

# Capítulo 1

## Introducción

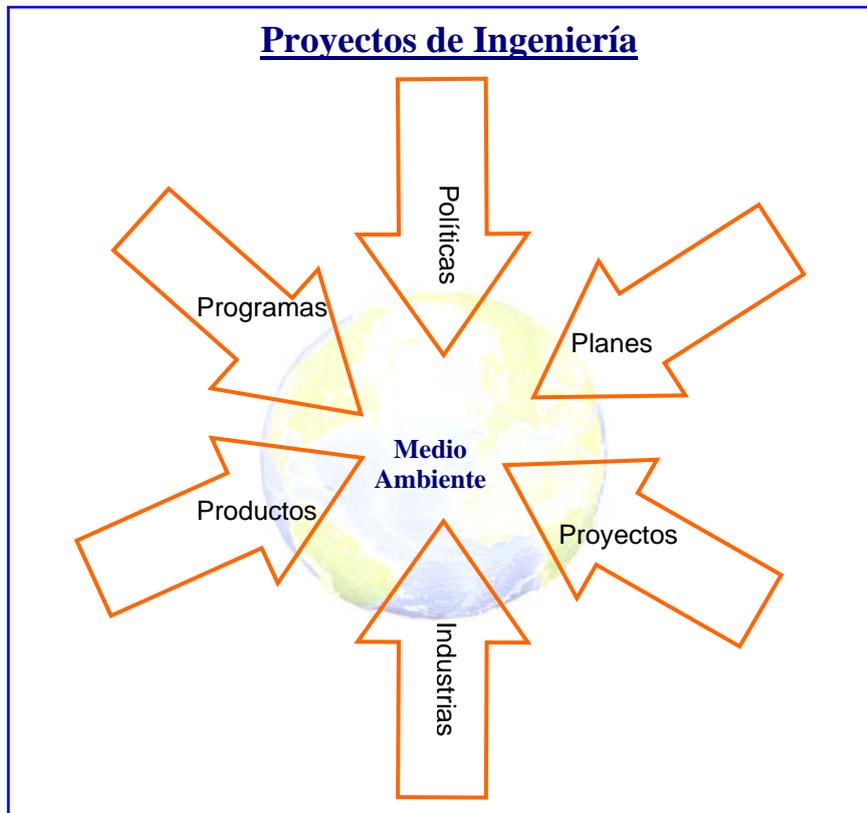
---

El desarrollo de esta tesis se enmarca en el convenio de colaboración para la impartición del *Programa de Doctorado de Proyectos de Ingeniería*, establecido entre la **Universidad Politécnica de Valencia (UPV)** en Valencia, España y las instituciones mexicanas **Dirección General de Educación Superior Tecnológica (DGEST)** y la **Secretaría de Educación Pública (SEP)** además de la colaboración financiera de los convenios entre la **Subsecretaría de Educación Superior (SES)** del gobierno de México y la **Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES)** de México. Por todo ello, el trabajo de investigación que aquí se presenta es producto de las coincidencias en los intereses de estas instituciones por mantener líneas de mejoramiento académico conjunto.

Es importante también destacar, de manera particular, el interés compartido entre el **Departamento de Proyectos de Ingeniería (DPI)**, donde se inscribe el programa de doctorado en “**Proyectos de Ingeniería e Innovación**” de la UPV, y el **Instituto Tecnológico de Mexicali** (institución perteneciente a la DGEST) en áreas de investigación coincidentes como las que se plantean en éste documento, ya que en el marco de sus actividades académicas, se propicia la

investigación en temáticas relevantes en el acontecer mundial y en especial por problemáticas tan preocupantes como la contaminación ambiental y la búsqueda de su mejor entendimiento mediante la utilización de una serie de metodologías de análisis de datos entre las que se incluyen herramientas de estadística avanzada y de inteligencia artificial.

Lógicamente, en la actualidad, es difícil concebir *proyectos de ingeniería* en cualquiera de sus líneas de trabajo (ejemplos: rediseño o desarrollo de nuevos productos, políticas, planes, programas, plantas productivas y proyectos en general) que no contemplen apartados importantes dedicados a aspectos de integración ecológica e impacto ambiental. Ver Figura 1.



**Figura 1.** Este esquema muestra las áreas generales que conforman las líneas de investigación en Proyectos de Ingeniería, donde todas tienen necesariamente espacios de trabajo coincidentes con la protección, respeto y mejoramiento ambiental.

Además, resulta indispensable considerar aspectos como: el desarrollo e implementación de políticas de control y mejoramiento ecológico para las industrias, el ecodiseño en el desarrollo de nuevos productos, los instrumentos de información, alerta y prevención de la contaminación ambiental, la formalización y normalización de sistemas de bancos de datos de contaminantes ambientales que favorezcan predicciones o proyecciones futuras de niveles de contaminación para su posible control, por mencionar sólo algunos aspectos.

Aunada a la relevancia de las herramientas de análisis utilizadas y a la problemática ambiental modelizada, otra característica considerada de interés en este trabajo tiene que ver con las zonas de estudio, localizadas éstas en la franja transfronteriza entre México y Estados Unidos.

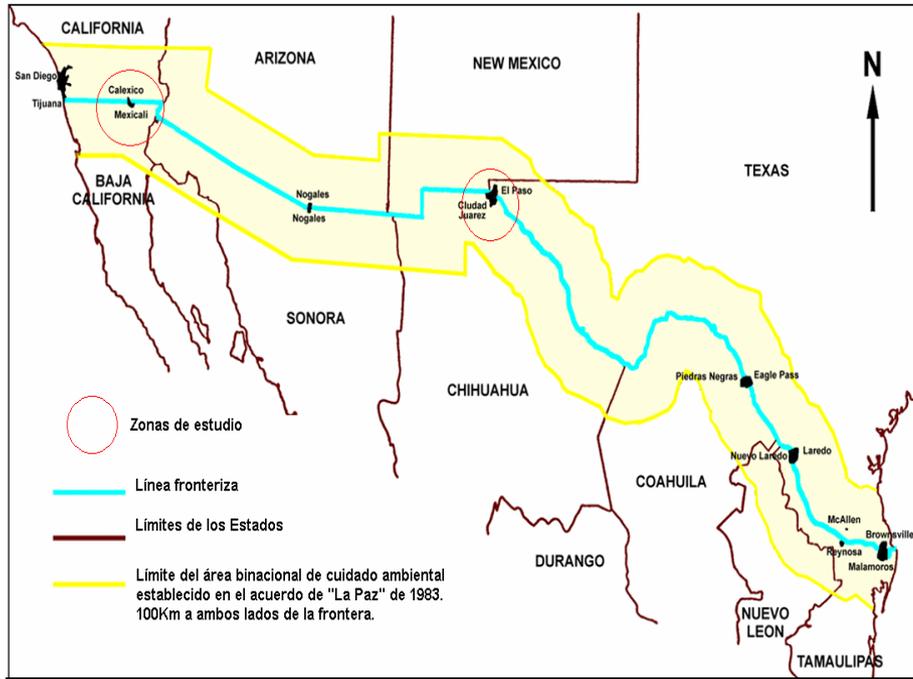
Generalmente las problemáticas ambientales en vecindades con distintos países exigen un carácter internacional, al ser habitual que los problemas ecológicos tengan consecuencias zonales amplias que difícilmente respetan fronteras naturales y mucho menos fronteras políticas. La situación de vecindad propicia, incluso exige el desarrollo y participación en tratados y esfuerzos conjuntos para frenar, delimitar y/o revertir daños ambientales.

Estados Unidos y México, comparten una larga franja divisoria de aproximadamente 3 153 kilómetros de longitud. Se tiene delimitada una área binacional de protección ambiental que abarca 100 kilómetros hacia dentro en ambos lados de la línea divisoria, de acuerdo a lo establecido en el *Tratado de La Paz* de 1983, tal como se muestra en la Figura 2.

Esta área binacional de protección ambiental presenta diversidad de características naturales, sociales, económicas y culturales que propician el desequilibrio ambiental para ambos países y hacen de ella un foco de atención en el diseño de políticas y programas gubernamentales binacionales de control en ese sentido.

Concretamente, se eligieron dos zonas transfronterizas con marcados problemas en la calidad del aire y consideradas por la *Agencia de Protección Ambiental de Estados Unidos* (US-EPA) y la *Secretaría de Medio Ambiente y Recursos Naturales de México* (SEMARNAT) como áreas que exceden los límites máximos en sus contaminantes, (US-EPA<sup>a</sup>, Green Book, 2007), lo que exige la atención de ambos países.

Una de ellas es la vecindad formada por Mexicali en Baja California, (México) y Calexico en California (Estados Unidos). La segunda zona estudiada está en la vecindad existente entre Ciudad Juárez, en Chihuahua (México) y El Paso en Texas (Estados Unidos). Ambas zonas están señaladas por un círculo en la Figura 2.



**Figura 2.** Mapa seccionado de México y Estados Unidos. Los círculos señalan las zonas fronterizas estudiadas. La franja entre las líneas superior e inferior delimita el área binacional de cuidado ambiental (100 km a cada lado de la línea fronteriza) establecida en el Convenio de La Paz de 1983. (Figura adaptada de un mapa proporcionado por la Agencia de Protección Ambiental, US-EPA).

En las zonas mencionadas, se trabajó en el desarrollo de modelos de predicción de contaminantes críticos en esas áreas. La modelización se ha aplicado exitosamente a diferentes objetivos, tales como la gestión de actividades administrativas, de comunicación, optimización de procesos y por supuesto para la predicción de *fenómenos no lineales* como es la aplicación principal que se busca en esta tesis.

Los modelos permiten representar de una manera práctica y manejable un gran número de fenómenos o comportamientos para ayudar a un mejor entendimiento y administración de los mismos, brindando un firme soporte para la toma de decisiones, y así manejar con cierto grado de certeza situaciones que gracias a su función como predictores, nos informan de manera anticipada la presentación de posibles eventos ya sean positivos, adversos, o simplemente informativos; desde un simple pronóstico de ventas, hasta delicadas predicciones en distintas áreas de la medicina, incluyendo también las predicciones de fenómenos atmosféricos que ayuden a alertar a la población y optar por acciones que eviten daños a nuestro entorno y seres vivos, siendo esto último uno de los fines del trabajo que aquí se presenta.

Las técnicas existentes para el desarrollo de modelos son numerosas. El grado de complicación para una adecuada modelización normalmente está relacionado con la naturaleza determinística o estocástica del fenómeno, así como la cantidad de parámetros requeridos para su correcta descripción, ya sea de un proceso o del fenómeno que se busca modelizar.

Como se ha comentado en párrafos anteriores, los trabajos de modelización de esta tesis se enfocan a problemáticas referentes a la calidad del aire. Estos fenómenos tienen como características esenciales que además de la aleatoriedad de las variables involucradas, se requiere de un gran número de ellas para la adecuada modelización del contaminante en particular.

Estas características de complejidad, combinadas con la relevancia del tema de la calidad del aire y protección medio ambiental propician el interés científico y la necesidad de continuar desarrollado esfuerzos al respecto, tratando de mejorar de manera continua tanto las metodologías, técnicas o herramientas de modelización, como en el desarrollo de modelos de otros contaminantes no menos importantes, que estas herramientas y modelos sean cada vez sean más robustos al momento de su utilización.

El campo de trabajo y lo que hay por hacer es también abundante, y ello motiva para continuar con investigaciones en esa línea, que seguramente formarán parte de publicaciones futuras.

Para el trabajo que atañe a esta tesis, se desarrollaron estudios extensos de distintas temáticas sobre calidad del aire, sus tipos de contaminantes, situación actual y características generales de las zonas de estudio que benefician la contaminación del aire. Además se estudiaron las técnicas existentes para la modelización, y en especial, la modelización de contaminantes atmosféricos, así como técnicas de evaluación del desempeño de los mismos.

Se desarrolló una búsqueda exhaustiva de los datos de contaminantes y factores meteorológicos para la aplicación de las técnicas de modelización seleccionadas, aprovechándose los resultados de los mismos para realizar un comparativo general de sus desempeños de acuerdo a las herramientas aplicadas. Lo anterior se detalla en los Capítulos 2, 3 4 y 5.

Como resultado de los estudios anteriores fue posible el desarrollo de ponencias internacionales como: *“Modellization of Atmospheric Pollution Using Artificial Intelligence Techniques”*, *“Comparación y análisis de técnicas estadísticas simples y de Inteligencia Artificial usadas en la modelización del ozono troposférico”*, *“Una revisión de las Experiencias en la Aplicación de la Minería de Datos en la Dirección de Proyectos”* para los VIII, X y XI Congresos Internacionales de Ingeniería de Proyectos, respectivamente. Además de la presentación de charlas relacionadas con el tema de tesis en eventos también de nivel internacional desarrollados en México, en ciudades como Chihuahua en el estado de Chihuahua, Mexicali y Tijuana, ambas ciudades del estado de Baja California.

Lo anterior, adicional a las publicaciones para la revista internacional *Environmental Modelling and Software* de los artículos *“Neural Network Prediction Model for Fine Particulate Matter (PM2.5) on the US-México Border in El Paso Texas and Ciudad Juárez (Chihuahua)”* y *“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)”*.

En el Capítulo 6, se expone un resumen de lo más relevante de los trabajos y publicaciones mencionadas en el párrafo anterior, elaborando como Capítulo 7, una discusión de resultados.

En el Capítulo 8 se presentan las conclusiones finales sobre los hallazgos durante el desarrollo de la investigación y que algunas de ellas dan pie para continuar trabajando en esa línea en futuras investigaciones como se detalla en el Capítulo 9.

Finalmente, el Capítulo 10 recopila la bibliografía y referencias más importantes empleadas durante la elaboración de esta tesis.



# Capítulo 2

## Marco de referencia

### 2.1 Fundamentos Teóricos. Calidad de Aire y modelos

#### 2.1.1 Calidad del Aire

Cuando se habla de una mala calidad del aire significa que están presentes en el aire sustancias que tienen efectos negativos sobre los seres vivos y el ecosistema. Hay que tomar en cuenta que de manera natural ya hay presencia en el aire de elementos nocivos, aunque estos son poco significativos para causar efectos adversos. En ese sentido se considera que el aire puro normalmente está compuesto aproximadamente de un 78% de nitrógeno, un 21 % de oxígeno y el 1% de gases como el dióxido de carbono, argón, neón, helio, entre otros, además de vapor de agua, ozono y partículas suspendidas en porcentajes sumamente bajos (Brimblecombe P.,1996). La relación entre el número de moléculas de estos gases con respecto al número total de moléculas de aire seco no varía sustancialmente y se puede considerar independiente de la situación geográfica y del tiempo (Jacobson, 1999), por lo que todo cambio de las características establecidas como *normales* del aire pueden también considerarse como contaminación.

## 2.1.2 Fuentes de contaminación

Se tiene asumido que los cambios en la calidad del aire pueden suceder con intervención o no del ser humano, por lo que las fuentes de generación de contaminantes pueden clasificarse como *fuentes antropogénicas* (originadas por el ser humano) o *fuentes biogénicas* (causadas por fenómenos naturales). Del primer tipo podemos encontrar ejemplos como la combustión causadas por el hombre, procesos industriales o cambios de uso de suelos, mientras que del segundo tipo causas claras son las erupciones volcánicas, los incendios provocados por los rayos, las tormentas de arena, la descomposición de plantas, o las concentraciones de desechos de animales de ganadería, por citar algunos casos.

En la literatura también existen diversas sub-clasificaciones. Por ejemplo, como *fuentes móviles* se definen aquellas que incluyen diversas formas de transporte como es el caso de automóviles, camiones y aviones. Las *fuentes fijas* o *estacionarias* son debidas a generadores de contaminantes no movibles, al menos por largos períodos de tiempo, como plantas de energía y parques industriales. Una *fente puntual* se refiere a una fuente en un punto fijo, tal como un incendio, una chimenea o un tanque de almacenamiento que emite contaminantes. Finalmente, una *fente de área* se refiere a una serie de fuentes pequeñas que en conjunto pueden afectar la calidad del aire en una región.

## 2.1.3 Contaminantes Criterio

Los contaminantes más comunes relacionados con la calidad del aire y la salud humana son identificados como “*Contaminantes Criterio*”, llamados así porque fueron objetos de estudios de evaluación de daños al ser vivo y al entorno, y publicados por diversos investigadores como “*Criterios de Calidad del Aire*”.

Así que, prácticamente desde su creación, en 1970, la *Agencia de Protección Ambiental de Estados Unidos* (US-EPA, por sus siglas en inglés), identificó y reguló seis contaminantes principales que interfieren con la

salud o el bienestar de las personas, o bien que producen efectos negativos (<http://www.epa.gov/air/urbanair/6poll.html>) sobre el medio ambiente y que han sido adoptados como tal por otros países, incluido México.

Los “*Contaminantes Criterio*” son:

1. Ozono (O<sub>3</sub>).
2. Material particulado (PM's)\*.
3. Dióxido de nitrógeno (NO<sub>2</sub>).
4. Monóxido de carbono (CO).
5. Dióxido de sulfuro (SO<sub>2</sub>).
6. Plomo (Pb).

\* también conocidas como Partículas Suspendingas o en suspensión

Cuatro de los seis contaminantes criterio nombrados anteriormente, son directamente producidos por la combustión fósil (carbón, gasolina, diesel, aceite o gas natural) o bien madera. Estos son: el NO<sub>2</sub>, (principal precursor de O<sub>3</sub>), CO, y las PM's.

Se establecen para los seis contaminantes dos tipos de regulaciones: “*límite primario*” que tiene el objetivo de proteger la salud humana y “*límite secundario*” cuyo fin es proteger daños medio ambientales y daños a propiedades como obras de arte expuestas al aire libre, monumentos y otras edificaciones. Estas regulaciones se han ido ajustando de acuerdo a resultados de nuevas investigaciones y los límites que actualmente se manejan se pueden revisar en la Tabla 1 que se muestra en la página 30 de este documento.

### **2.1.3.1 Ozono (O<sub>3</sub>)**

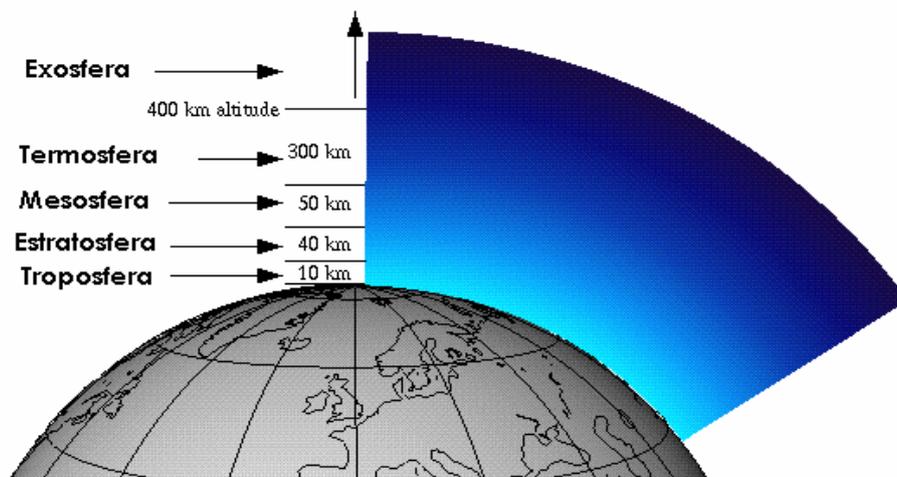
El ozono es un compuesto gaseoso natural de la atmósfera que se produce en presencia de luz solar. Oxida materiales no inmediatamente oxidables por el oxígeno gaseoso formado principalmente por reacciones atmosféricas de hidrocarburos y óxidos de nitrógeno bajo la influencia de la luz solar. De acuerdo a estudios del cambio climático, el ozono constituye el tercer gas efecto invernadero más importante después del dióxido de carbono y el metano. (IPCC, 2007).

## Capítulo 2

El 90% de su concentración se distribuye en la estratosfera, por lo que al ozono localizado en esa zona se le identifica como “*ozono estratosférico*” u “*ozono bueno*”.

El “*ozono estratosférico*” forma una capa situada entre los 25 y 35 Km. de altura, cuya función es filtrar la luz ultravioleta que llega a la tierra, por lo que esta función resulta beneficiosa al ser humano y a la naturaleza en general.

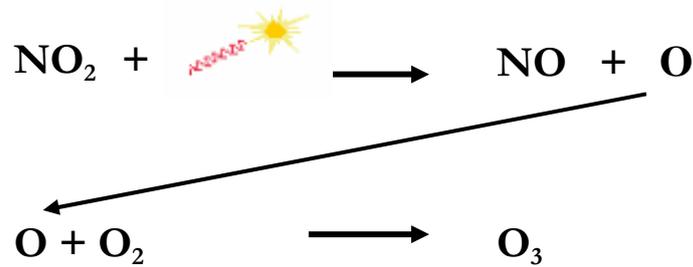
El 10% restante es el “*ozono troposférico*”, (también identificado como “*ozono malo*”) situado aproximadamente desde los 0 hasta los 10 Km. de altura con respecto a la tierra (Ver Figura 3). El estudio del “ozono malo” es el que conforma una parte importante en este trabajo de tesis.



**Figura 3.** Ubicación de las distintas capas de la atmósfera. En los primeros 10 Km. a partir de la superficie de la tierra se forma el *ozono troposférico*, considerado dañino. (Figura adaptada de: [www.eduspace.esa.int](http://www.eduspace.esa.int)).

Por su naturaleza reactiva principalmente a la radiación solar, es un contaminante generado esencialmente durante el día y con la presencia de corrientes de aire. Los impactos más importantes pueden ocurrir en sitios muy alejados (cientos de kilómetros) desde las fuentes de emisión de los contaminantes primarios que lo originaron, por esta razón con frecuencia se ven afectadas zonas rurales y suburbanas.

Para la formación del O<sub>3</sub>, el NO<sub>2</sub> absorbe radiación solar y se produce su fotodisociación en NO (monóxido de nitrógeno) y O (oxígeno). La molécula de oxígeno de la reacción se combina con una molécula de O<sub>2</sub> y forma O<sub>3</sub>. Este proceso se muestra en la Figura 4.



**Figura 4.** Proceso fotoquímico para la formación del ozono.

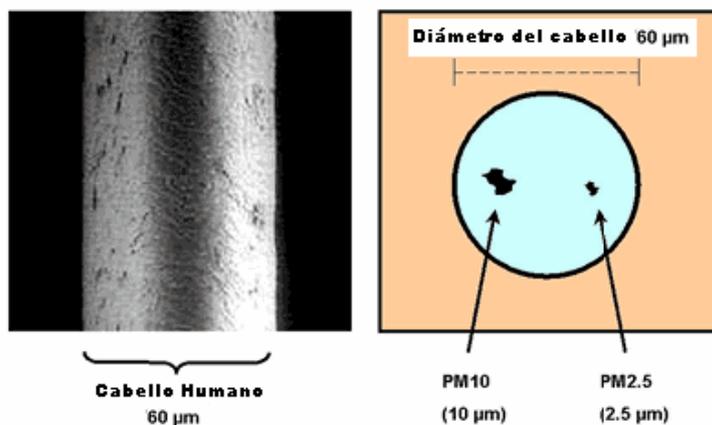
En ausencia de Compuestos Orgánicos Volátiles (COVs), el O<sub>3</sub> de la reacción actúa con el NO regenerando compuestos iniciales de NO<sub>2</sub> y O<sub>2</sub>. Cuando están presentes COVs, su oxidación da origen a radicales libres RO<sub>2</sub> que reaccionan con el NO para producir NO<sub>2</sub> sin el consumo de O<sub>3</sub>, lo que propicia su acumulación.

