen un comportamiento parecido; por esta razón en el mes de diciembre, que es cuando cambia el comportamiento de la estación ACO (es decir, se disminuyen los niveles de concentración de ozono), en ese caso la estación ACO no tiene el mismo comportamiento de sus vecinos (éstos tienen niveles altos de concentración de ozono), por eso se tiene un mayor ERA. Por lo tanto se observa que es en algunos días del mes de diciembre donde se obtienen los *peores ajustes* de la estación ACO.