



UNIVERSIDAD DE SANTIAGO DE CHILE
FACULTAD DE CIENCIA
DEPARTAMENTO DE FÍSICA

**DESARROLLO DE UN MODELO PARA
PRONOSTICAR CONCENTRACIONES
EXTREMAS DE PM2.5 EN SANTIAGO**

GIOVANNI ANGELO SALINI CALDERÓN

Profesor Guía:
Dr. Patricio Pérez Jara

*Tesis para optar al grado
de Doctor en Ciencias con
mención en Física*

SANTIAGO – CHILE

2009

**DESARROLLO DE UN MODELO PARA
PRONOSTICAR CONCENTRACIONES
EXTREMAS DE PM2.5 EN SANTIAGO**

GIOVANNI ANGELO SALINI CALDERÓN

Este trabajo de Graduación fue preparado bajo la supervisión del profesor guía Dr. Patricio Pérez J., del Departamento de Física de la Universidad de Santiago de Chile y ha sido aprobado por los miembros de la comisión calificadora

Dr. Patricio Pérez Jara

Dr. René Garreaud Salazar

Dr. Ernesto Gramsch Labra

Dr. Marcelo Mena Carrasco

Dr. Ulrich Raff Biggemann

.....
Director Departamento de Física
Dr. Álvaro San Martín Ulloa

© Giovanni Angelo Salini Calderón.

Se autoriza la reproducción parcial o total de esta obra, con fines académicos, por cualquier forma, medio o procedimiento, siempre y cuando se incluya la cita bibliográfica de este documento.

**DESARROLLO DE UN MODELO PARA
PRONOSTICAR CONCENTRACIONES
EXTREMAS DE PM2.5 EN SANTIAGO**

GIOVANNI ANGELO SALINI CALDERÓN

Trabajo de Graduación presentado a la Facultad de Ciencia,
en cumplimiento parcial de los requerimientos exigidos
para optar al grado de Doctor en Ciencias con mención en
Física

SANTIAGO – CHILE

2009

Dedicatoria

*“... El Amor nunca deja de ser; pero las profecías se acabarán, y cesarán las lenguas, y la ciencia acabará. Porque en parte **conocemos**, y en parte **profetizamos**, más cuando venga lo **Perfecto**, entonces lo que es en parte se acabará...”*

- 1 Corintios 13, 8 - 10

Dedico este trabajo de tesis a mi querida esposa *Ester* y a mis amados hijos(as) *Isabella*, *Estefano* y *Giordano*, por todo el sacrificio que debieron realizar y soportar durante todo este período de mi perfeccionamiento. Muchísimas gracias por todo.

Agradecimientos

Agradezco en primer lugar y muy especialmente a mi profesor guía, Dr. Patricio Pérez Jara, por brindarme su apoyo, motivación, enseñanza (sabiduría) y paciencia, introduciéndome en esta interesante área de investigación aplicada y bien actual.

No puedo dejar de agradecer a la *Comisión Nacional de Investigación en Ciencia y Tecnología (CONICYT)*, por su gran apoyo al concederme la beca N° 21060428 del período marzo 2006-febrero 2009 inclusive, para realizar mis Estudios de Doctorado y la *beca N° 23090028* (correspondiente al año 2009) de Término de Tesis. También mis agradecimientos por la beca de estadía y asistencia al XVI Simposio Chileno de Física efectuado en noviembre de 2008, en la UTFSM, Valparaíso.

Quiero agradecer también, a la *Dirección General de Graduados* de la *Universidad de Santiago de Chile* por su apoyo a través de la beca de arancel de esta carrera, durante todo el período del doctorado (cuando fue necesario) así como por las becas de estadía para presentación de trabajo al *26th International Symposium on Forecasting* de junio de 2006, Santander, España.

Un agradecimiento especial para los profesores Dr. Carlos Esparza y Dr. Sergio Plaza por su apoyo y ayuda.

Finalmente, dejo mi agradecimiento expreso a la *Dirección de Perfeccionamiento y Posgrado* de la *Universidad Católica de la Sma. Concepción*, por su apoyo y a la *Facultad de Ingeniería*, de dicha institución, por su confianza en mí, creyendo que podía llevar a cabo este proyecto, a pesar de todas las adversidades, pesimismo y dudas de creer que no sería capaz de llevar a buen término este proyecto antes de concederme la autorización final.

Tabla de Contenidos

| | |
|------------------------|------|
| Dedicatoria | v |
| Agradecimientos | vi |
| Tabla de Contenidos | vii |
| Índice de Tablas | ix |
| Índice de Figuras | x |
| Abreviaciones y Siglas | xi |
| Resumen | xiii |
| Summary | xv |
| Introducción | xvii |

Capítulo 1: *Introducción a la Contaminación Ambiental*

| | |
|---|----|
| 1.1 Fundamentos | 1 |
| 1.2 Los efectos nocivos de los aerosoles en la salud humana | 4 |
| 1.3 La contaminación ambiental en Santiago de Chile | 5 |
| 1.4 Inventario de emisiones para la ciudad de Santiago | 13 |
| 1.5 La calidad del aire, índice ICAP y normativa vigente | 17 |
| 1.6 La base de datos | 23 |
| 1.7 Investigación propuesta | 24 |

Capítulo 2: *Métodos para analizar series temporales no lineales*

| | |
|--|----|
| 2.1 Introducción a las series temporales | |
| 2.1.1 Fundamentos | 26 |
| 2.1.2 Series temporales no lineales | 31 |
| 2.1.3 Técnicas de modelamiento para series temporales no lineales | 32 |
| 2.2 Introducción a la Teoría del Caos | |
| 2.2.1 Fundamentos | 35 |
| 2.2.2 Exponentes de Lyapunov y Predictibilidad | 39 |
| 2.2.3 Otros parámetros para estudiar posible caos en series temporales no lineales | 43 |
| 2.3 Fundamentos de las Redes Neuronales Artificiales | |
| 2.3.1 La neurona biológica y su naturaleza bioeléctrica | 46 |
| 2.3.2 Estructura y Funcionamiento de un Sistema Neuronal Artificial | 50 |
| 2.3.3 Resumen de los modelos de RNA más importantes | 55 |
| 2.3.4 la Red de Perceptrón de Multicapas (MLP) | 61 |
| 2.3.5 Funcionamiento del Algoritmo | 65 |
| 2.4 Introducción al Algoritmo de Cluster Híbrido | |

| | |
|---|-----|
| 2.4.1 Fundamentos | 69 |
| 2.4.2 El algoritmo de cluster K-Means | 71 |
| 2.4.3 Descripción del modelo no paramétrico de cluster HCA | 72 |
| 2.5 Modelos numéricos deterministas, MM5, CMAQ, Chimere y WRF | 75 |
| | |
| Capítulo 3: <i>Aplicaciones a la Predicción y Pronóstico de series temporales de contaminación ambiental</i> | |
| 3.1 Introducción a la predicción de variables de contaminación ambiental. | 85 |
| 3.2 Determinación del exponente de Lyapunov de series temporales de contaminación ambiental | |
| 3.2.1 Introducción | 88 |
| 3.2.2 El caso de las cuatro estaciones de monitoreo | 88 |
| 3.2.3 El caso de la estación Parque O'Higgins | 91 |
| 3.2.4 Evidencia de la sensibilidad a las condiciones iniciales | 92 |
| 3.3 Aplicación de las RNA a la predicción de concentraciones de material particulado fino | 94 |
| 3.4 Hacia un modelo operacional de pronóstico de concentraciones de PM _{2.5} en la ciudad de Santiago | |
| 3.4.1 Introducción | 101 |
| 3.4.2 El caso de tres clases | 101 |
| 3.4.3 El caso de cuatro clases | 108 |
| 3.5 Pronóstico de PM _{2.5} incluyendo concentraciones de CO como dato de entrada | |
| 3.5.1 Introducción | 112 |
| 3.5.2 Resultados comparativos obtenidos para pronóstico de PM _{2.5} mediante algoritmo HCA sin incluir e incluyendo CO | 113 |
| | |
| Capítulo 4 <i>Conclusiones Generales</i> | |
| 4.1 Conclusiones generales y perspectivas | |
| 4.1.1 Conclusiones generales | 115 |
| 4.1.2 Perspectivas | 117 |
| | |
| Apéndice | |
| A.1 Resumen de valores de los componentes contaminantes para el inventario de emisiones de los años 2000 y 2005 | 118 |
| A.2 Exponentes de Lyapunov para los años 1994-1996 | 119 |
| Bibliografía | 120 |

Índice de Tablas

| | | |
|-------|---|-----|
| 1.1 | Potencial meteorológico de contaminación atmosférica (PMCA) | 11 |
| 1.2 a | Resumen Inventario de Emisiones año 2000 (ton/año) | 14 |
| 1.2 b | Resumen Inventario de Emisiones año 2000 (%) | 14 |
| 1.3 a | Resumen Inventario de Emisiones año 2005 (ton/año) | 15 |
| 1.3 b | Resumen Inventario de Emisiones año 2005 (%) | 15 |
| 1.4 | Normas de calidad del aire para PM10 y PM2.5 en diferentes países | 16 |
| 1.5 | Índice ICAP por material particulado respirable | 21 |
| 3.1 | Exponentes de Lyapunov para todo el año, 2000-2007 y las cuatro estaciones de monitoreo | 89 |
| 3.2 | Exponente de Lyapunov para el periodo de invierno y las cuatro estaciones de monitoreo | 90 |
| 3.3 | Diversos parámetros de las series de PM _{2.5} para meses de otoño-invierno en Estación P. O'Higgins | 91 |
| 3.4 | Diferentes parámetros en el estudio de las series de PM _{2.5} para meses de otoño-invierno en Estación P. O' Higgins | 92 |
| 3.5 | Clasificación de vectores Estación N (Invierno) | 93 |
| 3.6 | Funciones AMI y FNN para las tres bases de datos estudiadas | 98 |
| 3.7 | Valores NRMSE para distintos tipos de modelos RNA | 99 |
| 3.8 | Valores promedio de PM2.5 en $\mu\text{g}/\text{m}^3$ | 105 |
| 3.9 | Número de días pertenecientes a cada clase | 106 |
| 3.10 | Porcentaje de concordancia entre la clase observada y la pronosticada usando tres métodos diferentes | 107 |
| 3.11 | Error porcentual absoluto promedio para los tres métodos | 107 |
| 3.12 | Comparación entre tablas de contingencia para el año 2007 usando los algoritmos HCA y MLP como métodos de pronóstico | 109 |
| 3.13 | Número de días pertenecientes a cada una de las clases | 110 |
| 3.14 | Porcentaje de acuerdo entre la clase pronosticada y la observada | 111 |
| 3.15 | Porcentaje de error promedio de los valores numérico pronosticados de 24HMAPM2.5 comparado con valor numérico observado | 111 |
| 3.16 | Error porcentual absoluto promedio con dos ecuaciones de pronóstico para el modelo HCA | 112 |
| 3.17 | Comparación entre el porcentaje de aciertos el pronóstico de PM2.5 incluyendo y no incluyendo CO | 114 |

Índice de Figuras

| | |
|--|-----|
| 1.1 Relación tamaño cabello humano y material particulado | 3 |
| 1.2 Regiones anatómicas del tracto respiratorio en el ser humano | 4 |
| 1.3 Episodio de contaminación en Londres, 1962 | 5 |
| 1.4 Cuenca de Santiago con la Cordillera de la Costa al Oeste y la de Los Andes al Este de la ciudad | 6 |
| 1.5 Composición del parque de vehículos livianos RM (fuente CONAMA-RM) | 7 |
| 1.6 Mapa de presiones típico para Sudamérica. A es por altas presiones y B por bajas presiones. La ciudad de Santiago se sitúa a 33,5° sur y 70,8° oeste | 8 |
| 1.7 Localización de la red de monitoreo MACAM II, en la ciudad de Santiago | 12 |
| 1.8 Microbalanza gravimétrica TEOM en la red MACAM II | 22 |
| 2.1 Bosquejo de una función de autocorrelación | 31 |
| 2.2 Información Mutua Promedio para una s. de t. dada | 33 |
| 2.3 Método de los falsos vecinos más cercanos | 35 |
| 2.4 El caso de dos trayectorias en el espacio de las fases | 40 |
| 2.5 Serie original y serie sustituta con el análisis de espectro de potencia e histograma | 45 |
| 2.6 Esquema de una neurona biológica | 47 |
| 2.7 Esquema de una neurona artificial | 54 |
| 2.8 Representación de la función escalón o función signo | 58 |
| 2.9 Esquema de una Red de Multicapas o Perceptrón de Multicapas | 60 |
| 2.10 Clasificación de modelos de RNA más importantes | 61 |
| 2.11 Superficie de error en el espacio de los pesos | 63 |
| 2.12 Esquematización de la función de Fermi | 64 |
| 2.13 Superficie de Error | 66 |
| 2.14 Tasa de aprendizaje | 67 |
| 2.15 Tasa de error versus número de PE's escondidos | 68 |
| 2.16 Ejemplo de mallas en modelos de mesoescala | 80 |
| 3.1 Función AMI para la serie temporal PM _{2.5} del año 1994 | 94 |
| 3.2 Falsos vecinos más cercanos para la serie temporal PM _{2.5} del año 1996 | 95 |
| 3.3 Arquitectura de la mejor red con valor RNMSE 0.5156 | 100 |

Abreviaciones y Siglas

AMI Average Mutual Information - Información Mutua Promedio
AR Autoregresivo
ARMA AutoRegressive Moving Average
ARIMA AutoRegressive Integrated Moving Average
BP Back propagation – Propagación hacia atrás
BS Black smoke – humo negro
CENMA Centro Nacional de Medio Ambiente
CMAQ Community Multiscale Air Quality
CO Monóxido de Carbono
CO₂ Dióxido de Carbono o Anhídrido Carbónico
CONAMA Comisión Nacional del Medio Ambiente
COV Componentes Orgánicos Volátiles (VOC Volatil Organic Compound)
DICTUC Dirección de Investigación en Ciencia y Tecnología de
la Universidad Católica de Santiago
EDP Ecuación Diferencia Parcial
EML Estación de Monitoreo La Florida
EMM Estación de Monitoreo Las Condes
EMN Estación de Monitoreo Parque O`Higgins
EMO Estación de Monitoreo Pudahuel
EMEP European Monitoring and Evaluation Programm
EPA-USA Environmental Protection Agency - Agencia de Protección
Medioambiental de los EEUU
FNN False Nearest Neighbours - Falso Vecino más Cercano
GMT Greenwich Mean Time basado en el meridiano 0° de esta localidad
HCA Hybrid Clustering Algorithm – Algoritmo de Clases Híbrido
H₂SO₄ Ácido Sulfúrico
ICAP Índice de Calidad del Aire por material Particulado
MA Moving Average – Promedio móvil
MACAM Red oficial de Monitoreo Automático de Calidad del Aire y
Meteorología de la ciudad de Santiago
MLP Multi Layer Network – Red Neuronal de Multicapas
MM5 Modelo de Mesoescala de 5^{ta} generación
NH₃ Amoníaco
NO₂ Dióxido de Nitrógeno
NO_x Óxidos de Nitrógeno
NRMSE Normalized Root Mean Squire Error – La raíz cuadrada del error
cuadrático promedio normalizado
O₃ Ozono

OMS Organización Mundial de la Salud (WHO – World Health Organization)
PAN Peroxiacetyl Nitrate – Nitro Peróxidos de Acetil
PE Processing Element – Elemento de Procesamiento, nodo o neurona
PM Particulate Matter – Material Particulado
PMCA Potencial Meteorológico de Contaminación Ambiental
RNA Red Neuronal Artificial
s. de t. Serie de tiempo
SEREMI de Salud Secretaría Regional Ministerial de Salud
SESMA Servicio de Salud Metropolitano del Ambiente
SO_x Óxidos sulfurosos
TEOM Microbalanza Gravimétrica (Tappered- Element Oscillating
Microbalance)
UCT Tiempo Universal Coordinado Nuevo estandar universal en reemplazo
del GMT. Aquel se basa en los relojes atómicos, mientras éste por el sol
y las estrellas. En la práctica UTC = GMT
USA - EEUU Estados Unidos de América
WHO World Health Organization (ver OMS)
WRF Weather Research and Forecasting - Investigación y Pronóstico del
tiempo
24HMAPM2.5 Promedio móvil de 24 horas de material particulado de 2,5 µm
de diámetro

Resumen

Considerando que el material particulado fino PM2.5 acapara actualmente el interés de los organismos encargados de proteger la salud de la población, hemos analizado la predictibilidad de las series temporales de dicho contaminante medidas en cuatro estaciones de monitoreo de la ciudad de Santiago entre los años 2001 y 2007. Se evaluaron diversos parámetros propios del análisis de series temporales no lineales como tiempo de rezago, la dimensión de encaje, los exponentes de Lyapunov y de Hurst, la dimensión de correlación y la de capacidad, así como la función de autocorrelación. Se encontró que las series en estudio tienen un comportamiento caótico y que una predicción en base a la información contenida en ellas sólo es posible para unas pocas horas hacia el futuro. Con los valores calculados de tiempo de rezago y dimensión de encaje se implementaron 3 modelos predictivos: un modelo de regresión lineal, una red neuronal de multicapas y un modelo de persistencia. Se encontró que el modelo MLP con una capa oculta fue el que mejor pronosticó la concentración de PM2.5 seis horas hacia el futuro. Sin embargo, de acuerdo a lo esperado, el error promedio fue relativamente alto.

Con la idea de mejorar la calidad de la predicción y desarrollar un modelo que pueda ser utilizado como una herramienta operacional en la gestión ambiental se agregaron variables meteorológicas medidas y pronosticadas como input. Apuntando a pronosticar tanto el valor numérico de la concentración de PM2.5 como su clase dentro de rangos predefinidos, en este caso se compararon tres tipos de modelo: un algoritmo de cluster, un modelo lineal y una red neuronal tipo MLP. Se encontró que para valores extremos, el algoritmo de cluster (llamado HCA) es el más preciso.

Finalmente se propuso una mejora al pronóstico de $PM_{2.5}$, al incluir concentraciones de CO como input, lo cual es una manera de tomar en cuenta el efecto de las fuentes móviles.

Summary

In this work we have analyzed the predictability of the time series for fine particulate matter PM_{2.5}, which were measured in four monitoring stations in the city of Santiago between 2001 and 2007. Several parameters from non linear time series analysis were calculated like time delay, embedding dimension, Lyapunov exponents, Hurst exponent, correlation dimension, capacity dimension and autocorrelation function. It was found that the series under study show chaotic behavior and that forecasting based on the content of the series is possible only for a few hours into the future. With the calculated values of time delay and embedding dimension three forecasting models were implemented: a linear model, a multilayer neural network (MLP) and a persistence model. It was found that with the MLP model with one hidden layer the best prediction of PM_{2.5} concentrations six hours into the future was obtained. However, according to what was expected, prediction errors were relatively high.

With the idea to improve the quality of the forecasting and in order to develop a model that could be used as an operational tool for environmental management, measured and forecasted meteorological variables were added as input. Pointing towards the forecasting of numerical values of PM_{2.5} concentrations and the class according to pre defined ranges, in this case three types of models were compared: a clustering algorithm, a linear model and an MLP. It was found that for extreme values, the clustering algorithm (named HCA) was the most accurate.

Finally it was proposed an improvement of the PM_{2.5} forecasting, by including CO concentrations as input, which is a way to take into account the effect of moving sources.

Introducción

Esta tesis está organizada de la siguiente forma:

En el capítulo 1 se presenta una introducción general al problema de la contaminación atmosférica en la ciudad de Santiago de Chile. Se explica luego cuáles son los efectos nocivos de los aerosoles en la salud de la población, haciendo hincapié en la peligrosidad de las partículas finas (PM_{2.5}). A continuación se describen las condiciones geográficas, climáticas y de fuentes emisoras que determinan la calidad del aire en Santiago, así como la red de estaciones de monitoreo de contaminantes. Paso seguido se presenta el Inventario de Emisiones para la ciudad de Santiago. En seguida se presenta la normativa vigente para material particulado, que corresponde a aquellas de diámetro menor que 10 μm (PM₁₀). La base de datos a usar así como la investigación propuesta en esta tesis cierra este capítulo.

En el capítulo 2 se describe un conjunto de herramientas para el estudio de las series temporales no lineales, especialmente en lo que se refiere a pronóstico de valores futuros. Se presenta en primer lugar una introducción a la teoría del Caos. Luego se describen los modelos de redes neuronales artificiales y los algoritmos de clustering. Finalmente, se cierra el capítulo con una descripción de los modelos deterministas de pronóstico de calidad del aire más conocidos, los cuales consideran en forma explícita la evolución de la meteorología y el comportamiento de las emisiones de contaminantes en una región.

En el capítulo 3 se presentan aplicaciones de las herramientas descritas en el capítulo 2 para el análisis y pronóstico de series temporales

de concentraciones de material particulado fino ($PM_{2.5}$), de acuerdo a los datos medidos en 4 estaciones de monitoreo de la ciudad de Santiago entre los años 2001 y 2007. Después de caracterizar dichas series en términos de tiempo de rezago, dimensión de encaje, exponentes de Lyapunov, dimensión de capacidad, exponente de Hurst, etc., se compararon tres modelos orientados a pronosticar valores futuros de $PM_{2.5}$ basados exclusivamente en la información contenida en la serie. Luego se analiza la capacidad de pronóstico de tres modelos que integran la información de todas las estaciones además de datos de meteorología y emisiones.

Finalmente, en Conclusiones se presentan los logros alcanzados en este trabajo, así como los planes para futuros trabajos.

OBJETIVO

El objetivo del presente trabajo es diseñar un modelo que permita pronosticar concentraciones extremas de material particulado fino ($PM_{2.5}$), algunas horas hacia el futuro, en la atmósfera de la ciudad de Santiago.

HIPÓTESIS

- Las series de tiempo de material particulado fino son no lineales y podrían presentar un comportamiento caótico, afectando la predictibilidad de valores futuros.
- Fijando parámetros con la ayuda de técnicas de análisis de series temporales no lineales podemos construir un modelo de redes neuronales que pronostique valores futuros de $PM_{2.5}$ con mayor precisión que modelos lineales y de persistencia.
- El incluir variables meteorológicas como input y combinando la información de todas las estaciones mejora la predicción, algunas horas en adelante, respecto de considerar sólo como input a la

variable de concentración de material particulado fino y cada estación por separado.

- Con la ayuda de un método de clustering podemos pronosticar mejor valores extremos de concentraciones de PM_{2.5}.
- El incluir una variable de contaminación que sea representativa de las fuentes móviles como input, puede mejorar el pronóstico de concentraciones de material particulado fino.

METODOLOGÍA

La metodología que será empleada en esta investigación, se describe a continuación.

Los datos de series temporales de material particulado PM_{2.5}, en la ciudad de Santiago, así como información de otros contaminantes y datos meteorológicos necesarios para implementar el modelo propuesto, en cierto periodo de tiempo, se recopilan a través de la red MACAM II, operada por el Ministerio de Salud, los cuales quedan registrados en el servidor AIRVIRO.

Se procederá a hacer un análisis de las series temporales adquiridas, para determinar linealidad, no linealidad, etc. Además, tratando a las series como sistemas dinámicos se aplicarán los siguientes test estándar: se determinará la información mutua promedio (AMI) para determinar el lag o rezago asociado a la(s) serie(s) temporal(es); la determinación de los falsos vecinos más cercanos (FNN) para calcular la dimensión de encaje, el cálculo de la función de autocorrelación, el exponente de Lyapunov más grande (para determinar la presencia de caos), entre otros posibles test a usar.

Con los parámetros así encontrados se implementarán tres modelos predictivos, uno de regresión lineal, un modelo neuronal (tipo MLP) y uno de persistencia, para comparar su capacidad de predecir varios pasos en adelante.

A continuación se agregarán variables exógenas como input a las series temporales de material particulado $PM_{2.5}$ existentes a tres modelos predictivos: uno de regresión lineal, otro neuronal tipo MLP y una nueva metodología de clustering denominado HCA y así confrontar su capacidad de hacer predicción.

Posteriormente se incluirán las concentraciones de CO (representativas de las emisiones de las fuentes móviles) como input en uno de los modelos señalados en el párrafo anterior para ver si mejora su capacidad de pronóstico.

Capítulo 1

Introducción a la Contaminación Ambiental

“If you are not a part of the solution, you are a part of the problem”

- Anonymous

1.1 Fundamentos

El estudio de la contaminación atmosférica ha adquirido una importancia vital en todo el mundo [1,2]. Estudios recientes dan cuenta que se estaría deteriorando el medio ambiente y, aún más, debido a ello estaría produciéndose un notable cambio climático a nivel planetario [3,4].

En primer lugar, en este trabajo se entenderá por contaminación del aire cuando se libera una sustancia al medio ambiente y en cantidades suficientemente grandes como para producir daño al ser humano y a otras especies, en cuanto a la reducción de su desarrollo y productividad, así como a lo material. Esta contaminación resulta de cualquier condición atmosférica, en la que ciertas sustancias alcanzan concentraciones superiores a su nivel considerado normal.

Se puede comprobar que las concentraciones presentes en el aire provienen de dos tipos de fuentes emisoras: las naturales (que proceden de fenómenos naturales como los volcanes, incendios forestales, el polvo atmosférico, entre otros) y las artificiales (o antropogénico) que se producen debido a las actividades humanas diarias, como por ejemplo industrias, chimeneas de los hogares, quemas agrícolas, tubo de escape de los vehículos, los vertederos, actividad minera, plantas térmicas, a carbón, petróleo o gas, etc. Cabe hacer notar, que para la contaminación natural

este proceso se ha venido registrando durante miles de años en el planeta y la atmósfera siempre ha sido capaz de adaptarse a estos cambios con mínimas complicaciones. El segundo se ha incrementado debido al crecimiento sostenido de la población y de la actividad de ésta.

Los principales agentes contaminantes de la atmósfera se pueden clasificar en contaminantes primarios y secundarios. Lo primero corresponde a sustancias emitidas por la fuente contaminante y vertidas directamente a la atmósfera, siendo las más comunes el monóxido de carbono (CO), los hidrocarburos (NO_x), los óxidos sulfurosos (SO_x), dióxido de carbono (CO₂), los compuestos orgánicos volátiles (COV), el material particulado PM, en general la mayoría de los hidrocarburos y las partículas, y otros que se presentan más raramente. Los contaminantes primarios conservan su identidad desde que son emitidos hasta que son medidos por medio de algún instrumento. Los contaminantes secundarios se forman en el aire a partir de reacciones químicas o fotoquímicas entre los contaminantes primarios y los compuestos que están en la atmósfera. Entre ellos NO₂, O₃ (ozono), SO₃, HNO₃, H₂SO₄, los nitroperóxidos de acilo conocidas por la sigla inglesa PAN (Peroxyacetyl nitrate)¹ y la mayoría de los iones NO₃⁻ y SO₄²⁻.

A su vez, las fuentes contaminantes se pueden clasificar, independientemente de su origen, en primer lugar como fuentes fijas (o estacionarias) como por ejemplo, las industrias, quemas agrícolas, leña residencial, y en segundo lugar como fuentes móviles, que incluye por ejemplo, los automóviles particulares, el transporte público, camiones, aviones, etc. [5].

De entre los elementos contaminantes de especial interés son los aerosoles, consistentes de material particulado, el cual típicamente a su vez está constituido por 38% de polvo en el medio ambiente, 11% de polvo

¹ Potente irritante de los ojos y de la respiración presente en el smog fotoquímico.

antropogénico, 10% nitrato, 8% amonio, 8% sulfato, 8% carbono orgánico, 7% carbono vegetal, 7% cloro y 3% otros, según fuente CENMA [6]. En la figura 1.1 se muestra, a modo de comparación, el tamaño del material particulado de 2,5 μm y 10 μm respecto del diámetro de un cabello humano promedio.

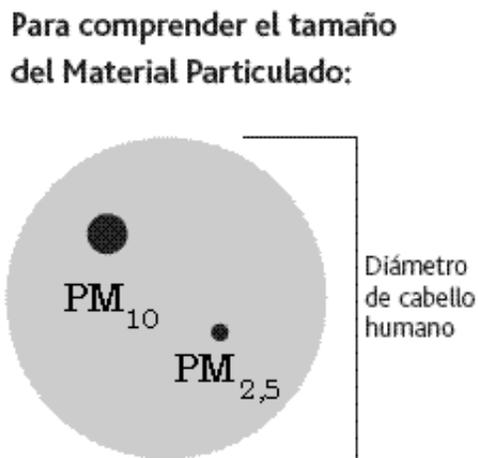


Figura 1.1: Relación tamaño cabello humano y material particulado.

El material particulado de diámetro igual o menor a 10 μm (PM_{10}) ha captado el interés de los organismos de salud, debido a su capacidad de penetrar en el sistema respiratorio humano. Sin embargo, el material particulado con un diámetro igual o menor a 2,5 μm ($\text{PM}_{2,5}$) es el que más daño estaría provocando a la salud humana [7], debido a que puede llegar fácilmente hasta los pulmones.

En el PM_{10} podemos distinguir una fracción gruesa, aquella con diámetro entre 2,5 μm y 10 μm y una fracción fina que coincidiría con el $\text{PM}_{2,5}$.

Fuentes de la fracción fina de PM_{10} son la combustión residencial, los procesos industriales y los vehículos gasolineras y Diesel (pesados y livianos). La fracción gruesa proviene principalmente del polvo en suspensión generado por el tráfico, la construcción y quemas agrícolas y de leña.

1.2 Los efectos nocivos de los aerosoles en la salud humana

A nivel de los organismos de salud, existe una especial atención a las sustancias químicas que tienen propiedades carcinogénicas o mutagénicas (que contienen los contaminantes orgánicos atmosféricos). Ello fundamentalmente debido a que está afectando gravemente la salud humana [8].

Un factor importante en el efecto en la salud, es el tamaño de las partículas, debido al grado de penetración y permanencia que ellas tienen en el sistema respiratorio humano. La mayoría de las partículas cuyo diámetro sea mayor a $5\ \mu\text{m}$ se depositan en las vías aéreas superiores (nariz) y en la tráquea y los bronquios. Aquellas cuyo diámetro es menor a dicho valor tienen mayor probabilidad de depositarse en los bronquiolos y alvéolos.

Se ha establecido que por cada $50\ \mu\text{g}/\text{m}^3$ de aumento de las concentraciones de 24 horas de PM10 en promedio se verifica un aumento de 3% de la mortalidad [1]. Con respecto al PM2.5, el análisis de datos en Santiago concluye que un incremento del promedio de 24 horas por sobre $70\ \mu\text{g}/\text{m}^3$ tiene efecto notorio en el número de consultas de urgencias respiratorias infantiles [1].

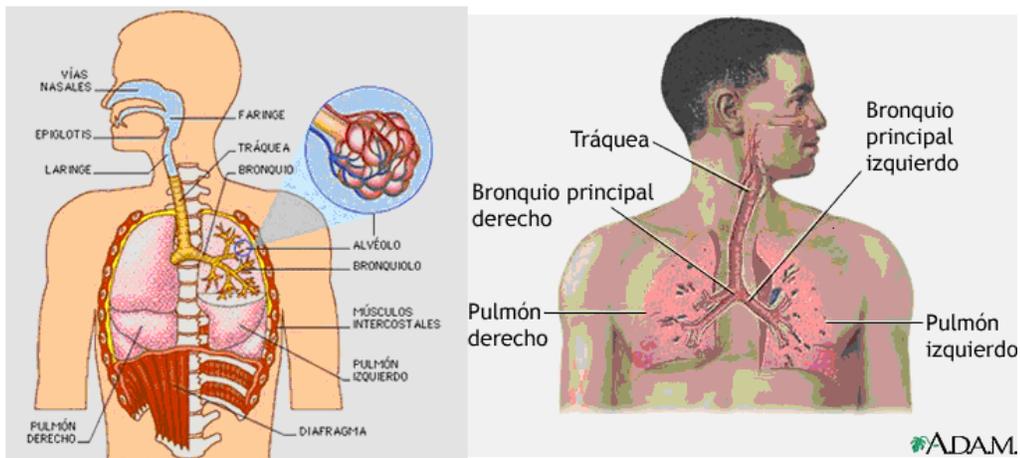
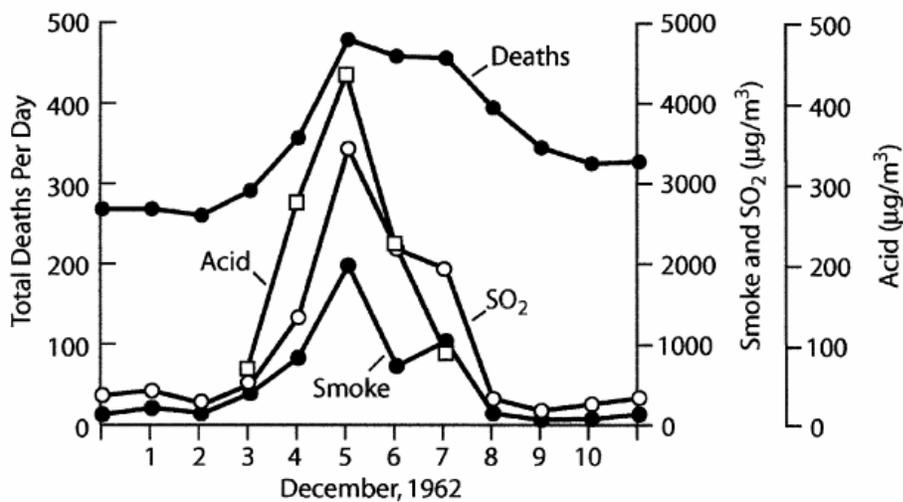


Figura 1.2: Regiones anatómicas del tracto respiratorio en el ser humano.