### 2.1.3.2 Partículas Suspendidas (PM's)

Son una mezcla de partículas sólidas y líquidas (vapor) que se encuentran en el aire emitidas directamente desde la fuente (*formación directa*) o bien son formadas por la condensación de contaminantes gaseosos (*formación indirecta*). Forman una compleja mezcla de partículas sólidas y vapores con sustancias orgánicas e inorgánicas en suspensión, incluyendo frecuentemente sustancias ácidas como sulfatos y nitratos. Todos estos compuestos en el aire tienen la capacidad de ser absorbidos vía nariz/boca y pasar por el tracto respiratorio, por lo que afectan a más personas que cualquier otro contaminante (Ver Figura 5).

Además es conocido que el potencial de daños a la salud a causa de las partículas suspendidas está directamente relacionado con el tamaño de éstas. Entre más pequeño sea su diámetro aerodinámico, más dañinas están consideradas, ya que penetran con mayor facilidad por el tracto respiratorio llegando a los pulmones, donde permanecen depositadas en los alvéolos pulmonares.

Existen estudios que demuestran relación entre el incremento de muertes o asistencias al médico por problemas de tipo respiratorio y cardiovascular cuando los niveles de partículas suspendidas son altos (USEPA, 2007; [www.epa.gov/asthma](http://www.epa.gov/asthma)). Además, la contaminación por partículas suspendidas es asociada con reducción de visibilidad (USEPA, 2007; [www.epa.gov/visibility](http://www.epa.gov/visibility)).



**Figura 5.** Referencia gráfica de las dimensiones de las partículas suspendidas respecto al diámetro de un cabello humano (Fuente: Air Resources Board de California).

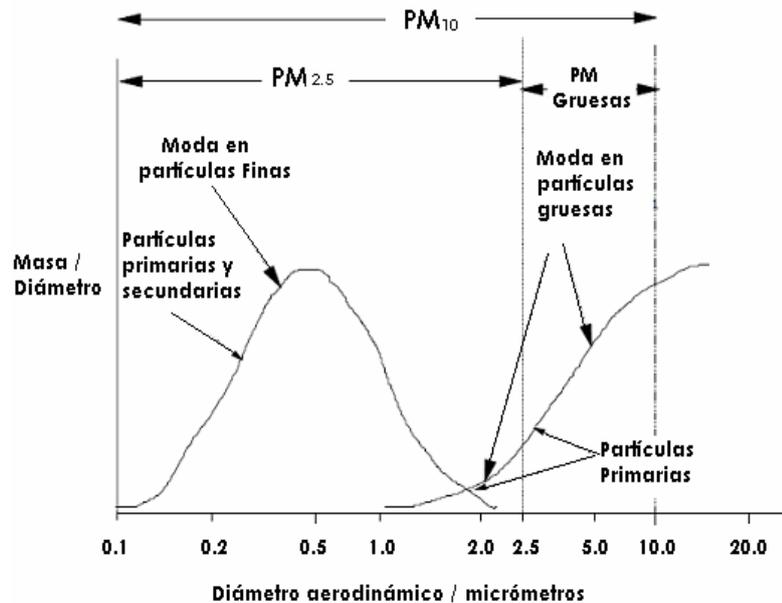
### 2.1.3.2.a Clasificación de las partículas en suspensión:

En primera instancia, las partículas suspendidas son reconocidas en función de su diámetro aerodinámico. La Figura 6 muestra gráficamente las características de las partículas de acuerdo a sus diámetros.

- $PM_{10}$  (diámetro  $< 10 \mu m$ ): *Partículas gruesas*. Permanecen menos tiempo en la atmósfera y pueden depositarse cerca de la fuente de

emisión debido a su peso, por lo que no viajan lejos de su fuente de emisión.

- $PM_{2.5}$  (diámetro  $< 2.5 \mu m$ ): *Partículas finas*. Ayudadas por corrientes de aire, pueden desplazarse grandes distancias y alejarse de la fuente de emisión y, también por su diámetro tan pequeño, tienen mayor facilidad permanecer suspendidas por tiempos mucho más largos que las *partículas gruesas* y mayor probabilidad de ser absorbidas por el ser humano durante la respiración.



**Figura 6.** Clasificación de las partículas suspendidas de acuerdo a su diámetro aerodinámico. (Figura adaptada de Wilson y Suh, 1997).

La EPA, en su documento “*Emission Inventory Guidance for Implementation of Ozone and Particulate Matter National Ambient Air Quality Standards and Regional Haze Regulations*” (USEPA, 2003a) establece las siguientes clasificaciones de las partículas suspendidas:

- *Partículas primarias*: Aquellas que entran a la atmósfera como una emisión directa desde las diferentes fuentes, las cuales a su vez pueden ser clasificables en *partículas filtrables* y *partículas condensables*. Las *partículas filtrables* tienen la característica que se

emiten directamente hacia la atmósfera en estado sólido o líquido y éstas pueden ser retenidas en un filtro específico para esa función, a diferencia de las *partículas condensables*, las cuales se emiten en estado gaseoso pero inmediatamente cambian a estado líquido o sólido una vez que son expulsadas a la atmósfera.

- *Partículas secundarias* tienen la peculiaridad de ser formadas en la atmósfera como consecuencia de las reacciones químicas y su posterior condensación o liquefacción. Esto normalmente sucede lejos de la fuente donde fueron emitidas.

### 2.1.3.3 Dióxido de nitrógeno (NO<sub>2</sub>)

Los óxidos de nitrógeno (NO<sub>x</sub>) generados por el hombre se producen por la oxidación de nitrógeno atmosférico presente en los procesos de combustión a altas temperaturas. El contaminante generado en forma primaria es el monóxido de nitrógeno (NO), parte del cual rápidamente se oxida y se convierte a dióxido de nitrógeno (NO<sub>2</sub>). Ambos óxidos, liberados a la atmósfera, participan activamente en un conjunto de reacciones fotoquímicas que en presencia de hidrocarburos reactivos, generan ozono (O<sub>3</sub>). Además, en su proceso de transformación, este contaminante forma nitratos, es decir, sales que pueden ser transportadas en el material particulado respirable y que en presencia de humedad, forman ácidos. Estos ácidos son una parte importante del material particulado secundario (PM<sub>2.5</sub>), que tiene efectos nocivos en la salud tal y como se ha especificado en párrafos anteriores.

Es imprescindible el control de ésta sustancia gaseosa, contribuye tanto a la formación de ozono como a la de material particulado fino. La principal fuente antropogénica de óxidos de nitrógeno, conocidos como NO<sub>x</sub> que incluye óxido nítrico (NO) y dióxido de nitrógeno (NO<sub>2</sub>), es el transporte automotor y cualquier otro tipo de combustión de hidrocarburos.

### 2.1.3.4 Monóxido de carbono (CO)

El monóxido de carbono es un gas inodoro, incoloro, inflamable y altamente tóxico. Es un producto de combustión incompleta y ocurre

cuando el carbono en el combustible se oxida parcialmente más que los óxidos del dióxido del carbono ( $\text{CO}_2$ ). Su aspiración reduce el flujo del oxígeno en el torrente sanguíneo y es particularmente peligroso para las personas con problemas de salud del corazón y vías respiratorias.

### **2.1.3.5 Dióxido de sulfuro ( $\text{SO}_2$ )**

El dióxido de azufre es el principal causante de la lluvia ácida que en la atmósfera es transformado en ácido sulfúrico. Es liberado en procesos de combustión que combustibles como el carbón, el petróleo, el diesel o el gas natural contienen compuestos azufrados. También los procesos metalúrgicos liberan ciertas cantidades de este gas, debido a que se emplean frecuentemente los metales en forma de sulfuros. En la naturaleza, el dióxido de azufre se encuentra en las proximidades de los volcanes y las erupciones pueden liberar cantidades importantes de este compuesto químico.

El dióxido de azufre es un gas irritante y tóxico. Afecta sobre todo las mucosidades y los pulmones provocando ataques de tos. Si bien éste es absorbido principalmente por el sistema nasal, la exposición de altas concentraciones por cortos períodos de tiempo puede irritar el tracto respiratorio, causar bronquitis y congestionar los conductos bronquiales de quienes sufren de problemas respiratorios (asma).

### **2.1.3.6 Plomo (Pb)**

El Plomo existe de forma natural en el ambiente generalmente en cantidades muy pequeñas, pero las concentraciones significativas son el resultado de las actividades humanas.

Por ejemplo, la utilización del plomo en gasolinas. Durante la combustión se generan sales de plomo (cloruros, bromuros, óxidos). Estas sales de plomo entran en el medio ambiente a través de los tubos de escape de los coches. Las partículas grandes se precipitan al suelo o en la superficie de las aguas, mientras que las partículas pequeñas viajarán largas distancias a través del aire y permanecerán más tiempo en la atmósfera. El plomo en suspensión caerá de nuevo sobre la tierra

## Capítulo 2

mediante la lluvia o la brisa. Este ciclo causado por la producción humana es mucho más extendido y dañino que el ciclo natural.

Otro ejemplo de la contribución humana a la contaminación por este metal, es gracias a la corrosión de las tuberías fabricadas con mezclas de materiales con plomo, frecuentemente utilizadas en los sistemas de drenaje. De esta manera, este contaminante también puede terminar en el agua y suelos.

Otras actividades humanas, como la combustión del petróleo, procesos industriales, corrosión de pinturas que contienen plomo, combustión de residuos sólidos, que también contribuyen.

El plomo es un elemento químico particularmente peligroso, y se puede acumular en organismos individuales, pero también entrar en las cadenas alimenticias causando graves problemas de salud (Krishnayya et al., 1986; Guidotti et al., 2007).

## 2.2 Regulaciones de contaminantes criterio

### 2.2.1 Normas Oficiales Mexicanas

Las Normas Oficiales Mexicanas tienen regulados los siguientes contaminantes, para evaluar la calidad del aire como medida de protección a la salud de la población (Tabla 1).

Contaminante	Valores límite			Normas Oficiales Mexicanas
	Exposición aguda		Exposición crónica (Para protección de la salud de la población susceptible)	
	Concentración y tiempo promedio	Frecuencia máxima aceptable		
Ozono (O <sub>3</sub> )	0.11 ppm (1 hora) (216 µg/m <sup>3</sup> )	1 vez al año	-	NOM-020-SSA1-1993(2)
	0.08 ppm (8 horas)	Quinto máximo en un año		
Monóxido de carbono (CO)	11 ppm (8 horas) (12595 µg/m <sup>3</sup> )	1 vez al año	-	NOM-021-SSA1-1993(1)
Dióxido de azufre (SO <sub>2</sub> )	0.13 ppm (24 horas) (341 µg/m <sup>3</sup> )	1 vez al año	0.03 ppm (media anual)	NOM-022-SSA1-1993(1)
Dióxido de nitrógeno (NO <sub>2</sub> )	0.21 ppm (1 hora) (395 µg/m <sup>3</sup> )	1 vez al año	-	NOM-023-SSA1-1993(1)
Partículas suspendidas totales (PST)	210 µg/m <sup>3</sup> (percentil 98 promedio de 24h)	24 horas	-	NOM-024-SSA1-1993(3)
Partículas con diámetro menor a 10 µm (PM10)	120µg/m <sup>3</sup> (percentil 98 promedio de 24h)	24 horas	50 µg/m <sup>3</sup> (promedio anual)	NOM-025-SSA1-1993(3)
Partículas con diámetro menor a 2.5µm (PM2.5)	60µg/m <sup>3</sup> (percentil 98 promedio de 24h)	24 horas	15µg/m <sup>3</sup> (promedio anual)	NOM-025-SSA1-1993(3)
Plomo (Pb)	-	-	1.5 µg/m <sup>3</sup> (promedio en 3 meses)	NOM-026-SSA1-1993(1)

Fuentes: (1) Diario Oficial de la Federación del 23 de diciembre de 1994.  
 (2) Diario Oficial de la Federación del 30 de octubre del 2002.  
 (3) Diario Oficial de la Federación del 26 de septiembre del 2005.

**Tabla 1.** Estándares mexicanos de contaminantes criterio.

## 2.2.2 Normas de Estados Unidos y California.

La EPA regula los seis contaminantes criterio. Establece límites de concentraciones y tiempos de exposición para la protección de efectos adversos al ser humano (*estándar primario*) o al entorno (*estándar secundario*). Estas regulaciones se muestran en la Tabla 2.

Contaminante	Tiempo promedio	Estándar de California	Estándar Federal	
			Primario	Secundario
Ozono O <sub>3</sub>	1-h <sup>(7)</sup>	(0.09 ppm California, - 180 µg/m <sup>3</sup> )	0.12 ppm	igual
	8-h <sup>(6)</sup>	(0.070 ppm -137 µg/m <sup>3</sup> California)	0.08 ppm (157µg/m <sup>3</sup> )	igual
PM <sub>10</sub>	24 h <sup>(3)</sup>	50 µg/m <sup>3</sup>	150 µg/m <sup>3</sup>	igual
	Media Anual <sup>(2)</sup>	20 µg/m <sup>3</sup>	50 µg/m <sup>3</sup> · revocado <sup>(2)</sup>	igual
PM <sub>2.5</sub>	24 h <sup>(5)</sup>	(igual para California)	35 µg/m <sup>3</sup>	igual
	Media Anual <sup>(4)</sup>	12 µg/m <sup>3</sup> California	15 µg/m <sup>3</sup>	igual
Monóxido de Carbono CO	8-hour <sup>1</sup>	9 ppm (10 mg/m <sup>3</sup> )	9 ppm (10 mg/m <sup>3</sup> )	Ninguno
	1-hour <sup>1</sup>	20ppm(23 mg/ m <sup>3</sup> )	35 ppm (40 mg/m <sup>3</sup> )	Ninguno
Dióxido de Nitrógeno NO <sub>2</sub>	Media Anual	0.030 ppm (56 µg/m <sup>3</sup> )	0.053ppm.(100µg/m <sup>3</sup> )	igual
	1-h	0.18 ppm (338 µg/m <sup>3</sup> )	---	igual
Dióxido de Sulfuro SO <sub>2</sub>	Media Anual	---	0.03 ppm	---
	24-h <sup>(1)</sup>	0.04 ppm (105 µg/m <sup>3</sup> )	0.14 ppm	---
	3-h <sup>(1)</sup>	---	No hay	0.5 ppm (1300 µg/m <sup>3</sup> )
	1-h	0.25 ppm (655 µg/m <sup>3</sup> )	No hay	---
Plomo Pb	Media x mes	1.5 µg/m <sup>3</sup>	---	Igual
	Media x 3meses	--	1.5 µg/m <sup>3</sup>	---

<sup>(1)</sup> No debe exceder más de una vez por año.

<sup>(2)</sup> Debido a falta de evidencia de problemas de salud ligadas a este límite de exposición de partículas gruesas, la agencia decidió revocar el estándar anual de PM<sub>10</sub> en 2006 (efectivo a partir de Diciembre 17, 2006).

<sup>(3)</sup> No debe exceder mas de una vez por año, en un periodo de 3 años.

<sup>(4)</sup> Para cumplir este estándar, el promedio de 3 años de las concentraciones ponderadas anuales medias de PM<sub>2.5</sub> de uno o múltiples monitores orientados a la comunidad no deben exceder de 15.0 µg/m<sup>3</sup>.

<sup>(5)</sup> Para cumplir este estándar, el promedio de 3 años del percentil 98avo. de concentraciones de 24-h en cada monitor orientado a la población no debe exceder de 35 µg/m<sup>3</sup> (efectivo desde Diciembre 17, 2006).

<sup>(6)</sup> Para cumplir este estándar, el promedio de 3 años, de los cuatro valores diarios más altos en promedios máximos de 8h de las concentraciones de ozono medidos en cada monitor dentro de un área durante cada año no debe exceder 0.08 ppm.

<sup>(7)</sup> En Junio 15 de 2005 la EPA revocó el estándar de 1h en todas las áreas, excepto en áreas designadas en: [Early Action Compact \(EAC\) Areas](#).

**Tabla 2.** Límites de los contaminantes criterio establecidos por la Agencia de Protección Medioambiental de Estados Unidos (USEPA) y el Consejo de Recursos del Aire de California (Air Resources Board, ARB).

## 2.2.3 Normas de la Organización Mundial de la Salud

Es importante agregar que recientemente la Organización Mundial de la Salud (OMS, 2006) ha publicado “*Directrices sobre la calidad del aire*” para la protección de la salud pública mundial y entre otra información, recomienda nuevos límites para los contaminantes que más afectan a la salud, los cuales se resumen en la Tabla 3.

Contaminante	Nuevos límites
PM <sub>10</sub>	20 µg/m <sup>3</sup> media anual, 50 µg/m <sup>3</sup> media en 24 h
PM <sub>2.5</sub>	10 µg/m <sup>3</sup> media anual, 25 µg/m <sup>3</sup> media en 24 h
O <sub>3</sub> *	100 µg/m <sup>3</sup> media en 8h, 0.05 ppm
NO <sub>2</sub>	40 µg/m <sup>3</sup> media anual, 200 µg/m <sup>3</sup> media en 1 h
SO <sub>2</sub>	20 µg/m <sup>3</sup> media en 24 h, 500 µg/m <sup>3</sup> media en 10 min.

\* El límite recomendado antes (media de 120 µg/m<sup>3</sup> en 8 h) se ha reducido a 100 µg/m<sup>3</sup>, dado que recientemente se han establecido asociaciones concluyentes entre la mortalidad diaria y concentraciones de ozono inferiores a 120 µg/m<sup>3</sup>.

**Tabla 3.** Nuevos límites para cualquier región, recomendados por la OMS en octubre del 2006 para PM<sub>10</sub>, PM<sub>2.5</sub> y O<sub>3</sub>.

## 2.3 Calidad del aire en la frontera México–Estados Unidos

Los contaminantes que están presentes en el aire de manera natural, el viento determina su dirección y dispersión. Por esta razón las líneas físicas divisorias entre una región y otra, o entre un país y otro, no representan barreras para intercambios de contaminantes en el ambiente. Hay países vecinos con zonas territoriales contiguas, que presentan graves problemas de contaminantes atmosféricos, y que además

preservan diferentes culturas, gobiernos y economías como lo son los Estados Unidos y México.

### 2.3.1 Convenios Binacionales

Desde los años sesenta se empieza a dar una gran importancia a aspectos ambientales en el mundo, y a pesar de que existía ya de alguna manera cooperación binacional entre México y Estados Unidos, no había una formalización de las acciones desarrolladas y estas eran aisladas.

Es hasta la firma del "*Convenio para la Protección y el Mejoramiento del Medio Ambiente en la Región Fronteriza*", mas conocido como "*Convenio de La Paz*" firmado en 1983, en la ciudad de La Paz, Baja California; (La Paz Agreement., 1983) donde se marca un hito en los esfuerzos de cooperación y se establecen compromisos concretos para el trabajo en el aspecto ambiental en la frontera.

Es precisamente mediante el convenio mencionado que se definió la "Franja fronteriza" de México y los Estados Unidos como: "*El área situada hasta cien kilómetros (62.5 millas) de ambos lados de las líneas divisorias terrestres y marítimas entre las Partes*". Esta franja puede observarse en la Figura 7, delimitada por la zona sombreada y líneas punteadas.

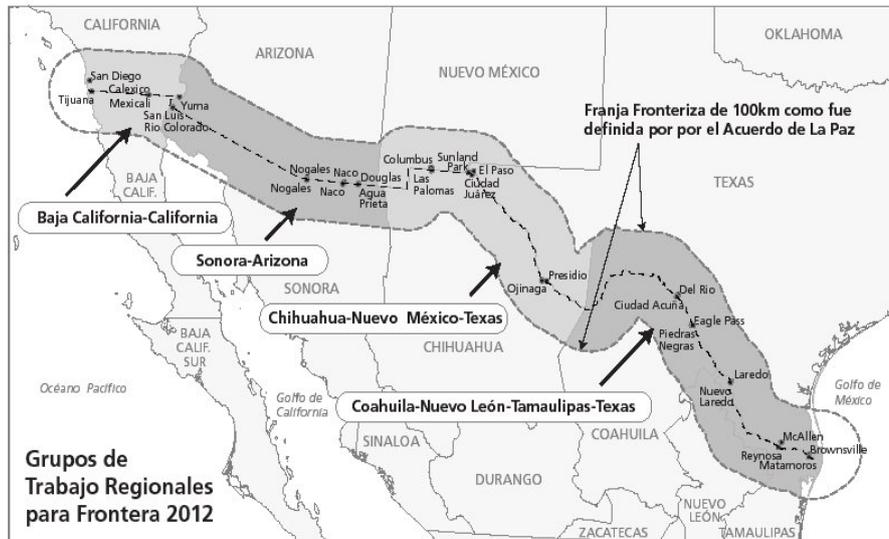
Posterior al Convenio de La Paz se trabajó con el programa "*Frontera XXI*", (SEMARNAP, 1996); el cual finalizó el año 2000 y como continuación de los trabajos medioambientales binacionales se inicia el programa "*Frontera 2012*" ([www.semarnat.gob.mx/frontera2012](http://www.semarnat.gob.mx/frontera2012)) que está vigente actualmente.

En el mapa de la Figura 7 están identificados los denominados "*grupos de trabajo regionales*"; estos se reúnen para planear y desarrollar las actividades además de formalizar a acuerdos conjuntos dentro del actual programa "*Frontera 2012*".

Los *grupos de trabajo regionales* son:

- Baja California – California.
- Sonora – Arizona.

- Chihuahua –Nuevo México –Texas.
- Coahuila – Nuevo León-Tamaulipas-Texas.



**Figura 7.** Las flechas en mapa que identifican la franja de protección ambiental binacional y los grupos de trabajo del programa “Frontera 2012”.

## 2.4 Modelos de predicción de Calidad del Aire

Es posible definir un modelo como la idealización de un sistema real, usado generalmente para plantear un problema, de manera que actúe como una representación simplificada de la realidad con el objetivo de permitir la manipulación y análisis del mismo, para lograr un mejor entendimiento del comportamiento del sistema.

Los tipos de modelos existentes van desde prototipos que representan al sistema a escalas manejables, seguros o de manera económica, hasta representaciones vía ecuaciones matemáticas o simulaciones computacionales. La elección del modelo a utilizar será el que más se ajuste a los objetivos del análisis. En este caso, se hace uso de modelos matemáticos y simulaciones computacionales para describir el

comportamiento de contaminantes atmosféricos que puedan servir para realizar predicciones de sus concentraciones. (US-EPA, 2003).

En el caso de los modelos de calidad del aire, normalmente usan técnicas matemáticas para simular los procesos físicos y químicos que afectan a los contaminantes en su dispersión y transformación en la atmósfera.