En general, altas concentraciones de material particulado puede ocasionar trastornos a la salud de la población como síntomas respiratorios severos, irritación de ojos y nariz, silicosis, asbestosis, agrava el asma y las enfermedades cardiovasculares, aumento en la frecuencia de cáncer pulmonar, muertes prematuras [9,10].

En la figura 1.2 se muestra la estructura básica del tracto respiratorio [7]. El aire ingresa y sale del tracto respiratorio a través de las vías aéreas extra-torácicas (nariz y boca).



December 1962, London pollution episode.

Figura 1.3: Episodio de contaminación en Londres², 1962 (de referencia [7]).

En la figura 1.3 se observa un ejemplo de episodio crítico que aconteció en la ciudad de Londres, Inglaterra, en el año 1962 en la cual se puede ver que en el quinto día del episodio creció la cantidad de muertes al aumentar los tres contaminantes (BS, SO₂ y ácido) [7].

1.3 La contaminación ambiental en Santiago de Chile

La ciudad de Santiago de Chile, posee una población de alrededor de 6 millones de habitantes que representa el 40% de la población del país, en

² En esta figura **acid** se refiere al ácido sulfúrico H₂SO₄

una superficie de alrededor de 135 Km². Situada en el centro del país y a una altura de 520 m, aproximadamente, sobre el nivel del mar, está rodeada de dos sistemas montañosos: la cordillera de Los Andes y la cordillera de La Costa. Vea figura 1.4



Figura 1.4: Cuenca de Santiago con la Cordillera de la Costa al Oeste y la de Los Andes al Este de la ciudad.

Su clima es de tipo mediterráneo, siendo los meses diciembre a febrero calurosos y secos, alcanzando temperaturas máximas alrededor de 35 °C, a la sombra. La altitud por sobre el nivel del mar crece de Poniente a Oriente.

Debido a su topografía y a las condiciones meteorológicas imperantes, especialmente en el otoño y el invierno, se observa por lo general una mala dispersión, tanto horizontal como vertical de los contaminantes emitidos por las numerosas fuentes en la ciudad. Las emisiones tienen también una tendencia al aumento por la creciente densidad poblacional, lo cual lleva asociado un aumento de las fuentes fijas y móviles. El parque automotor, por ejemplo, muestra una expansión significativa durante los últimos años.

En la figura 1.5, se muestra la evolución temporal del número de vehículos livianos, desde el año 1970 hasta el año 2010, comprobándose el rápido aumento (fuente: CONAMA-RM) [11].

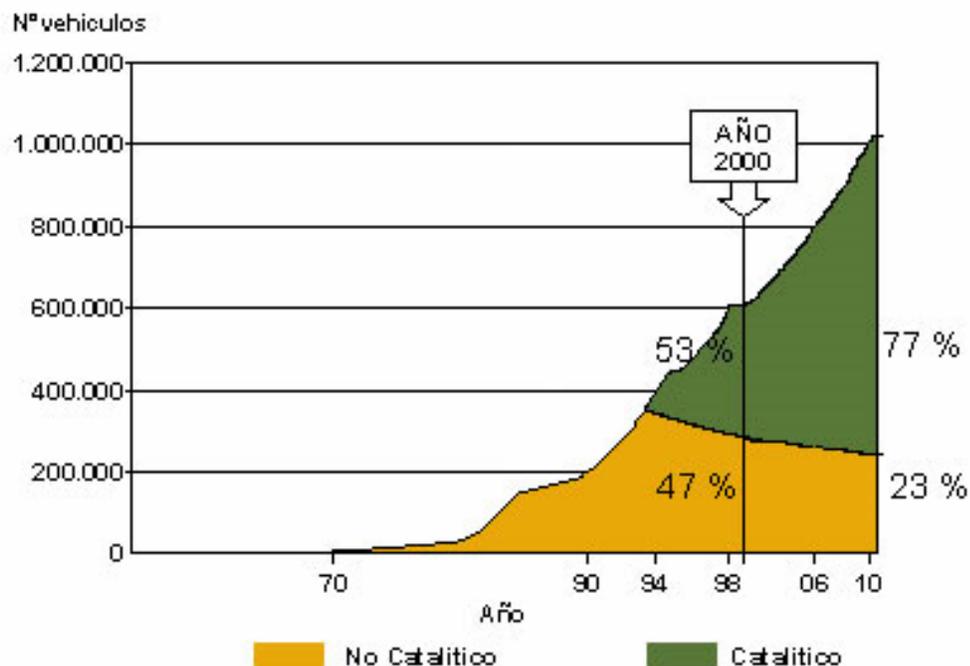


Figura 1.5: Composición del parque de vehículos livianos RM (fuente [11]).

Para ilustrar las condiciones meteorológicas imperantes en la ciudad de Santiago y sus alrededores podemos referirnos a la figura 1.6.

En esta figura podemos observar un mapa típico de las presiones a nivel del mar (en unidades en exceso de 1000 hPa) para la región de Sudamérica, en la cual A es por altas presiones y B por bajas presiones. La baja presión cercana a Santiago tiene su mínimo en la costa chilena, a aproximadamente 200 km al nor-oeste de la capital de Chile. Ante esta situación desciende aire sobre la región de Santiago el cual comprime la capa de mezcla, con el resultado que quedan atrapados los contaminantes emitidos en la ciudad. Este tipo de configuración se observa más frecuentemente y con mayor intensidad en el periodo de otoño-invierno.

Existen dos fenómenos meteorológicos que obstaculizan la adecuada dispersión de los contaminantes. Estos son la inversión térmica de subsidencia y la radiativa.

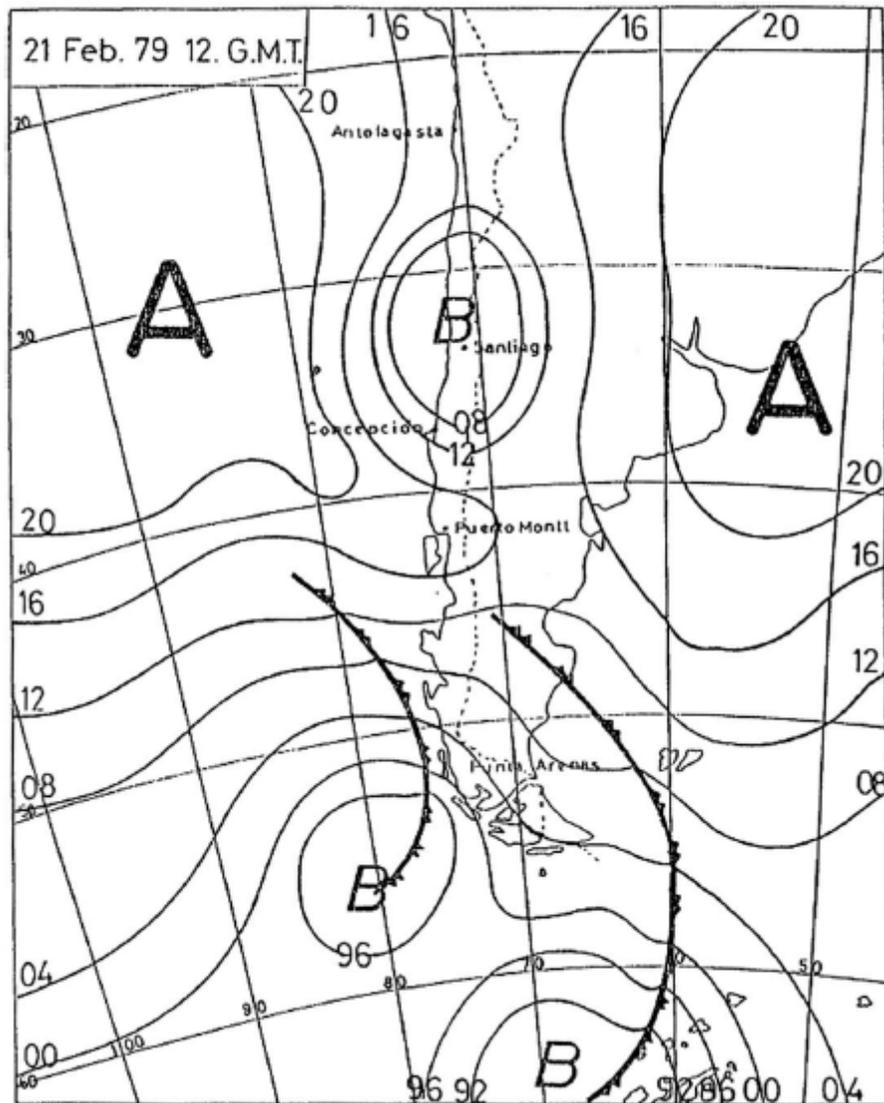


Figura 1.6: Mapa de presiones típico para Sudamérica. A es por altas presiones y B por bajas presiones. La ciudad de Santiago se sitúa a $33,5^\circ$ sur y $70,8^\circ$ oeste.

Se define inversión térmica como aquel fenómeno en el que se asienta una capa de aire caliente por sobre masas de aire de menor temperatura, impidiendo así el flujo del aire en dirección vertical (corriente ascendente), estabilizando las condiciones e impidiendo, por tanto, la dispersión vertical de los contaminantes.

Por inversión térmica de subsidencia se entenderá aquel fenómeno que se presenta por el predominio de altas presiones en el Pacífico (anticiclón), lo cual va acompañado por el descenso de aire sobre la región metropolitana, el cual al comprimirse se calienta rompiendo la tendencia del gradiente típico de temperatura. Se forma así una capa de inversión cuya base se puede situar a alturas menores a los 700 m de altitud por sobre el nivel del mar. En cambio, la inversión térmica radiativa se produce por el enfriamiento de la tierra en períodos fríos –especialmente noches de invierno-, afectando a las masas de aire inmediatamente contiguas a ella, que quedan debajo de las capas más cálidas.

Se ha logrado establecer que en la región de Santiago se dan dos tipos de condiciones sinópticas ligadas directamente con los episodios de alta contaminación atmosférica, a las cuales se les denomina tipo A y tipo BPF [1].

Los episodios tipo A se caracterizan por la entrada de masas de aire de altas presiones y altas temperaturas en altura (denominada dorsal cálida), en la troposfera media y sobre la zona central reemplazando a una vaguada en altura (que es lo contrario a una dorsal, área de bajas presiones y temperaturas), cuyo eje presenta normalmente una orientación noroeste-sureste. La formación de una vaguada costera³ se produce a niveles bajos, entre el centro de altas presiones fría al otro lado de la cordillera de Los Andes (alta fría migratoria) y un centro de altas presiones sobre el pacífico subtropical, el cual está sincronizado a las condiciones de altura mencionada más arriba.

Son estos dos sistemas de altas presiones los que obstaculizan el desplazamiento de la dorsal y la vaguada, las cuales sólo pueden moverse en la dirección norte-sur. Aún más, toda esta situación produce un viento

³ Que se define como el área de bajas presiones situada frente a las costas centrales de Chile, que al desplazarse hacia el Este, obliga a descender a las masas de aire desde la ladera andina occidental.

superficial del este (desde la cordillera hacia la costa), provocando que los máximos de concentraciones de PM_{10} , se den hacia el sector poniente de la cuenca metropolitana (Pudahuel, Cerro Navia, Lo Prado). Durante los días que dura este tipo de episodio se pueden apreciar cielos despejados y alta variación de temperatura: días cálidos y noches heladas, durante los días que dura este tipo de episodio.

Los episodios tipo BPF son generados por la irrupción de una vaguada desde la costa hacia el interior y el flujo de aire ascendente en la troposfera media, el cual acompaña a un sistema frontal débil, que va perdiendo energía a medida que se acerca a la cuenca de Santiago. En este tipo de episodio se puede apreciar un centro de bajas presiones al este de la cordillera y una alta fría migratoria desplazada sobre el océano Atlántico, esta condición estando asociada, usualmente, a una cobertura nubosa prefrontal del tipo media y alta con bajo factor de ventilación.

El Potencial Meteorológico de Contaminación Atmosférica (PMCA) [12], por su parte, es un descriptor netamente meteorológico, siendo su valor inversamente proporcional al factor de ventilación, entendiéndose por factor de ventilación el producto del espesor de la capa mezcla superficial por el viento medio dentro de la capa.

Se pueden producir también episodios mixtos, alternándose los episodios de los dos tipos, A y BPF, con períodos de 24 horas.

En la tabla 1.1, se describe las características, condiciones meteorológicas y categorías para este índice (Fuente: CENMA – Universidad de Chile). El PMCA aplica sólo a la cuenca de Santiago y en la época que comprende desde Abril hasta Agosto, asociándose a concentraciones de material particulado.

Tabla 1.1: Potencial meteorológico de contaminación atmosférica (PMCA)

| Categoría del PMCA (Nº de denominación) | Condiciones de ventilación/dispersión de los contaminantes | Características |
|--|--|---|
| 1 Bajo | Muy buenas | <ul style="list-style-type: none"> • Ausencia de inversión térmica de subsidencia. • Sistemas frontales activos. • Marcada inestabilidad. |
| 2 Regular/Bajo | Buenas | <ul style="list-style-type: none"> • Inversión térmica de subsidencia débil y/o elevada. • Sistemas frontales de regular actividad o inestabilidad ligera, • Advecciones generalizadas de aire húmedo y nubosidad baja. • Bajas segregadas o núcleos fríos. |
| 3 Regular | Regulares | <ul style="list-style-type: none"> • Predominio anticiclónico normal del período invernal. • Ausencia de precipitaciones. • Advecciones débiles de nubosidad baja al sector poniente de la cuenca. |
| 4 Regular/Alto | Malas a críticas | <ul style="list-style-type: none"> • Presencia de configuraciones tipo A o BPF. • Índice de circulación zonal medio o alto. |
| 5 Alto | Críticas | <ul style="list-style-type: none"> • Presencia de configuraciones tipo A o BPF. • Bajo índice de circulación zonal. |

De la tabla 1.1, de tipificación del índice PMCA, se concluye que las configuraciones asociadas a las categorías 1,2 y 3 (Bajo, Regular-Bajo y Regular, respectivamente), corresponden en general a condiciones de no-episodios. En cambio, las categorizadas como 4 y 5 (Regular-Alto y Alto) corresponderían a condiciones de episodios de alta contaminación. Se puede comprobar que este índice es más bien de tipo cualitativo que cuantitativo. Y por tanto, es en cierta medida subjetivo.

A partir del año 1988 funciona en Santiago y de manera permanente, una red de monitoreo automático de contaminantes atmosféricos y parámetros meteorológicos (red MACAM), con la cual se han obtenido series temporales que contienen la información que puede ser motivo de diversos estudios.

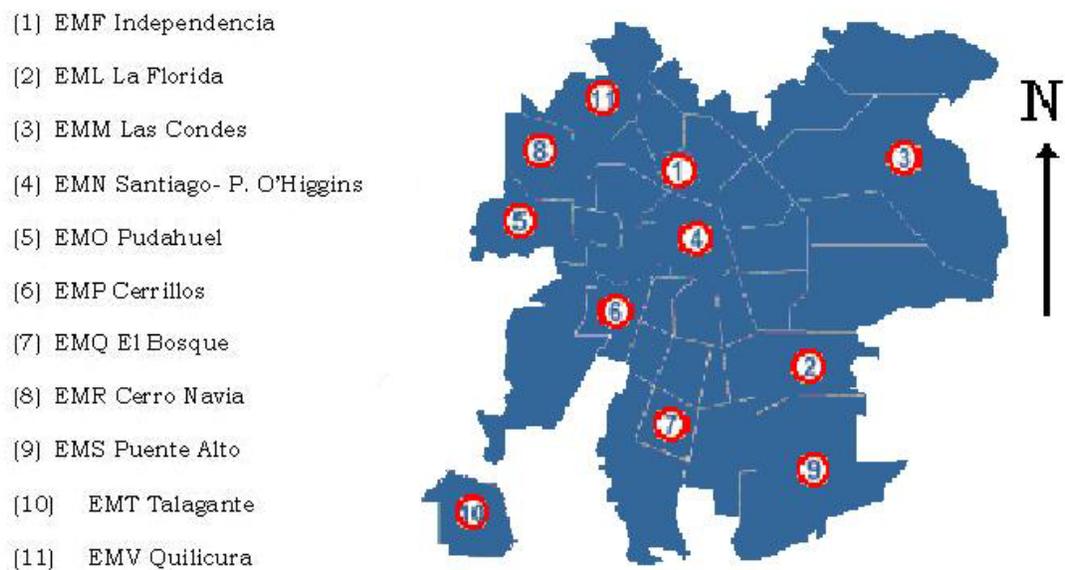


Figura 1.7: Localización de la red de monitoreo MACAM II, en la ciudad de Santiago.

De las estaciones de monitoreo que conforman actualmente la RED MACAM II, hasta hace 2 años, sólo cuatro monitoreaban $PM_{2.5}$.

En la figura 1.7, se muestra la ubicación de las estaciones de monitoreo de la red MACAM II. Las estaciones analizadas en este estudio son: 2, 3, 4 y 5.

1.4 Inventario de emisiones para la ciudad de Santiago

Los inventarios de emisiones se consideran como una herramienta fundamental para la adquisición de información vinculada al problema de la contaminación ambiental y de calidad del aire, siendo estos inventarios el resultado de la recolección y validación de las emisiones producidas, tanto de tipo antropogénico como biogénico, en la ciudad o región en estudio.

El primer inventario de emisiones de la ciudad de Santiago se realizó al año 1989 por encargo de la Intendencia de la región Metropolitana. En la construcción de este tipo de inventario es la componente de las fuentes móviles, en comparación a las fuentes fijas (puntuales y de superficie), la que presenta mayores dificultades en su estimación.

El inventario más reciente fue elaborado por el DICTUC de la Pontificia Universidad Católica de Chile por encargo de la CONAMA-RM [11] y fue dado a conocer el año 2005. Este correspondió a una actualización de un inventario elaborado el año 2000 por el Centro Nacional del Medio Ambiente (CENMA) [6] dependiente de la Universidad de Chile.

El inventario de emisiones [11] se compone de información que proviene de (a) fuentes fijas puntuales (o estacionarias), (b) fuentes fijas superficiales (residencial, comercial, quemados agrícolas, otras), (c) fuentes móviles en ruta (buses, camiones, autos particulares, comerciales, taxis, motocicletas), (d) fuentes móviles fuera de ruta (para aeropuertos, como Tobalaba, A. Merino B., Las Condes y tractores agrícolas), (e) emisiones de polvo fugitivo (en la construcción, demolición, reconstrucción, preparación terrenos agrícolas). La información del actual inventario incluye: material particulado PM_{10} , $PM_{2.5}$, CO (monóxido de carbono), NO_x (óxidos nitrosos), COV (compuestos

orgánicos volátiles), SO_x (óxidos sulfurosos), NH₃ (amoníaco), valores medidos en unidades de toneladas métricas por año.

Tabla 1.2a: Resumen Inventario de Emisiones año 2000 (ton/año)

| | PM ₁₀ | CO | NO _x | COV | SO _x | NH ₃ | Total |
|---------------|------------------|--------|-----------------|-------|-----------------|-----------------|--------|
| Estacionarias | 2597 | 16014 | 8149 | 56264 | 5991 | 28458 | 117473 |
| Móviles | 2425 | 175586 | 46650 | 24664 | 2197 | 933 | 252455 |
| Total | 5022 | 191600 | 54799 | 80928 | 8188 | 29391 | 369928 |

En la tabla 1.2a, se presentan un resumen de los valores de estos componentes contaminantes para el inventario de emisiones del año 2000 [11]. Y en la tabla 1.3a el mismo resumen pero para el inventario del año 2005, al cual se le ha añadido el componente contaminante PM_{2.5} no considerada en el inventario anterior [11].

Tabla 1.2b: Resumen Inventario de Emisiones año 2000 (%)

| | PM ₁₀ | CO | NO _x | COV | SO _x | NH ₃ | Total |
|---------------|------------------|-------|-----------------|-------|-----------------|-----------------|-------|
| Estacionarias | 51.7 | 8.4 | 14.9 | 69.5 | 73.2 | 96.8 | 31.8 |
| Móviles | 48.3 | 91.6 | 85.1 | 30.5 | 26.8 | 3.2 | 68.2 |
| Total | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

En la tabla 1.2b se aprecia que el aporte total de las emisiones a la atmósfera debido a las fuentes móviles fue de 68,2%.

De éstas el CO contribuye con el 69,6% del total de las fuentes móviles. En tanto que los NO_x lo hacen con un 18,5%. A su vez, los COV contribuyen con un 47,9% del aporte debido a todas las fuentes fijas. Se puede apreciar también que el amoníaco (NH₃) aporta con un 7,7% respecto del total de todas las emisiones. Y es este, justamente, un componente fundamental para la conformación de los aerosoles más finos (PM_{2.5}) siendo importante localizar el origen de sus fuentes.

Tabla 1.3a: Resumen Inventario de Emisiones año 2005 (ton/año)

| | PM ₁₀ | CO | NO _x | COV | SO _x | NH ₃ | Total |
|---------------|------------------|---------------|-----------------|---------------|-----------------|-----------------|---------------|
| Estacionarias | 2937 | 23706 | 13815 | 82940 | 13147 | 31825 | 168370 |
| Móviles | 1750 | 213596 | 32010 | 19673 | 168 | 1049 | 268246 |
| Total | 4687 | 237302 | 45825 | 102613 | 13315 | 32874 | 436616 |

En cambio, de la tabla 1.3b se observa que el aporte total de las emisiones a la atmósfera, debido a las fuentes móviles, fue de 61,2%. De éstas el CO contribuye con el 79,3% del total de las fuentes móviles. En tanto que los NO_x lo hacen con un 11,9%.

Tabla 1.3b: Resumen Inventario de Emisiones año 2005 (%)

| | PM10 | PM2.5 | CO | NOX | COV | SOX | NH3 | Total |
|---------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Estacionarias | 62.7 | 68.1 | 10.0 | 30.1 | 80.8 | 98.7 | 96.8 | 38.8 |
| Móviles | 37.3 | 31.9 | 90.0 | 69.9 | 19.2 | 1.3 | 3.2 | 61.2 |
| Total | 100 |

Al comparar las tablas 1.2b y 1.3b, observamos que las fuentes móviles han aumentado su aporte contaminante absoluto, pero han disminuido su aporte porcentual. Esto debido probablemente, al rápido crecimiento del parque automotor, por un lado, y a las mayores exigencias de calidad de los combustibles en los últimos años por otro⁴.

El inventario de emisiones es una herramienta importante para cuantificar la magnitud del problema ambiental de una determinada región y sus actualizaciones permiten evaluar el impacto de las medidas de descontaminación implementadas.

Las metas de calidad del aire en una ciudad, región o país se fijan usualmente en términos de una norma. Lo que se desprende del inventario de emisiones de Santiago, además de las mediciones de las estaciones de

⁴ En el apéndice A.1 se presentan dos tablas más relativo a estos dos Inventarios de Emisiones.

monitoreo permite comprobar que las concentraciones de PM_{2.5} superan con frecuencia la actual norma vigente en los Estados Unidos de América (USA).

La ciudad de Santiago, sin embargo, actualmente no contempla una norma para este tipo de material particulado. Como se menciona en la sección 1.5 en Chile existe una normativa vigente para material particulado PM₁₀, la cual es similar a la que se aplica actualmente en USA, a través de la EPA [13].

En la tabla 1.4 se muestran las normas de PM10 y PM2.5 para algunos países, incluido Chile.

Tabla 1.4: Normas de calidad del aire para PM10 y PM2.5 en diferentes países

| | PM10 | | PM2.5 | | CO |
|------------------------|--------------------------|--------------------------|--------------------------|--------------------------|------------------------|
| | <i>Máx Prom</i> | <i>media Arit.</i> | <i>Máx Prom</i> | <i>media Arit.</i> | <i>Máx Prom</i> |
| | <i>24 hr</i> | <i>Anual</i> | <i>24 hr</i> | <i>Anual</i> | <i>8 hr</i> |
| | $\mu\text{g}/\text{m}^3$ | $\mu\text{g}/\text{m}^3$ | $\mu\text{g}/\text{m}^3$ | $\mu\text{g}/\text{m}^3$ | mg/m^3 |
| OMS | 50 | 20 | 25 | 10 | 10 |
| EPA-USA (NAAQS) | 150 | | 35 | 15 | 10 |
| CANADA | | | 30 | | |
| AUSTRALIA | | | 25 | 8 | |
| EEA (CEE) | 50 | 40 | | 25* | 10 |
| España | 50 | 40 | | 25 | 10 |
| México | 120 | 50 | 65 | 15 | 12,6 |
| Argentina | 150 | 50 | 35 | 15 | 10 |
| Perú | 150 | 50 | --- | --- | 10 |
| Chile | 150 | 50 | 63** | 25** | 10 |

* Aplicado desde Enero de 2005. Para el año 2015 se espera reducirlo a 20.

** A aplicar a partir del año 2012.

El primer caso corresponde a los valores propuestos por la *Organización Mundial de la Salud* (organismo perteneciente a las *Naciones Unidas*). Se observa que la norma de PM10 de la *Comunidad Económica Europea* (CEE) es considerablemente más estricta que la vigente en Chile y EEUU. Por otro

lado, la norma de PM_{2.5} de EEUU es notoriamente más estricta que la propuesta para el año 2012 en Chile.

1.5 La calidad del aire, índice ICAP y normativa vigente

La calidad del aire en Chile se define en términos del promedio móvil (o MA de moving average) de 24 horas de material particulado de diámetro menor o igual a 10 micrómetros, es decir, 24MA de concentración de PM₁₀. Las concentraciones son medidas en unidades de $\mu\text{g}/\text{m}^3$.

Al ser estas partículas lo suficientemente pequeñas, pueden penetrar el sistema respiratorio humano y si se sobrepasa el límite fijado por la Organización Mundial de la Salud (OMS o WHO [14]) de $50 \mu\text{g}/\text{m}^3$ (máximo promedio de 24 horas), se estaría en situación de peligro para la salud de las personas. Aunque la norma en la mayoría de los países se ha establecido para partículas tipo PM₁₀, estudios recientes de epidemiología han sugerido normar partículas de tamaño menores que $2,5 \mu\text{m}$ (PM_{2.5}) [15].

Cada año en el período de invierno Santiago sufre un aumento repentino en los niveles de ciertos contaminantes emitidos a la atmósfera durante un corto período de duración, siendo estas situaciones conocidas como *Episodios Críticos de Contaminación Atmosférica*. El origen de estos episodios se debería a la convergencia de varios factores meteorológicos, los cuales impedirían una buena ventilación de la cuenca de la ciudad de Santiago, interactuando esto con emisiones de contaminantes antes y durante el evento.

Es el decreto Supremo N° 59 de 1998 (correspondiente a la normativa vigente), el que fijó los criterios para decretar o no un episodio crítico de contaminación. Y se refiere sólo al material particulado PM₁₀, a través del *Índice de Calidad del Aire*, ICAP [16].

La normativa exige la existencia de un modelo de pronóstico de concentraciones de PM₁₀, de manera que el episodio se puede establecer

por constatación o en base a los resultados del modelo. La primera metodología que fue implementada en Chile para pronosticar la calidad del aire para material particulado PM₁₀ se hizo a través de la *Resolución N° 12.612 del año 1998, del Ministerio de Salud*, para ser aplicada en la región Metropolitana y aparecido el día 3 de Julio de 1998 en el *Diario Oficial*. En esta resolución se establecen consideraciones y condiciones de confiabilidad del modelo predictivo de la calidad del aire (desarrollado por CONAMA) [17]. Además, allí se describieron los diversos aspectos que determinaron su construcción y posterior operación. Aún más, se estableció detalladamente las variables involucradas en el modelo, la situación de calidad del aire; proyección de emisiones; condiciones de dispersión, y tendencia de las concentraciones. En un anexo, incluido en esta resolución, se indica las tipificaciones de configuraciones meteorológicas asociadas a las categorías de Potencial Meteorológico de Contaminación Atmosférica PMCA.

Este primer modelo de predicción para Santiago [17], se construyó mediante una superposición lineal de efectos en la forma de correcciones a un valor de concentración basal. La predicción consistió en estimar a las 17:30 hrs. de cada día, la máxima del promedio móvil de 24 hrs. de PM10 por cada estación de monitoreo, para las 24 horas siguientes, independiente de cual sea la hora en que se dé dicho valor máximo. Cabe destacar que este modelo fue orientado exclusivamente a PM10, debido a que las situaciones de emergencia detectadas a la fecha en la Región Metropolitana se debieron a las altas concentraciones de este material particulado. Además, este modelo estuvo orientado a la predicción de los niveles de concentración de PM10 registrados en las estaciones del sector sur-poniente de Santiago, en donde se registraron los valores más altos de material particulado.

Por su parte, con la resolución del Servicio de Salud Metropolitano del Ambiente (SESMA) del año 1998, se dio inicio a una marcha blanca del segundo modelo de pronóstico de calidad del aire, para Santiago,

denominado CASSMASSI [18], basado en las modificaciones en cuanto a la metodología de pronóstico del modelo de CONAMA del año 1998. En el año 1999 se validó el modelo desde el 1 de Abril hasta el 17 de septiembre.

Fue el año 2000 que comenzó a operar de manera definitiva y hasta ahora, estando aún vigente. Cabe consignar que, dado el periodo dentro de un año calendario entre el primero de Abril y el 17 de Septiembre para ser construido y luego validado, su uso esta restringido al mismo periodo señalado anteriormente.

La metodología de pronóstico de concentraciones de PM_{10} debido a CASSMASSI, está basada en algoritmos de cálculo que se desarrollaron a partir de las aplicaciones de técnicas estadísticas de regresión múltiple, dirigida a hallar relaciones entre posibles variables predictivas y la variable a predecir.

Las variables que incluyen los predictores son de tipo meteorológicas observadas (altura, temperatura, humedad relativa, vientos observados a niveles estándar de presión; altura, presión, temperatura de la base y tope de la capa de inversión térmica cercana a la superficie; variables derivadas de las anteriores –espesores, diferencias con el día anterior); variables relacionadas con emisiones (día de la semana, índice de día feriado/no feriado, factor de cambio de concentraciones según día de la semana usado en el modelo de CONAMA, factor de cambio de concentraciones según día de la semana específico para cada estación de monitoreo); variables de calidad del aire de PM_{10} (concentración máxima de 24 hrs. observada el día anterior, concentraciones promedio de 24 hrs. observada a las 10:00 del día actual, cambio de concentraciones medias de 24 hrs. observadas a las 10 hrs. –en la última hora, últimas 3 hrs., últimas 6 hrs.-); variable a pronosticar para cada una de las estaciones de monitoreo (concentración máxima de 24 hrs. de PM_{10} esperada para el día siguiente).

En sí, el modelo de pronóstico Cassmassi pronostica el valor máximo de concentración promedio de 24-hrs. de PM_{10} para el periodo de 00 a 24 horas

del día siguiente, en unidades de $\mu\text{g}/\text{m}^3$, en cada una de las estaciones que monitorean este contaminante.

La concentración pronosticada para el día siguiente se calcula mediante ecuaciones que son diferentes para cada estación de monitoreo. Las variables que se requieren para el cálculo de las ecuaciones se obtienen de información relacionada con el cambio esperado de las concentraciones según el día de la semana, las concentraciones de PM_{10} medidas en la red MACAM II, la información meteorológica de altura que se obtiene a través de un aparato de radiosonda de la Dirección Meteorológica de Chile y de las condiciones meteorológicas a escala sinóptica y regional observadas y pronosticadas para la región.

A continuación se presenta como un ejemplo, la ecuación de regresión multivariada que se usa para pronosticar concentraciones de PM_{10} en la estación Pudahuel [19]

$$Y_0 = 39.4V_0 + 0.33C_0 + 2.06T_{0\sigma} + 0.21D_0 - 21.7 \quad [1.1]$$

En esta ecuación, Y_0 es el máximo del promedio móvil de 24-hrs. de PM_{10} esperado para el día siguiente, en $\mu\text{g}/\text{m}^3$. V_0 es la estabilidad atmosférica pronosticada para el día siguiente (PMCA). C_0 es el promedio de 24-hrs. de PM_{10} medido a las 10:00 a.m. del día presente en la estación Pudahuel. T_0 es la temperatura en $^{\circ}\text{C}$ al nivel 925 hPa medido a las 12 UTC del día presente, de acuerdo a los resultados del sondeo que se realiza en Rocas de Santo Domingo situada a 80 Km hacia el oeste de Santiago. D_0 es el cambio, en las últimas 24 horas, de altura de 500 hPa (en metros) medido a las 12 UTC del día presente por medio del mismo sondeo.

Estas dos últimas variables dan información importante acerca de la intensidad de la inversión térmica esperada para Santiago en las siguientes horas.

Las aplicaciones operacionales de esta metodología de pronóstico para cada estación monitorea, considera dos algoritmos. El primero incluye al índice PMCA pronosticado para el día siguiente. El segundo está basado sólo en observaciones (del día actual y del día anterior). De esta forma, si el primer algoritmo no se puede aplicar por información insuficiente, se usa el segundo algoritmo.

Tabla 1.5: Índice ICAP para material particulado respirable

| ICAP | Categoría ICAP | PM ₁₀ µg/m ³ (24hr) | Nivel | Episodio |
|------------------------|----------------|---|-------|----------------|
| 0 – 100 Bueno | 0 | 0 | 0 | - |
| 101 – 200 Regular | 100 | 150 | 0 | - |
| 201 – 300 Malo | 200 | 195 | 1 | Alerta |
| 301 – 400 Crítico | 300 | 240 | 2 | Pre-emergencia |
| 401 – 500 Peligroso | 400 | 285 | 2 | Pre-emergencia |
| > 501 Excede | 500 | 330 | 3 | Emergencia |

El ICAP constatado o el que origina el modelo Cassmassi, son el indicador para que la autoridad competente determine si se está en presencia o no de un episodio catalogado como crítico. Este índice lleva el promedio de concentraciones de PM₁₀, a una escala conveniente. En USA también se utiliza una escala para determinar si se está o no en presencia de un episodio crítico.

Así, habrá un episodio crítico cuando se alcanzan niveles que superen el índice 200 de ICAP. En otras palabras, cuando se superen los niveles de concentración de 195 µg/m³. A continuación vienen las tipificaciones de malo, crítico, peligroso o excede, de acuerdo a la tabla 1.5.

De acuerdo a la actual normativa vigente para material particulado, un índice ICAP de 200, es considerado malo para la salud de la población y corresponde a $195 \mu\text{g}/\text{m}^3$ (media aritmética diaria). Esta última situación la Comisión Nacional del Medio Ambiente, CONAMA, la ha denominado situación de “alerta”. Si se supera los $240 \mu\text{g}/\text{m}^3$ ó 300 ICAP o es menor a 500 ICAP, se denomina “pre-emergencia”. Si es mayor a 500 ICAP, es denominado “emergencia”.

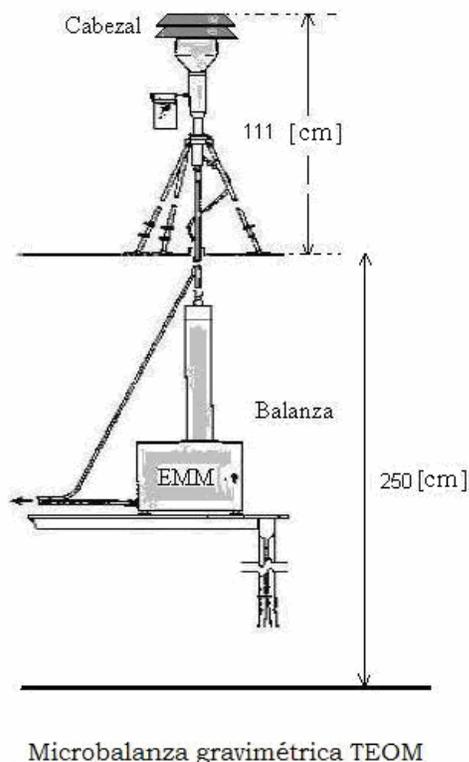


Figura 1.8: Microbalanza gravimétrica TEOM en la red MACAM II

El material particulado grueso y el material particulado fino se miden actualmente mediante una microbalanza que funciona mediante cambios de frecuencia y que permite captar directamente el nivel de concentración que existe en el lugar en estudio donde se toma la muestra. Esta microbalanza denominada TEOM es un monitor que se comporta como un instrumento

gravimétrico el cual toma aire del medio ambiente a través de un filtro y a una proporción de flujo constante.

Este dispositivo pesa este filtro constantemente y calcula las concentraciones de masa en un rango cercano al tiempo real de 10 minutos. Además este instrumento calcula la acumulación de masa total en el filtro, cada 30 minutos y cada una hora. Otros contaminantes se miden usando analizadores de gases que registran los valores recolectados cada cinco minutos. En la figura 1.8 se muestra un diagrama esquemático del instrumento TEOM usado en la red MACAM II.

Para $PM_{2.5}$, en promedio, las concentraciones más altas se observan en la estación EMN (Parque O'Higgins) y las más bajas en la estación EMM (Las Condes).

1.6 La Base de datos

La base de datos usada en este estudio fue mayoritariamente obtenida en parte desde un servidor desarrollado por una empresa Sueca que trabaja para el Gobierno Chileno y denominada AIRVIRO [20]. Desde ahí se puede obtener, entre otros datos, promedios horarios de PM_{10} y $PM_{2.5}$; CO, dirección del viento, velocidad del viento, humedad relativa, temperatura, etc. Esta información existe desde el año 1997 para todas las estaciones de la red MACAM II. Este estudio se enfoca en el pronóstico de $PM_{2.5}$ para los años entre 2004 y 2007. Durante esos años, sólo 4 estaciones de monitoreo midieron concentraciones de $PM_{2.5}$, aquellas ubicadas en La Florida, Las Condes, Parque O'Higgins y Pudahuel.

Además, se utilizaron valores pronosticados y constatados del PMCA, variable discreta estimada por meteorólogos de CENMA y USACH dos veces por día, a las 10:00 a.m. y 10:00 p.m.

Para el desarrollo de un modelo preliminar de pronóstico de PM_{2.5} descrito en la sección 3.3 se utilizaron datos correspondientes a la Red de monitoreo MACAM I, previa a la MACAM II.

1.7 Investigación propuesta

Debido a lo señalado previamente, en el sentido que no existe actualmente una normativa vigente para el material particulado más fino (PM_{2.5}), se pretende desarrollar un modelo que permita hacer pronóstico con algunas horas de anticipación de este tipo de material particulado, de modo que sea posible prever especialmente las situaciones de episodios de alta contaminación. Como ya se mencionó, el sector transporte es una importante fuente de contaminación del aire. Debido a ello se pretende estudiar cómo incorporar el efecto de las fuentes móviles de contaminación.

La aplicación de metodologías nuevas para el análisis de series temporales no lineales ha ido en aumento [21,22,23,24]. El estudio de estos métodos, como lo son la información mutua promedia, la función de autocorrelación y la determinación de falsos vecinos más cercanos [25,26], ha sido útil para fijar algunos parámetros en aplicaciones de la nueva metodología de las redes neuronales artificiales (RNA) [27] para el estudio de series temporales no lineales. Se verá su potencial aplicación a la predicción de valores futuros de series temporales de material particulado fino.

Estudios recientes relativo al material particulado PM_{2.5}, en que se consideran series temporales no lineales, han sido realizado para la ciudad de Santiago sin caracterizar el tipo de fuente (fija o móvil) [28]. En el caso de fuentes móviles, se han hecho estudios que involucran la metodología de redes neuronales artificiales [29], pero en que sólo se realiza predicción y no pronóstico de varios pasos en adelante.

Por otro lado, una nueva metodología, para el análisis y pronóstico de series temporales, ha aparecido recientemente, conocido como Hybrid Clustering

Scheme (HCA: *Hybrid Clustering Algorithm*) [30]. Lo usual en este tipo de algoritmo de cluster en series temporales, es descubrir a un conjunto de grupos de datos con características similares por su proximidad, de acuerdo a una métrica predefinida. Se estudiará si es posible utilizar el algoritmo HCA para detectar concentraciones extremas de PM2.5 dado que en dicho rango las redes de multicapa tienden a perder precisión.

Capítulo 2

Métodos para analizar series temporales no lineales

“There is no happiness except in the realization that we have accomplished something”

- Henry Ford

“Science is a wonderful thing if one does not have to earn one’s living at it”

- Albert Einstein

2.1 Introducción a las series temporales

2.1.1 Fundamentos

Una series de tiempo⁵ (s. de t.) se puede definir como [31] *una secuencia de datos u observaciones de una variable física (o de otro tipo, como económica, por ejemplo), tomados a intervalos de, Δt , representado como un conjunto de variables discretas o continuas*. Matemáticamente hablando, una s. de t. se define como un conjunto de mediciones de cierto fenómeno o experimento, registrados secuencialmente en el tiempo.

Sean estas observaciones:

$$x(t_1), x(t_2), \dots, x(t_N) \quad / \quad t \in T \subseteq \mathbb{R} \quad [2.1]$$

siendo $x(t_i)$ el valor de la variable x en el instante t .

Si $T = \mathbb{Z}$, se dice entonces que la s. de t. es discreta. Y si $T = \mathbb{R}$, se dice que es continua.

⁵ O serie temporal o también, serie cronológica.

Si $t_{i+1} - t_i = k \quad \forall i = 1, 2, \dots, n - 1$, se dice que la serie es equiespaciada. En caso contrario, es no equiespaciada.

Normalmente estas series temporales están basadas en valores medidos que portan ruido. Los valores de estas series generalmente contienen una parte sistemática (determinista) y otra componente estocástica (aleatoria), que representa a la interferencia ruidosa, la cual a su vez provoca fluctuaciones estadísticas alrededor de los valores deterministas [32].

No obstante, no se puede observar ninguna de estas partes en las series reales. Por lo que los métodos de pronóstico (*forecast*) tratan de aislar la parte sistemática. Así, los pronósticos están basados en la parte sistemática de la serie. Sin embargo, la parte aleatoria determina la forma de la distribución. Por lo tanto, uno de los problemas que intenta resolver las s. de t. es el de la predicción.

Se puede afirmar, entonces, que el propósito del análisis de una s. de t. es comprender la variabilidad de éstas; identificar las oscilaciones regulares e irregulares de estas series; describir las características de estas oscilaciones y estudiar los procesos físicos – entre otros -, que dan origen a cada una de estas oscilaciones.

Aún más, se puede afirmar que los tres objetivos principales en el análisis de las series temporales, son: comprender la dinámica subyacente en la serie; pronosticar eventos futuros, y controlar eventos futuros vía intervención [33].

Por lo general, el análisis de las series cronológicas se refiere a estimar los factores (o componentes) que producen el comportamiento general o patrón de las series; y el uso de estas estimaciones para predecir el comportamiento futuro de la serie.

Al analizar una s. de t. dada, primero se enfoca en estudiar su estructura interna (autocorrelación, tendencia, estacionalidad, etc.), para alcanzar una mejor comprensión del proceso dinámico mediante el cual la serie de datos fue generada. Vale decir, se necesita identificar el ciclo regular (autocovarianza, análisis armónico); estimar la importancia de estos ciclos (análisis de espectro de potencia) y aislar o remover estos ciclos (filtrado).

Básicamente existen dos enfoques para el análisis de series temporales: el enfoque en el dominio temporal y en el dominio espectral. El primero está basado en el uso de la función de covarianza para las series temporales, la cual se realiza a través del enfoque de Box-Jenkins [34], que está, a su vez, basado en el modelamiento lineal de fenómenos estacionarios. En cambio, el segundo approach está basado en la función de densidad espectral y en el análisis de Fourier.

La mayoría de las s. de t. son complejas y todas contienen por lo menos una de las siguientes componentes: linealidad, tendencia (trend) o tendencia secular; variación cíclica (periódica, estacionalidad o variación estacional - *seasonality*), estacionaria, ruido, y variación irregular o caótica.

Una serie temporal es lineal si valores futuros de esta serie pueden expresarse como una función lineal de alguno o todos sus valores previos.

La tendencia se entiende como el movimiento continuo de una variable durante un período de tiempo extendido. En otras palabras, es la conducta a largo plazo de la variable durante un período de longitud prolongado, y refleja la dirección general de la s. de t. la cual puede ser ascendente o descendente.

Por su parte, la variación cíclica o estacionalidad son las variaciones periódicas, en el nivel general de la actividad, durante un período relativamente prolongado. Series de tiempo periódicas con componentes periódicas son periódicas. Matemáticamente, se puede decir que la serie presenta variación estacional si existe un número s tal que:

$$x(t) = x(t + ks) \quad [2.2]$$

Se puede afirmar que una serie temporal es estacionaria si su media y varianza son constantes en el tiempo y la autocovarianza depende sólo del rezago temporal.

El ruido aleatorio puede estar presente en toda o parte de la serie temporal.

Por último, la variación caótica se produce por sucesos inusuales, que producen movimientos sin un patrón discernible. Tales movimientos son únicos y probablemente no ocurrirán de nuevo de la misma manera.

En términos estadísticos, una s. de t. es estacionaria cuando el proceso estocástico subyacente está en un estado particular de equilibrio estadístico. Vale decir, cuando la probabilidad conjunta de $x(t)$ e $y(t - \tau)$ sólo depende de τ y no de t .

Un modelo clásico para una s. de t., supone que una serie $x(t_1), x(t_2), \dots, x(t_N)$ puede ser expresada como una suma o producto de tres componentes: tendencia, estacionalidad y un término de error aleatorio [38].

Existen tres modelos de s. de t. que, generalmente, se acepta como buenas aproximaciones a los componentes de los datos observados. Estas son:

$$(1) \text{ Aditivo: } x(t) = T(t) + E(t) + \epsilon(t) \quad [2.3a]$$

$$(2) \text{ Multiplicativo: } x(t) = T(t) \cdot E(t) \cdot \epsilon(t) \quad [2.3b]$$

$$(3) \text{ Mixto: } x(t) = T(t) \cdot E(t) + \epsilon(t) \quad [2.3c]$$

con $x(t)$ siendo la s. de t. observada en el instante t ; $T(t)$ la componente de tendencia; $E(t)$ la componente de estacionalidad y $\epsilon(t)$ la componente aleatoria.

Una suposición usual es que $\varepsilon(t)$ sea una componente aleatoria o ruido blanco, con media cero y varianza constante.

Una forma de visualizar la tendencia, es mediante el suavizamiento de la serie:

$$Z(t) = F(x(t)) \quad [2.4]$$

en que F se denomina *Filtro lineal*. El filtro lineal más usado es el *Promedio Móvil* (moving average, MA). Con este filtro se busca eliminar las componentes estacionales y estocásticas (ruidosas).⁶

La función autocovarianza se usa para estimar los periodos dominantes en las s. de t. Se define por autocovarianza, como la covarianza de una variable con ella misma, en algún otro instante de tiempo, medido por un cierto *lag* o rezago temporal, τ .

A partir de la definición de autocovarianza, se puede definir la *autocorrelación* como la autocovarianza normalizada. En la figura 2.1 se puede ver un bosquejo de una función de autocorrelación típica.

Si el *lag* es pequeño, la autocorrelación es aún positiva para muchas variables físicas. Esto significa que existe alguna “persistencia” en las variables. Luego, si existen N observaciones en secuencia, la función de autocorrelación nos dice que los datos significa decir a su vez que el grado de libertad es menor que el número de datos, N.

El análisis armónico se usa para identificar variaciones periódicas (o regulares) en una serie temporal. Así, el espectro de potencia de una s. de t. se obtendrá al graficar la amplitud de los armónicos como una función de k .

Entre los modelos de s. de t. lineales más populares [38], están los modelos AR (autorregresivos) y los modelos MA; los modelos ARMA, son frecuentemente usados para modelar estructuras dinámicas lineales, para

⁶ Se suaviza cuando existen muchos cambios bruscos o movimientos irregulares.

describir relaciones lineales entre variables retardadas (lagged), y servir como medios para pronóstico lineal. Un caso particularmente útil es el ARIMA, el cual contiene a todos los modelos anteriores, incluyendo el ARMA estacionario.

Típica Función Autocorrelación



Figura 2.1: Bosquejo de una función de autocorrelación

El primer modelo (AR) se define como:

$$X_t = b_1 \cdot X_{t-1} + \dots + b_p \cdot X_{t-p} + \epsilon_t \quad [2.5]$$

en que el último término es un ruido blanco.

El modelo MA se define como:

$$X_t = \epsilon_t + a_1 \cdot \epsilon_{t-1} + \dots + a_q \cdot \epsilon_{t-q} \quad [2.6]$$

Finalmente un proceso ARMA se define como:

$$X_t = b_1 \cdot X_{t-1} + \dots + b_p \cdot X_{t-p} + \epsilon_t + a_1 \cdot \epsilon_{t-1} + \dots + a_q \cdot \epsilon_{t-q} \quad [2.7]$$

2.1.2 Series temporales no lineales

En la naturaleza, los fenómenos asociados a ella, así como en diversas disciplinas, las series cronológicas no siempre son lineales, sino más bien son series temporales no lineales [36,37,38]. Por lo que las herramientas definidas anteriormente, en 2.1.1, no servirán para el estudio y análisis de este tipo de s. de t.

Es así que el comportamiento determinista y no lineal ha sido observado en diversas series temporales desde la física, la economía, la fisiología, la biología, etc. Una característica común de estos datos, es que ellos son altamente complejos y exhiben una correlación no lineal [39].

En las últimas décadas, la comprensión de sistemas caóticos y otros no lineales, llegó a un punto donde la falla de los modelos lineales era evidente, en apariencia, para describir muchos procesos dinámicos. Sin embargo, los métodos lineales siguen siendo potentes cuando se analizan los datos. Esto porque la teoría lineal está altamente desarrollada, a diferencia de los métodos no lineales en la teoría de los sistemas no lineales, la cual está incompleta y aún en desarrollo, comparado con la elegancia y completitud que tienen los algoritmos de modelado lineal. Sin embargo, los métodos lineales convencionales son, en general, inapropiados para modelar y hacer predicción de las series temporales que han sido originadas desde procesos no lineales. Por ejemplo, las series cronológicas que surgen desde procesos caóticos, exhiben típicamente un espectro de Fourier continuo, el cual es opuesto a las frecuencias discretas de modelos lineales. Por lo que estos métodos son generalmente ineficientes. Durante los pasados años, muchas técnicas de modelamiento y análisis de s. de t. no lineales han sido sugeridas [39]. Una característica común de estos métodos, es la reconstrucción del *espacio de estado* o *espacio de las fases*, de un sistema dinámico seguido del teorema de *encajamiento* (embedding), debido a Takens [40].

2.1.3 Técnicas de modelamiento para series temporales no lineales

Mediante técnicas de modelación de sistemas dinámicos [21,22,23,24,25], se ha logrado caracterizar diferentes tipos de series temporales en la naturaleza. Para ello, corrientemente se emplean como herramientas matemáticas la función AMI (*Información Mutua Promedio* o *Average Mutual Information*) y la función FNN (*Falsos vecinos más Cercanos* o *False Nearest Neighbours*).

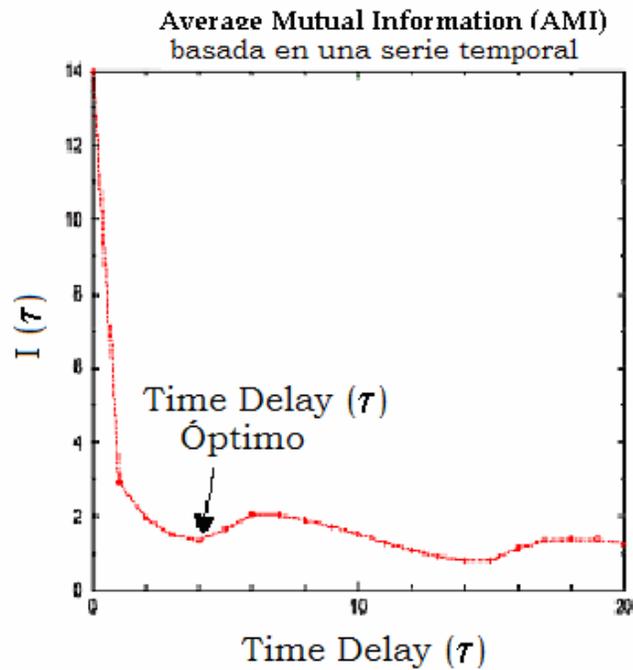


Figura 2.2: Información Mutua Promedio para una s. de t. dada.

La función AMI, $I(\tau)$, nos entrega el promedio de la información binaria que se transmite entre datos espaciados en un intervalo de tiempo T . La información mutua, también conocida como *información conjunta*, es una medida de la dependencia entre variables. Es además una medida de carácter general, que tiene en cuenta correlaciones de todo tipo (y no solo lineales). Así, si las variables son independientes entre sí, $I(\tau)$ será cero y su valor será tanto mayor cuanto mayor sea la correlación entre ambas. Esta función tiene su propio mínimo en cierto instante, simbolizado por τ , que corresponde al espaciamiento que implica una correlación apropiada entre datos sucesivos para permitir un mejor modelamiento. En la figura 2.2 se puede observar este primer mínimo. Se considera un buen estimador del tiempo a este primer mínimo hallado por el método AMI, para efectuar la reconstrucción del atractor en el espacio de las fases [26,39,40].

Por otro lado, la función FNN, permite hallar la dimensión de encaje $d_e(\tau)$, correspondiendo a la mínima cantidad necesaria de variables dinámicas para modelar al sistema. En nuestro caso, corresponde a la cantidad de datos pasados consecutivos necesarios para poder hacer una predicción, un cierto intervalo de horas hacia el futuro. Para esta función, si su curva decrece rápidamente, pero nunca se va a cero, entonces se tiene indicios de que se estaría en presencia de ruido.

Más explícitamente, la dimensión encajada de un sistema dinámico es una dimensión global entera que da el número de coordenadas necesarias para desplegar (*unfold*) su trayectoria, en el espacio de fases. Ver figura 2.3. Para el cálculo de esta dimensión de encaje se necesita tener el rezago temporal, τ , para cada serie de tiempo, la que es obtenida a partir de la función AMI. Una vez que se tiene este valor, se escoge una dimensión d para algún punto $s(k)$. Con ello será suficiente conocer el vector:

$$y(k) = (s(k), s(k + \tau), \dots, s(k + (d - 1) \cdot \tau)) \quad [2.8]$$

para poder predecir la evolución del sistema dinámico. Entonces, en el espacio d -dimensional, el vector $y(k)$ tiene un vecino más cercano $y^{NN}(k)$, el cual es el vector con la mínima distancia euclidiana a él. Si d es menor que d_e , es muy probable que este vecino más cercano sea falso; vale decir, que corresponde a una proyección desde el espacio dimensional d_e real.

Este método requiere colocar dos valores umbrales a algún valor más arbitrario, los cuales se usan para determinar los falsos vecinos. Cabe señalar que este método es sensible a la tasa de muestreo de la serie de tiempo. Por último, se puede decir que este método aunque es intuitivo y fácil de implementar, no es directo su uso e interpretación.

La función AMI se define como [21,41]:

$$I(\tau) = \sum_{s(n), s(n+\tau)} P(s(n), s(n + \tau)) \log_2 \left[\frac{(Ps(n), s(n + \tau))}{(Ps(n) \cdot s(n + \tau))} \right] \quad [2.9]$$

siendo $P(s(n))$ la probabilidad de obtener el valor s en el instante n , y $P(s(n), s(n + \tau))$ es la probabilidad conjunta de conseguir un valor dado de la variable s en el instante $n + \tau$, dado el valor en el instante n .

El método de los Falsos Vecinos más Cercanos (FNN)

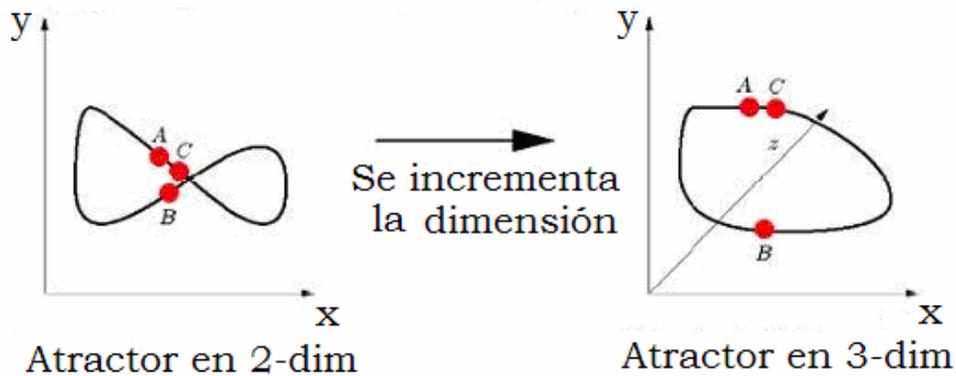


Figura 2.3: Método de los falsos vecinos más cercanos.

Por su parte, la función FNN se define como sigue: el número de falsos vecinos próximos para un d dado es el número de veces que la ecuación [2.10] supera un umbral dado [21,41,42].

$$\sqrt{\frac{R_{d+1}^2(k) - R_d^2(k)}{R_d^2(k)}} = \frac{|s(k + d \cdot \tau) - s^{NN}(k + d \cdot \tau)|}{R_d(k)} \quad [2.10]$$

en que R_d^2 es el cuadrado de la distancia euclidiana entre los puntos vecinos más cercanos, vistos desde un espacio en d dimensión; y lo mismo para R_{d+1}^2 .

2.2 Introducción a la teoría del Caos

2.2.1 Fundamentos

El gran poder de las ciencias descansa en la capacidad de relacionar *causa* y *efecto*. Las Leyes de la Gravitación debida a *Isaac Newton* permiten

predecir con miles de años de anticipación los eclipses astronómicos. Sin embargo, existen otros fenómenos de la naturaleza cuya predicción no es tan simple [43,44]. Por ejemplo el clima, ya que los movimientos de la atmósfera a pesar de obedecer las leyes de la física en la misma medida que lo hacen los planetas del Sistema Solar, las predicciones meteorológicas se hacen aún en términos de probabilidades. Al igual que este caso, hay otros fenómenos en la naturaleza que presentan aspectos impredecibles. Y esto debido a que no hay una relación clara entre causa y efecto, por lo que se dice que estos fenómenos poseen elementos *aleatorios*.

La *Teoría del Caos* [45,46,47] es aquella rama de las matemáticas y la física que trata ciertos tipos de comportamientos impredecibles de los sistemas, los que pueden clasificarse, básicamente, como estables, inestables y caóticos [48,49,50].

Para poder clasificar el comportamiento de un sistema dinámico como caótico, deberá tener las siguientes propiedades:

- Debe ser sensible a las condiciones iniciales.
- Debe ser transitivo, y
- Sus órbitas periódicas deber formar un conjunto denso en una región compacta del espacio de fases.

La sensibilidad a las condiciones iniciales está directamente relacionada con el *exponente de Lyapunov* [51,52,53], que se verá en más detalle en la sección siguiente, y es una cantidad que caracteriza el radio de separación de dos trayectorias que eran al principio infinitesimalmente próximas. A modo de ilustración de esta sensibilidad a las condiciones iniciales se menciona el llamado *efecto mariposa*, en donde el aleteo de las alas de una mariposa podría crear sutiles cambios en la atmósfera, los que durante el transcurrir del tiempo podrían modificarse hasta hacer que ocurra un tornado.

Por su parte, el término transitividad quiere decir que existen muchas órbitas densas, en el espacio de las fases. Esto implica que un sistema dinámico no podrá descomponerse en dos subsistemas independientes entre sí. En otras palabras, a pesar que las características de sensibilidad a las condiciones iniciales señala que las órbitas que se hallan infinitamente cercanas se alejan en el tiempo, dichas órbitas pasarán por todas las regiones que componen el atractor extraño, a largo plazo, formando un único sistema dinámico.

Además, y a modo de aclaración se entenderá por caos no la ausencia de orden, sino más bien cierto orden de características impredecibles, pero que se puede describir en forma concreta y precisa.

El comportamiento caótico está ligado a lo que se conoce como *atractores extraños*, los cuales pueden llegar a tener una enorme complejidad, como por ejemplo el atractor de Lorenz, el cual surge del modelo tridimensional del sistema climático [54].

Con el advenimiento de la *Teoría del Caos* se hizo un sorprendente descubrimiento: algunos sistemas deterministas muy simples pueden generar comportamientos aleatorios. A esta aleatoriedad se le ha denominado *Caos*. Surge, entonces, una aparente paradoja: *el caos es determinista*. Así, en principio el futuro está completamente determinado por el pasado, pero en la práctica las pequeñas incertidumbres se agrandan, de tal forma que si bien el movimiento es *predecible a corto plazo, no lo es a largo plazo*. Aún más, se encontró que había *orden en el caos*. Bajo el comportamiento errático subyacen formas geométricas [55].

Por otro lado, el determinismo inherente en el caos muestra que muchos fenómenos aleatorios son más predecibles que lo que se había creído. El caos permite, entonces, hallar orden en sistemas tan diversos como la atmósfera, el corazón, etc.

Una propiedad básica del caos es que cualquier efecto, por pequeño que parezca, adquiere rápidamente proporciones macroscópicas. Otra propiedad básica lo constituye el crecimiento exponencial de los errores debido a la dinámica caótica. Así, se entenderá por caos a la *ocurrencia de eventos aperiódico aparentemente aleatorio en un sistema determinista*.

El marco general del que emerge el caos es la teoría de los *sistemas dinámicos*, el cual consta de dos partes: la noción de *estado*, que es la información esencial de un sistema; la segunda es la *dinámica* la cual es una regla que describe cómo evoluciona el estado en el tiempo [56]. Esta evolución se representa en un *espacio de configuraciones* o *espacio de las fases*, concepto abstracto cuyas coordenadas son las componentes del estado. Así, las coordenadas de este espacio variarán de acuerdo al sistema en cuestión.

En el caso de un sistema mecánico, las coordenadas serán posiciones y velocidades.

La evolución temporal de un sistema dinámico puede desarrollarse de modo continuo o discreto. En el primer caso se hablará de *flujo* y en el segundo de una *aplicación*.

Se puede recurrir a la dinámica, o *ecuaciones de movimiento*, para saber cómo evoluciona un sistema desde su estado inicial. En el caso de sistemas simples (como el péndulo simple sin rozamiento), puede suceder que las ecuaciones admitan una solución explícita. Vale decir, una fórmula que exprese cualquier estado futuro en función del estado inicial. Luego, una solución explícita suministra un atajo, un algoritmo simple que precisa sólo del estado inicial y el tiempo final para predecir el futuro sin pasar por estado intermedios [57].

Sin embargo, el comportamiento impredecible de los sistemas dinámicos caóticos no puede expresarse mediante una solución explícita. Por lo tanto, no existen atajos para predecir su comportamiento futuro.

En términos generales, y considerando la dinámica de un fenómeno de la naturaleza, cuando las ecuaciones y relaciones son lineales, entonces se pueden encontrar una solución explícita o “resolver” el sistema. Sin embargo, si las ecuaciones son de carácter no lineales en general no se pueden resolver estas ecuaciones.

Si tomamos como ejemplo el caso de la fricción, al introducir no linealidad en las ecuaciones de movimiento de algún sistema simple provocará dificultades en su resolución. Sin fricción, como sabemos, la cantidad de fuerza requerida para acelerar un cuerpo se puede expresar como una relación lineal: *Fuerza = masa · aceleración*.

Por lo tanto, la fricción complica las cosas, ya que la cantidad de energía cambia, dependiendo de cuán rápido el objeto se esté moviendo.

La no linealidad, por lo tanto, cambia las reglas deterministas dentro de un sistema y hace difícil de predecir qué es lo que va a suceder un tiempo más adelante [58].

2.2.2 Exponentes de Lyapunov y Predictibilidad

Además de las técnicas para analizar s. de t. no lineales descritas en la sección 2.1.3, a saber, información mutua promedio y falsos vecinos más cercanos, existe otra más la cual es hallar el exponente de Lyapunov [59,60,61], con la adición de que este parámetro dará luz sobre la caoticidad o no de la serie temporal en estudio.