Basados en datos meteorológicos e información de las fuentes de emisión tales como tasas de emisión y parámetros de chimenea, estos modelos caracterizan la dispersión de los contaminantes primarios que son emitidos directamente en la atmósfera y en algunos casos también las reacciones químicas que ocurren en ésta, para formar contaminantes secundarios. Finalmente, generan como resultado una concentración ambiental del contaminante simulado.

La aplicación de un modelo, es una tarea compleja que demanda la integración sistemática de una gran cantidad de información generada por diferentes fuentes, tales como: inventario de emisiones, meteorología, topografía, usos de suelo, etc.

Sin embargo, modelizar de manera efectiva un fenómeno de contaminantes atmosféricos podría ser una potente herramienta de análisis, que tiene un amplio espectro de aplicación que va desde la evaluación de estrategias de control de emisiones y su impacto en la calidad del aire, soporte en la toma de decisiones de carácter ambiental, hasta la generación de información científica valiosa para comprender mejor la dinámica de la atmósfera y la contaminación en un lugar o región determinada.

El caso de los modelos de predicción de la calidad del aire son de gran utilidad para:

- Realizar notificaciones a favor de la salud pública.
- Planeación y gestión de políticas ambientales.
- Planificar programas para tomar el control en episodios críticos.
- Predecir los cambios en las concentraciones ambientales.
- Detectar cambios en los patrones de emisión.

## 2.4.1 Resultados buscados en los modelos de predicción

Las necesidades detectadas para los modelos de predicción pueden variar de acuerdo a la región, pero al final de cualquier necesidad se podrá concluir que todas tendrán relación con la conservación, la seguridad y la salud de la población. A partir de estos principios, se pueden establecer algunas características más deseadas, detectadas en un modelo de predicción y que bien pudiera encajar con lo que consideramos como características deseadas de predicción de cualquier otro contaminante.

- **Límites mínimos de error.** Los modelos actualmente existentes, manejan índices de error suficientemente grandes, que dan resultados poco fiables.
- ***Predicciones de corto, mediano y largo plazo.*** Dependiendo de necesidades específicas, el usuario podría necesitar los resultados de la predicción como parte de sus planes, en intervalos de tiempo de horas, días o semanas.
- **Predicciones multizonales.** Con ello se busca que el área geográfica que abarque la predicción sea amplia y suficientemente eficiente, para poder ser aplicada a toda el área metropolitana de una ciudad, o entre ciudades cercanas.
- **Predicciones de valores máximos del contaminante.** Este tipo de predicción sirve de ayuda para mantener el control de los niveles más altos de contaminantes a vigilar y poder tomar acciones en torno a ello. Uno de los principales fallos de los modelos de predicción actuales es que los valores más altos de las concentraciones no son predichos correctamente.

## 2.5 Características de la Zona de Estudio

Las poblaciones que forman parte de este estudio son Mexicali (México) y Calexico (Estados Unidos) pertenecientes a los grupos de trabajo de la región “*Baja California-California*” y las poblaciones de Ciudad Juárez (México) y El Paso (Estados Unidos), éstas últimas pertenecen a los grupos de trabajo de la región “*Chihuahua-Nuevo México-Texas*”, de acuerdo al tratado binacional vigente “*Frontera 2012*”.

Ambas zonas de estudio (Mexicali-Calexico y Ciudad Juárez-El Paso) tienen un clima semidesértico aunque los límites de temperaturas alcanzadas tanto en invierno como verano difieren significativamente.

### 2.5.1 Región Mexicali-Calexico

Mexicali es la capital del estado mexicano de Baja California y está ubicada en el noroeste de México en latitud  $32^{\circ}43'N$  a  $30^{\circ}52'$  Norte, longitud  $114^{\circ}42'$  Oeste a  $115^{\circ}56'$  O, con una altitud media de 3 metros sobre el nivel del mar. De acuerdo al censo de población del 2005 (INEGI, 2005), Mexicali contaba con 855962 habitantes (Ver Figura 8).



Figura 8 Ciudades de Mexicali y Calexico (Fuente: Google Map, 2007).

Calexico en cambio, con una ubicación en latitud 32° 40' 42" Norte, longitud 115° 29' 53" Oeste (32.678246, -115.497997), en su censo más reciente realizado el año 2000, tenía registrados 27,109 habitantes, y se esperaba que para el 2005 llegara a los 36,005. (Census Bureau, 2007 <http://quickfacts.census.gov>).

Ambas poblaciones comparten un clima cálido-seco, con temperaturas extremas que van desde los - 5° C a los 53° C a la sombra y con escasa precipitación anual (INE, 2005).

En esta región existen diversos detonantes de la mala calidad del aire en la región. Los vehículos de transporte son el principal generador de contaminantes debido a la gran cantidad de vehículos de carga y de uso personal existentes, muchos de ellos con una mala mantención y ajustados de forma que sus niveles de generación de contaminantes son elevados.

Además, operan tres plantas generadoras de energía: la planta termoeléctrica de Cerro Prieto, que abastece a la ciudad además de exportar energía a Estados Unidos, y otras dos de reciente creación con capital extranjero (INTERGEN y SEMPRA) que exportan en su totalidad a Estados Unidos la energía generada (CCAAN, 2002).

De acuerdo a lo anterior, se observa la escasez de agua, la disposición ilegal de basuras (tanto industriales como de la población), el transporte de todo tipo de mercancías hacia ambos lados de la frontera, las emisiones de los automóviles (muchos de ellos en mal estado), las calles sin pavimentar, el uso de pesticidas, fertilizantes y quemas en amplias zonas agrícolas en ambos valles de la frontera (Valle de Mexicali y Valle Imperial) y ladrilleras que queman sus productos sin controles ambientales. También es importante incluir que esta zona sirve de vía para la migración legal e ilegal, lo que influye en el crecimiento poblacional y una mala planeación de servicios públicos (<http://scerp.org>, 2005).

## 2.5.2 Región Ciudad Juárez-El Paso

Ciudad Juárez es la capital del estado Chihuahua, estado más grande de la República Mexicana. Coincide además que, Ciudad Juárez es la ciudad fronteriza más poblada, con 1,313,338 habitantes (INEGI, 2005), y se encuentra localizada en Latitud:  $31^{\circ}07'38$  y  $31^{\circ}48' 0$  Norte y Longitud:  $106^{\circ}06' 57$  y  $106^{\circ}98' 44$  Oeste. Ver Figura 9.

El Paso, Texas registró en el año 2006 una población de 736,310 habitantes (U.S. Census Bureau, 2006 Population Estimates).

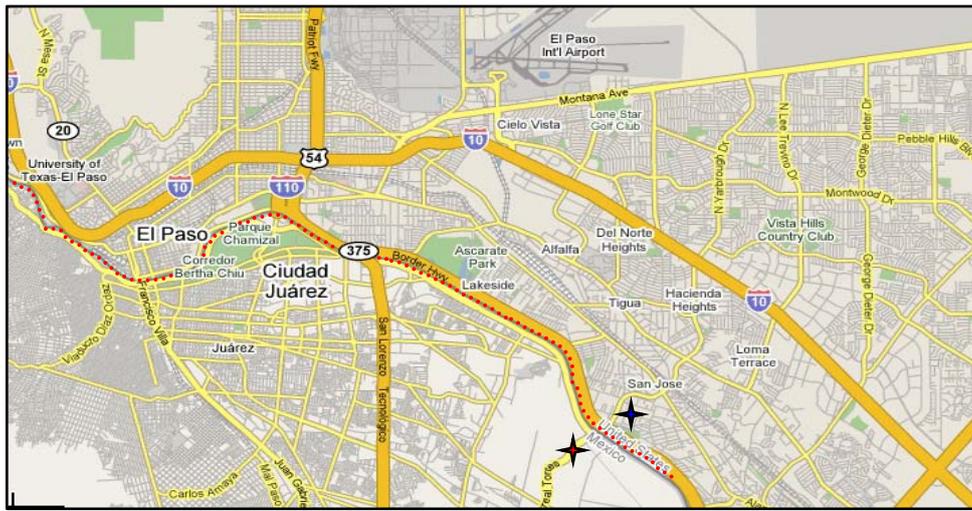


Figura 9. Mapa de la región Ciudad Juárez-El Paso. Fuente: Google Map 2007.

Los problemas que favorecen la generación de contaminantes en el aire son similares a los de la zona Mexicali- Calexico pero, debido a la gran diferencia poblacional en ambas zonas, la problemática se agudiza, ya que tanto Ciudad Juárez como El Paso son ciudades altamente industriales con mayor número de industrias a ambos lados de la frontera y con mayores índices de pobreza.

## 2.6 Contaminantes criterio modelados.

### 2.6.1 Partículas en suspensión y ozono

Las *partículas en suspensión* y *ozono* son los contaminantes con más altos niveles registrados en las zonas de estudio y que han encendido “*focos rojos*” por daños a la salud en la frontera de Estados Unidos y México (USEPA, AQS- Air Quality System, 2005). La Organización Mundial de la Salud, establece este tema como preocupante en distintas partes del mundo (OMS, 2006).

Frecuentemente la contaminación en el aire se estudia en términos de problemas locales y en escasas ocasiones con una visualización global. Sin embargo el rápido crecimiento de la población, los cambios de uso de suelos, la urbanización de las regiones, pueden expandir el alcance de los problemas de la calidad del aire regional o local y alterar significativamente la composición atmosférica de grandes regiones. El ozono y las pequeñas partículas en suspensión pueden permanecer suficiente tiempo en la atmósfera y tienen una amplia influencia en regiones lejanas a sus fuentes de generación, lo que contribuye al cambio climático.

Debido principalmente a las causas anteriormente descritas, estos contaminantes criterio (el *ozono* y las *partículas en suspensión*) fueron seleccionados para ser estudiados y modelizados en este trabajo.



# Capítulo 3

## Objetivos

---

Los trabajos desarrollados en la presente tesis doctoral tienen como objetivo general aportar mejoras a los métodos y herramientas de análisis de fenómenos no lineales multivariantes, especialmente de los fenómenos ambientales que intervienen en la calidad del aire. Para ello se aplicarán las herramientas y metodologías que brinda la estadística avanzada enfocada al análisis multivariante y de Minería de Datos.

Los objetivos específicos son:

- Conocer y comprender el comportamiento del fenómeno ambiental de acuerdo a datos históricos del contaminante y sus factores involucrados.
- Predecir a corto plazo las concentraciones de los contaminantes bajo estudio.
- Evaluar la efectividad de distintas herramientas de modelización no lineal multivariante, como Redes Neuronales y otras herramientas de Inteligencia Artificial.

- Proporcionar una herramienta que pueda ser de utilidad para mejorar el control, planificación y toma de decisiones referidas a medidas de seguridad para la salud de la población.

Es pertinente agregar que la selección estratégica de la zona de estudio también se considera un aspecto importante. La franja fronteriza entre México y Estados Unidos presenta aspectos ambientales que comprometen la salud y los ecosistemas que comparten, principalmente sus ciudades vecinas con muy diferentes políticas ambientales y económicas.

Como parte de los objetivos se consideró un estudio para la selección de las regiones que evidencien la mala calidad del aire, y que adicionalmente dispusieran de la cantidad y calidad de datos indispensables para una adecuada modelización.

Las regiones consideradas de interés fueron objeto además de un detallado análisis para determinar los contaminantes críticos predominantes: partículas suspendidas y ozono. Los trabajos de estudio y modelización de esta tesis, llevan la tarea de trabajar primordialmente con ellos, buscando aportar a su entendimiento y control.

Para el desarrollo de modelos de material particulado (partículas suspendidas) se seleccionó la región de vecindad entre las ciudades: El Paso, Texas, ubicada en el lado estadounidense y Ciudad Juárez, Chihuahua localizada en el lado mexicano. Para el desarrollo de modelos de ozono se trabajó con la región vecindada entre Calexico California (Estados Unidos) y Mexicali, Baja California (México).

Después de evaluar las técnicas pertinentes para el caso de los contaminantes de estudio, se busca mejorar la precisión de las predicciones respecto a modelos similares existentes desarrollados en otras zonas del mundo de acuerdo a la literatura reciente.

El alcance de los objetivos presenta la posibilidad de valorar la utilidad de los modelos expuestos en los sistemas de predicción y de diagnóstico, tanto para la evaluación del estado de un territorio (Estudios de Impacto

Ambiental), como para el empleo por medios de protección civil en la gestión de advertencias a la población.

Además es importante mencionar que la calidad de las predicciones se medirá de acuerdo a los criterios estandarizados contenidos en el denominado “*Model Validation Kit*”, que establece un marco de referencia común para evaluar el desempeño de modelos estadísticos y de dispersión atmosférica.

Las medidas a utilizar para evaluar la calidad de modelos desarrollados se describen en la Tabla 4.

Descripción	Formula
Error Cuadrático Medio	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (C_p - C_o)^2}$
Error de Cuadrático Medio Normalizado	$NMSE = \frac{Media(C_o - C_p)^2}{Media(C_o)Media(C_p)}$
Coefficiente de correlación	$R = \frac{Media((C_o - Media(C_o))(C_p - Media(C_p)))}{\sigma_{C_o} \sigma_{C_p}}$
Desviación Media Geométrica	$MG = e^{(Media(\ln(C_o)) - Media(\ln(C_p)))}$
Bias Fraccional	$FB = 2 \frac{Media(C_o) - Media(C_p)}{Media(C_o) + Media(C_p)}$
Varianza Geométrica	$VG = e^{Media((\ln(C_o) - \ln(C_p))^2)}$
Factor 2 (FAC2)	$0.5 \leq C_o/C_p \leq 2$
Varianza Fraccional	$FV = 2 \frac{\sigma_{C_o} - \sigma_{C_p}}{\sigma_{C_o} + \sigma_{C_p}}$

Índice de concordancia	$d_2 = 1 - \frac{\sum_{i=1}^n (C_p - C_o)^2}{\sum_{i=1}^n ( C_p - \text{Media}(C_p)  +  C_o - \text{Media}(C_o) )^2}$
Media del Error Absoluto	$MAE = \frac{1}{n} \sum_{i=1}^n  C_o - C_p $
Media del Error de Bias	$MBE = \frac{1}{n} \sum_{i=1}^n (C_o - C_p)$

**Tabla 4.** Medidas de Rendimiento del Paquete de Validación de Modelos o “Model Validation Kit”, (Chang et al, 2005 ). En las ecuaciones descritas,  $C_p$ = Representa *valores pronosticados*,  $C_o$  = *Valores observados*, *Media* = *Media aritmética*.

# Capítulo 4

## Hipótesis

---

Considerando que los procesos ambientales que influyen en la calidad del aire son fenómenos altamente no lineales y multifactoriales se pretende construir modelos de predicción, a corto plazo, de contaminantes atmosféricos. Debe tenerse en cuenta que las técnicas usuales de generación de modelos (útiles en mecánica de medios continuos y para el estudio de procesos con menor incertidumbre en su comportamiento) resultan inviables para el caso de estudio que se aborda en esta tesis.

No son viables las técnicas usuales de deducción de modelos debido a que las ecuaciones diferenciales conducen a modelos poco realistas. Sobre todo por el desconocimiento de las condiciones de entorno y la alta sensibilidad que presentan los fenómenos atmosféricos, aún con las pequeñas perturbaciones que sucedieran.

Se considera la existencia de una estrecha dependencia entre los contaminantes a modelizar y las variables meteorológicas utilizadas.

Además de los precursores tomados en cuenta para la generación de los contaminantes modelados, hay otros parámetros potenciales explicativos que no han sido medidos, pero se asume que la explicación principal es

proporcionada por las variables que se presentan y que han sido utilizadas.

Se asume la existencia y uso de series de datos estadísticamente independientes. Estos datos son aplicados en las técnicas de Minería de Datos, herramientas viables para el desarrollo de los modelos de predicción para fenómenos multivariantes.

Para efectos de reconocimiento y estudio de los fenómenos atmosféricos a modelar se consideran igualmente viables los modelos de predicción obtenidos con la utilización de técnicas de “*caja blanca*”. Como los modelos lineales, regresión generalizada, y persistencia, además de los modelos de predicción desarrollados con técnicas consideradas de “*caja negra*” como Redes Neuronales. Cada una de dichas herramientas representan ciertos riesgos ya conocidos, así como las desventajas y ventajas esperadas de acuerdo a trabajos similares analizados.

Los modelos paramétricos y semiparamétricos (de “*caja blanca*”) son relativamente sencillos y rápidos de desarrollar, por no requerir de equipo dedicado ni considerable tiempo por parte del investigador. Se diferencia de las técnicas de “*caja negra*” porque el esfuerzo computacional es considerable y exige una importante dedicación y dominio de la herramienta por parte del investigador para obtener resultados aceptables.

La ventaja de estos últimos, recae en que logran incorporar en sus modelos propuestos la información mas característica del fenómeno modelizados debido a que la herramienta “*aprende*” de las “*experiencias pasadas*” las cuales están integradas en sus almacenes de datos referente al fenómeno en análisis con que son alimentadas.

Debido a dicho aprendizaje, los márgenes de error en la predicción son bastante menores a los márgenes de error en las predicciones obtenidas con las técnicas de “*caja blanca*”. Para esta tesis serán de gran utilidad las técnicas de *caja blanca* porque representaran los márgenes máximos de error esperados en los modelos a desarrollar. Aunque hay que tener en consideración, que los resultados en la predicción que se obtienen con

un modelo cualquiera como el usado, nunca pueden ser mejores que la calidad de los datos utilizados.

Se toma en consideración la hipótesis del interés de conocer con anticipación los valores máximos diarios de los contaminantes, y no necesariamente la predicción de los valores horarios de los mismos. Esto se debe a la importancia de detectar posibles episodios dañinos y proteger la salud de la población y demás seres vivos.

También se asume la alta correlación existente entre las concentraciones máximas de los contaminantes como el ozono de las partículas finas ( $PM_{2.5}$ ) de un día específico, respecto a la concentración máxima del día siguiente.

Además, se establece que los modelos desarrollados con Técnicas de Inteligencia Artificial y en especial con Redes Neuronales Artificiales, dada la naturaleza no lineal de los datos a modelar, pueden adaptarse más eficientemente a la naturaleza de la información obtenida, que los modelos desarrollados con técnicas paramétricas y semiparamétricas.

Una última hipótesis que pretende corroborar es que las concentraciones de las primeras horas del día de las variables predictoras para las partículas suspendidas y el ozono, son determinantes de las concentraciones máximas de éstos contaminantes a predecir para el resto del día.



# Capítulo 5

## Materiales y Métodos

---

Como se ha comentado en capítulos anteriores, el propósito de este estudio es el desarrollo de modelos de predicción de dos contaminantes criterio que en distintas regiones del mundo y específicamente para la zona fronteriza México-US. Demostrar los efectos críticos para la salud y medioambiente. Para ello, se aplicaron y evaluaron técnicas estadísticas tradicionales de modelización paramétrica y semi-paramétrica, además de técnicas no paramétricas.

Las técnicas paramétricas a las que se hace referencia son *Persistencia* y *Regresión Lineal* y, por la colinealidad típica entre las variables ambientales se juzgó importante incluir la técnica semiparamétrica de *Regresión Ridge*.

Como herramientas no paramétricas se utilizaron distintas técnicas de inteligencia artificial como redes neuronales *Perceptrón Multi Capa* (MLP), *Perceptrón Multi Capa Cuadrática* (SMLP), *Función de Base Radial* (RBF) y *Redes Elman*, además de *Máquina de Vectores Soporte* (Support Vector Machine, SVM), herramienta que esta siendo aplicada recientemente en la comunidad científica con resultados eficientes en el modelado de distintos fenómenos.

## **5.1. Métodos aplicados para el desarrollo de los modelos de predicción**

### **5.1.1. Persistencia**

Es considerado el más sencillo. Consiste en suponer que la medición registrada en un intervalo de tiempo actual, será la misma en el intervalo de tiempo siguiente. Por ejemplo: el valor de ozono registrado de manera real a las diez de la mañana será la predicción para las once de la mañana. Después el valor real registrado a las once de la mañana corresponderá a la predicción para las doce del día y así de manera consecutiva. O bien, el comportamiento real del día de hoy de las concentraciones de ozono, será la predicción para el día de mañana.

Este método es considerado el menos exacto, debido a la complejidad que supone explicar el comportamiento de la generalidad de los contaminantes en el aire, sin embargo es interesante su utilización ya que hay una asociación estadística positiva con las concentraciones pasadas y futuras de los contaminantes atmosféricos (Wilks, 1995). Esto sucede porque los valores de concentraciones grandes de un contaminante, tiende a ser precedido por otro valor grande, y una concentración baja, tiende a ser precedida por valores bajos. Con frecuencia, se aprovechan los resultados de este modelo como punto de partida o límite máximo de error en la predicción.

En el caso del modelo de predicción del ozono, las predicciones resultantes del modelo de persistencia fueron usadas como una variable predictora más, presuponiendo la influencia del valor de las concentraciones de ozono en un intervalo de tiempo anterior respecto al valor de la concentración de ozono en el de tiempo actual.

El método de persistencia tiene la ventaja de no necesitar grandes esfuerzos estadísticos, matemáticos o computacionales, pero dada la naturaleza no lineal de formación de contaminantes en el aire, es necesario que se complemente con otros métodos de predicción más robustos.

### 5.1.2. Regresión Lineal Multivariante

El método de regresión, especialmente cuando nos referimos a la Regresión Lineal Multivariante, es una aplicación estadística bastante usual en problemas de predicción de los contaminantes ambientales y los distintos parámetros meteorológicos que intervienen.

Este método, describe la relación y corrobora el nivel de dependencia entre diversos parámetros que intervienen en la formación los contaminantes en estudio (Cassmassi, 1987; Hubbard y Cobourn, 1997; Ryan, 1994; Dye et al, 1996).

La ecuación general de la Regresión Multivariante corresponde con:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}, \quad i=1,2,\dots,n, \quad \text{Ecuación (1)}$$

$Y_i$  = Variable independiente. Variable a predecir.

$\beta_0$  = Constante.

$\beta_p$  = Coeficientes (pesos)

$X_i$  = Variables independientes. Parámetros de predicción

$n$  = Número de variables predictoras.

Existen diversos programas computacionales que resuelven modelos de relación multivariante y son relativamente sencillos de usar. Pero, normalmente, el margen de error en la predicción se incrementa conforme se incrementa el número de variables en la ecuación de regresión. Es conveniente tomar en cuenta que éste método es más eficiente para predecir valores medios y tiende a sub-predecir los valores máximos, mientras que exagera las concentraciones más bajas (EPA-454/R-99-009, 1999).

Se trata de una gran desventaja del método, porque concentraciones altas de contaminantes están directamente relacionadas con efectos en la salud. Es preferible que el método tienda a “sobre-predecir”, antes que pronosticar concentraciones por debajo de la realidad.

### 5.1.3. Regresión Ridge

El cálculo de los coeficientes de correlación lineal con mínimos cuadrados, cuando existen relaciones de colinearidad entre las variables involucradas, normalmente produce resultados con valores inestables (Faraway, J. 2002). En los contaminantes ambientales es común que se presenten altas correlaciones entre sí. Lamentablemente, la existencia de una alta correlación entre las variables predictoras es una característica que la *Regresión Lineal* no resuelve de manera eficiente, lo que puede llevar a predicciones menos precisas.