El exponente de Lyapunov de un sistema dinámico se define [62,63,64] como una cantidad que permite detectar el grado de separación entre dos trayectorias inicialmente muy próximas entre sí, en el espacio de las fases.

Supongamos que un sistema dinámico evoluciona de acuerdo a la ecuación

$$x_{n+1} = f(x_n), \quad f \in C^1 \quad [2.11]$$

Considerar un punto próximo a un atractor del sistema dinámico. Ver figura 2.4. Sea este punto denotado por $x_n + \delta x_n$. Si hacemos una expansión en serie de Taylor de la ecuación [2.11] se obtiene

$$\delta x_{n+1} = \frac{df}{dx} \delta x_n + \dots \quad [2.12]$$

Se define

$$e^\lambda = \left| \frac{\delta x_{n+1}}{\delta x_n} \right| = \left| \frac{df}{dx} \right| \quad [2.13]$$

que en la literatura se denomina *número local de Lyapunov* [52,64].

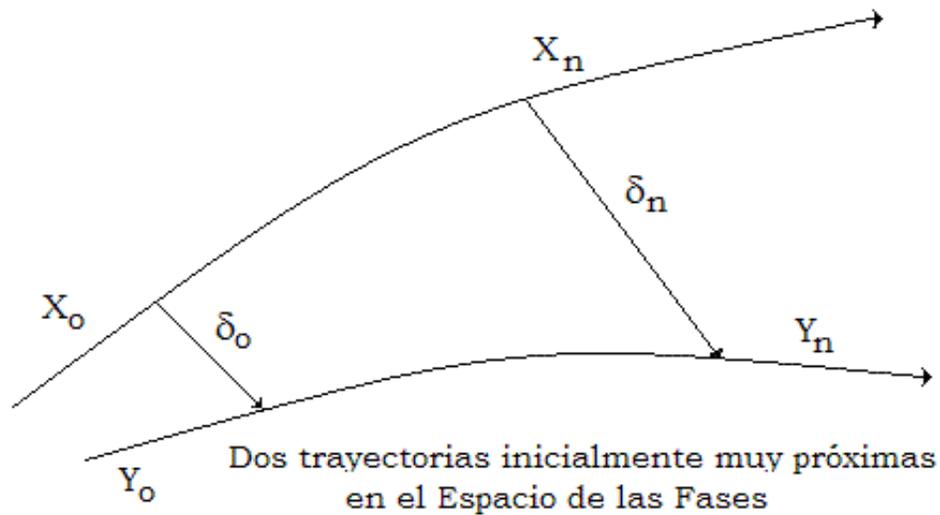


Figura 2.4: El caso de dos trayectorias en el espacio de las fases
 Por lo tanto, en exponente local de Lyapunov se define como

$$\lambda = \log \left| \frac{df}{dx} \right| \quad [2.14]$$

Ya que la expresión $\left| \frac{df}{dx} \right|$ sobre la órbita usualmente no es constante, se promedia $\langle \log \left| \frac{df}{dx} \right| \rangle$ sobre muchas iteraciones.

El parámetro que aparece en esta ecuación, λ , es el *exponente de Lyapunov*. Supongamos que el sistema en estudio tiene dimensión N . Esta dimensión es también la del *Espacio de las Fases* [45,47,52]. Entonces, habrá N exponentes de Lyapunov (se dice que hay *un espectro completo de exponentes de Lyapunov*). Por su parte, si se considera que los primeros D de estos exponentes son cero, el atractor deberá tener dimensión D . El entero D es algunas veces denominado la *dimensión topológica* [52].

Es común referirse sólo al más grande de los exponentes, pues determina una cota máxima para la *predictibilidad* del sistema dinámico en estudio [65].

Considerar el caso de una dimensión. Si se parte de dos trayectorias bien próximas, por ser x_0 e y_0 con separación inicial $\delta_0 = y_0 - x_0$, entonces divergen como:

$$\delta_n \equiv y_n - x_n = f^n(y_0) - f^n(x_0) = \delta_0 e^{\lambda_n n} \quad [2.15]$$

Despejando λ_n , se obtiene:

$$\lambda_n = \frac{1}{n} \ln \left(\frac{\delta_n}{\delta_0} \right) = \frac{1}{n} \ln \left(\frac{f^n(y_0) - f^n(x_0)}{y_0 - x_0} \right) \quad [2.16]$$

Cuando δ_0 es muy pequeño, se puede aproximar λ_n por:

$$\lambda_n \approx \frac{1}{n} \ln (f'^n(x_0))$$

En esta última ecuación f'^n el término en paréntesis en ecuación [2.16], que luego de hacer la aproximación para λ_n se transforma en una derivada.

Pero como

$$f'^n(x_0) = \prod_{i=0}^{n-1} f'(x_i)$$

se tendrá finalmente

$$\lambda_n \approx \frac{1}{n} \ln f^{(n)}(x_0) = \frac{1}{n} \sum_{i=0}^{n-1} \ln f'(x_i) \quad [2.17]$$

En la ecuación [2.17] λ_n si $n \rightarrow \infty$ es el exponente de Lyapunov, cuando este límite existe.

Los exponentes de Lyapunov son parámetros que intentan evaluar el *horizonte de predictibilidad* del sistema (*predictability*). En otras palabras, hasta cuánto más allá podemos predecir, con cierta precisión, si ocurrirá determinado suceso [65,66].

Se puede hacer una estimación numérica de los exponentes de Lyapunov a partir de una serie temporal. Si el sistema dinámico en estudio posee dimensión d , determinada por los métodos descritos en 2.1.3, entonces el sistema será caótico si al menos uno de los d exponentes de Lyapunov es positivo. Si, además, estos exponentes se ordenan de mayor a menor y el exponente máximo es positivo, será prueba adicional de la existencia de caos. Esto hace que exista un horizonte de predictibilidad finito. La suma de todos ellos, además, debe ser negativa para que haya un atractor estable. Si la suma no es negativa, entonces el sistema tenderá al azar. Por el contrario, los sistemas con todos los exponentes cero o negativos son lineales, es decir, son predecibles.

Cuando se tiene que el exponente de Lyapunov más grande es positivo, en una serie temporal, entonces indicará trayectorias disipativas con bifurcación y multiperiodicidad de tal modo que la serie de tiempo no podrá ser predicha a largo plazo. No obstante, su máxima escala de tiempo predecible T_f se puede estimar usando la siguiente relación con el máximo exponente de Lyapunov L_{E_1} :

$$T_f = \frac{1}{L_{E_1}} \quad [2.18]$$

en que T_f tiene la misma unidad que Δt , que es la separación entre dato y dato.

Huang [67] señaló que dentro del rango de máxima escala temporal predecible, el error sistemático de predicción es insensible a la variación de la longitud del paso en adelante de la predicción. Sin embargo, fuera de este rango el error sistemático puede aumentar mucho. Luego, T_f está definido como un índice para medir predictibilidad de sistemas caóticos [65,66].

Existen diferentes métodos de cálculo para determinar estos exponentes de Lyapunov, siendo el más simple el debido a Wolf *et al.* [68,69].

2.2.3 Otros parámetros para estudiar posible caos en series temporales no lineales

Los siguientes parámetros permiten hacer un estudio complementario a los ya señalados anteriormente y son la dimensión de correlación, la dimensión de la capacidad, la función correlación, el exponente de Hurst, la entropía y los datos sustitutos.

La dimensión de la capacidad (similar a la dimensión de Hausdorff⁷), se calcula dividiendo sucesivamente el espacio de las fases con la dimensión encajada, d_e , en hipercubos iguales y graficando el logaritmo de la fracción de hipercubos que son ocupados con puntos de la data versus el logaritmo de la dimensión lineal normalizada del hipercubo. La dimensión está bien definida sólo si existe un plateau en este gráfico [38,52]. Al calcular la dimensión de la capacidad para valores de dimensión de encaje creciente, la dimensión de la capacidad debería incrementarse también, pero eventualmente se saturará en el valor correcto. Sin embargo, una dimensión de capacidad más grande que 5 implicará, esencialmente, datos aleatorios.

Por su parte, la dimensión de correlación (d_c) [70,71] de un sistema dinámico es una medida de su complejidad y permite discriminar entre un sistema determinista, uno aleatorio y uno caótico. Básicamente consiste en

⁷ Esta dimensión es la definición formal para determinar la dimensión fractal de un objeto e indica qué tanto ocupa el espacio que la contiene..

que al ir recorriendo con cada iteración los puntos, a través de la base de datos, se toma un nuevo punto y una esfera hiperdimensional de dimensión encajada d_e y radio r se centra en cada punto. Se calcula entonces la fracción de puntos subsiguientes en el registro dentro de esa esfera para varios valores de r y se hace un gráfico del logaritmo de este número versus el logaritmo del radio. Una dimensión de correlación apropiada requiere un *plateau* sobre este gráfico. A medida que la dimensión de encaje crece, la dimensión de correlación debería incrementarse, pero eventualmente saturarse en el valor correcto. Si d_c es igual a 1, indicará que la serie temporal posee un comportamiento oscilatorio periódico; por su parte, si d_c es igual a 2 indicará que la serie posee oscilaciones cuasi-periódicas; si d_c es mayor que 2 será indicio que la serie posee comportamiento caótico; y finalmente, si d_c es igual, aproximadamente, a 5 implicará esencialmente datos aleatorios [52].

La entropía (o entropía informática), por su parte, la cual es una medida del grado del desorden, es la suma de todos los exponentes positivos de Lyapunov (en base e), siendo el algoritmo de cálculo de este parámetro el debido a Procaccia y Grassberger [72]. Su recíproco es aproximadamente el tiempo sobre el cual es posible hacer una predicción que tenga sentido – llamado horizonte de predictibilidad- [52].

El exponente de Hurst [52], H , es un número que indica el grado de influencia del estado presente sobre el estado futuro de la serie temporal. También se puede comprender como el grado de similitud del fenómeno en cuestión con el *movimiento Browniano* o *camino aleatorio*. Los valores de H varían entre 0 y 1. Si $H > 0,5$ implica series de tiempo persistentes (correlación positiva); vale decir, caracterizadas por efectos de memoria de largo plazo. Por ejemplo, si $H = 0,7$ entonces hay una probabilidad del 70% que el siguiente elemento de la serie temporal exhiba la misma tendencia

que la del elemento actual⁸. Si $H=0,5$ el sistema es aleatorio (correlación nula o ruido blanco). Finalmente, el sistema es antipersistente (correlación negativa) si $H < 0,5$, el cual significa que el sistema cubre menos distancia que el aleatorio, en el caso de una partícula errática.

Una gran diferencia entre la información mutua y la dimensión de correlación, es que la primera puede emplearse tanto a secuencias de números como a secuencias de símbolos. En cambio, la segunda sólo puede aplicarse a secuencias numéricas. Usualmente, la información mutua cae a cero más rápidamente que la función correlación. La información mutua sirve de medida efectiva de complejidad, siendo sus propiedades y especialmente su generalidad (no está restringido a correlaciones lineales) especialmente apta para el estudio de series temporales no lineales que presentes casticidad.

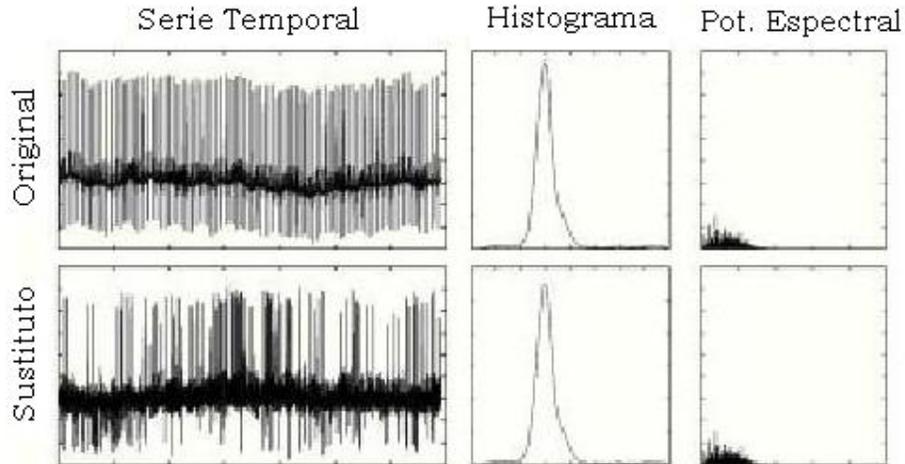


Figura 2.5: Series original y serie sustituta con el análisis de espectro de potencia e histograma

Existen diferentes métodos para generar series de datos sustitutos (surrogate) [52,73], permite hacer un estudio más refinado de series supuestamente deterministas. Lo que justamente hace es “barajar” los datos

⁸ Se ha encontrado que las s. de t. persistentes son las que más comúnmente se hallan en la naturaleza y en los mercados de capital y de economía.

colocándolos al azar y se vuelve a hacer el mismo análisis de los parámetros anteriores. En otros términos, se trata de conseguir una serie de datos aleatorios que “imiten” determinadas propiedades de la serie original en el dominio del tiempo (como la varianza y la amplitud intercuartil), en el dominio de las frecuencias (por ejemplo, la potencia espectral) o en ambas.

En la figura 2.5 se muestra un ejemplo de cierta serie temporal a la cual se le ha determinado su histograma y espectro de potencia tanto a la serie original como a la serie sustituta. En estas series sustitutas se trata de conservar las propiedades lineales de la serie original, destruyendo a la vez toda estructura no lineal con la con el proceso de barajado de los datos. Por su parte, el método de Schreiber y Schmitz [38] permite conseguir series sustitutas que presentan el mismo histograma y prácticamente la misma potencia espectral que la serie original.

Estas metodologías han sido ya aplicadas a la predicción series de tiempo caóticas mediante redes neuronales artificiales [74].

2.3 Fundamentos de las Redes Neuronales Artificiales

2.3.1 La neurona biológica y su naturaleza bioeléctrica

Las Redes Neuronales Artificiales (RNA) están inspiradas en las neuronas biológicas del cerebro humano [75] el que consiste de unas cien mil millones de neuronas que conectadas entre sí son capaces de realizar tareas complejas como memoria, asociación, aprendizaje, clasificación, percepción, atención, comunicación, razonamiento, inspiración, conciencia, entre otros atributos del sistema; esto gracias a las aproximadamente 10^{15} interacciones interneuronales que se producen en total.

La unidad “básica” del cerebro humano es la neurona -célula nerviosa especializada- y más aún del *Sistema Nervioso Central* (SNC). Las neuronas se comunican entre sí a través de la sinapsis. Cada neurona tiene alrededor

de 60.000 puntos de contacto o sinapsis, que le permite comunicarse con otras neuronas.

La señal generada por la neurona y que es transportada a lo largo del axón es un impulso eléctrico [75]. Ver figura 2.6. En cambio, la señal que se transmite entre los terminales axónicos de una neurona y las dendritas de la neurona siguiente es de origen químico; más específicamente, se efectúa mediante las moléculas de ciertas sustancias químicas transmisoras denominadas *neurotransmisoras* que fluyen a través de unos contactos que se denominan *sinapsis*. La mayoría de las sinapsis ocurren entre el terminal axónicos de una neurona (que lleva un impulso eléctrico) y las dendritas de la neurona siguiente (que recibe el estímulo). Ver figura 2.6. La sinapsis se realiza en un espacio de entre 50 a 200 Angstrom, denominado justamente *espacio sináptico*.

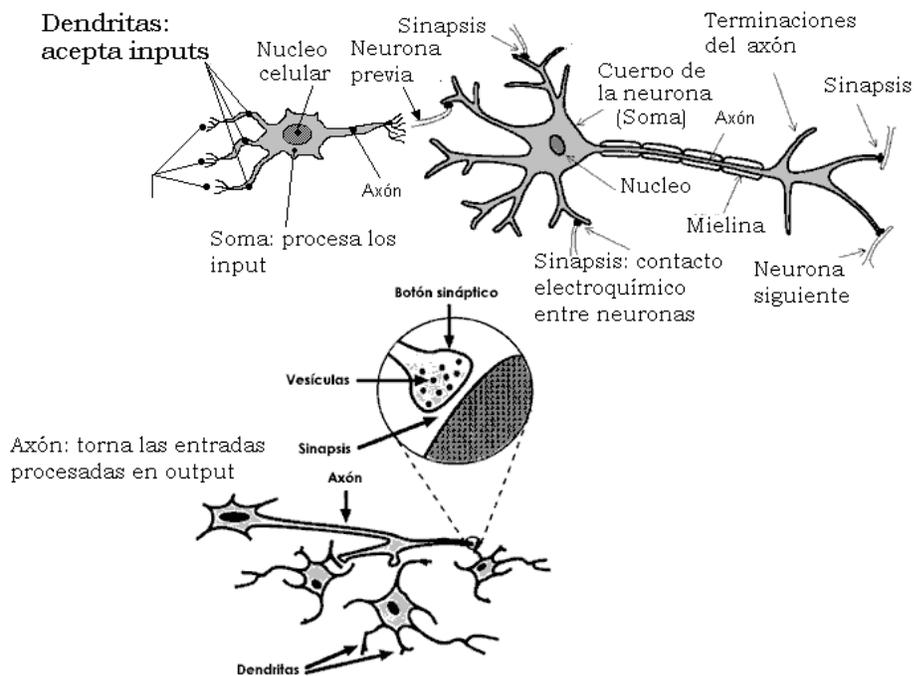


Figura 2.6: Esquema de una neurona biológica

El mecanismo “mediador” en la unión entre neuronas que transmiten el impulso de una neurona a la siguiente puede ser de naturaleza “eléctrica”

(sinapsis eléctrica) o “química” (sinapsis química). Esta última es mucho más abundante que la anterior, suponiendo la liberación del *neurotransmisor* desde el terminal axónico. Este neurotransmisor es vertido en el espacio de la separación sináptica.

La generación de estas señales eléctricas está relacionada con las propiedades de la membrana celular de la neurona. El medio interno, en la célula, es 10 veces más rico en potasio que el externo. Y el medio externo es 10 veces más rico en sodio que el interno. Es justamente esta diferencia de concentraciones entre el medio interno y el externo la que produce una diferencia de potencial de aproximadamente -60 mV a través de la membrana, y se denomina *potencial en reposo* de la célula nerviosa. La llegada de las señales procedentes desde otras neuronas hace aumentar el potencial levemente desde el valor en reposo. En otras palabras, se hace más positivo o menos negativo (despolarización). Este potencial modifica la permeabilidad de la membrana de la neurona receptora, de modo que cuando llega a cierto valor umbral comienza una entrada masiva de iones sodio invirtiendo la polaridad de la membrana. Esta inversión de voltaje de la cara interior de la membrana, a su vez, cierra el paso a los iones de sodio y le abre el paso a los iones de potasio hasta que se restablece el equilibrio en reposo.

Cuando la magnitud de la despolarización alcanza un valor aproximado de -40 mV, la neurona llega a un punto crítico que se denomina *valor umbral, nivel de disparo o potencial de acción* (inversión del voltaje) en el que la despolarización resultante se propaga lo largo del axón de la neurona receptora y provoca la emisión de los neurotransmisores en los terminales axónicos. A continuación viene un período refractario de alrededor 0,5 ms a 2 ms caracterizado por la inexcitabilidad de la neurona. Luego la neurona está lista para enviar otro impulso.

Existen dos tipos de sinapsis, la inhibitoria y la excitatoria. La primera dificulta la emisión de impulsos debido a que sus neurotransmisores tienden

a estabilizar el potencial de acción de la membrana. En cambio la segunda provoca la disminución del potencial en la membrana de la célula postsináptica, la que genera el impulso de la señal a mayor velocidad. Las neuronas reciben entradas de ambos tipos de sinapsis. En cada instante algunas de ellas van a estar activas y otras en reposo. Es justamente la suma de estos efectos excitatorios e inhibitorios la que determinará si la neurona será estimulada o no para emitir un impulso eléctrico.

Los impulsos nerviosos pueden alcanzar una velocidad de, aproximadamente 300 [Km/hr] por las distintas células del cerebro. Así, todo lo que percibimos sucedió algunas milésimas de segundos atrás, el tiempo que tardaron los impulsos en llegar al cerebro para que éste los interprete.

La capacidad de aprendizaje de nuestro cerebro así como su resistencia a las averías locales (robustez) se debe fundamentalmente al funcionamiento colectivo y simultaneo de las neuronas que lo componen (capacidad de procesamiento en paralelo) [76], estando organizadas éstas en redes altamente interconectadas. Por ejemplo, al efectuar un recuerdo nuestro cerebro realiza un gigantesco cálculo que si se compara con un ordenador, el cual haría la misma tarea de manera secuencial, le tomaría muchísimo más tiempo en realizar la misma tarea.

Pese a lo reducido de la velocidad de procesamiento de cada neurona, el *paralelismo* le permite capturar la imagen de un rostro al primer encuentro. Su capacidad de memoria es tan amplia que almacena las experiencias de una vida entera y es tan versátil que el recuerdo de una escena trae a colación asociaciones de imágenes visuales, sonidos, olores, sabores, sensaciones táctiles y emociones. La mayoría de las veces, los recuerdos se originan a partir de impresiones sensoriales.

La *Naturaleza* a través de la evolución del cerebro humano logró resolver el problema de cómo desarrollar un potentísimo computador tomando un

enorme número de unidades e interconectándolas entre sí más o menos al azar, en una gran extensión, siendo el sistema capaz de modificar las conexiones (propiedad de *adaptabilidad*), a través de un proceso de aprendizaje constituyendo así el cerebro humano un *Sistema Complejo*. Esta complejidad yace en las innumerables interconexiones y no en la variedad de los componentes. Se supone que esta variedad –tipos de células-, darían efectos de segundo orden comparados con los provenientes de la naturaleza colectiva.

Pero ¿cómo recordamos? Esto es posible gracias a las uniones o sinapsis mencionadas anteriormente. El cerebro puede cambiar de acuerdo a nuevos estímulos, los que ocasionan el crecimiento de nuevas sinapsis o mejora las ya existentes [76]. Luego, de esto se podría decir que la edad no sería un límite para el aprendizaje (en un individuo sano), ya sea de un oficio, de idiomas e incluso practicar deportes.

La palabra clave para esto es *Red*. Y es el comportamiento colectivo de las redes de neuronas la que permiten el pensamiento, no la actividad de una sola neurona.

2.3.2 Estructura y Funcionamiento de un Sistema Neuronal Artificial

Las *Redes Neuronales Artificiales* (RNA) son algoritmos para tareas cognitivas, tales como el aprendizaje y la optimización, las cuales están, en algún aspecto, basado en conceptos derivados de la investigación sobre la naturaleza del cerebro.

La investigación ha demostrado [77,78] cómo las redes que consisten de un gran número de elementos de proceso muy simples, las neuronas, pueden ser usados para efectuar tareas computacionales. En estas redes se unen cientos de unidades individuales de procesamiento a través de conexiones, imitando las propiedades del cerebro [79].

La información almacenada en una red neuronal no tiene una “dirección” de memoria como en los computadores usuales y no está localizada en un lugar preciso. La información está almacenada en “toda” la red, ¡y la noción misma de programa secuencial carece de sentido! [80].

Si esta información almacenada no se pierde al dañarse sólo unas pocas neuronas, entonces se dice que existe *Robustez y Conectivismo*. Distintos trozos de información o patrones se pueden almacenar en la misma red, los cuales se recuperan por *asociatividad*. En un computador convencional para recuperar la información se debe dar la dirección de la memoria. En el cerebro la información se halla bajo muchas direcciones sirviendo también un trozo parcial de ésta para recuperar toda la información. El patrón a ser recuperado actúa como un *atractor* – existiendo muchos en el cerebro –, hacia el cual convergen nuestros pensamientos.

Las neuronas son marcadamente más lentas en relación a los componentes de los computadores actuales. Aquellos operan en la escala temporal de 10^{-3} seg., mientras que los componentes lo hacen a 10^{-9} seg. Así, la neurona en sí misma es un pobre calculador. Sin embargo, debido a que es capaz de procesar información en paralelo –como se señaló anteriormente–, es un estimador muy rápido.

El modelo de una neurona en términos de pulsos eléctricos, que entran y salen, fue propuesto en 1943 por McCulloch y Pitts [81]. De allí se deduce que un modelo electrónico de una neurona se podría construir, no a partir de un microprocesador, sino de un chip de amplificador. Así, una red neuronal puede construirse por completo a partir de chips, no existiendo procesador central. Las propiedades *inteligentes* de una red se deben a su estructura e interconexiones y no son un producto de un programa escrito por el hombre.

Las neuronas artificiales como unidades individuales, son modeladas como sistema de dos estados: en reposo y activado, la transición ocurre

cuando el estímulo supera un cierto umbral. La información (o conocimiento) es guardada en la intensidad de las interacciones entre estas unidades.

Las siguientes características generales, en una RNA, pueden ser consideradas como básicas en su modelación [78,79]:

- Arquitectura de la red (disposición y número de nodos o “neuronas”)
- Tipos de conexión (simétrica o asimétrica)
- Tipo de dinámica o algoritmo de actualización de la red. Puede ser secuencial (asincrónico) o paralelo (sincrónico); estático o dinámico.
- Regla de aprendizaje (supervisado, no supervisado, etc.)
- Tipo de función de transferencia (lineal o no lineal)

Más detalladamente lo que se requiere es:

1. Disponer de un conjunto de N nodos o unidades de procesamiento (o bien PE de processing element)
2. Asociar a cada nodo una variable real de estado $S_i(t)$ (estado de activación)
3. Una regla de aprendizaje para establecer las conexiones entre los nodos, asociando un número real J_{ik} con la conexión entre el nodo i y el nodo k . Estos J_{ik} se conocen también como *sinapsis*, *Factor de peso*, o de *premio y castigo*. [82]
4. Con cada nodo i se asocia una constante real θ_i (llamada *umbral de activación* o *bias*)
5. Para cada nodo i se define una función de transferencia o activación $f_i[s_k, J_{ik}, \theta_i (k \neq i)]$, la cual determina su estado.

Convencionalmente los nodos se asocian con las neuronas, las conexiones con las *sinapsis* y las constantes θ_i con los *umbrales de activación*, como se muestra en la figura 2.7. Generalmente la función de transferencia es la misma para todos los nodos y tiene la forma $f(\sum_k J_{ik} \cdot s_k - \theta_i)$, siendo $f(x)$ la *función escalón* o una *función más suave*

de tipo sigmoidea⁹. Los *nodos de entrada* son aquellos que no reciben conexiones y los *nodos de salida* aquellos de los cuales no salen conexiones. En caso que en una red haya caminos cerrados (caminos que partiendo de una neurona y siguiendo las conexiones vuelvan a la misma neurona), se habla de redes con *retroalimentación* o *recurrentes (feedback)*. En caso contrario, se tendrá redes con *alimentación hacia adelante (feedforward)*.

Un caso especial de una red con retroalimentación es aquella en que cada una de las neuronas se conecta con todas las restantes. De acuerdo a la definición previa, aquí no se tiene nodos ni de entrada ni de salida. O bien todos ellos pueden ser considerados como de entrada y de salida. La función de transferencia asociará, por lo tanto, los estados de las neuronas de entrada consigo misma y se habla entonces de *redes autoasociativas*.

En redes alimentadas hacia adelante la salida es una función explícita de la entrada. Vale decir, la entrada se propaga a través de la red y produce la salida correcta. Contrariamente, las redes recurrentes siempre necesitan algún tipo de relajación para alcanzar un atractor.

En algunos problemas la elección de una intensidad de conexión apropiada para una red se puede calcular con anticipación. Pero esto no siempre es práctico puesto que a veces involucra una gran matriz de inversión o un gran problema de optimización.

Es mejor y más fácil adoptar una aproximación iterativa en la cual los pesos apropiados J_{ik} se hallan mediante el mejoramiento sucesivo desde un punto arbitrario de partida. Entonces, se dice que la red está *aprendiendo* una tarea.

⁹ Por ejemplo, la función $(1 + \exp(-x))^{-1}$ o “función de Fermi”

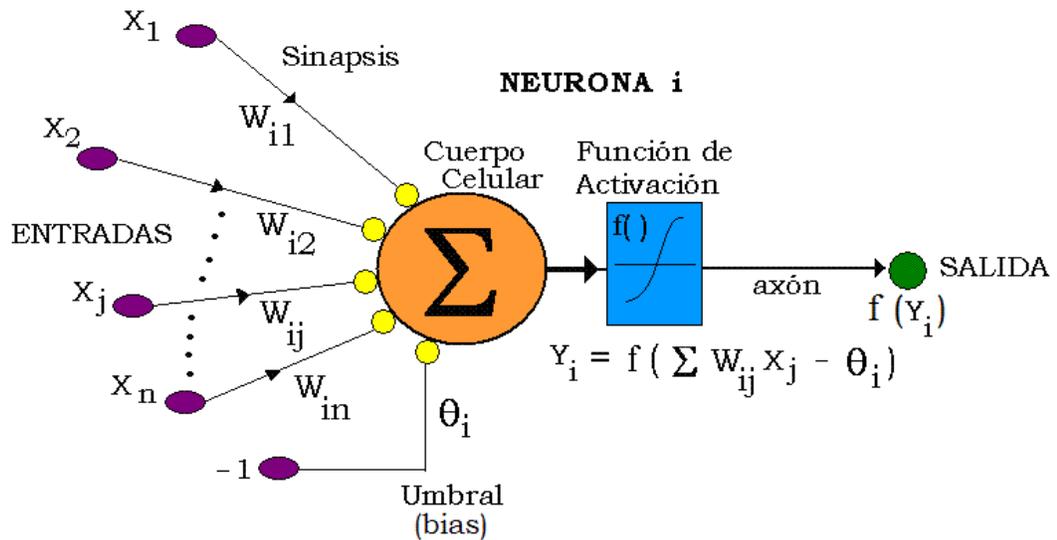


Figura 2.7: Esquema de una neurona artificial

Las redes neuronales se pueden clasificar como *estáticas* y *dinámicas*. Las primeras implementan transformaciones no lineales de la forma $u = G(x)$, con $x \in \mathbb{R}^n$ y $u \in [0, 1]^m$, con n y $m \in \mathbb{N}$, representando las dimensiones de x y u respectivamente. Ejemplos de estas redes son el perceptrón, las redes de multicapas, y las redes de base radial [83,84,85].

Las segundas son descritas mediante ecuaciones diferenciales o de diferencia (también llamados mapeos), y son de interés debido a que muchos sistemas que se modelan en el mundo real son sistemas dinámicos no lineales. Ejemplos de estas redes lo constituyen la red de Hopfield [86] y sus variantes, redes neuronales con retardo temporal, etc.

En cuanto a paradigmas de aprendizaje son ampliamente usados dos tipos: el *supervisado* y el *no supervisado*. En el primer caso, la salida de la red se compara con una respuesta deseada, y recibe retroalimentación sobre cualquier error (llamado a veces *aprendizaje con un profesor*: el profesor le dice a la red que la respuesta está correcta, o al menos –en aprendizaje con esfuerzo–, si están o no sus respuestas correctas).

En el segundo caso, no existe un profesor y no hay respuestas correctas ni erradas. Se le suministran sólo los datos de entrada, no recibiendo la red ninguna información que le indique cual es el output o salida. La red debe descubrir por sí misma categorías o características de interés en los datos de entrada. Para ello debe modificar los pesos para asociar a la misma unidad de salida los vectores de entrada más similares. Se suele decir que estas redes son capaces de *auto-organizarse*. En esta clase de aprendizaje, se conocen dos tipos de entrenamiento no supervisado: el competitivo/cooperativo y el hebbiano, siendo el primer caso orientado para detectar las similitudes de las entradas, y en el segundo caso se orienta más a la ordenación mediante clases (clustering).

Es de interés ensayar patrones de entrada que no estén dentro del conjunto de entrenamiento, para ver si la red puede generalizar exitosamente lo que ha aprendido; estos se denominan patrones de test.

2.3.3 Resumen de los Modelos de RNA más importantes

En las RNA se pueden dar una gran variedad de arquitecturas, cada una de ellas especificada por las interconexiones de nodos o PE's. También se pueden definir distintas operaciones ejecutadas por cada nodo y método de "entrenamiento" para optimizar el rendimiento de la red. De acuerdo a esto, podemos tener distintos modelos de redes [77,78,79].

Históricamente, la aparición de los modelos más conocidos fue como se explica a continuación.

Los matemáticos Warren McCulloch y Walter Pitts [81] en el año 1943 consideraron el caso en que cada neurona puede estar sólo en dos estados (1 ó 0 , o también -1 y $+1$), en el que el tiempo se mide en pasos discretos ($t = 1, 2, 3, \dots$) y la función de transferencia es la función escalón de tal manera que el estado de una neurona en el instante $t + 1$ está dada por:

$$S_i(t + 1) = \Theta\left(\sum_j J_{ij} \cdot S_j(t) - \theta_i\right) \quad [2.19]$$

Ellos mostraron que para una adecuada elección de las conexiones J_{ij} esta red es capaz de realizar cualquier computación imaginable, o sea que es una máquina de Turing¹⁰. Luego la neurona dispara si la suma de los pesos de las entradas $\sum_j J_{ij} \cdot S_j$ alcanza o excede el valor umbral θ_i , representado esquemáticamente en la figura 2.7.

En resumen, se modeló por primera vez en la historia a una neurona como un *elemento umbral lógico*. No obstante, ellos no especificaron cómo se podría formar las interconexiones necesarias entre las neuronas y cómo esto puede ocurrir a través del aprendizaje.

En el año 1949 el psicólogo Donal Hebb [82] sugirió que la conexión entre dos neuronas es *plástica* y cambia en proporción a la correlación de actividad entre la neurona presináptica y la postsináptica. Esta propiedad se usa en muchos otros modelos¹¹.

El *Perceptrón*, por su parte, fue propuesto en el año 1958 por Rosenblatt [83] como un prototipo de sistema auto-organizado, interactivo y dinámico. Este tipo de red es capaz de aprender, clasificar, generalizar y memorizar, reconociendo los patrones que se le presente a la red. El algoritmo de evolución de la red responde a la ecuación:

$$Y_r = \Theta\left(\sum_{i=1}^L J_{ri} \cdot x_i\right) \quad [2.20]$$

Para la clasificación de un patrón de entrada x , cada elemento r calcula un valor binario de salida Y_r , de acuerdo a la ecuación [2.20]. Acá los coeficientes J_{ri} , $i = 1, 2, \dots, r$, determinan el comportamiento del elemento r . El umbral θ_i se igualó a cero para simplificar. Así, durante la fase de entrenamiento (o aprendizaje) cada elemento ajusta sus coeficientes J_{ri} de tal forma que sólo reacciona a los patrones de entrada de su clase C_r con un

¹⁰ Relacionado con el test de Turing. Vea referencia [78,80]. Mas detalle se puede encontrar en Hopcroft, J.E., *Máquinas de Turing*, Investigación y Ciencia, 94 (1984).

¹¹ Conocida como regla de Hebb,: la conexión sináptica entre dos neuronas se mejora o refuerza cuando ellas están simultáneamente activas.

valor de salida $Y_r = 1$. Para que esto sea posible debe garantizarse una solución, es decir, deben existir pesos J_{ri}^* para el cual la ecuación [2.16] resuelve correctamente la clasificación del problema.

Existe un *Teorema de Convergencia* para el entrenamiento de un perceptrón y es el siguiente:

Dado un perceptrón simple (o elemental), un estímulo E y cualquier clasificador $C(E)$ para el cual existe una solución y suponiendo que estos estímulos son mostrados en una secuencia arbitraria y que cada uno de ellos se repite un número finito de veces, entonces si se comienza desde cualquier estado inicial, un procedimiento de corrección de error siempre garantizará la convergencia a una solución para $C(E)$ en un tiempo finito.

La función de transferencia para el perceptrón puede ser tanto una función lineal, en cuyo caso sería análogo a hacer una regresión lineal a un conjunto de datos, como una no lineal. En este último caso, se usa para el perceptrón la función escalón como se muestra en la figura 2.8.

La regla de aprendizaje del perceptrón es de tipo *supervisado*, lo que requiere que sus resultados sean evaluados y se efectúen las modificaciones necesarias si fuese necesario.

Sin embargo, fueron Minsky y Papert [87] quienes en el año 1969 advirtieron que el perceptrón era incapaz de resolver problemas que no se pudiesen separar linealmente, como el problema del OR excluyente o XOR.

Debido a esto hubo un lapso considerable de tiempo en que se estancó, aparentemente, el interés por la investigación en el área de las RNA.

En el año 1982 Hopfield [86], trabajando en ese entonces para IBM, diseñó una red en la que las neuronas simultáneamente activas refuerzan las sinapsis. Esta red permite recuperar información almacenada en ella a partir de un conocimiento parcial del estado inicial propiedad conocida como *memoria asociativa*. Este es un ejemplo de red recurrente.

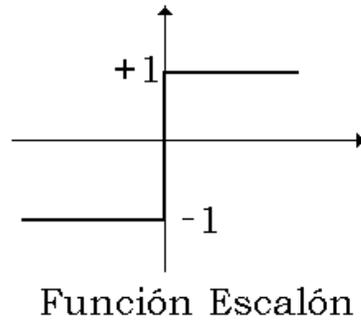


Figura 2.8: Representación de la función escalón o función signo.

Hopfield recurriendo a la *Física Estadística* modeló a las neuronas del sistema nervioso con la variable física llamada *espín*. Estos pueden “apuntar” hacia arriba o hacia abajo. Las neuronas se comunican entre sí “disparando” una señal (espín hacia arriba: + 1) o no (espín hacia abajo: - 1), a través de la sinapsis. Usando esta idea el modelo de Hopfield establece una correspondencia con una subclase de sistemas magnéticos (los *vidrios de espín*) [88]. La física ofrece una herramienta para estudiar estos modelos: la *Mecánica Estadística*. En el modelo de Hopfield, la forma como se enseña a una red a aprender es por medio de la llamada *regla de Hebb*, la cual, como se mencionó anteriormente, dice que la sinapsis entre neuronas son reforzadas cuando neuronas pre y post sinápticas están simultáneamente activas.

La dinámica, en este modelo, se actualiza de acuerdo a la relación:

$$S_i(t + \delta t) = \text{sgn}(h_i(t)) \quad [2.21]$$

en la cual el tiempo fluye en pasos discretos, siendo la neurona a actualizar escogida al azar entre todo el conjunto y de manera asincrónica.

En la ecuación [2.21], la función *sgn* es la función signo, que hace las veces de función de transferencia no-lineal, ver figura 2.8.

$$\text{sgn}(x) = \begin{cases} +1 & \text{si } x \geq 0 \\ -1 & \text{si } x < 0 \end{cases} \quad [2.22]$$

El término $h_i = \sum_j J_{ij} \cdot S_j$ ($j = 1, 2, \dots, N$) se asocia al *potencial de membrana* de la neurona i , con la simplificación adicional de que los umbrales se colocaron igual a cero.

Si escogemos p patrones para ser almacenados, la elección específica de las conexiones sinápticas se hace en base a un subconjunto de p de las 2^N posibles configuraciones de la red, en su conjunto. Representando estas configuraciones por medio de la variable ξ_i^μ , con $i = 1, 2, \dots, N$ y $\mu = 1, 2, \dots, p$ tiene:

$$J_j = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu \quad [2.23]$$

la cual se conoce como la *Regla de Hebb*.

Pero el modelo de *Hopfield* de memoria asociativa es menos que óptimo en muchos aspectos. Por ejemplo, es incapaz de almacenar patrones correlacionados. Otro problema que surge es el de la invariancia traslacional, siendo incapaz de reconocer la equivalencia de patrones que difieran sólo por un simple desplazamiento.

En el *modelo de Hopfield*, cada neurona está conectada a cada una de las otras neuronas. Por lo tanto, con respecto a sus conexiones, el modelo no tiene *estructura interna* y es *homogéneo*. No obstante, las redes neuronales biológicas están muchas veces diseñadas de manera estructuradas, como por ejemplo el *sistema visual*.

Justamente el modelo redes de multicapas, que se verá en la siguiente sección, está estructurado de esta manera. Este es un modelo que contrasta con el modelo de perceptrón y el de *Hopfield*, en cuanto a que posee una arquitectura alimentada hacia adelante y con un mínimo de tres capas;

contiene unidades *escondidas*, en una o más capas ocultas, que no están directamente conectadas a las señales de entrada o de salida, como se muestra en la figura 2.9. Luego, los estados de estas neuronas no son afectadas directamente por el *mundo externo*.

La regla de aprendizaje del perceptrón no puede aplicarse al caso de multicapas y surge el *problema de asignación de crédito*: si una tarea dada se efectuó de mal modo, no es claro cuál de los pesos son responsables para el mal resultado y cómo ellos han sido cambiados [89].

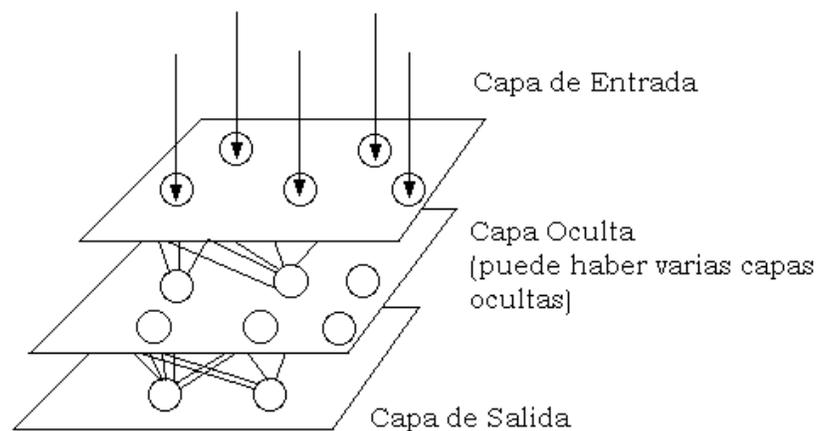


Figura 2.9: Esquema de una Red de Multicapas o Perceptrón de Multicapas.

El interés por las redes neuronales se ha estimulado con el descubrimiento de nuevos procedimientos de aprendizaje tal como el método de Kohonen [90] de formar *mapas topográficos* – o mapeos no lineales-. El interés surge debido a los estudios del centro motor en la corteza humana. Se sabe que las regiones neuronales que controlan el movimiento de varias partes del cuerpo humano se arreglan de lado a lado como sobre un mapa de la superficie del cuerpo humano [91]. Pero este mapa está extrañamente distorsionado; el tamaño de una región no corresponde al tamaño de aquella parte del cuerpo que la controla, sino más bien a la importancia y complejidad de sus actividades. Kohonen propuso un mecanismo por medio del cual una red neuronal puede organizarse a sí

misma de un modo tal que las neuronas se localizan en una región en el espacio de parámetros de estímulo externo, de acuerdo a su cantidad de actividad. Este modelo hace uso de un tipo distinto de aprendizaje, denominado *aprendizaje no supervisado*, a diferencia de los modelos anteriores.

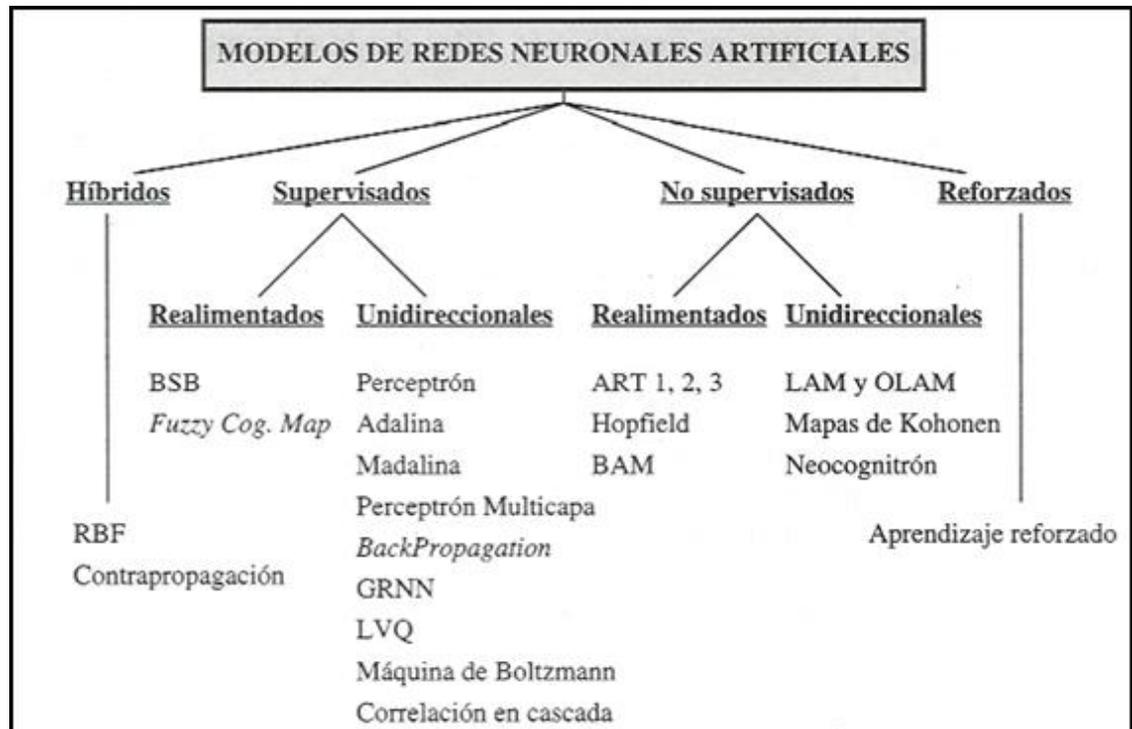


Figura 2.10: Clasificación de modelos de RNA más importantes.

Otros modelos que surgieron¹², son el modelo de Little, la máquina de Boltzmann, las redes de Hamming, la *Adeline* de Widrows, la *Memoria Asociativa Bidimensional* (o BAM) de Kosko [92], entre otros. En la figura 2.10 se muestra un resumen de diferentes modelos de RNA.

2.3.4 La Red Perceptrón de Multicapas (MLP)

Fue Werbos [95] quien obtuvo el algoritmo de entrenamiento capaz de obtener los valores óptimos de los pesos de estos sistemas neuronales.

¹² Por limitación de espacio, sólo se mencionan acá. Para más detalle ver [80,89,93].

Werbos y Parker [95] propusieron una arquitectura diferente para las redes neuronales de multicapas (red de perceptrón de multicapas o MLP de *multilayer perceptron*), y una regla de aprendizaje supervisado llamada la *regla delta generalizada*, que es una generalización de la regla Delta de aprendizaje. Esta regla busca variar los pesos de modo que se reduzca el error cometido por la red lo más rápidamente posible.

Con este algoritmo de aprendizaje, también denominado *algoritmo de propagación hacia atrás*, o *back-propagation*, Rumelhart y McClelland (1986) [84] lograron una buena aproximación para resolver este problema. La propagación hacia atrás provee un mecanismo por propagación de errores hacia atrás a través de la red, tal que las unidades escondidas son ajustadas de acuerdo a cómo muchas de ellas participaron en el nivel de error de la salida de la red.

Este algoritmo de entrenamiento demostró ser superior al de las redes no supervisadas o auto-organizadas en aprendizaje, en cuanto a relaciones funcionales complejas. Este método aunque efectivo, es lento, complejo y, biológicamente *no plausible*, pues las neuronas no pueden pasar información hacia atrás por sus axones. Es decir, el flujo de información en los sistemas biológicos reales, es asimétrico.

El trabajo seminal de Rumelhart Y Mc Clelland popularizó esta regla de aprendizaje o de entrenamiento y extendieron el tema sugiriendo múltiples aplicaciones. El objetivo del proceso de aprendizaje es ajustar los pesos de la red para que los patrones de entrada, introducidos en la red, permitan obtener las salidas deseadas. El aprendizaje, de tipo supervisado, supone la necesidad de disponer de pares de entrenamiento entrada-salida deseada. Este proceso se lleva a cabo mediante la minimización de una función error, que generalmente se corresponde con la suma de errores en las salidas al cuadrado.

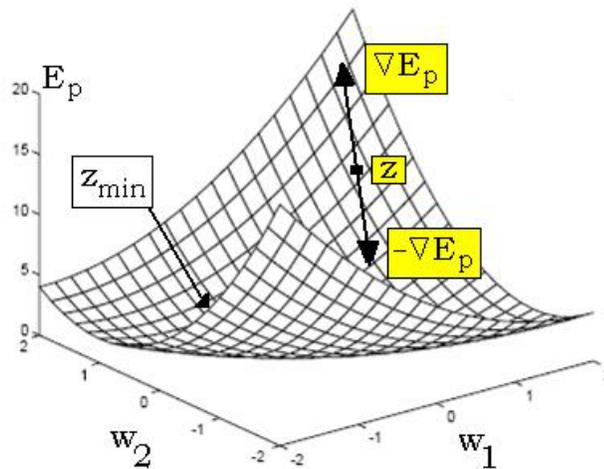


Figura 2.11: Superficie de error en el espacio de los pesos.

La superficie denominada de error la cual indica el error que se comete en la red neuronal para un determinado valor de los pesos, se representa en un espacio de $N + 1$ dimensiones en que N es el número de pesos de la red y el eje vertical (u ordenada) es el error. Ver figura 2.11. Cuando se asigna inicialmente unos valores determinados para los pesos de la red, el error corresponderá a un punto sobre esta superficie. A partir de este punto de la superficie de los pesos, se determina la dirección de la pendiente más pronunciada en dirección hacia abajo. Este método, denominado *del gradiente descendente* o *del descenso más pronunciado*, está basado en la búsqueda de la dirección (eje z en la figura 2.11) en la que una pequeña variación del vector de los pesos produce que decrezca el error más rápidamente.

En el perceptrón de multicapas (o *multilayer perceptron MLP*) los nodos se agrupan por capas conectándose todos los nodos de una capa con los de la siguiente capa y así sucesivamente (ver figura 2.9). Se trata de una red en la cual las señales se propagan únicamente hacia delante (feedforward).

La neurona i suma las entradas recibidas ponderadas por su peso asociado, de la siguiente manera:

$$net_i = W_{i1} \cdot I_1 + W_{i2} \cdot I_2 + \dots + W_{in} \cdot I_n \quad [2.24]$$

donde net_i es la suma lineal de los inputs (I_i) recibidos y n denota el número de nodos que estamos considerando. Esta suma resultante se hace pasar a continuación a través de una función de transferencia que, generalmente, es una función sigmoideal del tipo función de Fermi y bosquejada en la figura 2.12:

$$H_i = [1 + \exp(-net_i)]^{-1} \quad [2.25]$$

Esto se calcula para cada neurona de la capa oculta.

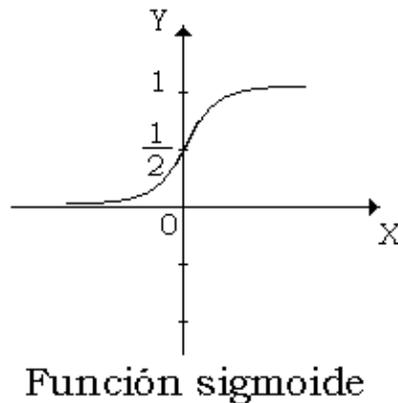


Figura 2.12: Esquemización de la función de Fermi.

Este resultado pasa a la capa de salida. Luego, la señal recibida por la neurona en la capa de salida k estará dada por:

$$net_k = W_{ki} \cdot H_i + \dots + W_{pi} \cdot H_p \quad [2.26]$$

donde p es el número de neuronas en la capa oculta. Este valor es de nuevo transformado en la neurona de la capa de salida cuyo resultado es el output definitivo O_j .

$$O_j = [1 + \exp(-net_i)]^{-1} \quad j = 1, \dots, n_s \quad [2.27]$$

siendo n_s el número de neuronas en la capa de salida.

2.3.5 Funcionamiento del algoritmo

El método para ajustar los pesos corresponde a la denominada regla delta generalizada [77,78,79,80,89]. Vale decir, los pesos se actualizan de forma proporcional a la delta o diferencia entre la salida deseada y la obtenida ($\delta = \text{sal. deseada} - \text{sal. real}$).

Dada una neurona (unidad j) y la salida que genera y_i , el cambio que se produce en el peso de la conexión que une la salida de dicha neurona con la unidad $N_i(W_{ji})$ para un patrón de aprendizaje p determinado está dado por:

$$\Delta W_{ji}(t + 1) = \alpha \cdot \delta_{pj} \cdot y_{pi} \quad [2.28]$$

siendo p el patrón de aprendizaje concreto, α la tasa de aprendizaje (que regula la velocidad de aprendizaje) y

$$\Delta W_{ji} = W_{ji}^{actual} - W_{ji}^{anterior} \quad [2.29]$$

En las redes MLP, en principio, no se puede conocer la salida deseada de las neuronas de las capas ocultas para determinar los pesos en función del error cometido, E_p . No obstante, al principio si se puede conocer la salida de las neuronas de la capa de salida. Luego, si se considera la unidad j de la salida, se tiene:

$$\delta_{pj} = (d_{pj} - y_{pj}) \cdot f'(net_j) \quad [2.30]$$

donde d_{pj} es la salida deseada j para el patrón p y net_j es la entrada neta que recibe la neurona j . Este término representa la modificación que hay que efectuar en la entrada que recibe la neurona j . Si la neurona no corresponde a la neurona de salida, el error que se produce estará en función del error que se cometa en las neuronas que reciban como entrada la salida de dicha

neurona; y es este procedimiento el que se denomina *algoritmo de propagación hacia atrás*.

Según esto, en el caso que j no sea neurona de salida, el error producido esta en función del error que se comete en las neuronas que reciben como entrada la salida de j :

$$\delta_{pj} = \left(\sum_k \delta_{pk} \cdot W_{kj} \right) \cdot f'(net_j) \quad [2.31]$$

siendo k el rango cubierto por todas aquellas neuronas a las que está conectada la salida N_j . Así, el error producido por una neurona oculta será la suma de los errores producido en las neuronas a las que está conectada la salida de ésta, multiplicada cada una de ella por el peso de la conexión.

El método de retropropagación del error, también denominado como del *gradiente descendente*, requiere un número importante de cálculos para lograr el ajuste de los pesos de la red.

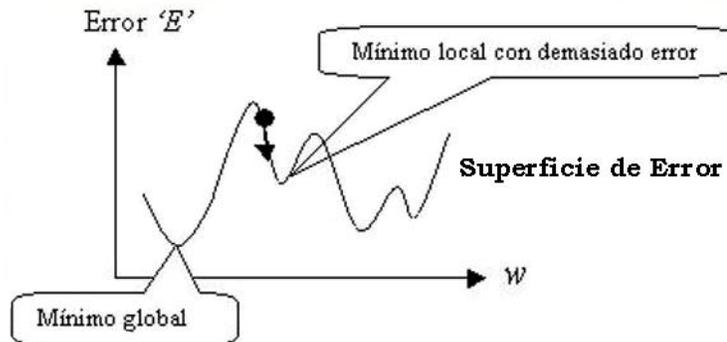


Figura 2.13: Superficie de Error

En la implementación del algoritmo se toma una amplitud de paso que viene dada por la *tasa de aprendizaje* α . Pero a mayor tasa será mayor la modificación de los pesos en cada iteración, con lo que el aprendizaje será más rápido, no obstante puede dar origen a oscilaciones. Rumelhart, Hinton y Williams [84] sugirieron añadir un término de momento, β , para filtrar estas

oscilaciones, en la expresión del incremento de los pesos. La expresión queda:

$$W_{ji}(t + 1) = W_{ji}(t) + \alpha \cdot \delta_{pj} \cdot y_{pi} + \beta \cdot (W_{ji}(t) - W_{ji}(t - 1))$$

O bien:

$$\Delta W_{ji}(t + 1) = \alpha \cdot \delta_{pj} \cdot y_{pi} + \beta \cdot \Delta W_{ji}(t) \quad [2.32]$$

Con este término de momento se logra la convergencia de la red en menor número de iteraciones, pues si en t el incremento de un peso era positivo y en $t+1$ también, entonces el descenso por la superficie de error en $t+1$ es mayor. Pero si en t el incremento era positivo y en $t+1$ negativo, el paso que se da en $t+1$ es más pequeño, lo cual es adecuado pues significa que se ha pasado por un mínimo y que los pasos deben ser menores para poder alcanzarlo.

El algoritmo *backpropagation* busca minimizar el error del sistema. Para ello se moverá por la superficie de la función de error dando pasos descendentes.

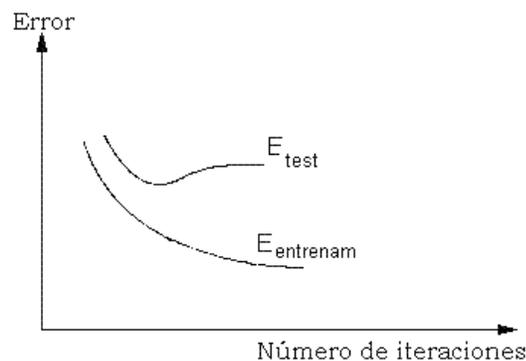


Figura 2.14: Tasa de aprendizaje.

Debido a este método de aprendizaje, el sistema podría eventualmente quedar atrapado en un mínimo local y alcanzar una solución que no fuese óptima, es decir con un elevado error como muestra la figura 2.13.

Existen diversos criterios heurísticos diseñados a fin de evitar este problema:

- a) Modificar el número de nodos en la capa oculta.
- b) Comenzar de nuevo el entrenamiento con un conjunto distinto de pesos.
- c) Agregar capas ocultas.
- d) Cambiar los parámetros de aprendizaje.

El fenómeno de “sobre-entrenamiento” consiste en que después de un cierto número de ciclos, el error de test comienza a crecer, mientras que el error por entrenamiento continúa decreciendo. Luego, el proceso de entrenamiento debería detenerse en este punto.

En la figura 2.14 se muestra el error en el aprendizaje, tanto para el caso de entrenamiento como el de test.

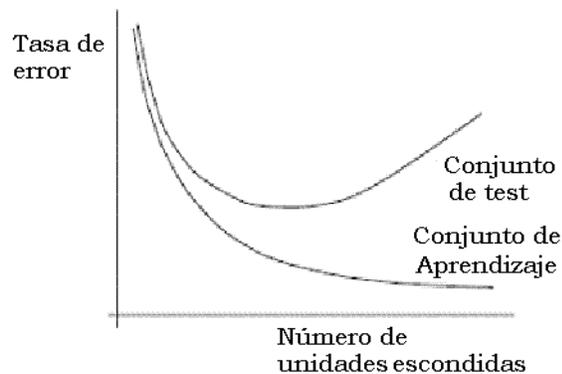


Figura 2.15: Tasa de error versus número de PE's escondidos.

Finalmente, la figura 2.15 muestra la tasa promedio del error por aprendizaje y la tasa promedio de error para test, como una función del número de unidades escondidas.

En este caso se observa que llega un momento en que la red no es capaz de generalizar, sino más bien sigue memorizando.

2.4 Introducción al Algoritmo de Cluster Híbrido

2.4.1 Fundamentos

La técnica de separar en clases (*clustering*) puede considerarse como el problema de *aprendizaje no supervisado* más importante, que trata de hallar una *estructura* en una colección de datos *sin rotular*.

Esta técnica consiste en particionar un conjunto de datos en colecciones de objetos de tal modo que dentro de cada partición los objetos sean lo más similares entre sí, y a su vez se diferencien de los objetos contenidos en otras particiones, es decir, sean disímiles a los otros objetos pertenecientes a otras clases. En la extensa literatura actual existen muchos algoritmos propuestos para efectuar el proceso de separar en clases [98].

Se define una clase como una colección de objetos que son similares entre sí y son disimilares a los objetos pertenecientes a otras clases.

El criterio de separación por clases puede entenderse de acuerdo a los dos criterios más comúnmente empleados: el criterio de la distancia y el conceptual. En el primer caso, dos o más objetos pertenecerán a una misma clase si ellos están “próximos” de acuerdo a una distancia dada (en este caso, distancia geométrica). Este tipo de clasificación se denomina *clustering basado en la distancia*.

En el segundo criterio de clasificación dos o más objetos pertenecerán a una misma clase si esta define un concepto *común* a todos aquellos objetos. En otras palabras, los objetos se agrupan de acuerdo a su ajuste con los conceptos que describen a dicha clase.

Se puede afirmar [98] que la única información que requieren los algoritmos de agrupación es la definición previa del vector de características. Algunos de estos algoritmos necesitan conocer, a lo sumo, también el número de clases.

Una vez que queda establecido este vector de características, los procedimientos de agrupación de clases reciben como datos de entrada a los objetos, los cuales han sido convertidos en vectores numéricos, a clasificar, de tal modo que a partir de estos datos de entrada el algoritmo sin supervisión y de manera autónoma agrupe esos vectores en clases (cluster), siendo esta la razón también porqué se les ha denominado *algoritmos de clasificación autoorganizada*.

Estas técnicas de cluster se usan cuando no hay conocimiento suficiente acerca de las clases en que se pueden distribuir los objetos de interés.

El término “cluster” es un concepto intuitivo y no tiene una definición matemática rigurosa. Los miembros de una clase debieran ser similares uno a otro y disímiles a los miembros de otra clase. Este algoritmo de clustering opera sobre un conjunto de datos sin rotular, produciendo en conjunto una partición sobre él, denotada por K.

Los algoritmos de clustering varían entre sí por el mayor o menor grado de reglas heurísticas que usan y, recíprocamente, por el nivel de procedimientos formales involucrados. Todos ellos se basan en el empleo sistemático de las distancias entre los vectores (los objetos a clasificar o agrupar) así como entre los cluster o grupos que se van haciendo y deshaciendo a lo largo del proceso de actuación del algoritmo concreto.

La distancia más comúnmente empleada es la distancia *Euclidiana* entre vectores, la cual se define como:

$$d_E(x_i, x_j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad [2.33]$$

en que E denota la clase y los subíndices i y j se refieren al vector, con $i \neq j$.

2.4.2 El algoritmo de cluster K-Means

El algoritmo *K-means* [97,98] es un algoritmo para clasificar o agrupar los objetos basados sobre los atributos o características en K número de grupos en que K debe ser un número entero positivo. El agrupamiento se hace minimizando la suma de los cuadrados de la distancia entre los datos y el centroide correspondiente del cluster. Así, el propósito de K-means es clasificar los datos en K grupos o clases, siendo necesario conocer *a priori* el número de clases existentes.

Es un algoritmo sencillo pero muy eficiente, siempre que el número de clases se conozca previamente con exactitud. El modo como opera este algoritmo, es como sigue.

Partiendo de un conjunto de datos a clasificar, este algoritmo realiza las siguientes operaciones:

1.- Al establecerse previamente K , se escoge al azar entre los elementos a agrupar K vectores, de forma que estos vectores constituirán los centroides de las K -clases. En otras palabras:

$$\alpha_1 : \mathbf{Z}_1(1); \quad \alpha_2 : \mathbf{Z}_2(1); \quad \dots \alpha_K : \mathbf{Z}_K(1) \quad [2.34]$$

2.- Como se trata de un proceso recursivo con un contador n , en la iteración genérica n se distribuyen todas las muestras $\{\mathbf{X}\}_{1 \leq j \leq p}$ entre las K clases, de acuerdo a la regla

$$\mathbf{X} \in \alpha_j(n) \quad \text{si} \quad \|\mathbf{X} - \mathbf{Z}_j(n)\| < \|\mathbf{X} - \mathbf{Z}_i(n)\| \quad \forall i = 1, 2, \dots, K/i \neq j$$

donde se ha indexado las clases y sus correspondientes centroides y la norma es usualmente la Euclidiana.

3.- Una vez redistribuidos los elementos a agrupar entre las distintas clases, se debe recalcular o actualizar los centroides de las clases. El cálculo de los nuevos centroides tiene por objetivo minimizar la cantidad:

$$\sum_{\mathbf{X} \in \alpha_i(n)} \|\mathbf{X} - \mathbf{Z}_i(n)\|^2; \quad i = 1, 2, \dots, K$$

Esta cantidad se minimiza usando la media aritmética de $\alpha_i(n)$:

$$\mathbf{Z}_i(n+1) = \frac{1}{N_i(n)} \sum_{\mathbf{X} \in \alpha_i(n)} \mathbf{X}; \quad i = 1, 2, \dots, K \quad [2.35]$$

siendo $N_i(n)$ el número de elementos de la clase α_i en la iteración n . En otras palabras, clasifica las n muestras de acuerdo a los centroides más cercanos, usando la distancia mínima euclidiana.

4.- El algoritmo termina si no se aprecien cambios en los centroides, vale decir, si ha alcanzado la posición estable siguiente:

$$\mathbf{Z}_i(n+1) = \mathbf{Z}_i(n) \quad [2.36]$$

De ser así, se termina el algoritmo. En caso contrario, se vuelve al paso 2.

Finalmente, si se verifica la ecuación [2.36] entonces el algoritmo calcula los centroides finales y termina.