*Regresión Ridge* (tradicionalmente se le conoce como “*Regresión Regularizada*” o bien *Regresión No Paramétrica*) es técnicamente una generalización de la *Regresión Lineal* pero incorpora relaciones no lineales en el análisis de regresión. La ventaja, es que resuelve mejor la presencia de multicolinealidad entre las variables predictoras.

El modelo de Regresión Rige utilizado en este estudio se muestra en la Ecuación 2.

$$\begin{aligned} \hat{y}_{nueva} &= \hat{\beta}^T x_{nueva} \\ &= \sum_i \alpha_i x_i^T x_{nueva} \end{aligned} \qquad \text{Ecuación (2)}$$

### 5.1.4. Redes Neuronales Artificiales:

Las redes neuronales artificiales son grupos interconectados por medios eléctricos, mecánicos o computacionales de neuronas artificiales cuyo modelo esta inspirado por el funcionamiento del sistema nervioso biológico.

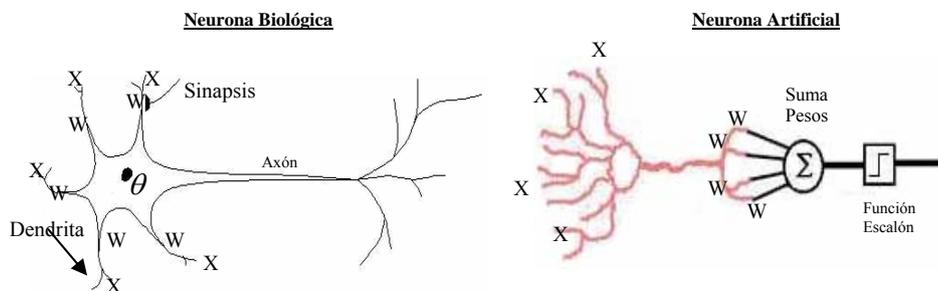
En una red neuronal artificial, cada nodo opera bajo principios análogos a una neurona biológica, por tanto, en una neurona biológica, cada sinapsis entrante a la neurona tiene un peso asociado a ella.

En la neurona artificial, cada nodo tiene un determinado número de líneas de entrada (análogo a las sinapsis de entrada de una neurona biológica) y una “función de activación” o también conocida como “función de transferencia” que determina cuando un nodo debe ser activado.

Cada red neuronal artificial debe ser “entrenada” para simular el comportamiento deseado usando un algoritmo de entrenamiento con una gran cantidad de datos destinados para ello. Una vez pasado el proceso de “entrenamiento” es necesario sea evaluado su “aprendizaje” por lo que se requieren de datos nuevos para probar que tan bien “aprendió”.

En una red neuronal artificial vista como técnica de inteligencia artificial, su algoritmo sigue la lógica de aprendizaje del cerebro humano. El cerebro trabaja con redes neuronales biológicas y la mayoría de las técnicas de Inteligencia Artificial, con redes neuronales artificiales.

El cerebro toma información de experiencias aprendidas para dar una respuesta a un evento, de forma análoga, una red neuronal artificial, alimenta con datos de entrada a cada neurona o perceptrón (variables predictoras) y estas son “entrenadas” para desarrollar “experiencia”, y poder tomar decisiones que serán reportadas en los datos de salida. Cada perceptrón es un modelo matemático de una neurona biológica (Ver Figura 10).



**Figura 10.** En esta figura se muestra una comparativa entre una *neurona biológica* (a la izquierda) y una *neurona artificial* o perceptrón (a la derecha). (Figura adaptada de [ese.stanford.edu](http://ese.stanford.edu)).

Como se puede observar en la Figura 10, la *dendrita* recibe señales eléctricas por los *axones* de otras neuronas que forman parte de la red. En el *perceptrón* (neurona artificial) la señal eléctrica esta dada por valores numéricos ( $X_i$ ).

Las señales eléctricas son moduladas en la *sinapsis* entre la *dendrita* y *axones*. El Perceptrón interpreta lo anterior multiplicando cada valor de entrada ( $X_i$ ) por determinados pesos ( $W_i$ ). Una neurona biológica responde o emite una señal de salida cuando el impulso eléctrico recibido es lo suficientemente fuerte como para sobrepasar un umbral establecido ( $\theta$ ). El Perceptrón simula este proceso vía el sumatorio de pesos de las entradas. La suma total representara la fuerza de la señal de entrada. Se procede a aplicar una función escalón al total calculado en el sumatorio para determinar el valor de salida. Las salidas de las neuronas biológicas alimentan a otras neuronas y los perceptrones alimentaran a otros perceptrones. (Minsky & Papert, 1988, [cse.stanford.edu](http://cse.stanford.edu)).

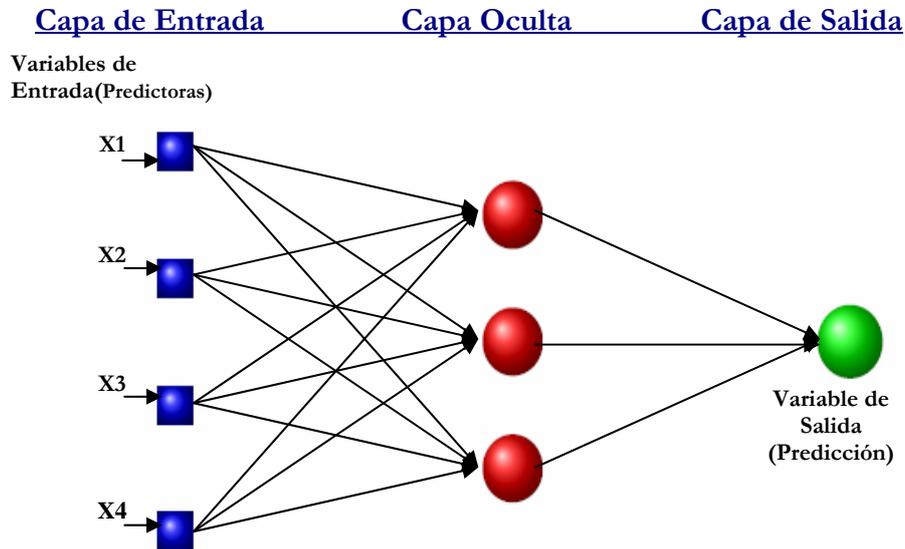
Para la modelización de contaminantes en el aire, se diseñan redes neuronales artificiales con varias capas de neuronas. A cada neurona en la entrada de la red, le corresponderá recibir los datos de un parámetro o variable predictora ( $X_i$ ) involucrado con el comportamiento o la formación del contaminante o variable que se pretende predecir. Durante el proceso de “entrenamiento” se busca identificar patrones y reconocer comportamientos de la base de datos que alimentó a la red, aunque estos sean complejos y multidimensionales (Ruiz-Suárez et al., 1995; Comrie, 1997; Gardner & Dorling, 1998).

Cabe mencionar que el esfuerzo computacional para desarrollar este tipo de modelos con redes neuronales artificiales es considerablemente mayor en comparación con el resto de las técnicas ya mencionadas.

### 5.1.5. Redes Perceptrón Multicapa (MLP)

Esta arquitectura de red neuronal es un aproximador universal de funciones, de modo que puede usarse para aproximarse a cualquier función continua en el espacio multidimensional real (Cybenko & Hornik, 1989). Es una red sin recurrencias y con un eficiente algoritmo

de retropropagación, lo que la hace relativamente sencilla de usar. Posee además una elevada capacidad de generalización y robustez, que minimiza la desventaja del considerable tiempo de entrenamiento que se requiere para el aprendizaje. Aunque es importante mencionar que este tiempo se incrementa conforme a la complejidad de la estructura de datos a modelar, como es el caso de los modelos de calidad del aire.



**Figura 11.** Red Perceptrón Multicapa (MLP) con cuatro neuronas de entrada (variables de entrada), 3 neuronas en la capa oculta y una neurona de salida (variable de salida). Alimentación hacia delante.

La Figura 11 muestra un ejemplo simple de la arquitectura de una red MLP, basada en una topología de red con conexiones hacia delante (feed-forward) y con tres tipos de capas: La *capa de entrada* (en éste caso con dos neuronas que actúan como almacén de datos de entrada con los valores de las variables predictoras), la *capa oculta* (con tres neuronas donde se realizan cálculos y ajustes de pesos y retroalimentación (*retropropagación*) comparando con los valores de la neurona de la *capa de salida*, (aprendizaje supervisado)) y la *neurona de salida* (que actúa como un almacén de datos de salida (predicciones)) (Gardner & Dorling, 1998).

Es común que este tipo de arquitecturas se diseñe con más de una capa oculta, aunque se ha demostrado que una red MLP con una sola capa oculta y funciones no lineales en las neuronas de dicha capa, puede aproximarse a cualquier función (Haykin, 1999).

La Ecuación 3, muestra la función de aproximación MLP con una capa oculta:

$$f(x_1, \dots, x_p) \approx F(x_1, \dots, x_p) = \sum_{i=1}^M \alpha_i \cdot \varphi\left(\sum_{j=1}^p w_{ij} \cdot x_j - \theta_i\right) \quad \text{Ecuación (3)}$$

La Ecuación 4, muestra la función de activación de propagación hacia atrás:

$$y_j(n) = \frac{1}{1 + \exp(-v_j(n))} \quad \text{Ecuación (4)}$$

### 5.1.6. Red Perceptrón Multicapa Cuadrática (SMLP)

Esta arquitectura de red fue creada por G.W. Flake (1998) y la denominó “Unidad Cuadrática Aumentada, Radialmente Extendida, Perceptrón Multicapa”, en inglés **S**quared **U**nit **A**ugmented, **R**adially **E**xtended, **M**ulti Layer **P**erceptron: **S**QUARE-**M**LP ó **S**MLP, aunque este último término: SMLP, es el mas manejado en la literatura.

SMLP es una arquitectura de red familiar a la estructura de *Modelos Aditivos Generalizados* (GAM), expuestos por Hastie (1990). Es una red neuronal híbrida que usa características de entradas lineales y no lineales resultantes de la expansión de la *Distancia Euclidiana* alrededor de un centro de un grupo de datos. La ecuación representativa propuesta por Flake (1998) para este tipo de redes es:

$$y = \sum_i w_i g\left(\sum_j u_{ij} x_j + \sum_k v_{ik} x_k^2 + a_i + b\right) \quad \text{Ecuación (5)}$$

Donde :

$$g(x) = \tan(x) \quad \text{Ecuación (6)}$$

ó bien :

$$g(x) = \frac{1}{(1 + e^{(-x)})} \quad \text{Ecuación (7)}$$

$g$  = Función de activación.

$\mathbf{x}$  = Vector de entrada.

$\mathbf{W}$  = pesos de la función de activación.

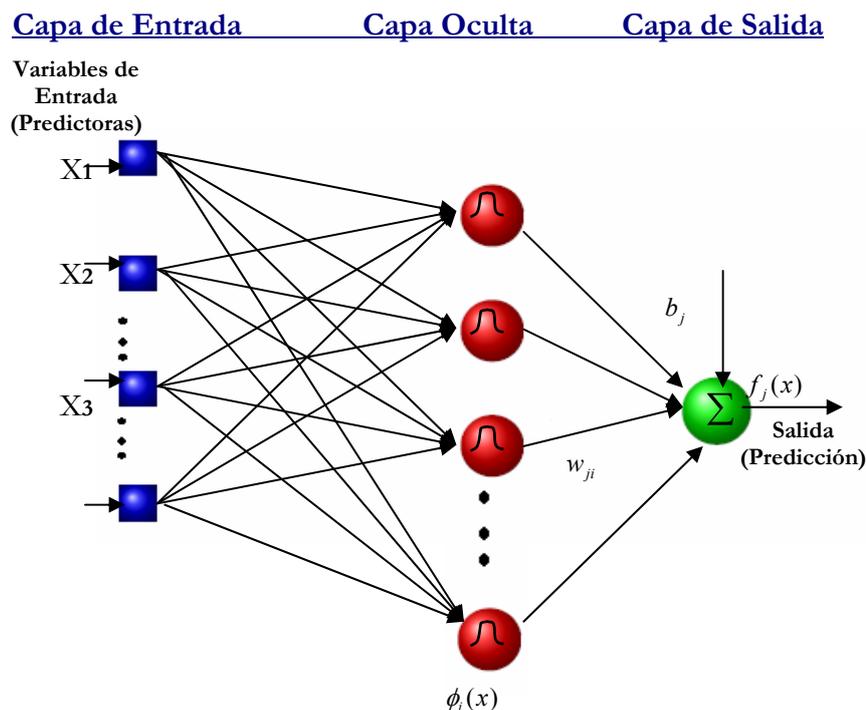
$b$  = valor base de la neurona de salida.

$a$  = Término lineal extraído con mínimos cuadrados.

Cada neurona en la capa oculta, tiene una función paramétrica de activación que puede estar cambiando de manera continua. Entradas cuadráticas se aproximan a las funciones radiales y entradas regulares a funciones de proyección (Flake, 1998).

### 5.1.7. Redes de Base Radial (RBF)

A diferencia de las *Redes Perceptrón Multicapa*, las *Redes de Base Radial* o redes RBF son arquitecturas que disponen siempre de tres capas: *capa de entrada*, *capa oculta* y *capa de salida*. Al igual que en MLP, cada neurona de una capa tiene conexión con todas las neuronas de la capa siguiente. La Figura 12 muestra un esquema típico de la arquitectura de una red RBF.



**Figura 12.** Esquema de una Red de Base Radial (RBF).

La red RBF tiene las características de que cada neurona de la capa oculta se activa con una función radial que define un centro para cada neurona de la capa oculta. Lo anterior hace que la neurona tome un elevado valor de activación cuando las entradas estén cerca de dicho centro y bajo cuando la distancia al centro sea alta. Generalmente, antes de la salida, se realizan transformaciones estadísticas basadas en funciones de distribución Gaussianas o lineales, por lo que las neuronas de salida frecuentemente actúan como sumadores. (Powell, 1985), (Broomhead & Lowe, 1988).

Moody y Darken (1989) definieron a la red RBF con la ecuación siguiente:

$$f(x) = \sum_{i=1}^n w_i \phi_i(x) + b \quad \text{Ecuación (8)}$$

$$x = [x_1, x_2, \dots, x_n]^T, w = [w_1, w_2, \dots, w_n]^T$$

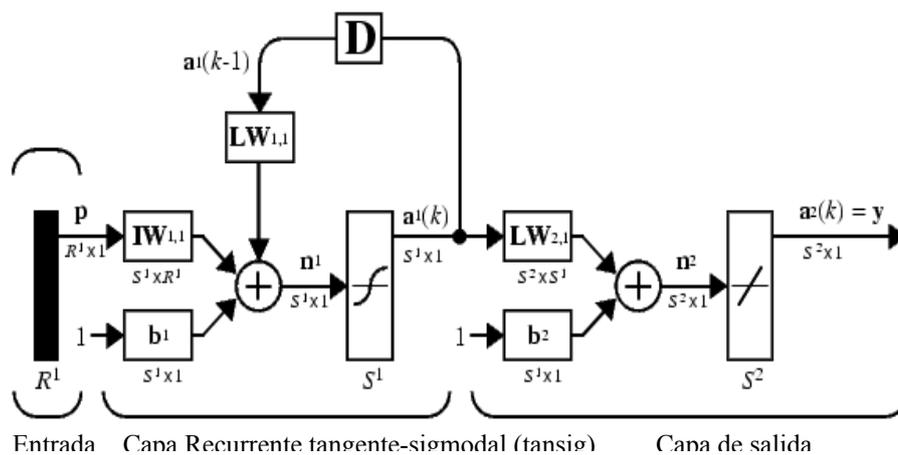
- $\phi$  = Función de activación no lineal.
- $x$  = Vector de entrada.
- $w$  = Pesos de la función de activación.
- $b$  = Valor base de la neurona de salida.

La Ecuación 9 muestra una función de activación gaussiana de base radial típicamente utilizada en RBF.  $\mu$  es el centro de la función y  $\sigma^2$  la amplitud de la misma.

$$\phi_i(x) = e^{-\frac{\|x-\mu\|^2}{2\sigma_i^2}} \quad \text{Ecuación (9)}$$

### 5.1.8. Redes Elman

Las redes Elman a diferencia de las otras arquitecturas de redes, son parcialmente recurrentes, por lo que presentan algunas conexiones hacia atrás. Algunas neuronas de entrada son utilizadas para recoger la información de la que disponían otras neuronas en el instante anterior. Llamadas también *neuronas de contexto*, dado que hace referencia al estado anterior de la red. Se realimentan las salidas de las neuronas de la última capa oculta de modo que, en cierto sentido, la red dispone de información acerca de la entrada del instante anterior (Elman, 1990).



**Figura 13.** Arquitectura básica de las redes Elman. Adaptada de : <http://www.mathworks.com/access/helpdesk/help/toolbox/nnet/>

$$a_1(k) = \tan \operatorname{sig} (IW_{1,1}p + LW_{1,1}a_1(k-1) + b_1) \quad \text{Ecuación (10)}$$

$$a_2(k) = \operatorname{purelin}(LW_{2,1}a_1(k) + b_2) \quad \text{Ecuación (11)}$$

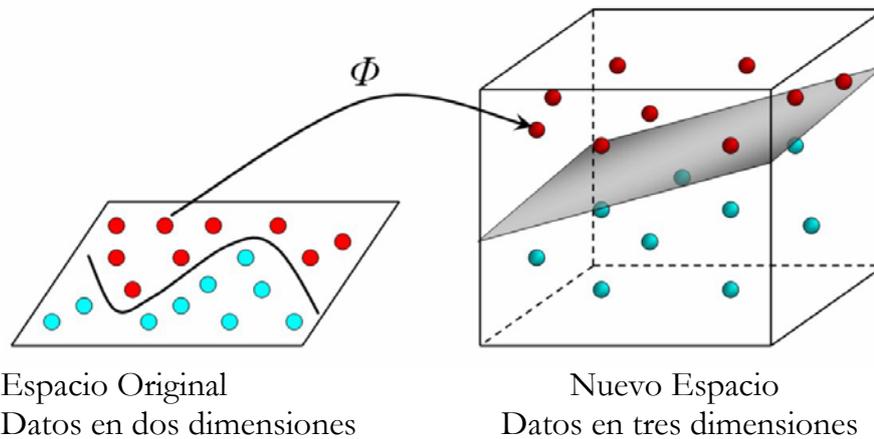
La capa oculta (recurrente) se estructura con neuronas con funciones de transferencia tangentes-sigmodales, por lo que se les llama “*neuronas tansig*”, y en la capa de salida sus neuronas trabajan con funciones de transferencia lineales, también llamadas “*neuronas purelin*”. Las Ecuaciones 10 y 11 muestran las funciones de transferencia mencionadas. Esta

característica favorece que la arquitectura de red se pueda aproximar a cualquier función, pero con la desventaja que la capa oculta tiende a requerir un gran número de neuronas, lo que incrementa la complejidad de la misma. (Works, 2007).

El aprendizaje se basa en el algoritmo de retropropagación, dado que a la hora de entrenar se desacoplan los bucles. De este modo, inicialmente se calculan las salidas y se ubican los datos para pasar a las neuronas de contexto. En el siguiente instante, se consideran estos datos como entradas a la red y se aplica de nuevo el algoritmo, repitiéndose el procedimiento (Ver Figura 13).

### 5.1.9. Máquina de Vectores Soporte (SVM)

Es una poderosa herramienta con fundamentos lineales para clasificación y regresión con aprendizaje supervisado (Cortes & Vapnik, 1995). Normalmente utiliza la técnica conocida como “kernel trick” para aplicar técnicas de clasificación lineal a problemas de clasificación no lineal (Ver figura 14).



**Figura 14.** SVM proyecta la información en espacios de mayor dimensionalidad, con el objetivo de convertir los datos en clases linealmente separables (Fuente: Martínez-de-Pisón et al., 2008).

SVM genera la salida lineal  $y(x)$ , trabajando dentro de un espacio formado por mapeos no lineales del vector de entrada  $x$  N-dimensional, dentro de un espacio K-dimensional usando una función no lineal  $\phi(x)$ .

K equivale al número de unidades (neuronas) en la capa oculta llamados “vectores soporte”. Las tareas de aprendizaje consisten en la minimización del error de la función  $y(x)$ , y al mismo tiempo busca mantener los pesos de la red al mínimo valor. La función del error dada por Vapnik (1998), es:

$$L_{\varepsilon}(d, y(x)) = \begin{cases} |d - y(x)| - \varepsilon & \text{para } |d - y(x)| \geq \varepsilon \\ 0 & \text{para } |d - y(x)| < \varepsilon \end{cases} \quad \text{Ecuación (12)}$$

$\varepsilon$  = nivel de exactitud esperada,  $d$  =destinación predeterminada,  $x$  =vector de entrada,  $y(x)$  =salida actual de la red bajo excitación de  $x$ .

La ecuación que describe a SVM esta dada por:

$$y(x) = \sum_{j=1}^k w_j \phi_j(x) + b = w^T \phi(x) + b \quad \text{Ecuación (13)}$$

$w = (w_1, \dots, w_k)^T$  Vector de pesos,  $b$ = bias,

$\phi(x) = (\phi_1(x), \dots, \phi_k(x))^T$  Función de activación.

En la búsqueda de la solución se introduce la función de Lagrange y multiplicadores de Lagrange  $\alpha_i$ , donde ( $i=1$ , hasta  $p$ ). La ecuación a SVM consiste en:

**Ecuación (14)**

$$\max \left\{ \sum_{i=1}^p d_i (\alpha_i - \alpha_i) - \varepsilon \sum_{i=1}^p (\alpha_i - \alpha_i) - \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p (\alpha_i - \alpha_i) (\alpha_j - \alpha_j) K(x_i, x_j) \right\}$$

*sujeta a las restricciones:*

$$\sum_{i=1}^p (\alpha_i - \alpha_i) = 0 \quad \text{Ecuación (15)}$$

$$0 \leq \alpha_i \leq C, \quad 0 \leq \alpha_i \leq C$$

$K(x_i, x_j) = \varphi^T(x_i) \varphi(x_j)$ , es el valor kernel de la función, de acuerdo al teorema de Mercer, definido por el grupo de datos de aprendizaje dados por  $x$  (Vapnik, 1998).

Básicamente, las SVM trabajan solamente con las instancias que están en las fronteras de decisión (denominados vectores soporte) de forma que, mediante programación cuadrática (QP), tratan de inducir separadores lineales o hiperplanos que maximicen la distancia mínima entre clases.

Las máquinas de vectores soporte aportan mejoras a los métodos clásicos de aprendizaje: la talla de la red no se fija desde un principio y se garantiza matemáticamente el máximo nivel de generalización (Martínez-de-Pisón et al., 2008).

## 5.2. Datos Utilizados

### 5.2.1. Fuentes de datos

Cada vez se disponen para la comunidad científica de un mayor número de relevantes y serias fuentes de información. A lo largo de la franja fronteriza México-Estados Unidos, en los últimos diez años se han instalado redes de monitoreo de contaminantes y condiciones atmosféricas con almacenamientos masivos de datos e información general relevante sobre los tipos de contaminantes presentes y característicos de cada región.