2.4.3 Descripción del modelo no paramétrico de cluster HCA

Esta propiedad es usada por varios algoritmos de pronóstico híbrido que, adicionalmente, emplean una técnica de función de aproximación para modelar las interacciones entre cada uno de los cluster.

El algoritmo que propusieron Sfetsos y Siriopoulos [30,99], que ellos denominaron *HCA* por *Híbrido Clustering Algorithm*, es un proceso dual e iterativo que genera sincrónicamente grupo de datos o clases y estima una relación lineal entre los datos que contienen cada una de las clases. Los datos son, entonces, asignados a cada clase basado en su distancia desde

el hiperplano que está definido desde las ecuaciones que gobiernan a cada una de las clases, cuya propiedad en común es que estos grupos se describen mejor debido a sus similares características o relaciones lineales.

Este modelo ha sido aplicado al pronóstico de series temporales de material particulado PM₁₀ [99], correspondiente a la ciudad de Helsinki. Se compara la capacidad de hacer pronóstico de esta metodología, con un modelo de regresión lineal y una red de perceptrón de multicapas. Se reportó una mejora de ese modelo en comparación con estos dos últimos.

Como se mencionara en la sección 2.1, el análisis de las s. de t. se usa para describir las características dominantes de un conjunto de datos organizados secuencialmente. Es frecuente que el análisis de estas s. de t. resulte desde una descripción del proceso a través de un número de ecuaciones que, en principio, combina el valor actual de la serie, y_t con las observaciones pasadas, y_{t-k} , modelando los errores, e_{t-m} y las variables exógenas, x_{t-j} . Luego, se puede expresar la forma generalizada de este tipo de proceso como:

$$y_t = f(y_{t-k}, x_{t-j}, e_{t-m} \text{ /para varios } k, j, m) \quad [2.37]$$

El algoritmo que propusiera Sfetsos [30], es un procedimiento iterativo que agrupa los datos basado en su distancia desde el hiperplano que mejor describe su relación.

Se implementa a través de una serie de pasos que se describen a continuación:

(Paso 1) Usando los coeficientes de correlación de las series en estudio, se estima la variable rezagada más importante:
 $d = [j_1, \dots, j_d]$.

(Paso 2) Se forma un vector de datos:

$$P_t = [y_t, y_{t...j1}, \dots, y_{t...jd}, X_{t-k}] \quad [2.38]$$

(Paso 3) Se selecciona el número de clases n_{cl} . Este número de clases inicial es un parámetro que lo fija el usuario. El único requisito es que cada clase contenga la suficiente cantidad de datos para que se implemente un modelo lineal simple.

(Paso 4) Se inicializa la clase cl . En lugar de colocar datos en las clases de manera aleatoria, se usa un algoritmo convencional de clustering (K-Means), ya que puede entregar un primer indicio sobre las características de las series.

(Paso 5) En cada clase, aplicar un modelo de regresión lineal a y_t usando como variables al conjunto d identificado en el Paso 1.

(Paso 6) Asignar los datos a una clase basada en su distancia (usualmente Euclidiana) desde el hiperplano $(d + 1) - dimensional$ resultante.

(Paso 7) Vaya al Paso 5 a menos que alguno de los procedimientos de término del algoritmo fuera alcanzado.

Se consideran los siguientes procedimientos de finalización:

- (a) Se alcanza el número máximo de iteraciones predefinido.
- (b) El proceso se detiene cuando todos los datos son asignados a la misma clase en el Paso 6, como en la iteración previa.

El número óptimo de clases está relacionado a la aplicación deseada del algoritmo HCA.

Se puede afirmar que el algoritmo HCA es un esquema completo de análisis de s. de t., que proporciona una salida dual, en el sentido que genera un grupo de datos cuya característica inicial es que ellos

“pertenecen” al mismo hiperplano y simultáneamente estima un modelo lineal que describe la relación existente entre las variables de la clase.

Luego, se obtiene un conjunto de n_{cl} ecuaciones lineales del tipo:

$$\hat{y}_{t,i} = a_{0,i} + \sum_{\text{para todo } d} a_{j,i} y_{t-d} \quad , i = 1, \dots, n_{cl} \quad [2.39]$$

Además, suministra el valor pronosticado para cada clase, $\hat{y}_{t,i}$ mientras se hallan las características de entrada y los coeficientes de regresión lineal.

2.5 Modelos numéricos deterministas MM5, CMAQ, Chimere y WRF

Un modelo mesoescalar es un modelo de predicción numérica de las condiciones meteorológicas que se encarga de estudiar a los fenómenos cuyas dimensiones espaciales van desde los kilómetros hasta miles de kilómetros; poseyendo ciclos de vida que abarcan desde los minutos hasta varios días.

La resolución horizontal y vertical, en estos modelos, es suficiente para pronosticar fenómenos meteorológicos de mesoescala, los cuales son a menudo producto del forzamiento de la topografía o de los litorales o están relacionados con la convección. Como ejemplo, tenemos fenómenos climáticos severos, incluidos los tornados y los sistemas, las brisas de mar, etc.

La mesoescala, en *Meteorología*, es el estudio de sistemas del tiempo atmosférico más pequeños que la escala sinóptica meteorológica, pero más grandes que la microescala y la escala de tormenta de los sistemas de nubes cúmulos.

Por su parte, la escala sinóptica en meteorología (también citada como escala grande o escala ciclónica) es una escala de longitud horizontal del orden de los 1.000 km o más. Esto corresponde con una escala horizontal típica de depresiones de media latitud. Muchas áreas de presiones altas y

bajas se leen en cartas meteorológicas como sistemas de escala sinóptica. La palabra sinóptico deriva de la palabra en idioma griego *sunoptikos*, significando *vistos juntos*.

Se denomina una alta resolución cuando en una misma área la distancia entre los puntos es pequeña por ejemplo para el caso del MM5 de 1 Km. En cambio una baja resolución se refiere a grandes distancias entre puntos por ejemplo 1 grado. Es el mismo principio que se aplica para la *resolución de las fotografías*. En el caso del modelo MM5 se trabaja con resoluciones de 1° (111 km), 50 km, 20 km, 10 km, 5 km, 1 Km. Al utilizar bajas resoluciones es posible simular sistemas a escala sinóptica y con las altas resoluciones se simulan sistemas mesoescalares.

Es evidente que a medida que se aumenta la resolución horizontal de un modelo se mejora la representación que de la orografía de la superficie de la Tierra tiene el modelo. Como el efecto de la orografía es muy importante en la atmósfera, los resultados de los modelos numéricos mejoran notablemente a medida que se pasa de resoluciones de 100 km a 50, o a 20 e incluso a 10 km.

Sin embargo, por debajo de los 10 km aproximadamente el simple hecho de aumentar la resolución horizontal y vertical ya no tiene por qué mejorar los resultados. La hipótesis hidrostática incluida en los modelos de ecuaciones primitivas empieza a perder su validez a esas resoluciones.

A este tipo de modelos usados en resoluciones de más de 5 km y que ya no incluyen la hipótesis hidrostática en su formulación, se les suele llamar modelos de mesoescala.

En el proceso de discretización debe elegirse una rejilla espacial en la que se representan los valores de las variables del modelo y en ella misma se resuelven las ecuaciones obteniéndose así valores de las variables

meteorológicas para un instante futuro en todos los nodos de la rejilla de integración.

Como ya hemos dicho antes el sistema de EDP sobre el que se formulan los modelos numéricos no es más que una simplificación de las ecuaciones que rigen la dinámica de fluidos particularizadas para el fluido atmosférico. La Dinámica de la Atmósfera consiste en el estudio de los movimientos que tienen lugar en el seno de la atmósfera terrestre y de los procesos de intercambio de energía que tienen lugar en ella. En esta formulación se desprecian las fuerzas moleculares y se tienen en cuenta las fuerzas del gradiente de presión, de Coriolis, la gravitatoria y la de rozamiento.

Las leyes físicas del movimiento y conservación de la energía rigen el desarrollo de las condiciones atmosféricas. Dichas leyes pueden expresarse por medio de una serie de ecuaciones matemáticas complejas que componen el núcleo de lo que se denomina *predicción numérica del tiempo*. Debido a que se utilizan para predecir lo que sucederá en el futuro, estas ecuaciones se conocen como *ecuaciones de pronóstico* (o de prognosis). Las variables de las ecuaciones representan distintos aspectos del tiempo (por ejemplo, viento, presión, etc.). Como estas ecuaciones determinan la manera en que las variables cambian con el tiempo, si conocemos el estado inicial de la atmósfera podemos resolver las ecuaciones para un momento en el futuro y obtener valores nuevos para esas variables. Esencialmente, esto explica de forma sencilla cómo los modelos numéricos generan sus pronósticos.

Los primeros modelos meteorológicos operativos sólo podían resolver características de escala sinóptica. Conforme han aumentado las capacidades de los recursos computacionales, se ha vuelto factible modelar y predecir fenómenos meteorológicos de mesoescala dentro de plazos razonables.

Estos modelos se ejecutan a distintas *resoluciones horizontales*, típicamente con un espaciado de malla inferior a los 30 Km. Debido a su resolución relativamente fina, estos modelos suelen ejecutarse para un área limitada y, por tanto, requieren información acerca de los límites de sus dominios. Si fuéramos a ejecutar estos modelos a escala global, un solo ciclo de ejecución tardaría tanto tiempo que los resultados dejarían de ser útiles para generar pronósticos a corto plazo.

En comparación con los modelos tradicionales de escala más grande, los modelos de mesoescala a menudo producen mejores pronósticos en regiones costeras y montañosas. Esto se debe a que aprovechan conjuntos de datos topográficos de alta resolución e información detallada de temperatura de la superficie del mar, cuando están disponibles.

Existe una relación directa entre la resolución horizontal de un modelo de predicción numérica del tiempo y el tamaño de las características meteorológicas que puede simular. Cuanto mayor la resolución, tanto más pequeñas las características que el modelo puede representar adecuadamente. Esta resolución está relacionada con la distancia entre los puntos de malla.

¿Qué espaciado de malla es necesario para resolver una característica meteorológica en particular? Normalmente, se requiere al menos cinco puntos de malla para definir una característica. La característica meteorológica más pequeña que se puede representar, incluso en el ámbito de un pronóstico a corto plazo, abarca entre cinco y siete puntos de malla. ¡Esto significa que un modelo que emplea una malla de 20 km de densidad es incapaz de pronosticar una característica de tamaño inferior a los 100 km! En terreno accidentado o montañoso, es posible que el modelo no logre identificar una característica forzada por la topografía incluso con una densidad de malla de 5 km. La resolución limita lo que los modelos de predicción numérica del tiempo pueden predecir.

Al aumentar la resolución, los recursos informáticos necesarios también aumentan, porque el modelo debe calcular los valores para un mayor número de puntos de malla. Como ilustra la figura 2.16, si dividimos por tres la distancia entre los puntos de la malla del modelo, la cantidad de puntos de la malla en la misma área aumenta a razón de nueve. Es más, a medida que reducimos el espaciado de malla normalmente reducimos el tiempo que transcurre entre los pasos intermedios del pronóstico, con el resultado de que se necesitan pasos intermedios adicionales para producir un pronóstico de la misma duración. No obstante, por varios motivos merece la pena satisfacer los mayores requerimientos de cómputo que exige una resolución más alta, entre otros porque se mejora la capacidad del modelo de representar el terreno. A su vez, esto mejora la exactitud de predicción de los fenómenos meteorológicos provocados o intensificados por la topografía por parte del modelo.

Igual que se necesita una resolución horizontal adecuada para representar distintos fenómenos atmosféricos, los modelos de predicción numérica del tiempo también se deben diseñar con la *resolución vertical* adecuada para que puedan pronosticar la estructura vertical y los efectos de una variedad de eventos meteorológicos. Es interesante observar que la razón entre la resolución horizontal y la resolución vertical debe ser coherente con la inclinación de los fenómenos meteorológicos de interés.

Para determinar la resolución *horizontal* óptima de un modelo se necesita suficiente detalle para asegurar que los eventos meteorológicos similares se pronosticarán con resultados equivalentes en casi cualquier lugar dentro del dominio de ejecución del modelo. Esto significa que se utilizará un espaciado de malla uniforme. Sin embargo, para determinar la *resolución vertical* óptima del modelo normalmente aprovechamos el hecho de que ciertos procesos atmosféricos suelen estar confinados en determinadas regiones verticales específicas de la atmósfera y, por tanto, tratamos de asignar una resolución vertical más alta a las zonas donde es más necesaria. Por

ejemplo, la resolución vertical debe ser muy fina (del orden de unos pocos hectopascales -hPa) cerca de la superficie de la Tierra. De esta forma el modelo puede capturar la transferencia de calor y humedad en la capa límite planetario producido por el calentamiento diurno de la superficie. Por otra parte, no se necesita este grado de resolución en la troposfera media (alrededor de 600 a 300 hPa), aunque sí es necesaria una mayor resolución cerca y debajo de la tropopausa, para poder predecir con exactitud la corriente en chorro.

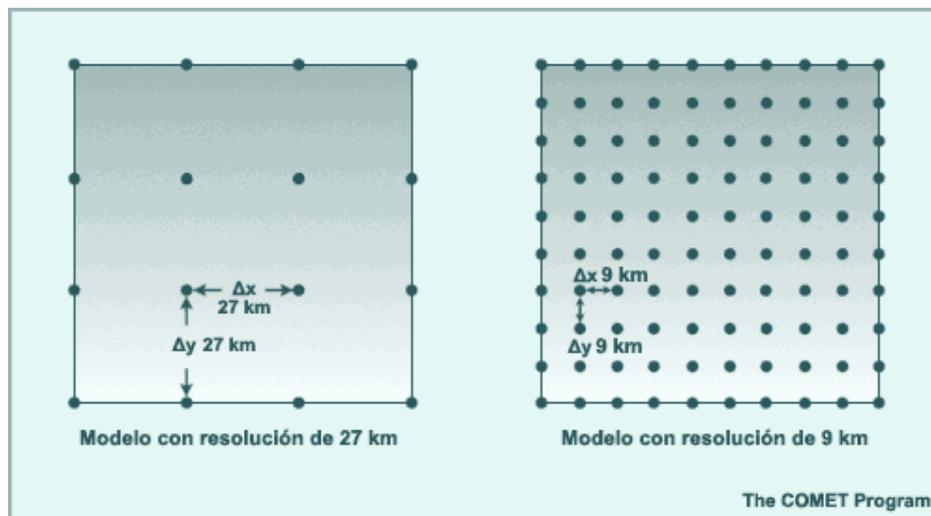


Figura 2.16: Ejemplo de mallas en modelos de mesoescala (ref. *The Comet Program*)

El uso de coordenadas sigma en lugar de la presión o la altura permite evitar el problema que surge cuando una superficie de presión o altura constante y el suelo se intersecan. En este ejemplo, una superficie de 920 hPa (y cualquier otra superficie debajo del nivel de 920 hPa) se intersecaría con una montaña, mientras que un nivel sigma de 0,92 sigue la forma del terreno. El relieve disminuye en los niveles sigma sucesivamente más altos.

Debido a que los niveles inferiores de un sistema de coordenadas sigma siguen la topografía, es fácil aumentar la resolución vertical cerca de la

superficie con sólo incrementar el número de niveles en esa zona. Con una mayor resolución en la atmósfera inferior, los modelos pueden definir mejor los procesos y características de la capa límite que contribuyen considerablemente a los aspectos del tiempo que se percibe, como los vientos, la turbulencia, la temperatura y la estabilidad en los niveles inferiores. Sin embargo, las coordenadas sigma tienen algunas desventajas. Por ejemplo, a veces pueden conducir a pronósticos de ondas excesivamente intensas y exageradas en sentido vertical a sotavento de las cordilleras, y pueden también no capturar adecuadamente las barreras topográficas cuando existe una inversión térmica debajo de la cima de una montaña.

La mayoría de los modelos de malla y todos los modelos espectrales de los conjuntos de herramientas de pronóstico numérico del tiempo operativo actuales son hidrostáticos. En contraste, muchos modelos de mesoescala, como MM5 y WRF, entre otros, no son hidrostáticos. Los modelos hidrostáticos presuponen la existencia de un equilibrio hidrostático, es decir, que el peso de la atmósfera, que ejerce presión hacia abajo, está en equilibrio con la fuerza del gradiente de presión, que empuja hacia arriba. Esta suposición acerca del equilibrio hidrostático es válida para los sistemas de escala sinóptica y global, y para *algunos* fenómenos de mesoescala.

Los procesos no hidrostáticos y sus efectos adquieren importancia cuando la longitud de una característica es aproximadamente igual a su altura. Como la altura de la mayoría de los fenómenos atmosféricos está limitada por la altura de la troposfera, esto representa un problema para las características de 10 km o menos de tamaño. Entre los posibles ejemplos de características meteorológicas importantes con procesos no hidrostáticos considerables cabe mencionar las tormentas convectivas, los frentes de racha y otras líneas de convergencia

Se requiere un modelo meteorológico de mesoescala para poder generar pronósticos de los parámetros meteorológicos tales como temperatura,

viento y humedad en una distribución espacial y temporal suficientemente densa para así determinar el transporte de partículas tal como el polen.

El modelo MM5 es un modelo numérico no hidrostático, limitado en el área (condición de frontera lateral) y de quinta generación, que se ha estado usando bastante para simular o predecir la circulación atmosférica a mesoescala. Fue desarrollado por la National Center for Atmospheric Research en conjunto con la University Corporation for Atmospheric Research de EEUU [100]. En esencia, el MM5 como modelo físico de predicción, resuelve en una malla las ecuaciones dinámicas básicas (conservación del momentum, conservación de la energía, conservación de la masa –ecuación de continuidad- y ecuación para los gases ideales) de pronóstico para las variables meteorológicas principales: altura geopotencial, 3 componentes de la velocidad del viento, temperatura, humedad relativa y presión.

El MM5 está compuesto de diversos programas escritos en Fortran [101]. Consta de cinco programas principales: Terrain el cual interpola los datos de elevación del terreno y vegetación (latitud y longitud) a la grilla de mesoescala; Regrid el cual lee los análisis y pronósticos meteorológicos y los interpola a la grilla horizontal del MM5; Rawins/little_R, que mejora el análisis de la etapa anterior, incorporando datos adicionales observados; InterPF que maneja la transformación de los datos requeridos del programa de análisis al modelo de mesoescala; y el MM5 el cual contiene la componente física, dinámica y numérica del sistema de modelaje [102].

Por su parte, el sistema de modelamiento fotoquímico numérico (modelo de transporte y transformaciones químicas) de calidad del aire (Community Multiscale Air Quality modeling system, CMAQ) de la US EPA [103], ha sido diseñado para aproximar la calidad del aire y modelar el comportamiento de diversos contaminantes, incluyendo ozono troposférico, partículas finas, tóxicos, deposición ácida, así como la degradación de la visibilidad. CMAQ se diseñó para que tuviese capacidades de multiescala de tal manera que

separe modelos que no son necesarios al modelamiento a escala regional y urbana (Byun and Ching). Además, este modelo fotoquímico puede acoplarse a uno meteorológico [104]. Así, el sistema MM5-CMAQ es un representante de los modelos de Tercera Generación de Sistemas de Modelización de la Calidad del Aire. Fue desarrollado en *Fortran 90* y al ser un sistema altamente modular, con una gran versatilidad y variedad de parametrizaciones el usuario puede incorporar nuevos procedimientos químicos o herramientas de solución numérica.

En general, el desarrollo de modelos para la calidad del aire requiere de un núcleo meteorológico acoplado con un modelo de transporte químico. En este sentido, el modelo de calidad del aire denominado CHIMERE (del Institut Pierre-Simon Laplace, INERIS, LISA, C.N.R.S.)[105], utiliza el inventario de emisiones EMEP (*European Monitoring and Evaluation Programm*¹³). Éste es un modelo fotoquímico de multiescala diseñado principalmente para proporcionar predicciones diarias de ozono, aerosoles y otros contaminantes y para realizar simulaciones a largo plazo para diversos escenarios de emisiones. Se puede acoplar con algún modelo meteorológico como el MM5 para hacer pronósticos de la calidad del aire de alguna región de interés. Por ejemplo, el modelo de pronóstico AQForecast acopla ambos modelos para realizar sus pronósticos en la península ibérica [105].

Finalmente, el modelo WRF (Weather Research and Forecasting) [106,107], es un modelo meteorológico que, al igual que el modelo MM5, es de mesoescala, pero utiliza ecuaciones de tipo Euler no hidrostáticas. Fue desarrollado por el Centro Nacional de Investigación Atmosférica de EU, con el cual es posible pronosticar con mayor exactitud la formación de fenómenos climatológicos, así como incendios y heladas. Es útil también para la predicción de acontecimientos climáticos que pongan en peligro la

¹³ EMEP es un programa dirigido políticamente y con sólido respaldo científico bajo la *Convención la Contaminación del aire de largo alcance más allá de las fronteras* para la cooperación internacional para resolver problemas más allá de las fronteras de la Unión Europea de Naciones.

producción agrícola, los recursos forestales, las actividades marítimas, así como las poblaciones situadas en zonas de riesgo. Este recurso informático puede utilizarse además para prever las condiciones del aire. El sistema WRF consiste del modelo mismo WRF, los preprocesadores para producir las condiciones de contorno inicial y lateral, tanto para datos reales como idealizados y pronósticos anidados de una sola vía, los postprocesadores para el análisis y visualización, y un programa tri-dimensional para la asimilación de datos variacional (3DVAR). Se puede acoplar a algún modelo fotoquímico numérico para la calidad del aire.

Todos estos programas requieren de una gran capacidad computacional y para efecto de pronóstico de calidad del aire requiere información detallada de las emisiones de contaminantes, en lo posible en línea cuantificando la contribución de todas las fuentes. En la práctica, siendo esto muy difícil de implementar, lo usual es alimentar los modelos con inventarios de emisiones que consideran datos promediados tanto temporal como espacialmente.

Capítulo 3

Aplicaciones a la predicción y pronóstico de series temporales de contaminación ambiental

“It doesn't say in the Bible that all laws of nature are expressible linearly”

- Enrico Fermi

“Chaos often breeds life, when order breeds habits”

- Henry Adams

3.1 Introducción a la predicción de variables de contaminación ambiental

Ha sido demostrado recientemente que el material particulado fino es más dañino para la salud humana que el material particulado grueso [108,109,110]. También está comprobado, como se mencionara en el capítulo 1, sección 4, que las partículas finas penetran el tracto respiratorio y tienen un efecto muy tóxico para el corazón [111] y los pulmones [112].

No obstante, muy pocos países cuentan con una norma para PM_{2.5}. Estados Unidos (USA) cuenta desde el año 1997 con una norma para 24-h de 65 $\mu\text{g}/\text{m}^3$ para este material particulado fino, siendo este valor recientemente disminuido a 35 $\mu\text{g}/\text{m}^3$. La norma anual se mantuvo en 15 $\mu\text{g}/\text{m}^3$. Por su parte, el Consejo Canadiense del Ministerio del Medioambiente impuso una norma para 24-h de 30 $\mu\text{g}/\text{m}^3$ en el año 2000. En México la norma actual para 24-h es de 65 $\mu\text{g}/\text{m}^3$, siendo la norma para el promedio anual de 15 $\mu\text{g}/\text{m}^3$. Finalmente, en Australia la norma vigente

para 24-h es de $25 \mu\text{g}/\text{m}^3$ y para el valor promedio anual esta fijado en $8 \mu\text{g}/\text{m}^3$. Ver tabla 1.4 para otros casos.

Hasta el momento Chile no cuenta con una norma de material particulado fino operativa, siendo las concentraciones de este material particulado excedidas frecuentemente, de acuerdo a los niveles máximos recomendados. Como ejemplo, en el año 2007 el máximo de 24-h de promedio móvil excedió el valor de $65 \mu\text{g}/\text{m}^3$ 79 días, y para 22 días fue mayor que $100 \mu\text{g}/\text{m}^3$. Estos valores excedidos ocurrieron durante el periodo otoño-invierno lo que indica un claro efecto estacional. Es justamente en este período que se observan fuertes inversiones térmicas, la que no ocurren en el periodo primavera-verano. Se puede afirmar que sobre la base de emisiones de $\text{PM}_{2.5}$ usual en una gran ciudad – debido al intenso tráfico, número de industrias en sus alrededores, etc.-, y su área geográfica rodeada por las montañas, es la meteorología la que domina la ocurrencia de eventos de muy altas concentraciones de material particulado fino en la atmósfera [113,114,115,116]. La importancia de la meteorología y la geografía de un lugar sobre las emisiones se ilustra bastante bien con el caso de Logan, UTAH, en Los Estados Unidos de América, cuya población es de unos 50.000 habitantes y donde, como una muestra representativa, las inversiones térmicas en el invierno de 2004 estuvieron fuertemente correlacionadas con las concentraciones de 24-h de material particulado fino, las cuales excedieron en nueve oportunidades el valor de $100 \mu\text{g}/\text{m}^3$ [117].

La posibilidad que ciertos días la población de una ciudad esté expuesta a concentraciones de $\text{PM}_{2.5}$ que estén en rangos considerados dañinos para la salud, implica contar con un eficiente sistema de pronóstico en línea y que pueda usarse como una herramienta de prevención. Por razones que quedarán claras más adelante en este capítulo nos concentraremos en el desarrollo de un posible modelo operacional de pronóstico de concentraciones de $\text{PM}_{2.5}$ que se enmarque dentro de la familia de modelos

llamados estadísticos, entre los cuales se encuentran modelos de regresión lineal, de redes neuronales y de clustering entre otros. Por ejemplo, el modelo presentado por Pérez *et al* [31] es una red de perceptrón multicapa (MLP) que predice promedios horarios de PM_{2.5} en el entorno de una estación de monitoreo de la ciudad de Santiago, desde una a 24 horas en adelante, donde los errores tienen un rango de 30% para horas más tempranas hasta 60% para las últimas horas del día. En este trabajo, sólo valores con rezago para el contaminante en estudio se consideró como entrada de la red neuronal, siendo demostrado que si se consideraban, además, variables meteorológicas se conseguía una mejora. También se demostró ahí que los valores pronosticados de PM_{2.5} usando una MLP fueron más precisos que los generados por un modelo de regresión lineal y de persistencia.

Por su parte, Ordieres *et al* [118] demostraron que una red neuronal supera a los modelos de regresión lineal y de persistencia a la hora de hacer predicción de concentraciones de PM_{2.5} en hasta 16 horas en adelante para un área urbana en la frontera México-USA.

También ha sido recientemente reportado un modelo de pronóstico de material particulado fino en una carretera Express [119], en que concentraciones promedio de una hora, un paso en adelante –equivalente a una hora-, se obtuvieron con un modelo lineal y una red MLP. Ellos concluyeron que con una elección apropiada de variables de entrada no existe una diferencia notable entre estos los dos modelos. Por su parte, un algoritmo híbrido de clases (HCA) se propuso para tareas de pronóstico, y se señala que puede superar a los modelos de redes neuronales. Este enfoque se usó para pronosticar material particulado grueso (PM₁₀), y los resultados mostraron una mejora en 10% por sobre los modelos de redes neuronales [120].

3.2 *Determinación del exponente de Lyapunov de series temporales de contaminación ambiental*

3.2.1 *Introducción*

Actualmente se han hecho diversos estudios acerca de posible caos en series de concentración de contaminación ambiental. Se ha encontrado, efectivamente, que series de este tipo presentan caos, gracias al cálculo de diversos parámetros que tipifican la caoticidad en series temporales no lineales [121,122,123,124]. Por ello se efectuó un análisis más minucioso para determinar si las series en estudio presentaban cierto grado de caoticidad.

3.2.2 *El caso de las cuatro estaciones de monitoreo*

Las estaciones en estudio fueron cuatro: Las Condes, Pudahuel, La Florida y Parque O'Higgins, que son las que registran mediciones de PM_{2.5} desde que comenzó a operar la red MACAM II.

Este estudio se efectuó para ver la predictibilidad para este tipo de material particulado. Se observó que todas poseen un exponente de Lyapunov positivo.

En la tabla 3.1 se muestran los valores máximos de los exponentes de Lyapunov en unidades de 1/hrs., durante todos los meses del año y para las cuatro estaciones de monitoreo. El valor promedio del exponente de Lyapunov para la estación Las Condes es de 0,445 (1/hrs.). Para la estación La Florida, 0,472 (1/hrs.); para la estación P. O'Higgins, 0,428 (1/hrs.), mientras que para la estación Pudahuel es 0,318 (1/hrs.).

De acuerdo a lo mencionado en la sección 2.2.2, ecuación 2.18, los límites de predictibilidad serían, respectivamente: 2,25 hrs., 2,11 hrs., 2,34 hrs. y 3,14 hrs.

Al ser caóticas, se caracterizan por una baja predictibilidad. Y eso es lo que justamente estaría comprobándose con los valores mostrados de la tabla 3.1, que las series de variables de contaminación contienen un alto grado de caoticidad, lo que implica a su vez que es imposible hacer una predicción confiable más allá de 3 horas.

En base estos resultados se puede argumentar que la precisión a conseguir en el pronóstico de PM2.5 con modelos deterministas (CMAQ, CHIMERE, etc.), no es muy alta, dado que para lograr precisión sería necesario disponer de información muy detallada de las fuentes emisoras. En la práctica en dicho tipo de modelos los datos de emisiones se ingresan considerando promedios espaciales y temporales, lo cual va en contra de una óptima estimación de condiciones de contorno.

Tabla 3.1: Exponente de Lyapunov (1/horas) para todo el año y las cuatro estaciones de monitoreo

| | 2000 | 2001 | 2002 | 2003 |
|--------------|---------------|---------------|---------------|---------------|
| Dim. Encaje | de=5 | de=5 | de=5 | de=5 |
| L. Condes | 0.428 ± 0.015 | 0.482 ± 0.015 | 0.413 ± 0.015 | 0.365 ± 0.013 |
| L. Florida | 0.504 ± 0.016 | 0.484 ± 0.015 | 0.394 ± 0.014 | 0.472 ± 0.016 |
| P. O'Higgins | 0.425 ± 0.014 | 0.399 ± 0.014 | 0.340 ± 0.014 | 0.429 ± 0.014 |
| Pudahuel | 0.351 ± 0.013 | 0.325 ± 0.014 | 0.276±0.013 | 0.351 ± 0.014 |
| | 2004 | 2005 | 2006 | 2007 |
| Dim. Encaje | de=5 | de=5 | de=5 | de=5 |
| L. Condes | 0.482 ± 0.015 | 0.462 ± 0.015 | 0.481 ± 0.015 | 0.446 ± 0.015 |
| L. Florida | 0.472 ± 0.015 | 0.503 ± 0.016 | 0.478 ± 0.015 | |
| P. O'Higgins | 0.456 ± 0.015 | 0.487 ± 0.015 | 0.456 ± 0.015 | 0.428 ± 0.015 |
| Pudahuel | 0.274 ± 0.013 | 0.365 ± 0.013 | 0.275 ± 0.013 | 0.325 ± 0.013 |

Aún más, el valor del exponente de Lyapunov permite determinar hasta cuántos pasos en adelante se pueden hacer un pronóstico confiable, basado sólo en los datos de la serie, como ya se mencionó.

Por su parte, en la tabla 3.2 se muestran los valores de los exponentes de Lyapunov en unidades de 1/hrs., considerando ahora sólo los meses de Otoño-Invierno y para las cuatro estaciones de monitoreo. Este cálculo se hizo pensando que es en ese período cuando se registran las concentraciones más altas de PM2.5.

El valor promedio del exponente de Lyapunov para la estación Las Condes, es de 0,442 (1/hrs.). Para la estación La Florida, 0,447 (1/hrs.); para la estación P. O'Higgins, 0,428 (1/hrs.), mientras que para la estación Pudahuel es 0,391 (1/hrs.).