De esta forma se pueden obtener las concentraciones, legislación actualizada, afectaciones a la salud de las personas y al medio ambiente, datos históricos de los contaminantes, datos meteorológicos y otros factores de cada zona. Además de investigaciones desarrolladas por grupos de trabajo, se dispone de bases de datos de reconocimiento científico abiertas a investigadores y personas interesadas en profundizar en esta problemática.

Actualmente, los monitores se encuentran instalados en sitios estratégicos de ambos lados de la frontera México-Estados Unidos. Estos registran de manera continua contaminantes existentes en el aire y otros factores meteorológicos.

Para el desarrollo de esta tesis, además de información proporcionada por la EPA (Environmental Protection Agency), ARB (Air Resources Board) y CALEPA (California Environmental Protection Agency), incluyendo las dependencias mexicanas SEMARNAT (Secretaría de Marina y Recursos Naturales), INE (Instituto Nacional de Ecología) e INEGI (Instituto Nacional de Estadística Geografía e Informática), entre otras instituciones. Se utilizó información recopilada en [http://www.arb.ca.gov/gismo/aqmis\\_v3/](http://www.arb.ca.gov/gismo/aqmis_v3/), página que pertenece a la Air Resources Board (ARB), que proporciona datos de las concentraciones de los principales contaminantes y parámetros relacionados con la calidad del aire de cada uno de los monitores instalados en California, además de otra información relevante.

De manera particular fue importante la descarga de información de Texas en: [http://www.tceq.state.tx.us/cgi-bin/compliance/monops/site\\_info](http://www.tceq.state.tx.us/cgi-bin/compliance/monops/site_info), que proporciona registros de monitores tanto de los monitores instalados en Texas, como los monitores de Ciudad Juárez (ciudad mexicana junto a El Paso, Texas).

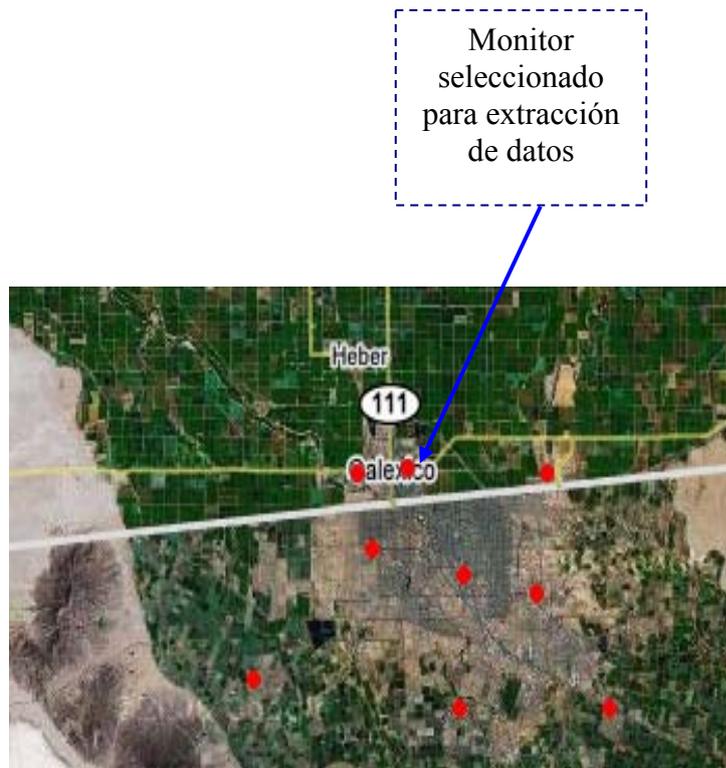
El estado de Baja California enlaza los monitores fronterizos existentes en las ciudades de Tijuana, Tecate, Rosarito y Mexicali en la página web siguiente: <http://aire.bajacalifornia.gob.mx/eng/site.cfm>, además de los siguientes enlaces: <http://www.epa.gov/air/data/geosel.html> y <http://www.epa.gov/air/data/reports.html>. Las páginas anteriores fueron de gran utilidad para la recopilación de información.

De manera exhaustiva, se buscaron y recopilaron en matrices los datos horarios disponibles de las fuentes antes citadas de los contaminantes atmosféricos y meteorológicos relacionados con el contaminante a modelar en cada zona de estudio. (PM<sub>2.5</sub> en la zona Ciudad Juárez- El Paso, Ozono en la región Mexicali- Calexico).

## 5.2.2. Zona Mexicali- Calexico

En la zona de estudio de Mexicali-Calexico se adoptó como principal generador de datos el monitor identificado con el número 060250005. Este monitor está situado geográficamente en latitud  $32^{\circ} 40' 34''$ , longitud  $115^{\circ} 28' 59''$  a 6 m. sobre el nivel del mar en la Ciudad de Calexico en California. Es administrado por la CARB (California Air Resources Board) en Estados Unidos.

En el mapa de la Figura 15 puede observarse la ubicación, siendo esta prácticamente sobre la línea de la frontera con Mexicali, Baja California.



**Figura 15.** Monitor fuente principal de datos atmosféricos y meteorológicos analizados para el desarrollo de modelos de predicción de ozono troposférico.

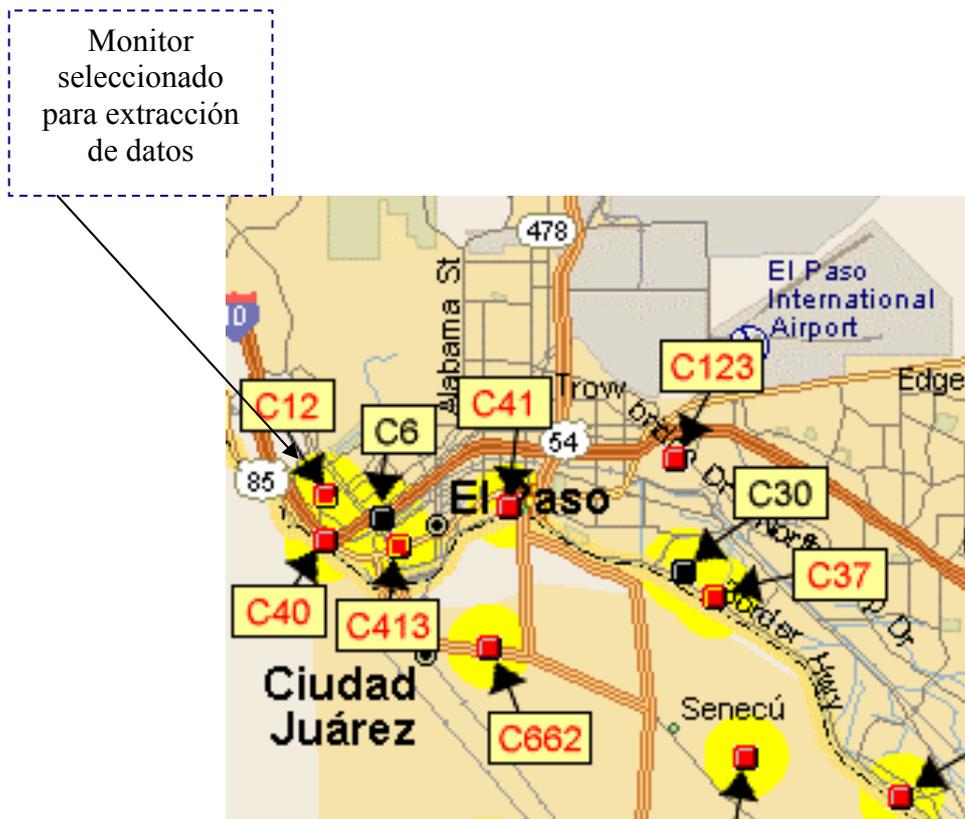
(Figura adaptada de:

[http://www.arb.ca.gov/qaweb/mapdemo/map\\_module.php](http://www.arb.ca.gov/qaweb/mapdemo/map_module.php))

### 5.2.3. Zona Ciudad Juárez-El Paso

En el caso de la zona de estudio de la frontera de Ciudad Juárez, Chihuahua y El Paso en Texas, el monitor principal de datos de contaminantes atmosféricos y parámetros meteorológicos está identificado con el número 48-141-0037, que también recibe el nombre de: “El Paso UTEP-C12”.

Esta estación de monitoreo esta localizada en El Paso, Texas en las coordenadas geográficas 31° 46' 06" Norte, 106° 30' 05" Oeste a 1158 m. sobre el nivel del mar y es administrado por EPA- Environmental Protection Agency (Ver Figura 16).



**Figura 16.** Monitor C12, fuente principal de datos atmosféricos y meteorológicos. (Mapa adaptado de material de Texas Commission on Environmental Quality, 2007).

### 5.3. Modelos de Predicción desarrollados

#### 5.3.1. Modelos de predicción de PM<sub>2.5</sub>

Para el desarrollo de los modelos de predicción de PM<sub>2.5</sub>, se recopilaron los datos disponibles en base horaria (24h al día) de las variables: *PM<sub>2.5</sub>*, *Temperatura*, *Humedad Relativa*, *Velocidad del Viento* y *Resultante de la Velocidad del Viento*, parámetros mayormente relacionados con las altas concentraciones del material particulado (Jorquera, 2001; Rodríguez et al., 2001; Yang, 2002).

A partir de esa primera matriz de datos, se utilizó la información de las primeras ocho horas del día, de los parámetros indicados en la Tabla 5, buscando la predicción del valor máximo de PM<sub>2.5</sub> para las 16 horas restantes de ese mismo día. Los datos recopilados fueron los correspondientes al periodo 2000 al 2002, adoptando la siguiente segmentación: las medias diarias de las primeras 8 horas del día de los años 2000 y 2001 usadas para los entrenamientos de los modelos con las diferentes metodologías, mientras que los datos correspondientes al año 2002 se utilizaron para la validación y pruebas.

El Paso (US) – Ciudad Juárez (México) Matriz de datos		
Parámetros Registrados	Símbolo	Unidades
Media de <i>PM<sub>2.5</sub></i> de las primeras 8h del día	Pm25m8	µg/m <sup>3</sup>
Máximo nivel de <i>PM<sub>2.5</sub></i> durante las primeras 8h del día	Pm25max8	µg/m <sup>3</sup>
Media de <i>Temperatura</i> de las primeras 8h del día	Tam8	°F
Media de <i>Humedad Relativa</i> de las primeras 8h del día	Hrm8	%
Media de <i>Velocidad del Viento</i> de las primeras 8h del día	Vvm8	m/s
Media de <i>Resultante de la Velocidad del Viento</i> de las primeras 8h del día	rdvm8	-
Seno de <i>Resultante de la Velocidad del Viento</i> de las primeras 8h del día (índice de la dirección del viento)	rdvsin8	-

**Tabla 5.** Parámetros utilizados para el desarrollo los modelos de predicción de las partículas menores a dos y media micras (PM<sub>2.5</sub>).

### 5.3.2. Modelos de predicción de Ozono

En la región de Calexico-Mexicali, se juzgó importante evaluar dos experimentos que se refieren a dos matrices de datos con rangos de información distintos: Una primera matriz (Matriz A) generada con los datos recabados durante 1999 al 2004 (los datos del 2001 no se incluyen por no existencia) de las variables siguientes: ozono, temperatura, dióxido de nitrógeno, monóxido nitrógeno, óxidos nitrógeno, resultante de la velocidad del viento, resultante de la dirección del viento, monóxido de carbono, presión barométrica, radiación solar.

La estructura final de la Matriz A, disponible para realizar el entrenamiento con las técnicas no paramétricas, se creó a partir de los datos medios del día anterior (24 horas antes) de las variables antes mencionadas y una variable adicional, el Ozono máximo del día anterior a la predicción (24h antes). Se buscó predecir la concentración máxima de ozono diario, tomando en cuenta los valores de las distintas variables predictoras del día anterior. En la Tabla 6 puede apreciarse la estructura de ambas matrices.

Una segunda matriz (*Matriz B*), fue creada con los valores medios de las primeras 6h del día (De 0 a 5h) durante los años 1997 al 2005 (a excepción del 2001 por no disponer de información suficiente). La matriz se formó por las siguientes variables: ozono, temperatura, dióxido de nitrógeno, monóxido de nitrógeno, óxidos de nitrógeno, resultante de la velocidad del viento, humedad relativa y monóxido de carbono. Se agregó una columna adicional dentro de ésta matriz que contenía los valores de las concentraciones máximas de ozono registradas en el día anterior. Para mayor claridad, ver la Tabla 6.

Región Calexico– Mexicali				
Matrices de datos				
VARIABLES REGISTRADAS	SÍMBOLO	UNIDADES	MATRIZ A	MATRIZ B
Monóxido de Carbono	CO	ppb	√	√
Monóxido de Nitrógeno	NO	ppb	√	√
Dióxido de Nitrógeno	NO <sub>2</sub>	ppb	√	√
Óxidos de Nitrógeno	NO <sub>x</sub>	ppb	√	√
Ozono	O <sub>3</sub>	ppb	√	√
Resultante de la velocidad del Viento	RWS	m/seg	√	√
Resultante de la Dirección del Viento	RWD	Rad.	√	--
Presión Atmosférica	BP	Pa	√	--
Temperatura	T	C	√	√
Humedad relativa	HR	%	-	√
Radiación solar	RS	watts/m <sup>2</sup>	√	--

**Tabla 6.** Parámetros Registrados utilizados para el desarrollo los modelos de predicción de ozono.

#### 5.4 Tratamiento de los huecos en los datos y total de datos analizados.

De acuerdo a las matrices de datos utilizadas, los registros con algún hueco en los datos (valores inexistentes) que se observaran en alguna de las variables de la tabla, fueron eliminados durante el procesamiento previo de los datos.

De esta forma se evitó el uso de técnicas de imputación para evitar acumulación de errores en los valores de las concentraciones (Dixon 1979). La Matriz A quedó conformada con 1343 instancias de datos completos, y la Matriz B se redujo a 2367 registros completos.

## 5.5 Las Herramientas

Esencialmente se usaron herramientas GPL (General Public License) que gozan de gran reconocimiento científico por distintas comunidades de investigadores por su versatilidad y aplicabilidad. Todas distribuidas de manera gratuita vía internet como el programa R, el cual fue utilizado en el procesamiento y evaluación de los datos, así como para desarrollar los modelos de *Regresión Lineal*, *Regresión Ridge*, y el modelo de *Máquina de Vectores Soporte* (SVM). El programa computacional *Stuttgart Neural Network Simulator* (SNNS) fue útil para el desarrollo de las Redes Neuronales Artificiales.

Las herramientas anteriores que son de acceso libre y gratuito. Están disponibles en <http://www.r-project.org/> (para R) y en <http://www-ra.informatik.uni-tuebingen.de/SNNS/> para redes neuronales. Herramientas proporcionadas por NODELIB también fueron útiles para el procesamiento de los datos.



# Capítulo 6

## Resúmenes de Publicaciones

---

Dado a que la opción seleccionada para la obtención del grado de doctor es de acuerdo al sistema de **“Tesis por Compendio de Publicaciones”**, se anexan a partir de la siguiente página los resúmenes bilingües de tres trabajos publicados.

## 6.1. Primer artículo publicado (en revista indexada)

doi : 10.1016/j.envsof.2004.03.010

### 6.1.a Resumen en inglés

#### **Neural network prediction model for fine particulate matter (PM<sub>2.5</sub>) on the US-México border in El Paso (Texas) and Ciudad Juárez (Chihuahua)**

Ordieres, J.B.<sup>a</sup>, Vergara, E.P.<sup>a</sup>, Capuz, S.<sup>b</sup>, Salazar, E.<sup>c</sup>

<sup>a</sup> Universidad de la Rioja, c/ Luis de Ulloa 20, 26004, Logroño, La Rioja, España

<sup>b</sup> Univ. Politécnica de Valencia, Camino de Vera, S/n. 46022, Valencia, España

<sup>c</sup> Inst. Tecnológico de Mexicali, Av. Tecnológico, S/n Col. Elías Calles, 21396 M., México

**Environmental Modelling & Software 20 (2005) 547-559**

#### **Abstract**

The daily average PM<sub>2.5</sub> concentration forecast is a leading component nowadays in air quality research, which is necessary to perform in order to assess the impact of air on the health and welfare of every living being. The present work is aimed at analyzing and benchmarking a neural-network approach to the prediction of average PM<sub>2.5</sub> concentrations. The model thus obtained will be indispensable, as a control tool, for the purpose of preventing dangerous situations that may arise. To this end we have obtained data and measurements based on samples taken during the early hours of the day.

Results from three different topologies of neural networks were compared so as to identify their potential uses, or rather, their strengths and weaknesses: Multilayer Perceptron (MLP), Radial Basis Function (RBF) and Square Multilayer Perceptron (SMLP). Moreover, two classical models were built (a persistence model and a linear regression), so as to compare their results with the ones provided by the neural network models. The results clearly demonstrated that the neural approach not only outperformed the classical models but also showed fairly similar values among different topologies.

Moreover, a differential behavior in terms of stability and length of the training phase emerged during testing as well. The RBF shows up to be the network with the shortest training times, combined with a greater stability during the prediction stage, thus characterizing this topology as an ideal solution for its

use in environmental applications instead of the widely used and less effective MLP.

**Keywords:** US-Mexico border; Air quality; Particulate matter; PM<sub>2.5</sub>; Neural Network Modeling; Multilayer Perceptron (MLP); Radial Basis Function (RBF); Square Multilayer Perceptron (SMLP)

### 6.1.b Resumen en español

#### **Modelo de Predicción con Redes Neuronales para material de partículas finas (PM<sub>2.5</sub>) en la Frontera US-México en El Paso (Texas) y Ciudad Juárez (Chihuahua)**

Ordieres, J.B.<sup>a</sup>, Vergara, E.P.<sup>a</sup>, Capuz, S.<sup>b</sup>, Salazar, E.<sup>c</sup>

<sup>a</sup> Universidad de la Rioja, c/ Luis de Ulloa 20, 26004, Logroño, La Rioja, España

<sup>b</sup> Univ. Politécnica de Valencia, Camino de Vera, S/n. 46022, Valencia, España

<sup>c</sup> Inst. Tecnológico de Mexicali, Av. Tecnológico, S/n Col. Elías Calles, 21396 M., México

**Environmental Modelling & Software 20 (2005) 547-559**

#### **Resumen**

El pronóstico promedio diario de las concentraciones de PM<sub>2.5</sub> es una parte importante en las investigaciones en calidad del aire, éste es necesario desarrollar para evaluar el impacto del aire en la salud y bienestar de cada ser vivo. El presente trabajo tiene el ánimo de analizar y comparar mejoras con redes neuronales artificiales la predicción del promedio diario de concentraciones de PM<sub>2.5</sub>. El modelo obtenido podría ser indispensable como herramienta de control, con propósitos de prevención de situaciones peligrosas a la salud que pudieran presentarse. Para este fin se han obtenido datos de concentraciones de las primeras horas del día.

Resultados de tres diferentes topologías de redes neuronales artificiales fueron comparados para identificar sus potenciales usos o bien sus fortalezas y debilidades. Las herramientas utilizadas fueron: Perceptrón Multicapa (MLP), Función Base Radial (RBF) y Perceptrón Multicapa Cuadrática (SMLP). Además, dos modelos clásicos fueron construidos (un modelo de persistencia y uno de regresión lineal), como comparativos de sus resultados respecto a los obtenidos con redes neuronales artificiales. Los resultados claramente

demonstraron que superan a los modelos clásicos, todas las topologías mantuvieron comportamientos similares en ese sentido.

Además se visualizaron durante las evaluaciones comportamientos diferenciados en términos de estabilidad y longitud de fase de entrenamiento. La red RBF fue la red que requirió menos tiempo de entrenamiento, combinando gran estabilidad durante la predicción, por lo que la caracterización de esta topología resulta ser la mejor solución para aplicaciones medio ambientales en lugar de la técnica ampliamente usada pero que menos efectiva en este estudio: MLP.

**Palabras clave:** Modelización; material particulado;  $PM_{2.5}$ ; Función de Base Radial (RBF); Perceptrón Multicapa Cuadrática (SMLP); Frontera US-México.

## 6.2 Segundo artículo publicado (en revista indexada)

doi : 10.1016/j.envsof.2007.11.009

### 6.2.a Resumen en Inglés

#### **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)**

E. Salazar-Ruiz<sup>a</sup>, J.B. Ordieres<sup>b,\*</sup>, E.P. Vergara<sup>b</sup>, S.F. Capuz-Rizo<sup>c</sup>

<sup>a</sup>Instituto Tecnológico de Mexicali, Av. Tecnológico, s/n Col. Elías Calles, 21396 Mexicali B.C. México

<sup>b</sup>Universidad de La Rioja, c/ Luis de Ulloa 20, 26004, Logroño, La Rioja, España.

<sup>c</sup>Universidad Politécnica de Valencia, Camino de Vera, s/n. 46022 Valencia, España

**Environmental Modelling & Software; in press available on line.**

(Received 13 August 2007; received in revised form 25 November 2007; accepted 28 November 2007)

#### **Abstract**

This study developed 12 prediction models using two types of data matrix (daily means and a selection of the mean for the first 6 h of the day). The Persistence parametric prediction technique was applied separately to these matrices, as well as semiparametric Ridge Regression and three non-parametric or artificial intelligence techniques: Support Vector Machine, Multilayer Perceptron and ELMAN networks.

The target was the prediction of maximum tropospheric ozone concentrations for the next day in the Mexicali and Calexico border area. The main ozone precursors and meteorological parameters were used for the different models. The proposals were evaluated using specific performance measurements for the air quality models established in the Model Validation Kit and recommended by the US Environmental Protection Agency.

Results with similar margins of error were obtained in various models developed in this study, and some of them have provided smaller margins of error than similar prediction models existing in the literature developed in

other regions. For this reason, we consider it feasible to apply the prediction models developed and they could be useful for supporting decisions in the matter of ozone pollution in the region under study, as well as for use in daily forecasting in this area.

Keywords: US-México border; Ozone neural network modeling; Multilayer Perceptron (MLP); ELMAN neural network; Support Vector Machine; Ridge Regression; Model Validation Kit (MvK); Transboundary air quality.

### 6.2.b Resumen en español

#### **Desarrollo y análisis comparativo de los modelos de predicción de ozono troposférico basados en herramientas lineales y de inteligencia artificial en Mexicali, Baja California (México), y Calexico California (US).**

**E. Salazar-Ruiz<sup>a</sup>, J.B. Ordieres<sup>b,\*</sup>, E.P. Vergara<sup>b</sup>, S.F. Capuz-Rizo<sup>c</sup>**

<sup>a</sup>Instituto Tecnológico de Mexicali, Av. Tecnológico, s/n Col. Elías Calles, 21396 Mexicali B.C. México

<sup>b</sup>Universidad de La Rioja, c/ Luis de Ulloa 20, 26004, Logroño, La Rioja, España.