Tabla 3.2: Exponente de Lyapunov para el periodo de invierno y las cuatro estaciones de monitoreo.

| | 2000 | 2001 | 2002 | 2003 |
|--------------|---------------|---------------|---------------|---------------|
| Dim. Encaje | de=5 | de=5 | de=5 | de=5 |
| L. Condes | 0.440 ± 0.024 | 0.453 ± 0.022 | 0.436 ± 0.025 | 0.455 ± 0.023 |
| L. Florida | 0.471 ± 0.023 | 0.452 ± 0.023 | 0.455 ± 0.023 | 0.452 ± 0.023 |
| P. O'Higgins | 0.447 ± 0.022 | 0.423 ± 0.023 | 0.417 ± 0.022 | 0.465 ± 0.023 |
| Pudahuel | 0.412 ± 0.021 | 0.398 ± 0.022 | 0.383 ± 0.022 | 0.411 ± 0.022 |
| | 2004 | 2005 | 2006 | 2007 |
| Dim. Encaje | de=5 | de=5 | de=5 | de=5 |
| L. Condes | 0.442 ± 0.022 | 0.414 ± 0.024 | 0.439 ± 0.023 | 0.446 ± 0.022 |
| L. Florida | 0.427 ± 0.022 | 0.443 ± 0.023 | 0.431 ± 0.022 | |
| P. O'Higgins | 0.466 ± 0.024 | 0.452 ± 0.023 | 0.464 ± 0.023 | 0.450 ± 0.022 |
| Pudahuel | 0.370 ± 0.021 | 0.389 ± 0.021 | 0.361 ± 0.021 | 0.401 ± 0.021 |

Nuevamente, de acuerdo a la ecuación 2.18, los valores de la predictibilidad serían: 2,27 hrs., 2,24 hrs., 2,34 hrs. y 2,56 hrs.,

respectivamente. Se observa que para Pudahuel, la estación con más episodios, su predictibilidad disminuye de 3,14 horas para datos de todo el año, a 2,56 horas considerando sólo otoño-invierno.

3.2.3 El caso de la estación Parque O'Higgins

Se estimó conveniente realizar un estudio más detallado de las series de PM2.5 en una de las estaciones de monitoreo de la red MACAM II. Se eligió la estación Parque O' Higgins, la cual muestra el promedio más alto de las concentraciones (ver tabla 3.8).

Para esta base de datos se efectuó un estudio de la posible presencia de caos y de otros parámetros de interés. Para ello, se consideraron sólo datos Otoño-Invierno (desde 1º de abril hasta 30 de septiembre).

Tabla 3.3: Diversos parámetros de las series de PM2.5 para meses de otoño-invierno en Estación P. O' Higgins

| Año | Dim. Capacidad | Sustituto | Dim. Correlación | Sustituto | Dim. Encaje |
|------|----------------|-----------|------------------|-----------|-------------|
| 2000 | 2,355 | 3,825 | 3,455 | 4,55 | 5 |
| 2001 | 2,115 | 3,063 | 3,601 | 4,476 | 5 |
| 2002 | 2,285 | 3,255 | 3,52 | 4,507 | 5 |
| 2003 | 2,503 | 3,573 | 3,671 | 4,643 | 5 |
| 2004 | 2,476 | 3,801 | 3,776 | 4,472 | 5 |
| 2005 | 2,17 | 3,834 | 3,744 | 4,526 | 5 |
| 2006 | 2,202 | 3,034 | 3,824 | 4,489 | 5 |
| 2007 | 2,194 | 3,33 | 3,943 | 4,496 | 5 |
| 2008 | 2,644 | 2,17 | 3,641 | 4,361 | 5 |

Se calcularon los siguientes parámetros: dimensión de capacidad, dimensión de correlación, ambas dimensiones con datos sustitutos, exponente de Hurst, entropía y predictibilidad. Ver tablas 3.3 y 3.4.

De acuerdo a lo mencionado en el capítulo 2, el hecho que las dimensiones de correlación y capacidad sean menores que cinco implica que las series no son aleatorias. Por otro lado, la predictibilidad calculada vía estimación de la entropía, siendo del mismo orden que la calculada vía

evaluación de exponentes de Lyapunov confirma que las series son caóticas.

Tabla 3.4: Diferentes parámetros en el estudio de posible caos para la estación P. O'Higgins

| Año | Dim. Correlación | Dim. Encaje | Fun. lag | Exp. Correl. | Exp. Hurst | Entropía | Máx. Predicción |
|------|------------------|-------------|----------|--------------|------------|----------|-----------------|
| 2000 | 2,361 | 2 | 16 | 64 | 0,265 | 0,525 | 1,9 |
| 2001 | 2,352 | 2 | 16 | 30 | 0,584 | 0,559 | 1,8 |
| 2002 | 2,316 | 2 | 16 | 128 | 0,257 | 0,489 | 2,0 |
| 2003 | 2,29 | 2 | 16 | 40 | 0,258 | 0,544 | 1,8 |
| 2004 | 2,357 | 2 | 16 | 128 | 0,247 | 0,416 | 2,4 |
| 2005 | 2,198 | 2 | 16 | 128 | 0,26 | 0,51 | 2,0 |
| 2006 | 2,336 | 2 | 16 | 74 | 0,241 | 0,488 | 2,1 |
| 2007 | 3,357 | 2 | 16 | 100 | 0,258 | 0,421 | 2,4 |
| 2008 | 2,305 | 2 | 16 | 542 | 0,185 | 0,405 | 2,5 |

Finalmente, el valor para la máxima predictibilidad (obtenido a través del recíproco de la entropía y en unidades de horas), resulta ser similar al encontrado con el método del exponente de Lyapunov. Vale decir, el pronóstico que se puede hacer es muy breve en el tiempo, entre 1,8 horas y 2,5 horas en adelante.

3.2.4 Evidencia de la sensibilidad a las condiciones iniciales

Para determinar cuan sensibles son las concentraciones de PM_{2.5} a las condiciones iniciales, se analizaron bases de datos conteniendo valores de los meses de Otoño-Invierno de variables meteorológicas (velocidad y dirección del viento, humedad relativa, temperatura), concentraciones de CO y el índice PMCA en la estación Parque O'Higgins. Se calcularon promedios diarios de estas variables para los años 2005 hasta 2008. Los días fueron separados en dos grupos. El primero fue el de aquellos casos en que el máximo del promedio de 24 horas de PM_{2.5} es menor o igual a 65 $\mu\text{g}/\text{m}^3$ y el segundo cuando dicho promedio es mayor que 65 $\mu\text{g}/\text{m}^3$. Los valores todas de las variables se mapearon al intervalo (0,1) y luego se calculó un vector centroide representativo de cada grupo en el cual cada una sus

componentes correspondía al valor promedio de cada una de las variables consideradas.

A continuación, se calculó la distancia euclidiana de todos los vectores a cada uno de los centroides. Si para un vector dado la menor de las dos distancias correspondía al centroide representativo de su grupo, se lo consideró clasificado correctamente. Si la menor distancia era a la del otro centroide se lo consideró erróneo. De esta manera se construyó la tabla 3.5.

Tabla 3.5: Clasificación de vectores Estación N (Invierno)

| | 2005 | | | 2006 | | |
|---------------------------------|-----------------|----------------|--------------|-----------------|----------------|--------------|
| | Correcto | Erróneo | Total | Correcto | Erróneo | Total |
| < 65 µg/m³ | 86% | 14% | 100% | 88% | 12% | 100% |
| > 65 µg/m³ | 85% | 15% | 100% | 87% | 13% | 100% |
| | 2007 | | | 2008 | | |
| | Correcto | Erróneo | Total | Correcto | Erróneo | Total |
| < 65 µg/m³ | 75% | 25% | 100% | 80% | 20% | 100% |
| > 65 µg/m³ | 89% | 11% | 100% | 86% | 14% | 100% |

Se puede observar que para el año 2005 por ejemplo, usando un criterio basado en la distancia, el 86% de los días en los cuales el promedio diario de PM_{2.5} es menor que 65 µg/m³ habría sido clasificado correctamente, mientras que el 14% de ellos habría sido clasificado en el grupo en el cual dicho valor es mayor que 65 µg/m³. Se puede argumentar que para un porcentaje apreciable de casos, condiciones similares dan lugar a concentraciones significativamente distintas de PM_{2.5}, lo cual indica que aún con la información provista por algunas variables exógenas, la predictibilidad de la serie de interés no será demasiado alta. El incorporar las concentraciones de monóxido de carbono como una de las componentes del vector es una manera de considerar el efecto de las fuentes móviles, que es un tipo de fuente que podría experimentar variaciones de un día para otro.

3.3 Aplicación de las RNA a la predicción de concentraciones de material particulado fino

Como una potencial aplicación de los conceptos mostrados en la sección 2.1.3 al pronóstico de valores futuros de concentraciones de material particulado fino, se ha intentado modelar los datos medidos en la estación Pudahuel de Santiago entre el 1° de Mayo al 30 de Septiembre (correspondiente a los meses de otoño e invierno) y entre los años de 1994, 1995 y 1996. Si bien estos datos corresponden mediciones de la red MACAM I, los exponentes de Lyapunov para estos años son del mismo orden de magnitud que los de la misma estación de la red MACAM II mostrados en la sección 3.2.2 (ver apéndice A.2). La idea es utilizar la información proporcionada por la función información mutua promedio (AMI) y el cálculo de la dimensión encajada para fijar la arquitectura de una red neuronal la cual a su vez nos permita pronosticar concentraciones de PM_{2.5} con algunas horas de anticipación [120].

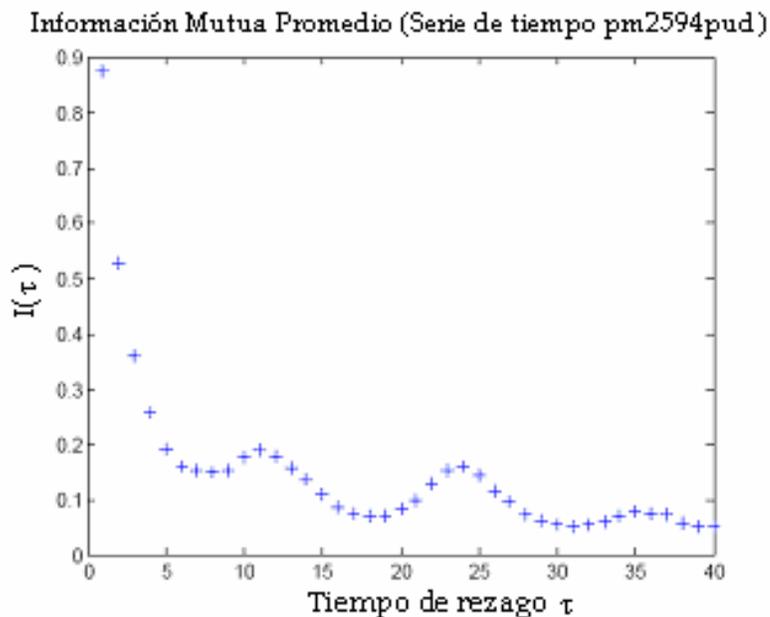


Figura 3.1: Función AMI para la serie temporal PM_{2.5} del año 1994.

En la figura 3.1 se muestra la función AMI para la base de datos del año 1994 (pm2594pud), en la cual se observa que el valor del tiempo de rezago óptimo es 7. Lo mismo se obtuvo para el año 1996 (pm2596). Para el año 1995 (pm2595) se obtuvo un tiempo de rezago de 8 [126].

A continuación, se calculó la función FNN, la cual permite determinar la dimensión encajada, d_e . Como se aprecia en la figura 3.2, el porcentaje de FNN cae a cero cuando la dimensión encajada ha sido alcanzada. En esta figura se muestra la función FNN (falsos vecinos más cercanos) para la base de datos pm2596, obteniéndose idéntica dimensión de encaje para las otras dos bases de datos. Se observa de esta figura, que los vecinos más cercanos caen a cero en $d_e = 4$.

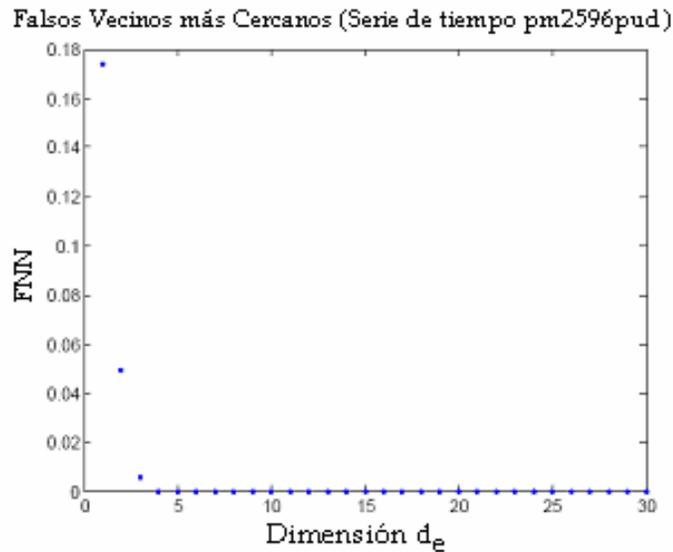


Figura 3.2: Falsos vecinos más cercanos para la serie temporal $PM_{2.5}$ del año 1996.

Retornando al ejemplo dado en 2.1.4, la base de datos se procesó usando los valores calculados de los parámetros τ y d_e . Específicamente, se seleccionaron cuatro entradas (las mismas que la dimensión de encaje), una única salida que corresponde al pronóstico de un paso en adelante que se quiere efectuar. Si bien el gráfico de la función AMI sugiere un $\tau = 7$, se usó

un $\tau = 6$, dado que siendo éste divisor de 24, permite tomar mejor en cuenta la periodicidad de los datos [101].

Los datos disponibles se agruparon en tres grupos, los cuales se generaron desplazando una ventana de tiempo cada 6 horas, sobre la serie completa. Así, de los 3.672 datos correspondientes a una determinada serie temporal, quedaron finalmente 122 datos por columna, de un total de 5 columnas siendo esta última la 06 a.m. del día que se quiere pronosticar (las cuatro primeras columnas corresponden a las 06 a.m., 12 p.m. 18p.m. y las 00 a.m., respectivamente, del día previo). Este procedimiento se efectuó con todas las bases de datos: pm2594, pm2595 y pm2596. Con todo esto, la nueva base de datos que surgió de la pm2594 se utilizó para entrenamiento; la nueva base de datos surgida de la antigua pm2595, se usó para test de validación durante el aprendizaje, y finalmente la nueva base de datos que emergió de la antigua pm2596, se usó para test [101].

Se diseñaron tres tipos de redes neuronales artificiales: un perceptrón, una red de mult capas con una sola capa escondida y otra red de mult capas, pero con dos capas ocultas. La capa de entrada para los tres tipos de redes, tenía cuatro neuronas. Mientras que la capa de salida tenía sólo una neurona. Los cuatro PE's de entrada se debe a que es la cantidad óptima que se obtiene a través de la función FNN, que corresponde a la dimensión encajada como se mencionó anteriormente, lo cual concuerda con [103]. Además, la neurona de salida se debe a que se requiere hacer sólo un pronóstico en adelante.

Es necesario señalar acá que el perceptrón simple resulta ser un modelo dinámico lineal al implementarse con una función de transferencia lineal. Por lo tanto considerar este caso permite comparar la red neuronal (de por sí no lineal) con un modelo lineal. Compararemos también los resultados con los obtenidos con un modelo simple de persistencia (la concentración de PM2.5 a las 6 am de mañana es igual a la de las 12 de la noche de hoy).

La red neuronal ha sido entrenada usando el método de propagación hacia atrás incluyendo el término de momentum (ver sección 2.3.4).

A pesar que la función AMI arrojó valores de 7 y 8 como tiempo de rezago, como se observa en la Tabla 3.1, se prefirió trabajar con un tiempo de 6. Se sabe que es el valor $\tau = 7$ que debemos emplear para calcular la función FNN. Sin embargo, se ha demostrado [100] que dicha función no va exactamente a cero. En cambio, si se elige $\tau = 6$ se consigue que FNN vaya exactamente a cero y con ello, logramos capturar la dinámica correcta de nuestras series temporales. El valor 6 es más conveniente, ya que siendo un divisor de 24 permite aprovechar mejor la base de datos. Por ello se ha escogido $\tau = 6$ como tiempo de rezago.

Como criterio de comparación de los modelos, se usó el error NRMSE – por normalizad root mean square error-, definido como:

$$NRMSE = \frac{\sqrt{\frac{1}{N} \sum_i (O_i - P_i)^2}}{\frac{\sum_i O_i}{N}} \quad [3.1]$$

en que O corresponde al valor real y P al predicho. El error en la predicción se calculó comparando el valor predicho por la red en la etapa de test, con el valor que efectivamente se presentó.

En la tabla 3.6, se tabulan los parámetros de tiempo de rezago, τ y la dimensión encajada, d_e para las tres bases de datos. Se puede observar que, a pesar que la función AMI para la segunda base de datos da un valor de rezago de 8 (y las restantes de 7), por las razones expuestas anteriormente se ha escogido el valor para AMI de 6, en las tres bases de datos.

Como se mencionó más arriba, el valor de la dimensión de encaje, para las tres series temporales, resultó ser de 4. Este dato sirve para determinar

el número de PE's en la capa de entrada, para el diseño de nuestra red neuronal.

Tabla 3.6: Funciones AMI y FNN para las tres bases de datos estudiadas

| | AMI(τ) | FNN (d_E) |
|--------|---------------|---------------|
| PM2594 | 7 | 4 |
| PM2595 | 8 | 4 |
| PM2596 | 7 | 4 |

En la tabla 3.7 se presentan los parámetros y resultados para las diferentes redes neuronales implementadas. Se aprecia que la “mejor” red neuronal fue aquella con una capa escondida y un número de 7 neuronas en dicha capa. Además, se muestra el caso del perceptrón simple, con función de transferencia lineal, así como función no lineal, siendo el mejor perceptrón aquel con un valor RNMSE de $0,5372 \pm 0,003$ (con función no lineal sigmoide). Para algunos modelos de red neuronal, se han analizados distintos valores de los parámetros que aparecen en la ecuación [2.32]. Además, se han incluido dos tipos de reglas de aprendizaje: la regla Delta-Generalizada (D-R) y una variante denominada Normalized Cumulative-Learning Rule (N-C), la cual para aliviar el problema de presentación estructurada del conjunto de datos en la entrada, acumula los cambios de los pesos sobre varias presentaciones de entrenamiento y, posteriormente, hace la actualización de todos a la vez, a diferencia de la regla delta usual, en la cual los pesos se van actualizando secuencialmente. Los cambios en los pesos para la regla delta, en el primer caso, son acumulados en el término de momentum estándar (ver ecuación [2.32]).

En resumen, podemos observar de la tabla 3.7 que una red de perceptrón simple con función de transferencia lineal es la que peor aprendió. Al cambiársele la función de transferencia a no lineal (tipo *sigmoide*), se aprecia que mejoró su tasa de aprendizaje. Por otro lado, la

red de multicapas con una capa escondida y entre 7 y 10 neuronas en esa capa, obtuvieron mejor aprendizaje, a diferencia de trabajo [114] en el cual un perceptrón superó a una red de multicapas en cuanto a predecir ciertas horas en adelante, en todas las situaciones. El valor para la mejor red de multicapas fue de $0,5156 \pm 0,047$.

Tabla 3.7: Valores NRMSE para distintos tipos de modelos RNA

| NNCO | β | R Apren | TF Trans | α | NRMSE |
|-------------------------|---------|---------|----------|----------|--------|
| Perceptrón simple | | | | | |
| 0 | 0 | N-C | L | 0,5 | 0,5505 |
| 0 | 0 | N-C | S | 0,5 | 0,5963 |
| 0 | 0 | N-C | S | 0,5 | 0,6087 |
| 0 | 1 | D-R | S | 0,5 | 0,5372 |
| 0 | 1 | N-C | L | 0,5 | 0,5385 |
| 0 | 0 | N-C | L | 0,9 | 0,549 |
| Red de Multicapas (MLP) | | | | | |
| 1 | 0 | N-C | S | 0,5 | 0,8375 |
| 1 | 0 | D-R | T | 0,5 | 0,5698 |
| 1 | 0 | D-R | T | 0,9 | 0,5499 |
| 1 | 0 | N-C | S | 0,9 | 0,5479 |
| 5 | 1 | D-R | S | 0,5 | 0,559 |
| 7 | 1 | D-R | S | 0,5 | 0,5156 |
| 10 | 0 | D-R | S | 0,5 | 0,5648 |
| 10 | 1 | D-R | S | 0,5 | 0,5259 |

L: Lineal; S: Sigmoide; T: Tanh
 NNCO: número de neuronas en capa oculta.
 R Apren: regla de aprendizaje.
 TF Trans: tipo de función de transferencia.
 hNRMSE: error de la raíz cuadrática promedio normalizado.
 N-C: regla de aprendizaje acumulativa normalizada.
 D-Regla delta generalizada.

También se probó una RNA con dos capas ocultas y algunas neuronas en cada capa oculta. No se incluyó el valor NRMSE debido a que no mejoró el rendimiento en comparación a los modelos anteriores.

Por su parte, mediante el método de persistencia, se obtuvo un valor NRMSE de $0,7649. \pm 0,050$.

En la figura 3.3 se muestra la arquitectura de la mejor red neuronal que obtuvo el mejor valor NRMSE, de acuerdo a la Tabla 3.7. Se usó una regla de aprendizaje delta generalizada, con función de transferencia no lineal tipo sigmoide.

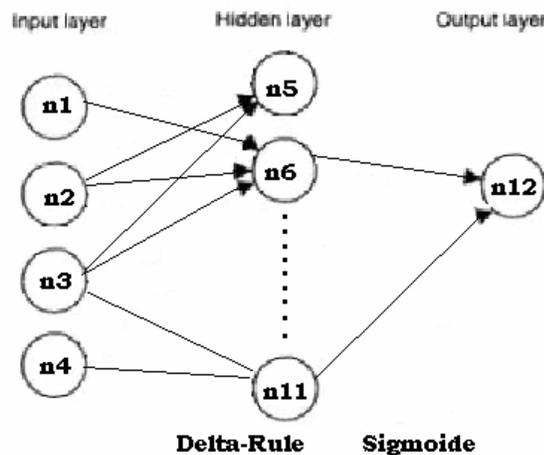


Figura 3.3: Arquitectura de la mejor red con valor RNMSE 0.5156.

En esta misma figura se muestran algunas conexiones entre los nodos de entrada y la capa oculta y entre esta última y el PE de salida.

Finalmente, podemos decir que ambos tipos de redes neuronales lograron superar al método de persistencias, siendo para el caso de red de multicapas, entre un 21,4% y un 32,4% mejor. En el caso del perceptrón simple, fue entre un 16,5% y un 28,2% mejor al método de persistencias.

Si bien hemos logrado demostrar que una red neuronal permite pronosticar valores futuros de concentraciones de PM2.5 mejor que un modelo lineal y uno de persistencia, los errores absolutos son relativamente altos, lo cual era de esperar considerando la presencia importante de caos en dichas series temporales y que estamos intentando pronosticar

basándonos exclusivamente en la información contenida en la misma serie. El esquema descrito probablemente tendría una mayor eficacia para el caso de series cuyo exponente máximo de Lyapunov sea considerablemente más pequeño. En la próxima sección mostraremos un esquema significativamente distinto de acuerdo al cual trataremos de compensar la baja predictibilidad de las series de PM_{2.5} incorporando como datos de entrada variables que contienen información meteorológica y de emisiones.

3.4 *Hacia un modelo operacional de pronóstico de concentraciones de PM_{2.5} en la ciudad de Santiago*

3.4.1 *Introducción*

Si bien actualmente el material particulado fino PM_{2.5} no está normado en Chile, la autoridad ambiental anunció recientemente que para el año 2012 debería entrar en vigencia una norma tanto para el promedio móvil de 24 horas como para el promedio anual. Dicha norma iría acompañada de un plan de contingencias, de acuerdo al cual, cuando el promedio de 24 horas supera ciertos límites se deberían implementar medidas restrictivas a vehículos e industrias. Para dicho efecto será necesario disponer de un modelo de pronóstico que permita estimar el máximo del promedio móvil de 24 horas de PM_{2.5} con un día de anticipación. A continuación se muestra un análisis de un conjunto de herramientas que podrían constituirse en la base de un modelo operacional de pronóstico de apoyo a la gestión de contingencias por PM_{2.5} en la ciudad de Santiago.

3.4.2 *El caso de tres clases*

En esta sección se presenta la comparación entre tres métodos de pronóstico con un día de anticipación del máximo del promedio móvil de 24 de concentraciones de PM_{2.5} de promedio móvil de 24 horas: regresión lineal (LR), red neuronal de multicapas (MLP) y una variante del algoritmo de clustering híbrido (HCA).

Como se mencionó en el capítulo 2, MLP es un método no lineal que ajusta los pesos de ciertas conexiones sobre la base de un conjunto de

patrones de entrenamiento, dada una función de activación. LR vendría siendo, en cambio, un caso particular de MLP en el cual la función de activación es de tipo lineal. Por su parte, HCA separa los datos en clases definidas por ciertos centroides. Después del entrenamiento los datos de entrada se distribuyen dentro de estas clases de acuerdo a su distancia a los centroides [128].

Los datos analizados fueron medidos en el periodo otoño – invierno entre los años 2001 y 2007, inclusive, en las estaciones de monitoreo L, M, N y O que son las únicas que registran datos de PM_{2.5} par dicho período. Los datos son promedios horarios, a partir de los cuales se se puede calcular el promedio móvil de 24 horas (24HMAPM_{2.5}).

Como se mencionara en el Capítulo 1, es en este periodo que la calidad del aire en la ciudad de Santiago llega a niveles críticos para la salud de la población, debido a que las condiciones meteorológicas se caracterizan por vientos débiles y frecuentes e intensas inversiones térmicas.

La idea es implementar una función multivariada [128,129]:

$$Y = F(x_1, \dots, x_n, z_1, \dots, z_m) \quad [3.2]$$

En esta relación Y es un cuadri-vector cuyas componentes son los máximos de 24HMAPM_{2.5} del día siguiente en 4 estaciones de monitoreo de Santiago; x_1, \dots, x_n son valores pasados de PM_{2.5} del día presente; z_1, \dots, z_m son variables exógenas medidas y pronosticadas.

Las variables de entrada (input) se seleccionaron después de hacer un análisis de correlaciones.

Una cantidad importante que va como entrada en los modelos es el PMCA descrito en el capítulo 1, la cual es una variable discreta que varía entre 1 y 5 y mide la ventilación de la región metropolitana [6]. Así, un valor de PMCA 1 se asigna a días con alta inestabilidad (por ejemplo, el paso de un frente activo). Por su parte, un valor de 5 corresponde a una fuerte

estabilidad atmosférica. En este caso, corresponde a situaciones con gran amplitud térmica bajo cielo claro y con baja altura de inversión térmica. El cálculo de esta cantidad requiere la evaluación de configuraciones sinópticas y de mesoescala, siendo pronosticada diariamente por meteorólogos de la Universidad de Santiago de Chile. Si en un determinado día se espera más de un valor de PMCA, se toma el peor de ellos (se evalúa cada 12 horas).

El objetivo en este caso, es pronosticar el máximo del promedio móvil de 24 horas de $PM_{2.5}$ para el día siguiente basado en la información disponible a las 20:00 hrs. del día de hoy [129].

Las variables de entrada (en total 15 inputs) para los tres modelos fueron:

- Promedio de una hora de $PM_{2.5}$ en cada estación a las 19:00 del día presente.
- Promedio de una hora de $PM_{2.5}$ en cada estación a las 20:00 del día actual.
- Promedio de 24 horas de $PM_{2.5}$ de cada estación de monitoreo a las 20:00 del día de hoy.
- Amplitud térmica en la ciudad para el día actual.
- Amplitud térmica pronosticada para el día siguiente.
- Índice de PMCA pronosticado para el día siguiente.

Las variables (promedio de 1-h de $PM_{2.5}$) dan información sobre las tendencias de este material particulado en la noche del día presente y ellos afectarán al 24HMAPM2.5 del día siguiente. Por su parte, el promedio de 24-h a las 20:00 indica la acumulación de material particulado fino en la atmósfera en las últimas horas del día. Las tres últimas variables son valores estimados de la ciudad para la parte central de Santiago.

La variación de temperatura entre los sitios de las estaciones de monitoreo son del orden típico de $1^{\circ}C$ y no se encontró que fuera relevante para la tarea de pronóstico de calidad del aire. Por otro lado, las amplitudes térmicas presentan una correlación importante con la intensidad de las

inversiones térmicas y el índice de PMCA, siendo representativos del área metropolitana completa, resumiendo la relevancia de la condición meteorológica general asociada con la calidad del aire.

Para el caso de una red tipo MLP, F en la ecuación [3.2] es una función no lineal particular, después que los parámetros ajustables se fijan durante la etapa de entrenamiento de la red. La no linealidad está presente siempre que la función de activación que propaga la información desde una capa a la siguiente es ella misma no lineal, siendo ésta típicamente una función de tipo sigmoide. Por ello un modelo lineal se puede ver como un caso particular de una red MLP en la cual su función de activación es una función lineal.

Ya que el propósito principal es pronosticar valores desconocidos en el futuro, se escogió entrenar los tres modelos con datos de los años previos. A continuación se generaron pronósticos para los años entre 2004 y 2007, inclusive. Para cada uno de estos años, los modelos de pronóstico fueron entrenados con los datos de los tres años previos. Así, para pronosticar el año 2004, se entrenaron los tres modelos con los años 2001, 2002 y 2003. Y así sucesivamente, para los años 2005, 2006 y 2007.

Los parámetros para un año dado se ajustan con la información para entrenamiento desde los tres años previos. Tanto para el caso de tres clases como para cuatro clases –que se verá en la siguiente sección–, no se usaron conjunto de datos como test de validación, para optimizar el número de pasos de entrenamiento, debido al hecho que se quiere implementar un modelo operacional, en cuyo caso los valores futuros se desconocen.

Uno de los objetivos del estudio es pronosticar en qué rango estarán las concentraciones del día siguiente. Se definen así clases.

La clase de días A: es aquella en que el máximo de $24\text{HMAPM}_{2.5}$ es menor que $65 \mu\text{g}/\text{m}^3$. La clase de días B corresponde a máximo entre 65

$\mu\text{g}/\text{m}^3$ y $100 \mu\text{g}/\text{m}^3$. Finalmente, la clase de días C corresponde al máximo más grande que $100 \mu\text{g}/\text{m}^3$.

También se ha implementado una variante del algoritmo HCA presentado en la sección 2.4.3 [128], comparando esta metodología con otras dos más tradicionales (regresión lineal y RNA).

Usualmente es el algoritmo HCA quien genera las clases. Pero en nuestro caso, las clases fueron definidas de una manera natural y de acuerdo a los valores de los máximos de promedio móvil de $\text{PM}_{2.5}$ entre las cuatro estaciones de monitoreo consideradas. Posteriormente se determinaron tres centroides, para cada una de las tres clases.

Tabla 3.8: Valores promedio de $\text{PM}_{2.5}$ en $\mu\text{g}/\text{m}^3$

| | Estación L | Estación M | Estación N | Estación O |
|----------|-------------------|-------------------|-------------------|-------------------|
| 2001(12) | 35 | 26 | 34 | 33 |
| 2001(5) | 44 | 31 | 48 | 48 |
| 2002(12) | 32 | 25 | 35 | 34 |
| 2002(5) | 43 | 30 | 50 | 51 |
| 2003(12) | 34 | 28 | 40 | 36 |
| 2003(5) | 48 | 34 | 60 | 55 |
| 2004(12) | 36 | 26 | 35 | 34 |
| 2004(5) | 47 | 30 | 48 | 48 |
| 2005(12) | 34 | 27 | 33 | 32 |
| 2005(5) | 41 | 29 | 42 | 41 |
| 2006(12) | 35 | 25 | 34 | 32 |
| 2006(5) | 46 | 31 | 48 | 47 |
| 2007(12) | 34 | 26 | 34 | 33 |
| 2007(5) | 46 | 33 | 50 | 50 |

Para los tres años de datos de entrenamiento, se calcularon los valores promedios de las 15 variables seleccionadas como entrada mencionada arriba dentro de las clases respectivas, A, B y C (tres centroides). Dentro de cada clase se generaron regresiones lineales que reproducen los valores de las cuatro variables de salida [129].

Una vez que se tiene los centroides para cada una de las clases, se puede realizar el test con los datos del año siguiente, asignando un vector dado a la clase con centroide a la mínima distancia Euclidiana desde él. Después que ha sido identificada la clase, se puede calcular el valor numérico pronosticado usando un algoritmo lineal válido para esa clase. Para un sistema de pronóstico operacional, sería deseable generar los valores lo más preciso posible para la clase de día siguiente y el valor numérico esperado del máximo de la concentración de $PM_{2.5}$.

En la tabla 3.8 se muestran los valores calculados para $PM_{2.5}$ tomado para el año completo (en el caso de (12)), y promedio tomado para los meses de invierno (para el caso de (5)), para las cuatro estaciones en estudio. Desde esta tabla se puede observar un evidente efecto estacional [129].

En la tabla 3.9 se muestra la cantidad de días que pertenecen a cada una de las tres clases, A, B o C.