<sup>c</sup>Universidad Politécnica de Valencia, Camino de Vera, s/n. 46022 Valencia, España

#### **Environmental Modelling & Software; disponible on line**

(Recibido 13 Agosto 2007; revisado 25 Noviembre 2007; aceptado 28 Noviembre 2007)

#### **Resumen**

Este estudio desarrolla doce modelos de predicción usando dos tipos de matrices de datos (medias diarias y una selección de la media de las primeras seis horas del día). Las técnicas paramétricas *Persistencia* y *Regresión lineal* fueron aplicadas, así como la herramienta semi-paramétrica *Regresión Ridge* y tres herramientas de inteligencia artificial: *Máquina de Vectores Soporte (SVM)*, *Perceptrón Multicapa (MLP)*, y *redes ELMAN*.

El objetivo principal fue la predicción de las concentraciones máximas de ozono para día siguiente. Los principales precursores del ozono y datos meteorológicos fueron usados en los diferentes modelos. Se tuvo también el propósito de evaluar las medidas de desempeño para los modelos de calidad

del aire establecidos en el llamado “Model Validation Kit” y recomendadas por la Agencia de Protección al Medio Ambiente de Estados Unidos.

Fueron encontrados resultados con márgenes similares de error para varios modelos desarrollados en este estudio, pero al menos uno de ellos mostró márgenes de error menores a predicciones similares de modelos desarrollados para otras regiones. Por esta razón, consideramos factible la aplicación de modelos de predicción aquí desarrollados y podrían ser útiles para dar soporte a decisiones en material de contaminación por ozono en la región de estudio y su uso para predicciones diarias.

**Palabras clave:** Frontera US-México, Modelización de ozono con redes neuronales, Perceptrón Multicapa (MLP), redes neuronales Elman, Máquina de Vectores Soporte (SVM), Regresión “Ridge”, Paquete de Validación de Modelos (MvK), Calidad del aire transfronteriza

## 6.3 Artículo publicado en Congreso Internacional

### 6.3.a Resumen en Inglés.

#### **Comparison and Analysis of Statistical and Artificial Intelligence Techniques for the Modelling of Tropospheric Ozone. Emphasis in the U.S-México Border.**

J.B. Ordieres Meré <sup>a</sup>, E. Salazar Ruiz <sup>bcp</sup>, S. Capuz Rizo <sup>c</sup>

<sup>a</sup>*Universidad de la Rioja,*

<sup>b</sup>*Instituto Tecnológico de Mexicali,*

<sup>c</sup>*Universidad Politécnica de Valencia.*

**X International Congress on Project Engineering, Valencia, Spain, 13-15 September, 2006, pp. 1534-1545**

#### **Abstract**

The situation with regard with the air quality in the border basin of Mexico - United States is real motive of attention for both nations. The continuous violation of the procedure and limits established of pollutants' concentrations puts in risk the human health and the development sustainable over that zone.

This paper does a evaluating about methods used for ozone prediction models, and the most important factors used for the study and pollutants modelling in the air in the border of Mexico with United States.

**Keywords:** tropospheric ozone, prediction model, artificial intelligence, air pollutants.

### 6.3.b Resumen en español.

#### **Comparación y Análisis de Técnicas Estadísticas Simples y de Inteligencia Artificial usadas en la Modelización de Ozono Troposférico. Énfasis en la Franja Fronteriza México-Estados Unidos**

J.B. Ordieres Meré <sup>a</sup>, E. Salazar Ruiz <sup>bc</sup>, S. Capuz Rizo <sup>c</sup>

<sup>a</sup> *Universidad de la Rioja*

<sup>b</sup> *Instituto Tecnológico de Mexicali*

<sup>c</sup> *Universidad Politécnica de Valencia.*

**X Congreso Internacional de Ingeniería de Proyectos. Valencia, España, 13-15 Septiembre, 2006, pp. 1534-1545**

#### **Resumen**

La situación respecto a la calidad del aire en la franja fronteriza de México-Estados Unidos es motivo de preocupación para ambas naciones. La continua violación de las normas y de los límites establecidos de las concentraciones de contaminantes pone en riesgo la salud humana y el desarrollo sustentable.

Esta ponencia tiene el propósito de evaluar y dar a conocer qué se investiga respecto a modelos de predicción de ozono, que factores se tomaron para el estudio y modelización de contaminantes en el aire como el ozono troposférico en la franja fronteriza de México con Estados Unidos.

**Palabras clave:** Ozono troposférico, técnicas de modelización, inteligencia artificial, contaminantes en el aire.



# Capítulo 7

## Discusión de Resultados

---

### 7.1 Modelo de partículas suspendidas $PM_{2.5}$

Éste estudio fue llevado a cabo a partir de los datos obtenidos en la zona *Ciudad Juárez, Chihuahua, México – El Paso, Texas, Estados Unidos*. Se aplicaron cinco técnicas de desarrollo de modelos de predicción:

- Dos técnicas de naturaleza *Paramétrica*: Persistencia y Regresión Lineal
- Tres técnicas *No Paramétricas*: Red neuronal Perceptrón Multicapa (MLP), Función de Base Radial (RBF) y Perceptrón Multicapa Cuadrático (SMLP).

Para una de las técnicas no paramétricas (RBF) se realizaron 5 modelos, cada uno con varianzas diferentes en sus funciones radiales. Por lo que finalmente se evaluaron 9 modelos de predicción: Siete modelos no paramétricos (5 con RBF, 1 con MLP, 1 con SMLP) y dos paramétricos.

Los resultados de los modelos paramétricos sirvieron como referencia con respecto a los resultados de los modelos no paramétricos. Esto se debe, a la naturaleza no lineal del problema y a la presencia de un gran número de las variables predictoras involucradas en el modelado (incluyendo entre ellas, el mismo contaminante a predecir).

Como datos de entrada se tomaron los valores de las primeras ocho horas del día (de 0h a 8h del un día de 24 h.) para determinar la concentración de  $PM_{2.5}$  de las restantes 16 horas del día.

Se tomó en cuenta que generando la predicción en las primeras 8 horas del día, se brindaba la oportunidad a que gran parte de la población pudiera ser informada, de la concentración que se obtendría del mencionado contaminante durante las horas pico y el resto del día.

Las variables predictoras involucradas fueron:

1. Velocidad del viento.
2. Resultante de la Dirección del viento.
3. Índice de la dirección del viento (seno de la dirección del viento).
4. Humedad Relativa.
5. Temperatura Ambiente.
6. Valor medio de  $PM_{2.5}$ .
7. Valor máximo de  $PM_{2.5}$ .

En los modelos de predicción con redes neuronales se agruparon los datos horarios del 2000 y 2001 para entrenamiento y los datos del 2002 como datos de testeo.

Los modelos se formaron con siete neuronas de entrada, dada las siete variables predictoras usadas, y una neurona de salida (correspondiente a la variable  $PM_{2.5}$  a predecir). El número de neuronas de la capa oculta se evaluó con el método de prueba y error (desde una neurona hasta 30)

seleccionando el número de neuronas que proporcionaba el menor error absoluto de entrenamiento (MAE).

El menor Error Medio Absoluto (MAE) se encontró con 18 neuronas en la capa oculta. La regla de aprendizaje utilizada fue la Retropropagación con Momento (BMP: Back Propagation with Momentum) requerida para el modelo MLP.

Para el modelo con SMLP se requirieron 20 neuronas en la capa oculta y para los modelos RBF se evaluaron los errores Error Medio Absoluto (MAE) de testeo usando distintas varianzas en la función gaussiana de activación de las neuronas. Estas varianzas fueron: 0.25, 0.5, 0.75, 1.5 y 3.5.

De acuerdo a las medidas de error obtenidas, se puede asumir que los modelos RBF fueron los más acertados, pero, si se observan los resultados calculados, se pueden ver resultados similares entre los distintos modelos de redes neuronales desarrollados (Tabla 7).

El RMSE del modelo SMLP con 20 neuronas en la capa oculta registra un valor de  $0.0028 \text{ PM}_{2.5} \mu\text{g m}^{-3}$  que, comparativamente con la misma medida de error del modelo RBF con 20 y 16 neuronas en la capa oculta es de  $0.0023 \mu\text{g m}^{-3}$ , por lo que la diferencia es de  $0.00005 \mu\text{g m}^{-3}$  entre ambos modelos.

Respecto al MAE, coincide la similitud de resultados entre el modelo RBF (varianza 3.5) con 10 neuronas en la capa oculta, y el modelo SMLP. Entre ambos modelos hay una diferencia MAE de  $0.00017 \mu\text{g m}^{-3}$ .

Es importante mencionar que una ventaja para la red RBF, es que requirió de menos tiempo de entrenamiento que el resto de los modelos de redes neuronales artificiales entrenadas.

<b>Medidas Desempeño</b>		
<b>RMSE</b>		
<i>Modelo</i>	<i>Topología</i>	<i>Resultado (<math>\mu\text{g m}^{-3}</math>)</i>
Persistencia	NA	0.41
Regresión Lineal	NA	0.27
MLP	7-18-1	0.00158
SMLP	7-20-1	0.00128
RBF Var=0,25	7-21-1	0.00132
RBF Var=0.50	7-20-1	0.00123
RBF Var=0.75	7-16-1	0.00123
RBF Var=1.5	7-10-1	0.00130
RBF Var=3.5	7-10-1	0.00127
<b>MAE</b>		
Persistencia	NA	4.41
Regresión Lineal	NA	0.02
MLP	7-18-1	0.00358
SMLP	7-20-1	0.00308
RBF Var=0,25	7-21-1	0.00321
RBF Var=0.50	7-20-1	0.00296
RBF Var=0.75	7-16-1	0.00293
RBF Var=1.5	7-10-1	0.00295
RBF Var=3.5	7-10-1	0.00291
<b>R<sup>2</sup></b>		
Persistence	NA	0.0806
Linear regression	NA	0.3983
MLP	7-18-1	0.3805
SMLP	7-20-1	0.3712
RBF Var=0,25	7-21-1	0.4611
RBF Var=0.50	7-20-1	0.4554
RBF Var=0.75	7-16-1	0.4451
RBF Var=1.5	7-10-1	0.4380
RBF Var=3.5	7-10-1	0.4143

**Tabla 7.** Medidas de rendimiento calculadas para los distintos modelos desarrollados para la predicción de PM<sub>2.5</sub>

Existen cada vez más trabajos de otros autores desarrollando modelos de predicción usando redes neuronales y cada uno propone distintas metodologías: usando diferentes medidas de rendimiento, diversos intervalos de datos, etcétera. Los modelos de predicción de  $PM_{2.5}$  existentes para otras regiones también presentan una diversidad de parámetros que se han tomado en cuenta para su formulación, así como temporalidad, cantidad y tipos distintos de datos, además de distintas medidas de eficiencia de los mismos.

Trabajos similares como los de Pérez et al. (2000), han desarrollado modelos no lineales para predecir  $PM_{2.5}$  usando redes neuronales en Santiago de Chile. Con errores de predicción de un rango del 30% al 60% en horas tempranas y hasta un 60% en horas de la tarde.

Pérez & Reyes (2002) desarrollan un modelo para predecir promedios diarios de concentraciones  $PM_{10}$  con 30h de anticipación, con una eficiencia de un 20% de error en la predicción.

Después de ser publicado el primer resultado de este trabajo de tesis, correspondiente precisamente a la modelización de  $PM_{2.5}$ , se han desarrollado nuevos modelos similares, como el de Salini et al (2006) que aplica también herramientas de redes neuronales (MLP con 7 neuronas en la capa oculta) para predecir  $PM_{2.5}$ , del cual obtiene un RNMSE de 0.5156. H. Niska et al. (2005) hace una evaluación de la integración de un modelo MLP para predecir de manera horaria varios contaminantes en el aire (entre ellos  $PM_{2.5}$ ), con modelos de predicción numérica de parámetros ambientales climáticos, los cuales presentan errores medios RMSE cercanos a  $4.6 \mu\text{g m}^{-3}$ , MAE de  $3.34 \mu\text{g}/\text{m}^{-3}$  y  $R^2$  de 0.44.

Entre otros resultados de distintos investigadores presentados al momento, en términos comparativamente son menos precisos con respecto a los modelos de la propuesta mejor de esta tesis que presenta errores RMSE de  $1.23 \mu\text{g m}^{-3}$ , MAE de  $2.93 \mu\text{g m}^{-3}$ , y  $R^2$  de 0.445 y corresponde al modelo para predicciones de medias diarias de  $PM_{2.5}$  con la técnica RBF.

## 7.2 Modelos de Predicción de Ozono Troposférico

Los modelos de predicción de ozono fueron desarrollados en la zona fronteriza Mexicali-Calexico. Se aplicaron en este caso, dos herramientas paramétricas, una semiparamétrica y tres no paramétricas:

- Dos de naturaleza *Paramétrica*: Persistencia y Regresión Lineal
- Una de naturaleza semiparamétrica: Regresión Ridge
- Tres *No Paramétricos*: Redes Neuronales Perceptrón Multicapa (MLP), Redes Elman y Máquinas de Vector Soporte (SVM).

En la región de Calexico - Mexicali, se juzgó importante evaluar dos experimentos que utilizaron dos matrices de datos con rangos de información distintos: Una de ellas se identificó como “*Matriz A*” con datos recabados durante 1999 al 2004 (los datos del 2001 fueron inexistentes) de las variables: ozono, temperatura, dióxido de nitrógeno, monóxido de nitrógeno, óxidos de nitrógeno, resultante de la velocidad del viento, resultante de la dirección del viento, monóxido de carbono, presión barométrica y radiación solar.

La estructura final de la *Matriz A* para efecto de los entrenamientos con las técnicas no paramétricas, partía de datos medios de un día anterior (24 horas antes) de las variables antes mencionadas, incluyendo una variable más: Ozono máximo de un día anterior a la predicción. Por lo que se buscó predecir la concentración máxima del ozono del día, tomando en cuenta los valores de las distintas variables predictoras de un día anterior.

Una segunda matriz denominada: “*Matriz B*”, estaba formada por los valores medios de las primeras 6 horas del día (De 0 a 5h) de los años 1997 al 2005 (a excepción del 2001 ya que no se disponía de información suficiente). La matriz tenía registradas las siguientes variables: ozono, temperatura, dióxido de nitrógeno, monóxido nitrógeno, óxidos nitrógeno, resultante de la velocidad del viento, humedad relativa y monóxido de carbono. Se agregó una columna adicional dentro de esta

matriz que contiene los valores de las concentraciones máximas de ozono registradas un día anterior.

De esta forma, la *Matriz A* se formó con 1343 registros completos mientras que la *Matriz B* quedó conformada por 2367 registros completos. La *Matriz A* se dividió en 850 datos para entrenamiento, 426 para testeo y 67 para validación. En cambio la *Matriz B* usó 1499 datos de entrenamiento, 750 datos para pruebas y 118 datos para validación.

Para determinar el número de neuronas en la capa oculta se evaluaron desde 1 hasta 29 neuronas, finalmente se seleccionaron 19 neuronas en la capa oculta para la *Matriz A* y 20 neuronas en la capa oculta para la *Matriz B* ya que registraron los mínimos errores de entrenamiento (RMSE de 0.047 ppb y 0.067 ppb para la *Matriz A* y la *Matriz B* respectivamente).

Al igual que en los modelos de las PM<sub>2.5</sub>, se usó como herramienta básica el software R para evaluación y procesamiento de los datos y para el desarrollo de los modelos Lineales y de Regresión Ridge. El programa “*Stuttgart Neural Network Simulator*” (SNNS) fue usado para el desarrollo de los modelos de redes neuronales. Además se usaron algunas herramientas construidas expresamente para la construcción de algunos componentes de los modelos usando librerías de NODELIB (Flake, 1998).

Cada entrenamiento de cada estructura de red usó un promedio de 0.35h, y se desarrollaron más de 350 diferentes estructuras. Las redes ELMAN, aunque no fueron las más precisas, requirieron aproximadamente un 41% menos de tiempo que otras redes para su entrenamiento.

A diferencia de lo que se realizó en el primer artículo, para evaluar la eficiencia del modelado, se decidió utilizar las medidas de rendimiento descrito en “Model Validation Kit” pues, como se menciona en capítulos anteriores, presenta una relación de medidas estandarizadas para evaluar los modelos de calidad del aire y dispersión de contaminantes que se consideró importante adoptar para esta investigación. Los resultados se muestran en las Tablas 8 y 9.

<b>RESULTADOS DE MODELOS USANDO MATRIZ "A"</b>						
<b>Medidas de desempeño</b>	<b>Persistencia</b>	<b>Lineal</b>	<b>R.Ridge</b>	<b>MLP</b>	<b>SVM</b>	<b>ELMAN</b>
RMSE	17.1547	14.0516	14.0568	<b>9.4303</b>	11.4345	10.8929
R	0.5351	0.6124	0.61243	<b>0.7417</b>	0.6050	0.6743
FB	-0.0004	0.0000	0.0000	<b>0.0153</b>	0.0028	0.0186
MG	0.9997	0.9624	0.9624	<b>1.0000</b>	0.9892	1.0000
NMSE	0.1295	0.0870	0.08697	<b>0.0430</b>	0.0624	0.0554
VG	1.1026	1.076	1.0769	<b>1.041</b>	1.0601	1.052
FAC2	0.9590	0.9799	0.9881	<b>1.0000</b>	1	1.0000
MAE	11.4430	9.8307	9.8379	<b>7.5261</b>	8.3845	8.2839
MBE	0.0208	0.0000	0.0000	<b>-0.6959</b>	-0.1308	-0.8599
d2	0.7221	0.7250	0.72497	<b>0.8511</b>	0.7561	0.8218
FV	-0.0009	0.4800	0.4800	<b>0.1209</b>	0.2837	-0.1031

**Tabla 8.** Resultado de los modelos de predicción de ozono usando Matriz A, propuestos en esta tesis.

<b>RESULTADOS DE MODELOS USANDO MATRIZ "A"</b>						
<b>Medidas de desempeño</b>	<b>Persistencia</b>	<b>Lineal</b>	<b>R.Ridge</b>	<b>MLP</b>	<b>SVM</b>	<b>ELMAN</b>
RMSE	19.5819	16.0386	16.0386	<b>13.7856</b>	14.6613	14.3106
R	0.5309	0.5976	0.52332	<b>0.6922</b>	0.6339	0.6012
FB	-0.0002	0.0000	0.0000	<b>-0.0190</b>	0.0169	-0.0429
MG	1.0000	0.9610	0.96104	<b>0.9431</b>	0.9884	0.9215
NMSE	0.1407	0.0891	0.08905	<b>0.0679</b>	0.0784	0.0766
VG	1.128	1.085	1.0855	<b>1.080</b>	1.0815	1.075
FAC2	0.9427	0.9768	0.9805	<b>0.9831</b>	0.9797	0.9718
MAE	13.0718	11.2663	11.2663	<b>10.2666</b>	10.1266	11.0234
MBE	0.0000	0.0000	0.0000	<b>0.0010</b>	-0.8858	0.0022
d2	0.7213	0.7115	0.7115	<b>0.7875</b>	0.7615	0.7282
FV	0.0004	0.5033	0.54883	<b>0.4312</b>	0.3606	0.2242

**Tabla 9.** Resultado de los modelos de predicción de ozono usando Matriz B, propuestos en esta tesis.

Los modelos paramétricos y semiparamétricos presentaron un comportamiento menos eficiente, como era de esperarse; aunque se

esperaba también que los modelos con Regresión Ridge fueran mejores que los de Regresión Lineal.

Los resultados fueron prácticamente los mismos con ambas técnicas, lo que nos indica que la colinealidad entre las variables predictoras fue suprimida durante el procesamiento de los datos tal y como lo muestra la gráficas de correlación de la Figura 17.

En tanto que, los resultados que proporcionan los modelos desarrollados con redes neuronales MLP coinciden con los trabajos de otros autores que han comparado métodos MLP con otras herramientas lineales y no lineales y que muestran a los modelos MLP como los más acertados.

Aunque existe un gran número de trabajos que han publicado temas similares al respecto, cabe mencionar trabajos como los de Gardner y Dorling (1999, 2000) y que utilizaron MLP y métodos de regresión en las predicciones de NOx y NO2 (1999), así como MLR, árboles de regresión.

Yi & Prybutok (1996) en su investigación concluyeron que MLP supera la metodología de Modelos Auto-regresivos Integrados de Medias Móviles (ARIMA) en la predicción del ozono.

En Kolehmainen et al. (2000), también se usaron redes neuronales híbridas y MLP para modelización de contaminantes en el aire encontrando que MLP era superior en la reproducción de comportamientos de contaminantes gaseosos en comparación con la reproducción de las partículas.

Lu et al. (2006), utilizaron dos niveles de clasificaciones de datos en conjunto con MLP para desarrollar 4 distintos modelos para cada una de 4 regiones distintas de Taiwan, encontrando como mejor modelo el de la estación de monitoreo Cutin; con  $R^2=0.70$ , IA= 0.91 y RMSE = 30 ppb.

Mientras que, Wang & Lu (2006), proponen el modelo *Híbrido Monte Carlo* para inicializar una red MLP con algoritmo de entrenamiento PSO (*Particle Swarm Optimization*) y finalmente obtiene errores RMSE que van

de 10.8868 ppb a 17.968 ppb en la región Tsuen Wan y de 15.4205 ppb a 30.5079 ppb en la región de Tung Cheng, ambas de Hong Kong.

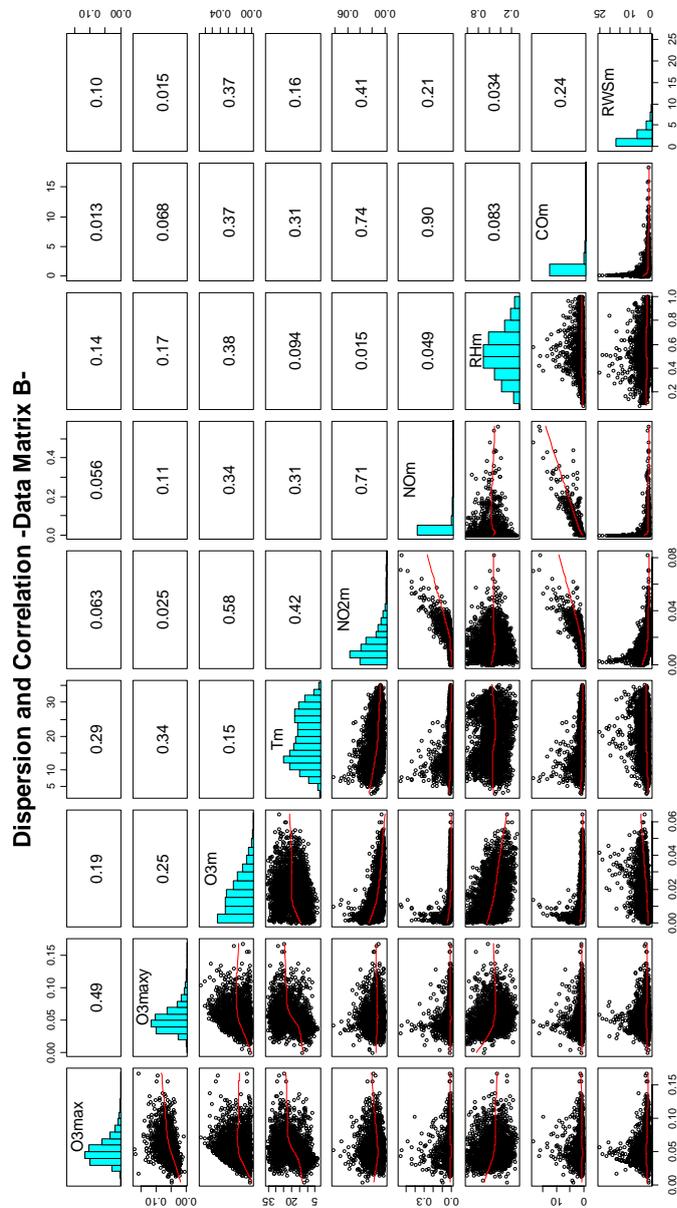


Figura 17. Gráficas de dispersión y correlación entre las distintas variables predictoras correspondientes a la Matriz B

Los modelos predicción de ozono desarrollados por el investigador Sousa et al. (2007) en Oporto, Portugal, estuvieron basados en *Redes Neuronales Artificiales* y *Análisis de Componentes Principales* (PCA) como entradas a la red para predecir ozono horario del día siguiente. El RMSE de entrenamiento obtenido fue de 15.37 ppb.

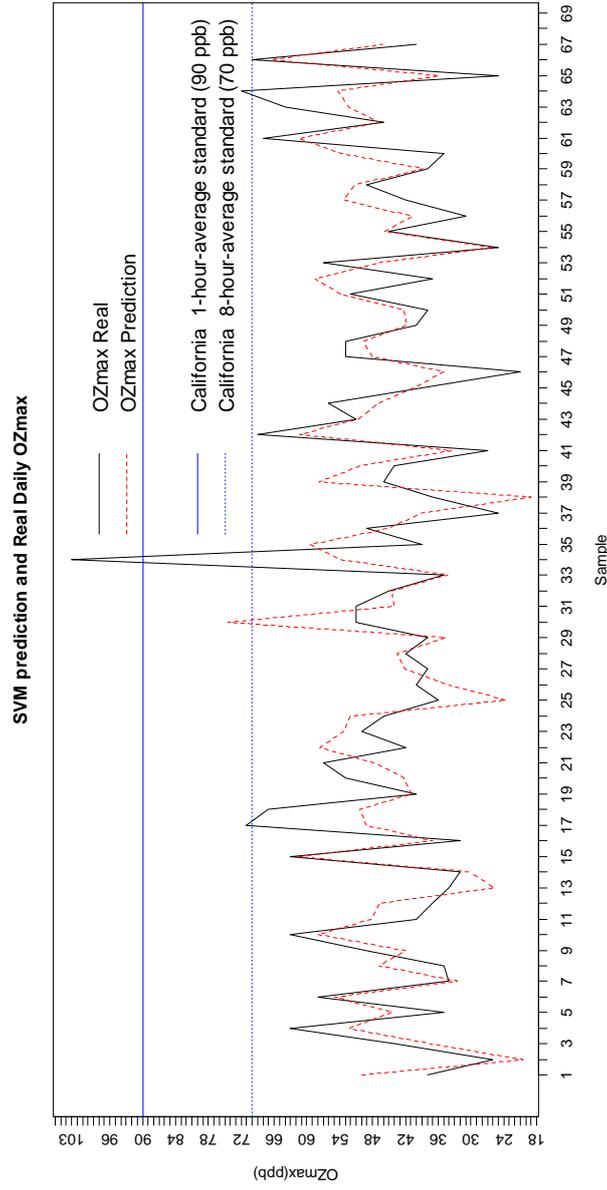
El comportamiento de los modelos desarrollados para esta tesis (mencionando solo los desarrollados con herramientas no paramétricas) correspondientes a la Matriz A, presentan un RMSE de 9.4303 ppb, para el mejor modelo con MLP y un RMSE de 11.4345 ppb para el peor que correspondió a SVM.

En el caso de los resultados de la Matriz B coincidió nuevamente como mejor herramienta las redes MLP con un RMSE de 13.7856 ppb y la peor SVM con RMSE de 14.6613 ppb, aunque su comportamiento general fue claramente peor que los modelos creados con la Matriz A.

Es evidente, que los resultados de modelos obtenidos con la Matriz B demuestran que, los datos de seis horas de anticipación no fueron suficientes para determinar eficientemente el comportamiento del ozono para el resto del día. Sin embargo los valores medios completos de los días anteriores resultan más eficaces para pronosticar el valor del ozono máximo del día siguiente.

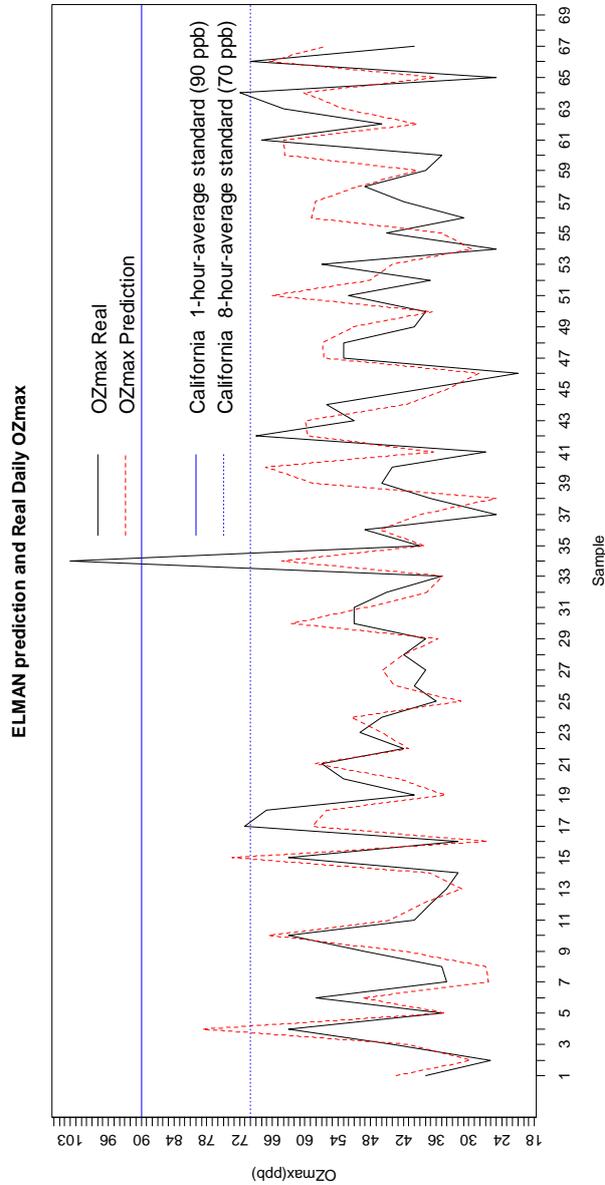
Aspectos adicionales que podemos observar en los comportamientos de los modelos desarrollados para esta tesis se plasman en las graficas de las figuras 18, 19 y 20; estas comparan datos reales y valores predichos resultantes de los distintos modelos desarrollados con las herramientas no paramétricas.

La gráfica de la Figura 18 compara valores reales y predichos con SVM. Muestra un desfase desde el inicio de la predicción y una clara tendencia a subestimar las concentraciones reales.



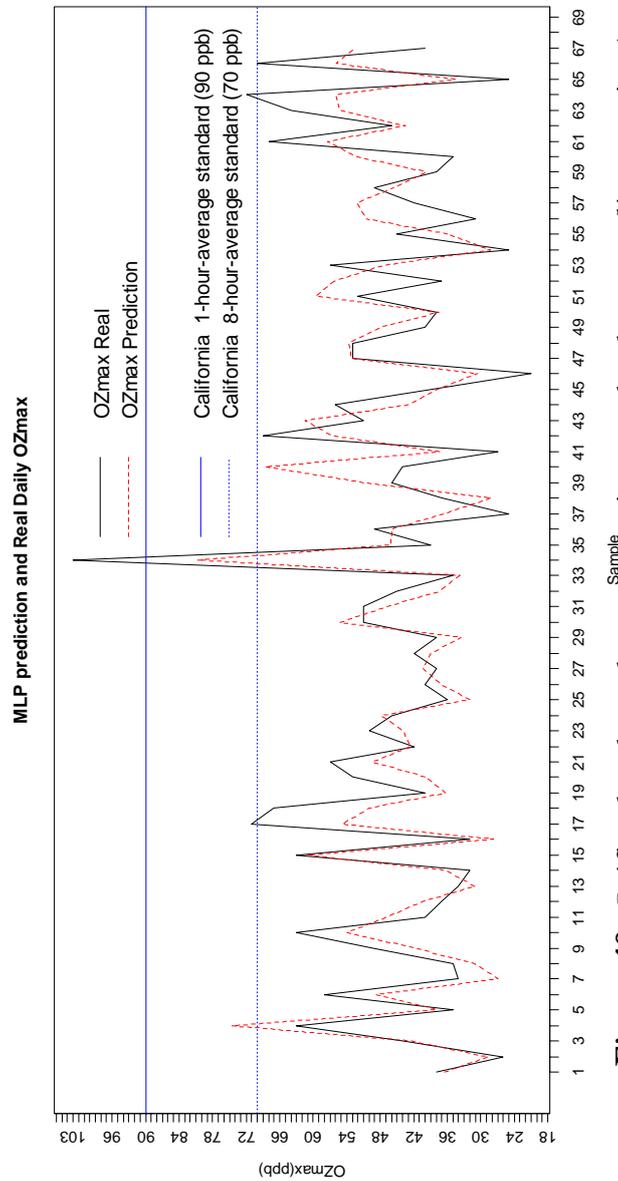
**Figura 18.** Gráfica de valores de concentraciones reales de ozono (línea continua) y sus correspondientes predicciones de acuerdo al modelo SVM (línea punteada).

El modelo ELMAN, inicia con buen comportamiento en sus estimaciones pero aproximadamente, después de la muestra 25, el margen del desfase se amplía considerablemente (Ver Figura 19).



**Figura 19.** Gráfica de valores de concentraciones reales de ozono (línea continua) y sus correspondientes predicciones de acuerdo al modelo ELMAN (línea punteada).

Aunque tiende a sobreestimar el valor de las concentraciones reales en sus predicciones, esto lo podríamos considerar como un punto a favor ya que estamos hablando de evitar daños a la salud y al medio ambiente, por lo que es preferible sobreestimaciones que subestimaciones de concentraciones de los contaminantes modelados en este trabajo.



**Figura 18.** Gráfica de valores de concentraciones reales de ozono (línea continua) y correspondientes predicciones de acuerdo al modelo MLP (línea punteada).

La gráfica del modelo MLP presenta un mejor comportamiento (como puede observarse al comparar la grafica de la Figura 20 respecto a las dos anteriores (Figuras 18 y 19) aunque no logra predecir con exactitud el pico más alto de la concentración de ozono.

De cualquier manera, si observamos la predicción del pico más alto de ozono, comparativamente hablando respecto a las predicciones resultantes de los dos modelos anteriores (SVM y ELMAN), el modelo MLP es el que más se aproxima en predicción a dicho valor máximo real.

Para las redes ELMAN se tenían mejores expectativas de eficacia, pero en este trabajo no lograron ser superiores en resultados a las redes MLP, aunque cabe mencionar que la diferencia no fue muy marcada.

En cambio, el tiempo aplicado de entrenamiento si fue substancialmente menor (41% menor) para las redes ELMAN, comparativamente con el tiempo requerido para el entrenamiento de las redes MLP.



# Capítulo 8

## Conclusiones

---

En las investigaciones iniciales se desarrollaron nueve modelos de predicción de PM<sub>2.5</sub> usando la información de 8 horas anteriores para predecir las 16 horas restantes del día en la región de Ciudad Juárez, Chihuahua, México – El Paso, Texas, Estados Unidos. De los nueve modelos de predicción desarrollados, siete eran con redes neuronales y dos paramétricos.

Los resultados de los modelos paramétricos tuvieron la función de servir de referencia respecto a los resultados de los modelos de Redes Neuronales, esto debido, a la naturaleza no lineal y multifactorial del contaminante a predecir. Se utilizaron 2 años de datos horarios para entrenamiento y datos horarios de un año como datos de prueba.

Al predecir después de las primeras 8 horas del día, se brinda la oportunidad de que gran parte de la población pueda ser informada de la concentración que se tendrá del mencionado contaminante durante las horas pico y el resto del día.

En este tipo de modelos se conocen las neuronas de entrada, que corresponden al número de parámetros o variables predictoras del modelo. Existe solo una neurona de salida, la cual representa a la variable que se busca predecir. Mientras que las neuronas en la capa oculta se determinan a través de prueba y error.

En el caso del modelo con MLP se usaron 18 neuronas para la capa oculta, 20 para el modelo SMLP. Para RBF varió el número de neuronas requeridas de acuerdo a las varianzas manejadas. Para varianzas 0.25, 0.50, 0.75, 1.5 y 3.5 se requirieron 21, 20, 16 y 10 neuronas en la capa oculta, de manera respectiva.

Se concluye que los modelos RBF son los más acertados, aunque con las diferencias relativamente pequeñas, pero si al observar los resultados calculados se pueden ver resultados similares entre los distintos modelos de redes neuronales desarrollados. Como punto a favor para la red RBF, se puede agregar que requirió de menos tiempo de entrenamiento, que el resto de los modelos de redes neuronales artificiales entrenadas.

Trabajos similares de otros autores como los de Pérez et al.(2000) en Santiago de Chile, con errores de predicción de un rango del 30% al 60% en horas tempranas y hasta un 60% en horas de la tarde. También Pérez & Reyes (2002) desarrollan un modelo para predecir promedios diarios de concentraciones  $PM_{10}$  con 30 h de anticipación, con una eficiencia de un 20% de error en la predicción. Modelos como el de Salini et al. (2006), que aplica MLP con 7 neuronas en la capa oculta para predecir  $PM_{2.5}$ , obtiene un RNMSE de 0.5156. Nuevos modelos como el de Niska et al. presentan errores medios RMSE cercanos a  $4.6 \mu\text{g m}^{-3}$ , MAE de  $3.34 \mu\text{g/m}^{-3}$  y  $R^2$  de 0.44.

La mejor propuesta de esta tesis presenta resultados de errores RMSE de  $1.23 \mu\text{g m}^{-3}$ , MAE de  $2.93 \mu\text{g m}^{-3}$ , y  $R^2$  de 0.445, corresponde a un

modelo para predicciones de medias diarias de  $PM_{2.5}$  con la técnica RBF.

En cuanto a los modelos de predicción de ozono que fueron desarrollados en la zona fronteriza Mexicali-Calexico, aplicaron las herramientas de Persistencia y Regresión Lineal, Regresión Ridge, Redes MLP, Redes ELMAN y Máquinas de Vector Soporte (SVM).

Se utilizaron dos matrices de datos con rangos de información distintos. La “*Matriz A*” formada de 1343 renglones de datos completos, datos medios de un día anterior (24 horas antes) recabados de 1999 al 2004 (sin datos del 2001). Se buscó predecir la concentración máxima de ozono del día, tomando en cuenta los valores de las distintas variables predictoras de un día anterior. La segunda matriz, (“*Matriz B*”), fue formada con 2367 registros completos de los valores medios de las primeras 6h del día (De 0 a 5h) durante los años 1997 al 2005 (a excepción del 2001 por inexistencia de información)

La *Matriz A* usa 850 datos para entrenamiento, 426 para pruebas y 67 para validación. En cambio la *Matriz B* usó 1499 datos de entrenamiento, 750 datos para testeo y 118 datos para validación.

Se seleccionaron 19 neuronas en la capa oculta para la *Matriz A* y 20 neuronas en la capa oculta para la *Matriz B* con errores mínimos de entrenamiento (RMSE de 0.047 ppb y 0.067 ppb para *Matriz A* y *Matriz B* respectivamente).

Cada entrenamiento de cada estructura de red usó un promedio de 0.35h, y se desarrollaron más de 350 diferentes estructuras. Las redes ELMAN, aunque no fueron las más precisas requirieron aproximadamente un 41% menos tiempo que otras redes para su entrenamiento.

## Capítulo 8

En lo que respecta a los resultados de error, se utilizó un grupo de medidas de desempeño que intenta estandarizar la evaluación de modelos de calidad del aire dispersión, conocido como: MvK (o Model Validation Kit)

Los modelos resultantes de la Matriz B, muestran que datos de seis horas de anticipación no fueron suficientes para determinar eficientemente el comportamiento del ozono para el resto del día, mas los valores medios de días completos anteriores resultan relativamente eficaces para pronosticar el valor del ozono máximo del día siguiente.

El mejor modelo con MLP, que corresponde a los datos modelados de la Matriz A, resulta un  $RMSE = 9.4303$  ppb y el peor con un  $RMSE = 11.4345$  ppb correspondió a SVM.

En el caso de los resultados de la Matriz B coincidió nuevamente como mejor herramienta MLP con un  $RMSE = 13.7856$  ppb y la peor SVM con  $RMSE = 14.6613$  ppb aunque su comportamiento general fue claramente peor que los modelos utilizando la Matriz A.

Otros autores que han comparado métodos MLP con otras herramientas lineales y no lineales encontrando a los modelos MLP los mas acertados. Gardner & Dorling (1999, 2000), que utilizaron MLP y métodos de regresión en las predicciones de NO<sub>x</sub> y NO<sub>2</sub> (1999), así como MLR, árboles de regresión. Al igual que Yi & Prybutok (1996) en su investigación concluyeron que MLP supera la metodología de Modelos Auto Regresivos Integrados de Medias Móviles (ARIMA) en la predicción del ozono.

Kolehmainen et al. (2000), encontró que MLP era superior en la reproducción de comportamientos de contaminantes gaseosos en comparación con la reproducción de las partículas.

Lu et al (2006), utilizaron dos niveles de clasificaciones de datos en conjunto con MLP para desarrollar 4 distintos modelos para cada una de 4 regiones distintas de Taiwan, encontrando como mejor modelo el de la estación de monitoreo Cutin; con  $R^2=0.70$ , IA= 0.91 y RMSE = 30 ppb.

En 2006, Wang & Lu proponen el modelo *Híbrido Monte Carlo* para inicializar una red MLP con algoritmo de entrenamiento PSO (*Particle Swarm Optimization*) y finalmente obtiene errores RMSE que van de 10.8868 ppb a 17.968 ppb en la región Tsuen Wan y de 15.4205 ppb a 30.5079 ppb en la región de Tung Cheng, ambas de Hong Kong.

Sousa et al (2007) en Oporto, Portugal, desarrollaron modelos para predecir ozono basados en *Redes Neuronales Artificiales* y *Análisis de Componentes Principales* (PCA) como entradas a la red para predecir ozono horario del día siguiente. El RMSE de entrenamiento obtenido fue de 15.37 ppb.

Con los antecedentes de los trabajos anteriores mencionados en el Capítulo 7 correspondiente a Discusión de Resultados, se puede concluir que los resultados de los modelos de predicción propuestos para esta tesis representan una alternativa viable para las regiones de estudio. Tanto los modelos de predicción para PM<sub>2.5</sub>, como los propuestos para predecir ozono, mejoran de manera substancial los resultados encontrados por otros investigadores de la literatura formal del momento.

En cuanto a las técnicas de modelado, se coincidió en que las Redes Neuronales Artificiales MLP son las que mejor simulan el comportamiento futuro de los contaminantes como el ozono mas no las PM<sub>2.5</sub> pues, aunque sus resultados no fueron del todo erróneos, las herramientas RBF y SMLP resultaron ser más eficientes.

## Capítulo 8

Se tenían mejores expectativas para las redes ELMAN, pero en este trabajo no lograron ser superior en resultados a las redes de MLP, aunque la diferencia no fue muy marcada. En cambio, el tiempo aplicado de entrenamiento si fue substancialmente menor (41% menor) para las redes ELMAN comparativamente con el requerido para el entrenamiento de las redes MLP.

Los modelos de Inteligencia Artificial están siendo utilizados como una excelente alternativa de herramienta de reproducción de fenómenos no lineales, como es el caso de los fenómenos ambientales y de muchos otros.

A pesar del considerable trabajo que generalmente representa, el beneficio se obtiene al lograr modelos de gran calidad en la reproducción de su comportamiento.

# Capítulo 9

## Desarrollos futuros

---

A pesar del optimismo por los resultados encontrados en las investigaciones desarrolladas tanto en Ciudad Juárez- El Paso con los modelos de predicción de PM2.5 y los modelos de predicción de ozono en la región Mexicali-Calexico, también se tiene conciencia de que el margen de error en los modelos de predicción resultantes aún es susceptible de ser mejorado.

Una mayor cantidad de datos disponibles, así como mejoras en la metodología de análisis, selección y agrupación de los datos pueden también dar oportunidad a mejorar lo hecho anteriormente.

También sería interesante también desarrollar y/o evaluar metodologías para recuperar o sustituir datos inexistentes que sean lo suficientemente eficaces como para no afectar el desempeño del modelo a formular.

Otra problemática que facilitaría evaluar el desempeño de los modelos similares desarrollados por distintos investigadores del tema, sería estandarizar de alguna manera las medidas de rendimiento, de manera

que el conjunto de la comunidad científica utilizara las mismas unidades de medida y considerara los mismos intervalos de tiempo.

Además, para fines de salud pública y mejoramiento de la calidad de vida, debemos ser en extremo cuidadosos con modelos que muestren tendencias a sub-predecir concentraciones máximas de contaminantes. Estos presentan mayores probabilidades de no detectar episodios de concentraciones fuera de los márgenes máximos permitidos.

En estas aplicaciones se prefiere que los modelos de calidad del aire tiendan a sobre-predecir para efectos de protección a la salud, por citar un aspecto importante que aún se debe mejorar.

Otro punto pendiente de estudio en profundidad es determinar qué tan eficiente es la predicción puntual de un contaminante en diversos puntos geográficos de la zona de estudio o qué otros factores ya sean atmosféricos, climáticos o de temporalidad podrían influir para mejorar las predicciones, entre otros aspectos.

Lo anterior muestra la pauta para futuros trabajos en ese sentido, por lo que como continuación de los trabajos a desarrollar en esta línea, posteriores a esta tesis, se trabaja en la generación de modelos de predicción multizonales de contaminantes en el aire y como caso de aplicación se trabajará en la misma zona del modelo de predicción de partículas  $PM_{2.5}$ , correspondiente a la región norte central de la república mexicana: Ciudad Juárez y la ciudad vecina estadounidense de El Paso, Texas.

# Capítulo 10

## Referencias

---

- Aguirre-Basurko E, Ibarra-Berastegi G., Madariaga I., 2006 Regression and multilayer perceptron-based models to forecast hourly O3 and NO2 levels in the Bilbao area. *Environmental Modelling & Software* 21 (2006) 430–446
- Bishop, C.M., 1995. *Neural Networks for Pattern Recognition*, Oxford. Oxford University Press.
- Brimblecombe P.(1996). *Air composition & chemistry*. Cambridge; New York: Cambridge University Press SERIES Cambridge environmental chemistry series.
- California Air Management Board; 2005. *Ozone and Ambient Air Quality Standards*  
<http://arb.ca.gov/research/aaqs/caaqs/ozone/ozone.htm>
- Castejón Limas, M., Ordieres Mere, J.B., De Cos Juez, F.J., Martínez De Pisón, F.J., 2001 *Control de Calidad, Metodología para El Análisis Previo a la Modelización de Datos en Procesos Industriales*. Universidad de La Rioja.
- CCAAN, CCE, CEC, CCA. Comisión para la Cooperación Ambiental de América del Norte. *Tres países. Un medio ambiente*.  
<http://www.cec.org/home/index.cfm?varlan=espanol>.
- Census Bureau, 2006, *Population Estimates*. Calexico city, California

- Census Bureau, rev. Mayo 2007. <http://quickfacts.census.gov>
- Chameides, W.L., Stedman, D.H., 1977. Tropospheric Ozone: Coupling transport and photochemistry. *J Geophys Res* 82 (12), pp. 1787-1794.
- Chaloulakou, A., Saisana, M., Spyrellis, N., 2003. Comparative assessment of neural networks and regression models for forecasting summer time ozone in Athens. *The Science of the Total Environment* 313, 1-13.
- Chang, J.C., and S.R. Hanna, 2004: Air quality model performance evaluation. *Meteorol. and Atmos. Phys.*, 87, 167-196.
- Chang J. C. and Hanna S. R., 2005. Model Validation Kit (MvK). BOOT Statistical Model Evaluation Software Package. National Environmental Research Institute Ministry of the Environment . Denmark
- Chen, L., Sandhu, H.S., Angle, R.P., McDonald, K.M., Myrick, R.H., 2000. Rural particulate matter in Alberta, Canada. *Atmospheric Environment* 34, 3365-3372.
- Chow, J.C., Watson, J.G., 2001. Zones of representation for PM10 measurements along the US/México border. *The Science of the Total Environment* 276, 49-68.
- Clapp, L.J., Jenkin, M.E., 2001. Analysis of the relationship between ambient levels of O<sub>3</sub>, NO<sub>2</sub> and NO as a function of NO<sub>x</sub> in the UK . *Atmospheric Environment* 35 (36), pp. 6391-6405
- Cox, W.M., and J.A. Tikvart., 1990: A statistical procedure for determining the best performing air quality simulation model. *Atmospheric Environment.*, 24A, 2387-2395.
- Davis, D.D., Orendovici, T. 2006. Incidence of ozone symptoms on vegetation within National Wildlife Refuge in New Jersey, USA . *Environmental Pollution* 143 (3), pp. 555-564.
- Dong Wang, Wei-Zhen Lu. Ground-level ozone prediction using multilayer perceptron trained with an innovative hybrid approach. *Ecological modelling* 198 (2006) 332-340. [www.elsevier.com](http://www.elsevier.com)
- Dixon, J.K., 1979. Pattern recognition with partly missing data *IEEE Transactions on Systems Man and Cybernetics* SMC-9 10, 617-621.
- European Commission., 1994. The evaluation of models of heavy gas dispersion. Model evaluation group seminar. Office for Official Publications of the European Communities, L-2985, Luxemburg.

- Fabian, P., Pruchniewicz, P.G., 1977. Meridional distribution of ozone in the troposphere and its seasonal variations. *J Geophys Res* 82 (15), pp. 2063-2073
- Filleul, L., Cassadou, S., Medina, S., Fabres, P., Lefranc, A., Eilstein, D., Le Tertre, A., Ledrans, M., 2006. The relation between temperature, ozone, and mortality in nine French cities during the heat wave of 2003. *Environmental Health Perspectives* 114 (9), pp. 1344-1347.
- Fischer, P., Hoek, G., Brunekreef, B., Verhoeff, A., Van Wijnen, J., 2003. Air pollution and mortality in the Netherlands: Are the elderly more at risk?. *European Respiratory Journal, Supplement* 21 (40), pp. 34S-38S.
- Flake, G.W., 1998. Square Unit Augmented, Radially Extended, Multilayer Perceptrons. In G. Orr, K.-R. Müller, and R. Caruana, editors, *Tricks of the Trade: How to Make Algorithms Really Work*, LNCS State-of-the-Art-Surveys. Springer-Verlag
- Frost, G.J., Trainer, M., Allwine, G., Buhr, M.P., Calvert, J.G., Cantrell, C.A., Fehsenfeld, F.C., Williams, E.J., 1998. Photochemical ozone production in the rural southeastern United States during the 1990 Rural Oxidants in the Southern Environment (ROSE) program. *Journal of Geophysical Research D: Atmospheres* 103 (D17), pp. 22491-22508
- Fuller, G.W., Carslaw, D.C., Lodge, H.W., 2002. An empirical approach for the prediction of daily mean PM10 concentrations. *Atmospheric Environment* 36, 1431-1441.
- Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multi-layer perceptron)- a review of applications in the atmospheric sciences. *Atmospheric Environment* 32, 2627-2636.
- Gardner, M.W., Dorling, S.R., 2000a. Statistical surface ozone models: An improved methodology to account for non-linear behaviors. *Atmospheric Environment* 34 (1), pp. 21-34.
- Gardner, M.W., Dorling, S.R., 2000b. Meteorologically adjusted trends in UK daily maximum surface ozone concentrations. *Atmospheric Environment* 34 (2), pp. 171-176.
- Gauvin, S., Reungoat, P., Cassadou, S., Dechenaux, J., Momas, I., Just, J., Zmir, D., 2002. Contribution of indoor and outdoor environments to PM2.5 personal exposure of children VEST study.

- The Science of the Total Environment 297, 175-181.
- Guidotti, TL, Calhoun, T., Davies-Cole, J.O., Knuckles, M.E., Stokes, L., Glymph, C., Lum, G., Ragain, L. et al, (2007). Elevated lead in drinking water in Washington, DC, 2003-2004: The public health response. *Environmental Health Perspectives* 115 (5), pp. 695-701.
- Guicherit, R., Roemer, M., 2000. Tropospheric ozone trends. *Chemosphere - Global Change Science* 2 (2), pp. 167-183.
- Hastie, T.J., Tibshirani R.J., 1990. *Generalized Additive Models*. Chapman & Hall, London.
- Haykin, S., 1999. *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, Englewood Cliffs, NJ.
- Hien, P.D., Bac, V.T., Tham, H.C., Nhan, D.D., Vinh, L.D., 2002. Influence of the meteorological conditions on PM<sub>2.5</sub> and PM<sub>2.5-10</sub> during the monsoon season in Hanoi, Vietnam. *Atmospheric Environment* 36, 3473-3484.
- Hien, P.D., Binh, N.T., Truong, Y., Ngo, N.T., 1999. Temporal variations of source impacts at the receptor, as derived from air particulate monitoring data in Ho Chi Minh City, Vietnam. *Atmospheric Environment* 33, 3133-3142.
- Ho, S.L., Xie, M., Goh, T.N., 2002. A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. *Computers & Industrial Engineering* 42, 371-375.
- Hopke, P.K., 1985. *Receptor modeling in environmental chemistry*. New York. Wiley, 1985, 319 pp.
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feed forward networks are universal approximators. *Neural Networks* 2, 359-366.
- Hornik, K., 1993. Some new results on neural network approximation. *Neural Networks* 6, 1069-1072.
- H.Niska, M.Rantamaki, T.Hiltunen, A.Karppinen, J.Kukkonen, J.Ruuskanen, M.. Kolehmainen. Evaluation of an integrated modelling system containing a multi-layer perceptron model and the numerical weather prediction model HIRLAM for the forecasting of urban airborne pollutant concentrations *Atmospheric Environment* 39 (2005) 6524–6536
- Hsin-Chung Lu, Jen-Chieh Hsieh, Tseng-Shuo Chang Prediction of

- daily maximum ozone concentrations from meteorological conditions using a two-stage neural network. *Atmospheric Research* 81 (2006) 124–139.
- Ihaka, R., Gentleman, R., 1996. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*. 299-314.
- INE, Instituto Nacional de Ecología, 2005, Dirección en Investigación sobre Calidad del Aire <http://www.ine.gob.mx/dgicurg/calaire/index.html> .
- INEGI, Instituto Nacional de Estadística Geografía e Informática, 2005, <http://www.inegi.gob.mx/inegi/default.aspx>.
- IPCC, The Scientific basis. Technical summary. (<http://ipcc-wg1.ucar.edu/index.html>), 2007
- Jacobson, M.Z., 1999. Effects of soil moisture on temperatures, winds and pollutant concentrations in Los Angeles. *Journal of Applied Meteorology* 38, 607–616.
- Jorquera, H., Palma, W., Tapia, J., 2000. An intervention analysis of air quality data at Santiago, Chile. *Atmospheric Environment* 34, 4073-4084.
- Kleinman, L.I., 2000. Ozone process insights from field experiments - part II: Observation-based analysis for ozone production. *Atmospheric Environment* 34 (12-14), pp. 2023-2033
- Kolehmainen, M., Martikainen, H., Ruuskanen, J., 2001. Neural networks and periodic components used in air quality forecasting. *Atmospheric Environment* 35, 815-825.
- Kongtip P., Thongsuk W., Yoosook W. and Chantanakul S., 2006. Health effects of metropolitan traffic-related air pollutants on street vendors. *Atmospheric Environment*, Volume 40, Issue 37, December 2006, Pages 7138-7145.
- Krishnayya, N.S.R., Bedi, S.J. 1986. Effect of automobile lead pollution on *Cassia tora* L. and *Cassia occidentalis* L. *Environmental Pollution Series A: Ecological and Biological*, 40 (3), pp. 221-226.
- Kukkonen, J., Harkonen, J., Karppinen, A., Pohjola, M., Pietarila, H., Koskentalo, T., 2001. A semi-empirical model for urban PM10 concentrations, and its evaluation against data from an urban measurement network. *Atmospheric Environment* 35, 4433-4442.

- La Paz Agreement., 1983. Agreement between the U.S. and México on cooperation for the protection and improvement of the environment in the border area, 14 august 1983. La Paz Baja California Sur, México TIAS No.10827.
- Lee, D.S., Holland, M.R., Falla, N., 1996. The potential impact of ozone on materials in the U.K. *Atmospheric Environment* 30 (7), pp. 1053-1065
- Lenschow, P., Abraham, H.-J., Kutzner, K., Lutz, M., Preuß, J.-D., Reichenbacher, W., 2001. Some ideas about the sources of PM10. *Atmospheric Environment* 35 (Supplement No. 1), S23-S33.
- Lu, H.-C., 2002. The statistical characters of PM10 concentration in Taiwan area. *Atmospheric Environment* 36, 491-502.
- Magliano, K.L., Hughes, V.M., Chinkin, L.R., Coe, D.L., Haste, T.L., Kumar, N., Lurmann, F.W., 1999. Spatial and temporal variations in PM10 and PM2.5 source contributions and comparison to emissions during the 1995 integrated monitoring study. *Atmospheric Environment* 33, 4757-4773.
- Martínez-de-Pisón, F.J; C. Barreto, A. Pernía, F. Alba. Modelling of an Elastomer Profile Extrusion Process using Support Vector Machines (SVM). *Journal of Materials Processing Technology* 197 (2008) 161–169.
- Mary Lou Thompson, Joel Reynolds, Lawrence H. Cox, Peter Guttorp, Paul D. Sampson., 2000. A review of statistical methods for the meteorological adjustment of tropospheric ozone. *Atmospheric Environment* 35 (2001) 617-630.
- McClellan, O.R., 2001. Setting ambient air quality standards for particulate matter, University of New Mexico, 13701 Quaking Aspen Place NE, Albuquerque, NM 87111, USA, Toxicology 00, 119 pp.
- Minsky, M. L. and Papert, S. A., 1988. *Perceptrons: An introduction to Computational Geometry*. Expanded edition. Cambridge, MA: MIT Press.
- Masters, T., 1993. *Practical Neural Network Recipes in CCC*. Academic Press, San Diego.
- McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*, 2<sup>nd</sup> ed. , London. Chapman and Hall.
- McDonnell, W.F., Nishino-Ishikawa, N., Peterson, F.F., Chen, L.H., Abbey, D.E., 2000. Relationship of mortality with the fine and

- coarse fraction of long term ambient PM<sub>10</sub> concentrations in nonsmokers. *Journal of Exposure Analysis and Environmental Epidemiology* 10, 427-436.
- Minsky, M. L. and Papert, S. A. 1988. *Perceptrons: An introduction to Computational Geometry*. Expanded edition. Cambridge, MA: MIT Press.
- Morel, B., Yeh, S., Cifuentes, L., 1999. Statistical distributions for air pollution applied to the study of the particulate problem in Santiago. *Atmospheric Environment* 33, 2575-2585.
- Monks, P.S., 2000. A review of the observations and origins of the spring ozone maximum. *Atmospheric Environment* 34 (21), pp. 3545-3561.
- Mukerjee, S., 2001. Selected air quality trends and recent air pollution investigations in the US-México border region. *The Science of the Total Environment* 276, 1-18.
- Mukerjee, S., Shadwick, D.S., Smith, L.A., Somerville, M.C., Dean, K.E., Bowser, J.J., 2001. Techniques to assess cross-border air pollution and application to a US-Mexico border region. *The Science of the Total Environment* 276, 205-224
- M.W. Gardner, S.R. Dorling, 1998. Neural network modelling and prediction of hourly NO<sub>x</sub> and NO<sub>2</sub> concentrations in urban air in London. *Atmospheric Environment* 33 (1999) 709-719.
- Ordieres, J.B., Vergara, E.P., Capuz, R.S., Salazar, R.E., 2005. Neural network prediction model for fine particulate matter (PM 2.5) on the US-México border in El Paso (Texas) and Ciudad Juárez (Chihuahua). *Environmental Modelling and Software* 20 (5), pp. 547-559.
- Organización Mundial de la Salud (OMS), “Directrices sobre la calidad del aire”/Oct.2006  
<http://www.who.int/mediacentre/factsheets/fs313/es/index.html>
- Ostro, B.D., Eskeland, G.S., Sanchez, J.M., Feyzioglu, T., 1999a. Air pollution and health effects: a study of medical visits among children in Santiago, Chile. *Environmental Health Perspective* 107, 69-73.
- Ostro, B., Chesnut, L., Vichit-Vadkan, N., Laixuthai, A., 1999b. The impact of particulate matter and daily mortality in Bangkok, Thailand. *Journal of Air and Waste Management Association* 49,

100-107.

- Pérez, P.; J. Reyes. "Prediction of maximum of 24-h average of PM10 concentrations 30 h in advance in Santiago, Chile". *Atmospheric Environment*. Vol. 36, pp. 4555-4561. 2002.
- Pérez, P; A. Trier and J. Reyes. "Prediction of PM2.5 concentrations several hours in advance, using neural networks in Santiago, Chile". *Atmospheric Environments*. Vol. 34, pp. 1189-1196. 2000.
- Pérez, P., Reyes, J., 2001. Prediction of particulate air pollution using neural techniques. *Neural Computing and Applications* 10, 165-171.
- Pérez, P., Trier, A., 2001. Prediction of NO and NO2 concentrations near a street with heavy traffic in Santiago, Chile. *Atmospheric Environment* 35, 1783-1789.
- Podnar, D., Koračin, D., Panorska, A., 2002. Application of artificial neural network to modeling the transport and dispersion of tracers in complex terrain. *Atmospheric Environment* 36, 561-570.
- Querol, X., Alastuey, A., Rodriguez, S., Plana, F., Mantilla, E., Ruiz, C.R., 2001a. Monitoring of PM10 and PM2.5 around primary particulate anthropogenic emission sources. *Atmospheric Environment* 35, 845-858.
- Querol, X., Alastuey, A., Rodriguez, S., Plana, F., Ruiz, C.R., Cost, N., Massague´ , G., Puig, O., 2001b. PM10 and PM2.5 source apportionment in the Barcelona metropolitan area, Catalonia, Spain. *Atmospheric Environment* 35, 6407-6419.
- Quintero, N.M; 2006. Contaminación y medio ambiente en Baja California; (TD182.7/.B3 / C65 / 2006)
- Reich, S.L., Gómez, D.R., Dawidowski, L.E., 1999. Artificial neural network for the identification of unknown air pollution sources. *Atmospheric Environment* 33, 3045-3052.
- Rodríguez, S., Querol, X., Alastuey, A., Kallos, G., Kakaliagou, O., 2001. Saharan dust contributions to PM10 and TSP levels in Southern and Eastern Spain. *Atmospheric Environment* 35, 2433-2447.
- Rodríguez, S., Querol, X., Alastuey, A., Mantilla, E., 2002. Origin of high summer PM10 and TSP concentrations at rural sites in Eastern Spain. *Atmospheric Environment* 36, 3101-3212.

- Ruppert, D., Wand, M. P. and Carroll, R.J. (2003). Semiparametric Regression. New York: Cambridge University Press.
- Salcedo, R.L.R., Alvim Ferraz, M.C.M., Alves, C.A., Martins, F.G., 1999. Time-series analysis of air pollution data. Atmospheric Environment 33, 2361-2372.
- Salini, C.G. y P. Pérez Jara , 2006 Estudio Mediante Series Temporales de Contaminación Ambiental Mediante Técnicas de Redes Neuronales Artificiales.
- Scebba, F., Canaccini, F., Castagna, A., Bender, J., Weigel, H.-J., Ranieri, A., 2006. Physiological and biochemical stress responses in grassland species are influenced by both early-season ozone exposure and interspecific competition. Environmental Pollution 142 (3), pp. 540-548.
- SCERP, 2003, Southwest Consortium for Environmental Research and Policy, Monograph Series No.6:"The U.S.-Mexican border environment: Air Quality Issues along the U.S. -Mexican Border" .San Diego State University Press.
- SCERP, 2005, Southwest Consortium for Environmental Research and Policy. <http://scerp.org> .
- SCERP, 2007, Southwest Consortium for Environmental Research and Policy. Security, Development, and the Environment in the Binational U.S.-Mexican Border Region.
- SEMARNAP, 1996, Secretaría de Medio Ambiente Recursos Naturales y Pesca. Programa Frontera XXI, 1ª Edición. México, D.F. 1996, 250 pp.
- SEMARNAT, 2007. Secretaria de Medio Ambiente y Recursos Naturales, Frontera 2012. Programa Ambiental México-Estados Unidos. <http://www.semarnat.gob.mx/presenciainternacional/instrumentosambientales/Pages/frontera2012.aspx>
- Sillman, S., 1999. The relation between ozone, NO(x) and hydrocarbons in urban and polluted rural environments. Atmospheric Environment 33 (12), pp. 1821-1845.
- Sousa, S.I.V., Martins, F.G., Alvim-Ferraz, M.C.M., Pereira, M.C., 2007. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations.

- Environmental Modelling and Software 22 (1), pp. 97-103
- Tao, K.M., 1993. A Closer Look at the Radial Basis Function (RBF) Networks. Conference Record of the 27th Asilomar Conference on Signals, Systems and Computers, vol. 1. IEEE Comput. Soc. Press, Los Alamitos, CA, pp. 401-405.
- The MathWorks, Neural network toolbox for use with MATLAB®: user guide, 2007.  
<http://www.mathworks.com/access/helpdesk/help/toolbox/nnet/>
- Tiittaa, P., Raunemaa, T., Tissari, J., Yl-Tuomi, T., Leskinen, A., Kukkonen, J., Harkonen, J., Karppinen, A., 2002. Measurements and modelling of PM2.5 concentrations near a major road.
- Trainer, M., Parrish, D.D., Goldan, P.D., Roberts, J., Fehsenfeld, F.C., 2000. Review of observation-based analysis of the regional factors influencing ozone concentrations. Atmospheric Environment 34 (12-14), pp. 2045-2061.
- US- EPA and Secretaría de Desarrollo Urbano y Ecología . SEDUE., 1991. Integrated Environmental Plan for the Mexican-U.S. Border Area, First Stage, 1992-1994. Washington, DC.
- US-EPA, 1996. Air Quality Criteria for Particulate Matter. EPA/600/P-95/001F. US Environment Protection Agency, Washington, DC.
- US-EPA, 1998. US-México Border XXI Program. United States-México Border Environmental Indicators EPA909-R-98-001. US, Environment Protection Agency, Washington, DC.
- US-EPA, 1996. U.S.-México Border XXI Program Framework Document. EPA 160-R-96-003. October 1996. Washington, DC.
- US-EPA., 2000a. Summary of Selected Environmental Indicators From the U.S.-Mexico Border XXI Program: Progress Report 1996-2000. EPA 909-R-00-002. US Environment Protection Agency, Washintong, DC.
- US-EPA., 2000b. National Air Quality and Emissions Trends Report 1998. EPA 454-R-00-003. US Environment Protection Agency, Washintong, DC.
- US-EPA., 2000c. Latest Findings on National Air Quality: 1999 Status and Trends. EPA-454-F-00-002 . US Environment Protection Agency, Washintong, DC.

- US-EPA and Secretaría de Medio Ambiente y Recursos Naturales SEMARNAT., 2003. Border 2012/Frontera 2012: U.S.-México Environmental Program. Washington, DC.
- US-EPA; 2003 “Emission Inventory Guidance for Implementation of Ozone and Particulate Matter National Ambient Air Quality Standards and Regional Haze Regulations”.
- US-EPA; 2005. Air Quality System (AQS), Situación ambiental fronteriza. Reporte de indicadores.
- US-EPA; 2006. National Ambient Air Quality Standards (Clean Air Act) <http://www.epa.gov/air/criteria.html> .
- USEPA; 2007. Non-attainment Areas for Criteria Pollutants, <http://www.epa.gov/oar/oaqps/greenbk> .
- USEPA; 2007. What Are the Six Common Air Pollutants?. <http://www.epa.gov/air/urbanair/6poll.html>
- USEPA; 2007. Communities in Action for Asthma-Friendly Environments National Asthma Forum. [www.epa.gov/asthma](http://www.epa.gov/asthma).
- USEPA; 2007. Visibility. [www.epa.gov/visibility](http://www.epa.gov/visibility)
- Vega, E., Reyes, E., Sanchez, G., Ortiz, E., Ruiz, M., Chow, J., Watson, J., Edgerton, S., 2002. Basic statistics of PM<sub>2.5</sub> and PM<sub>10</sub> in the atmosphere of México City. *The Science of the Total Environment* 287, 167-176.
- Watson, J.G., Chow, J.C., 2001. Source characterization of major emission sources in the Imperial and Mexicali Valleys along the US/Mexico border. *The Science of the Total Environment* 276, 33-47
- Weschler, C.J., 2006. Ozone's impact on public health: Contributions from indoor exposures to ozone and products of ozone-initiated chemistry. *Environmental Health Perspectives* 114 (10), pp. 1489-1496.
- Wilks D.S. (1995). *Statistical methods in the atmospheric sciences*. Academic Press, San Diego, CA, 467.
- Wilson, W.E. and Suh, H.H., 1997. Fine particles and coarse particles: concentration relationships relevant to epidemiologic studies. *Journal of the Air and Waste Management Association* 47, pp. 1238–1249.

- Yang, K.-L., 2002. Spatial and seasonal variation of PM10 mass concentrations in Taiwan. *Atmospheric Environment* 36, 3403-3411.
- Yi, J., Prybutok, V.R., 1996. A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. *Environmental Pollution* 92, 349–357.