Tabla 3.9: Número de días pertenecientes a cada clase

| | CLASE A | CLASE B | CLASE C |
|----------|---------|---------|---------|
| 2004 (5) | 81 | 52 | 20 |
| 2005 (5) | 90 | 58 | 5 |
| 2006 (5) | 75 | 65 | 13 |
| 2007 (5) | 74 | 5 | 22 |

En la tabla 3.10 Se muestra el porcentaje de concordancia entre la clase pronosticada y la clase observada, para cada una de las tres clases, A, B y C, usando los tres métodos en estudio: regresión lineal –perceptrón lineal-, MLP y HCA.

Se puede observar que el algoritmo HCA fue más exacto en detectar situaciones de altas concentraciones (clase C), si se lo compara con los

otros dos métodos. Por ejemplo, en los años 2006 y 2007, HCA detecto un 66% y un 77%, respectivamente, que corresponden a más aciertos que los que reportan los otros algoritmos.

Tabla 3.10: Porcentaje de concordancia entre la clase observada y la pronosticada usando tres métodos diferentes

| | MLP | | | LINEAL | | | HCA | | |
|------|-----|-----|-----|--------|-----|------|-----|-----|-----|
| | A | B | C | A | B | C | A | B | C |
| 2004 | 84% | 55% | 67% | 75% | 69% | 605% | 81% | 62% | 65% |
| 2005 | 85% | 78% | 40% | 88% | 79% | 20% | 90% | 60% | 20% |
| 2006 | 82% | 55% | 31% | 84% | 57% | 31% | 84% | 43% | 62% |
| 2007 | 84% | 61% | 36% | 95% | 63% | 18% | 88% | 50% | 77% |

En la tabla 3.11 se observa el porcentaje de error promedio de los valores numéricos pronosticados de los máximos de 24HMAPM2.5, comparado con los valores numéricos observados. La columna *máx* (máximo) muestra el error entre los valores máximos pronosticados y los observados, independiente de la estación de monitoreo.

Tabla 3.11: Error porcentual absoluto promedio para los tres métodos.

| | MLP | | | | | LINEAL | | | | | HCA | | | | |
|------|-----|-----|-----|-----|-----|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | L | M | N | O | Máx | L | M | N | O | Máx | L | M | N | O | Máx |
| 2004 | 15% | 14% | 23% | 31% | 21% | 18% | 19% | 25% | 40% | 26% | 18% | 18% | 22% | 32% | 22% |
| 2005 | 16% | 18% | 20% | 22% | 17% | 17% | 19% | 19% | 21% | 15% | 15% | 16% | 21% | 22% | 18% |
| 2006 | 17% | 17% | 25% | 31% | 21% | 18% | 18% | 24% | 30% | 20% | 20% | 19% | 25% | 33% | 22% |
| 2007 | 18% | 18% | 20% | 26% | 19% | 19% | 21% | 19% | 23% | 20% | 18% | 19% | 19% | 26% | 20% |

De acuerdo a estos resultados, se observó que los tres métodos empleados pueden llegar a ser herramientas útiles para una gestión apropiada de la calidad del aire, en una ciudad con altos índices de contaminación ambiental.

No obstante, el algoritmo de clases parece ser más exacto en cuanto a pronosticar valores extremos y es por esta razón que es más eficiente para detectar clases de días A y C [128,129].

3.4.3 El caso de cuatro clases

Como la nueva norma para el promedio de 24 hrs. en USA es de $35 \mu\text{g}/\text{m}^3$ y la antigua norma era de $65 \mu\text{g}/\text{m}^3$, sería útil saber las veces que se ha excedido esta norma. Por ello esta vez se han creado cuatro clases: los días de clase A es cuando el máximo de 24HMAPM2.5 es menor que $35 \mu\text{g}/\text{m}^3$; los días de clase B son aquellos cuando el máximo de 24HMAPM2.5 está en el rango de $35 \mu\text{g}/\text{m}^3$ y $65 \mu\text{g}/\text{m}^3$; los días de clase C es cuando el máximo de 24HMAPM2.5 esta entre $65 \mu\text{g}/\text{m}^3$ y $100 \mu\text{g}/\text{m}^3$ y los días de clase D cuando el máximo de 24HMAPM2.5 es mayor que $100 \mu\text{g}/\text{m}^3$. Este último valor es considerado como crítico pues supera en cuatro veces al valor establecido por la OMS [14].

Teniendo en mente que estamos espacialmente interesados en situaciones con altas concentraciones de $\text{PM}_{2.5}$, en los modelos MLP se ha modificado del conjunto de datos para entrenamiento repitiendo los casos de días C una vez y repitiendo los casos de días D dos veces. Considerando, por su parte, que se incluye los datos entre Abril y Agosto para 3 años, los conjuntos para entrenamiento tienen del orden de 700 muestras de la relación entre la entrada y la salida. Es por esta razón que una red de tres capas con 20 unidades o PE's en la capa oculta es una elección apropiada para esta arquitectura. Además, el número de pasos de entrenamiento usado que produjo un error razonablemente bajo en el ajuste de los datos de entrenamiento, y en promedio no conduce a sobreentrenamiento de la red, era de 2×10^6 pasos.

Para los tres años de datos de entrenamiento, se calcularon los valores promedios de las 15 variables seleccionadas como entrada mencionada arriba dentro de las clases respectivas, A, B, C y D (cuatro centroides). Dentro de cada clase se generaron regresiones lineales que reproducen los valores de las cuatro variables de salida. Una vez que se tiene los centroides para cada una de las clases, se puede realizar el test con los datos del año siguiente, asignando un vector dado a la clase con centroide a la mínima

distancia Euclidiana desde él. Después que ha sido identificada la clase, se puede calcular el valor numérico pronosticado usando el algoritmo lineal válido para esa clase.

Para un sistema de pronóstico operacional, sería deseable generar los valores lo más preciso posible para la clase de día siguiente y el valor esperado del máximo de la concentración de $PM_{2.5}$.

Con el método HCA la clase es aquella de la mínima distancia entre el vector de 15 componentes de las entradas del año correspondiente al test y los cuatro centroides calculados con los datos de entrenamiento. Se observa que para el año 2004 la tasa más alta de las clases B y C identificados con los algoritmos lineal y MLP se compensa en algo con la baja tasa para las clases A y D, siendo en esta dos últimas clases el método HCA más preciso [22].

Tabla 3.12: Comparación entre tablas de contingencia para el año 2007 usando los algoritmos HCA y MLP como métodos de pronóstico

| 2007 | HCA | Pronosticado | | | | Total | % O | 2007 | MLP | Pronosticado | | | | Total | % O |
|-----------|-------|--------------|----|----|----|-------|-----------|-----------|-------|--------------|----|----|----|-------|-----------|
| | | A | B | C | D | | | | | A | B | C | D | | |
| Observado | A | 14 | 1 | 0 | 0 | 15 | 93 | Observado | A | 13 | 2 | 0 | 0 | 15 | 87 |
| | B | 17 | 34 | 7 | 0 | 58 | 59 | | B | 8 | 42 | 8 | 0 | 58 | 72 |
| | C | 2 | 11 | 28 | 15 | 56 | 50 | | C | 0 | 13 | 40 | 3 | 56 | 71 |
| | D | 0 | 0 | 5 | 17 | 22 | 77 | | D | 0 | 2 | 12 | 8 | 22 | 36 |
| | Total | 33 | 46 | 40 | 32 | 151 | 62 | | Total | 21 | 59 | 60 | 11 | 151 | 68 |
| % F | 42 | 74 | 70 | 53 | | | % F | 62 | 71 | 67 | 73 | | | | |

Para el año 2005 el método MLP parece levemente mejor, pero dado que durante este año hubo sólo cinco días de clase D, el 40% de acuerdo se debió a un caso adicional de concordancia con las observaciones si se compara con los modelos lineales y HCA. Más aún, los días D para este año fueron más bien atípicos. Por su parte, para los años 2006 y 2007 el método HCA dio la clasificación más precisa para las situaciones de concentración alta y baja. A su vez. El rendimiento del modelo HCA para la identificación de días D para estos años es notable (en términos relativos) [130].

Más detalle acerca del rendimiento de este modelo se obtiene calculando las tablas de contingencia. Por ejemplo, en la tabla 3.12 se muestra una tabla de contingencia para los modelos HCA y MLP, para el año 2007. Las columnas A, B, C y D muestran el número de días pronosticados de pertenecer a una clase dada contra la clase del día observado [130].

La columna 0% muestra el porcentaje de días observados por la clase que fue pronosticado pertenecer a esa clase. La cantidad %F entrega el porcentaje de días pronosticados que fue verificado que ocurrió. $100 - \%F$ debería ser, entonces, el porcentaje de falsos positivos.

Tabla 3.13: Número de días pertenecientes a cada una de las clases

| | CLASE A | CLASE B | CLASE C | CLASE D |
|----------|---------|---------|---------|---------|
| 2004 (5) | 19 | 49 | 52 | 20 |
| 2005 (5) | 13 | 55 | 58 | 5 |
| 2006 (5) | 13 | 60 | 65 | 13 |
| 2007 (5) | 15 | 58 | 56 | 22 |

Los números en la diagonal, correspondiente a los valores pronosticados para cada una de las clases, corresponden a los pronósticos exitosos para las clases. Por último, en la esquina inferior derecha aparece registrada en negrita el porcentaje total de concordancia entre las clases observadas y pronosticadas.

En la tabla 3.13 se muestra la clasificación para los días entre el 1 de abril y el 31 de agosto, desde el año 2004 al año 2007. Se hace notar que en un modelo operacional útil para la administración de la calidad del aire, es deseable tener un pronóstico confiable para las altas concentraciones [130].

La tabla 3.14 muestra los resultados del porcentaje de identificación de la clase correcta para cada uno de los tres métodos.

Tabla 3.14: Porcentaje de acuerdo entre la clase pronosticada y la observada

| | MLP | | | | LINEAL | | | | HCA | | | |
|------|-----|-----|-----|-----|--------|-----|-----|-----|-----|-----|-----|-----|
| | A | B | C | D | A | B | C | D | A | B | C | D |
| 2004 | 47% | 82% | 67% | 50% | 37% | 80% | 67% | 50% | 84% | 55% | 62% | 70% |
| 2005 | 46% | 75% | 40% | 20% | 46% | 76% | 76% | 40% | 77% | 60% | 64% | 20% |
| 2006 | 31% | 70% | 52% | 46% | 38% | 70% | 57% | 38% | 77% | 62% | 40% | 62% |
| 2007 | 87% | 72% | 71% | 36% | 67% | 69% | 64% | 41% | 93% | 59% | 50% | 77% |

La tabla 3.15 muestra, por su parte, el porcentaje de error promedio de los valores numéricos pronosticados en comparación con el valor numérico observado para el máximo de 24HMAPM2.5 por estaciones, de acuerdo a la relación:

$$PE = \left\langle \frac{|y_{ta} - y_{tp}|}{y_{ta}} \right\rangle \times 100 \quad [3.3]$$

donde y_{tp} es el valor pronosticado, y_{ta} es el valor actual y el paréntesis angular quiere decir promedio sobre todos los casos de test. La columna *max* compara el valor máximo pronosticado con el valor máximo observado, independiente de la estación. Se observa que no existe gran diferencia entre los errores obtenidos con los tres métodos [130].

Tabla 3.15: Porcentaje de error promedio de los valores numéricos pronosticados de 24HMAPM2.5 comparado con valor numérico observado.

| | MLP | | | | | LINEAL | | | | | HCA | | | | |
|------|-----|-----|-----|-----|-----|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | L | M | N | O | Máx | L | M | N | O | Máx | L | M | N | O | Máx |
| 2004 | 16% | 16% | 22% | 31% | 20% | 16% | 18% | 23% | 32% | 21% | 16% | 18% | 20% | 28% | 21% |
| 2005 | 17% | 19% | 21% | 23% | 18% | 17% | 18% | 19% | 23% | 16% | 17% | 20% | 19% | 20% | 17% |
| 2006 | 18% | 19% | 26% | 32% | 21% | 17% | 17% | 24% | 31% | 21% | 21% | 22% | 24% | 30% | 22% |
| 2007 | 18% | 18% | 19% | 25% | 19% | 19% | 18% | 20% | 25% | 20% | 18% | 19% | 19% | 25% | 20% |

Los errores de pronóstico con el método HCA parece levemente más grande, lo cual puede deberse a la diferencia de los resultados obtenidos con las ecuaciones de clases diferentes para el mismo vector de entrada, lo

cual afectará los casos de identificación de clase errónea que están en el límite entre clases.

Esto puede corregirse, en parte, usando dos en lugar de cuatro ecuaciones para el pronóstico del valor numérico, una para las clases A y B y otra para las clases C y D. En este caso, desde los resultados mostrados en la tabla 3.16, que muestra el error porcentual promedio de los valores numéricos pronosticados del los máximos de 24HMAPM2.5 para HCA usando dos ecuaciones para pronóstico, se observa una disminución en los errores de pronóstico, lo cual llega a ser ahora levemente más pequeño que los errores de pronóstico obtenido con los algoritmos lineales y MLP [130].

Tabla 3.16: Error porcentual absoluto promedio con dos ecuaciones de pronóstico para el modelo HCA

| | HCA | | | | |
|------|-----|-----|-----|-----|-----|
| | L | M | N | O | Máx |
| 2004 | 15% | 15% | 21% | 29% | 19% |
| 2005 | 15% | 18% | 19% | 20% | 16% |
| 2006 | 19% | 18% | 23% | 28% | 20% |
| 2007 | 17% | 18% | 18% | 23% | 18% |

3.5 **Pronóstico de $PM_{2.5}$ incluyendo concentraciones de CO como dato de entrada**

3.5.1 **Introducción**

El análisis de las mediciones en las estaciones de monitoreo de la red MACAM II, ha mostrado una alta correlación entre las concentraciones de $PM_{2.5}$ y monóxido de carbono (CO). El gas CO es inodoro, incoloro, sin sabor, tóxico y un gas muy estable el cual es el producto de la combustión incompleta en los motores de los vehículos [131,132,133,134]. Cuando este gas se inhala se combina con la hemoglobina de la sangre e inhibe su capacidad de transportar oxígeno a los tejidos del cuerpo humano y de los organismos vivos. Se ha demostrado que la inhibición en un 50% de la hemoglobina en una persona le provoca la muerte [108]. Las

concentraciones en el aire de CO, en la ciudad de Santiago, se producen fundamentalmente por las emisiones de los motores de los vehículos que circulan por las calles.

Las autoridades de la salud (MINSAL) en Chile han establecido un límite de seguridad de 10 mg/m^3 (o 9 partes por millón [ppm]) para la concentración promedia de 8-h de CO (CO8), la cual está de acuerdo con la norma de la EPA (USA) [13].

Así, si las concentraciones de CO exceden la norma, el aire es considerado dañino para la salud de la población. El máximo diario de CO8 excede los 9ppm varias veces en el año en las estaciones de monitoreo, especialmente en aquellas estaciones que están situadas en la zona céntrica y oeste de la ciudad.

En un trabajo debido a Pérez [134] se reportó un modelo basado en una red neuronal tipo MLP consistente de tres capas, para pronosticar los valores máximos del promedio móvil de 8-h de las concentraciones de CO para el día de mañana.

Los datos de la capa de entrada fueron: CO, la humedad relativa promedio, la temperatura máxima, la velocidad del viento promedio y el máximo de CO del día.

3.5.2 Resultados comparativos obtenidos para pronóstico de $PM_{2.5}$ mediante algoritmo HCA sin incluir e incluyendo CO

En esta sección se estudia una posible corrección al modelo de pronóstico para $PM_{2.5}$ basado en el algoritmo HCA. Se pretende establecer si agregando como datos de entrada las concentraciones de CO a las 19:00 hrs. y a las 20:00 hrs. de hoy, mejorará o no el pronóstico para $PM_{2.5}$ del día de mañana.

En la tabla 3.17 se muestran los resultados de considerar las concentraciones de CO a las 19:00 hrs. y las 20:00 hrs. como información

adicional en los datos de entrada en el algoritmo HCA usado en la sección 3.3.

Tabla 3.17: Comparación entre el porcentaje de aciertos en el pronóstico de PM2.5 incluyendo y no incluyendo CO

| AÑO | CON CO | | | SIN CO | | | |
|------|------------|-------|------------|------------|-------|------------|-----|
| | Favorables | Total | % Aciertos | Favorables | Total | % Aciertos | |
| 2004 | A: | 20 | 20 | 100 | 20 | 20 | 100 |
| | B: | 31 | 48 | 65 | 27 | 48 | 56 |
| | C: | 29 | 52 | 56 | 29 | 52 | 56 |
| | D: | 17 | 20 | 85 | 14 | 20 | 70 |
| | TOTAL | 97 | 140 | 69 | 90 | 140 | 64 |
| 2005 | A: | 12 | 13 | 92 | 11 | 13 | 85 |
| | B: | 25 | 55 | 45 | 31 | 55 | 67 |
| | C: | 31 | 58 | 53 | 33 | 58 | 57 |
| | D: | 2 | 5 | 40 | 1 | 5 | 20 |
| | TOTAL | 70 | 131 | 53 | 76 | 131 | 58 |
| 2006 | A: | 11 | 13 | 85 | 11 | 13 | 85 |
| | B: | 30 | 62 | 48 | 34 | 62 | 55 |
| | C: | 24 | 63 | 38 | 25 | 63 | 40 |
| | D: | 7 | 13 | 54 | 9 | 13 | 69 |
| | TOTAL | 72 | 151 | 48 | 79 | 151 | 52 |
| 2007 | A: | 14 | 14 | 100 | 14 | 14 | 100 |
| | B: | 37 | 73 | 51 | 36 | 73 | 49 |
| | C: | 24 | 52 | 46 | 28 | 52 | 54 |
| | D: | 12 | 12 | 100 | 11 | 12 | 92 |
| | TOTAL | 87 | 151 | 58 | 89 | 151 | 59 |
| 2008 | A: | 19 | 19 | 100 | 19 | 19 | 100 |
| | B: | 45 | 66 | 68 | 42 | 66 | 64 |
| | C: | 22 | 44 | 50 | 21 | 44 | 48 |
| | D: | 13 | 24 | 54 | 8 | 24 | 33 |
| | TOTAL | 99 | 153 | 65 | 90 | 153 | 59 |

Puede observarse que para cada año considerado en el estudio en dos casos (2004 y 2008) el pronóstico mejoró al usar como dato de entrada las concentraciones de CO. En uno fue prácticamente el mismo (año 2007) y en dos fue adverso (años 2005 y 2006). En todo caso, para aciertos a la clase D, que es la más peligrosa para la salud, la mayor parte de las veces hubo una mejoría al incluir la información de CO.

Conclusiones Generales

“The scientist is not a person who gives the right answers, (s)he’s one who asks the right questions”

- Claude Lévi-Strauss

4.1 Conclusiones Generales y Perspectivas

4.1.1 Conclusiones Generales

Se demostró que las series temporales de material particulado fino PM_{2.5} son caóticas. Esto implica necesariamente que la predictibilidad de estas series, basada sólo en los datos de la misma serie, es bastante pobre, no mayor que tres horas.

En un primer intento de desarrollar un modelo de pronóstico, el estudio de series temporales no lineales encontró en una red MLP con una capa escondida, a la mejor red para hacer pronóstico de concentraciones de material particulado fino PM_{2.5} algunas horas de anticipación, la que se obtuvo a través de la determinación de la dimensión de encaje y el tiempo de rezago.

Se puede afirmar que cuando efectos no lineales no son demasiado importantes en la modelación, las redes de mult capas no son significativamente mejores que un algoritmo lineal. Sin embargo, como ocurrió en este caso, cuando estos efectos no lineales pasan a ser importantes, las redes de mult capas son mejores en cuanto a su capacidad de predicción, respecto de los modelos lineales.

Si bien el modelo basado en una red neuronal resultó más preciso que un modelo básico de persistencia, debido en parte a la presencia de caos, los resultados no son lo suficientemente buenos para ser usados en un modelo operacional de pronóstico. Se estimó conveniente, para aumentar la capacidad de predicción, incorporar información adicional, tal como datos meteorológicos y de emisiones.

Es así como se estudiaron tres tipos de modelos estadísticos, uno lineal, una red neuronal tipo MLP y un algoritmo de clusters en los cuales los datos de entrada fueron las concentraciones horarias de $PM_{2.5}$ de las 19:00 y las 20:00 del día actual, amplitudes térmicas constatadas y pronosticadas y un factor de ventilación (PMCA) enfocados a pronosticar el rango en que se encontrará el máximo del promedio de 24 horas de $PM_{2.5}$ durante el día siguiente. Otra característica es que los tres modelos integran la información de las cuatro estaciones de monitoreo en un único algoritmo. Se lograron tasa de acierto bastante aceptable para un modelo operacional para la gestión de la calidad del aire en la ciudad. El algoritmo de clusters HCA parece ser más preciso para detectar situaciones de alta concentración de contaminantes. La razón de ello es que el método HCA no tiene la tendencia a suavizar los valores extremos comparado con el método lineal y el modelo MLP.

La información de emisiones está contenida implícitamente en las concentraciones horarias de $PM_{2.5}$ de las 19:00 y 20:00. Esta información se complementó incorporando datos de concentraciones de CO para estas mismas horas, considerando que este gas es emitido principalmente por fuentes móviles.

Debido al carácter caótico de las series de $PM_{2.5}$, es de esperar que los denominados modelos deterministas no consigan un grado muy alto de precisión. Esto porque ellos necesitan información detallada de las emisiones de la ciudad, la cual por su complejidad sólo puede ser ingresada

después de calcular promedios espaciales y temporales, lo cual deja fuera una consideración detallada de las condiciones iniciales.

4.1.2 Perspectivas

Para tomar en cuenta el efecto de las fuentes móviles al utilizar el algoritmo HCA, un estudio que puede llevarse a cabo es aquel que incluya como entrada las concentraciones de CO del día de mañana, suponiendo que estas se pueden pronosticar de forma independiente, lo cual con seguridad mejoraría el pronóstico de $PM_{2.5}$.

Es importante poder estudiar y diseñar nuevos modelos, distintos a los ya estudiados, para hacer pronósticos cada vez más precisos y poder detectar con buena anticipación situaciones de peligro para la población.

Un modelo combinado (híbrido) HCA-MLP podría ser una opción de tal manera que para los valores intermedios se utilice el valor pronosticado por el MLP y para valores extremos se asignen los rangos de acuerdo al HCA.

También sería útil e interesante poder determinar la equivalencia (si es que existe) entre el factor PMCA y el factor de Pasquill-Guifford, para ser utilizado en un modelo de dispersión de contaminante $PM_{2.5}$ cuando este es emitido principalmente por fuentes móviles.

Relacionado con esto último se necesita determinar cómo se relaciona el ángulo de la dirección del viento con la ecuación de dispersión para una fuente lineal de contaminante, debido a fuentes móviles.

APÉNDICE

A.1 Resumen de valores de los componentes contaminantes para el inventario de emisiones de los años 2000 y 2005

En la Tabla A.1 se presentan un resumen de los valores de los componentes contaminantes para el inventario de emisiones del año 2000. En esta ocasión los valores en porcentaje se determinan hacia la derecha para cada tipo de fuente. Así, puede apreciarse que para las fuentes móviles el mayor aporte proviene del contaminante CO (69,6%). En cambio, para las fuentes fijas está en primer lugar el contaminante COV con un 47,9%, NH₃ con 24,2% y CO con un 13,6%.

Tabla A.1: Resumen Inventario Emisiones año 2000 (%)

| | PM ₁₀ | CO | NO _x | COV | SO _x | NH ₃ | Total |
|---------------|------------------|------|-----------------|------|-----------------|-----------------|-------|
| Estacionarias | 2.2 | 13.6 | 6.9 | 47.9 | 5.1 | 24.2 | 100.0 |
| Móviles | 1.0 | 69.6 | 18.5 | 9.8 | 0.9 | 0.4 | 100.0 |

En la Tabla A.2 se presentan un resumen de los valores de los componentes contaminantes para el inventario de emisiones del año 2005. En esta ocasión los valores en porcentaje se determinan hacia la derecha para cada tipo de fuente. Así, puede apreciarse que para las fuentes móviles el mayor aporte proviene del contaminante CO (79,6%). En cambio, para las fuentes fijas está en primer lugar el contaminante COV con un 49,3%, NH₃ con 18,9% y CO con un 14,1%.

Tabla A.2: Resumen Inventario Emisiones año 2005 (%)

| | PM ₁₀ | CO | NO _x | COV | SO _x | NH ₃ | Total |
|---------------|------------------|------|-----------------|------|-----------------|-----------------|-------|
| Estacionarias | 1.7 | 14.1 | 8.2 | 49.3 | 7.8 | 18.9 | 100.0 |
| Móviles | 0.7 | 79.6 | 11.9 | 7.3 | 0.1 | 0.4 | 100.0 |

A.2 Exponentes de Lyapunov

En la Tabla A.3 se presentan los exponentes de Lyapunov para las tres series temporales de la red MACAM I (período 1994-1996), de la estación de monitoreo Pudahuel (EMO) y para dos valores de dimensión de encaje: 4 y 5. se puede observar que si se compara con los valores de la Tabla 3.1 ($d_e=5$) y la estación EMO, los valores de la Tabla A.3 están dentro del rango calculado para las cuatro estaciones de monitoreo y entre los años 2000 a 2007.

Tabla A.3: Exponente de Lyapunov (1/hora) para período de otoño-invierno

| | 1994 | 1995 | 1996 |
|-------------|---------------|---------------|---------------|
| Dim. Encaje | de=4 | de=4 | de=4 |
| Pudahuel | 0.663 ± 0.029 | 0.683 ± 0.029 | 0.555 ± 0.027 |
| Dim. Encaje | de=5 | de=5 | de=5 |
| Pudahuel | 0.445 ± 0.022 | 0.452 ± 0.022 | 0.376 ± 0.020 |

BIBLIOGRAFÍA

- [1] Morales, R.G.E. (editor), *Contaminación Atmosférica Urbana* (Editorial Universitaria, Santiago, Chile, 2006).
- [2] Molina, M.J. & Molina, L.T., *Megacities and atmospheric pollution*, *Journal of the Air & Waste Management Association*, **54**, pp. 1047-3289, (2004). Disponible en español con el título *Megaciudades y Contaminación Atmosférica*(2007) en: <http://www.abstractatus.com/espanol/mariomolina/>
- [3] A.P. Sokolov, P.H. Stone, C.E. Forest, R. Prinn, M.C. Sarofim, M. Webster, S. Paltsev, C.A. Schlosser, D. Kicklighter, S. Dutkiewicz, J. Reilly, C. Wang, B. Felzer, H.D. Jacoby. *Probabilistic forecast for 21st century climate based on uncertainties in emissions (without policy) and climate parameters*. *Journal of Climate* preprint (2009): 1 DOI: [10.1175/2009JCLI2863.1](https://doi.org/10.1175/2009JCLI2863.1)
- [4] Climate Change 2007, *Climate change impacts, adaptations and vulnerability*. IPCC WGII Fourth Assessment Report, 2007.
- [5] R. N. Colville et al., *The transport sector as a source of air pollution*, *Atmospheric Environment*, 35, pp. 1537-1565 (2001).
- [6] CENMA, *Análisis comparativo de condiciones meteorológicas asociadas a episodios de contaminación atmosférica, en Santiago, durante los períodos de otoño invierno de 1997, 1998, 1999 y 2000*, CENMA, proyecto meteorología y calidad del aire (Enero de 2001).
- [7] M. Foster y D.L. Costa eds. *Air Pollutants and the Respiratory Tract. Lung Biology in Health and Disease*, Vol. 204, (Taylor & Francis, Boca Raton, 2nd ed., USA, 2005). M. Lippmann: Air pollution, Public health and Regulatory Consideration, chap. 12 pp. 405-441.
- [8] J. Schwartz, F. Laden y A. Zanobetti, *The concentration-response relation between PM_{2.5} and daily deaths*, *Environmental Health Perspectives*, **110** (10), pp. 1025-1029, October (2002).
- [9] W. Dab, et al . *Short term respiratory health effects of ambient air pollution: results of the APHEA project in Paris*, *Journal of Epidemiology and Community Health*. **50** (Suppl. 1), pp. S42-S46 (1996).
- [10] O.G. Raabe, *Respiratory exposure to air pollutants* en: Swift DL, Foster WM. *Air pollutants and the respiratory tract*. New York, NY: Marcel Dekker, 1999:39-73.
- [11] Gobierno de Chile CONAMA-Región Metropolitana, *Actualización del inventario de emisiones de contaminantes atmosféricos en la región metropolitana 2005*. <http://www.conama.cl/rm>.
- [12] J. Rutllant, R. Garreaud. *Meteorological air pollution potential for Santiago, Chile: towards an objective episode forecasting*. *Environmental Monitoring and Assessment*, 35, pp. 223-244 (1995).
- [13] <http://www.epa.gov/>

- [14] OMS, *Guías de calidad del aire de la OMS relativas al material particulado, el ozono, el dióxido de nitrógeno y el dióxido de azufre. Actualización mundial 2005. Resumen de evaluación de riesgos.* (OMS, 2006). WHO, 2006. Air Quality Guidelines. Global Update 2005. Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide.
- [15] Ilabaca, M. et al, *Association between levels of fine particulate and emergency visits for pneumonia and other respiratory illnesses among children in Santiago, Chile*, Journal of the Air and Waste Managements Association. **49**, pp. 154-163 (1999).
- [16] R. Katz, M.I. Benítez. *Uso de un sistema de pronóstico para el control de la calidad del aire de Santiago.* Estudios Públicos, 77 (2000).
- [17] J. Cáceres T. tesis de memoria, (Primer) Modelo de Pronóstico para Santiago (CONAMA-RM, 1997).
- [18] J.C. Cassmassi, *Improvement of the forecast air quality of the knowledge of the local meteorological conditions in the metropolitan region.* Technical report Number 2. National Commission of the Environment Metropolitan Region, Santiago, Chile. 1999.
- [19] P. Pérez, *Air quality forecasting in a large city*, Air pollution XVI: WIT Transactions on Ecology and the Environment, vol. 116, editors C.A. Brieva y J.W.S. Longhurst, Wessex Institute of Technology, WIT Press, 2008, Skiathos, Grecia.
- [20] AIRVIRO: <http://146.155.5.120/iairviro/IAirviroMain.htm>
- [21] H. Abarbanel, *Analysis of observed chaotic data*, Institute for Nonlinear Science, (Springer-Verlag Telos, 1996, 1st ed.)
- [22] P. Bryant, R. Brown y H.D.I. Abarbanel, *Lyapunov exponents from observed time series*, *Phys. Rev. Letters*, **65**(13), pp. 1523-1526 (1990).
- [23] R. Brown, P. Bryant, y H.D.I. Abarbanel, *Computing the Lyapunov spectrum of a dynamical system from an observed time series*, *Phys. Rev. A*, **43**(6), pp. 2787-2806 (1991).
- [24] M.B. Kennel, R. Brown y H.D.I. Abarbanel, *Determining embedding dimensions for phase-space reconstruction using geometrical constructions*, *Phys. Rev. A*. **45**, pp. 3403-3411 (1992).
- [25] A. M. Fraser & H. L. Swinney, *Independent coordinates for strange attractors from mutual information*, *Phys. Rev. A.*, **33**, 1134-1140 (1986).
- [26] A.M. Fraser, *Information and entropy in strange attractors*, *IEEE Transactions on Information Theory*, **35**(2) pp. 245-262, (1989).
- [27] G. Salini & P. Pérez, *Estudio de series temporales de contaminación ambiental mediante técnicas de redes neuronales artificiales*, V Encuentro de Modelos Físicos y Matemáticos en Ingeniería, EMFIMIN 2005, Universidad de Santiago de Chile, 10-11 de Noviembre, pp.25, (2005).
- [28] P. Pérez, A. Trier and J. Reyes, *Prediction of $PM_{2.5}$ concentrations several hours in advance, using neural networks in Santiago, Chile*, *Atmospheric Environment*, **34**, pp. 1189-1196 (2000).
- [29] M. Khare & S. M. Nagendra, *Artificial neural networks in vehicular pollution modelling*, (Springer-Verlag, Berlin Heidelberg, 2007).

- [30] A. Sfetsos & C. Siriopoulos, *Time series forecasting with a hybrid clustering scheme and pattern recognition*, IEEE Trans. Syst. Man Cyber. - Part A Syst. and Man, **34**, (3) May, pp. 399-404 (2004).
- [31] Harnett y Murray. *Introducción al análisis estadístico*. Cap. 13: Series cronológicas y números índices. (Addison-Wesley Iberoamérica, México, 1987).
- [32] D. Peña S. *Estadística, modelos y métodos. Vol. 1 Modelos lineales y series temporales*. (Alianza Universidad Textos, 2da. Edición, Alianza Editorial S. A., Madrid, 1995).
- [33] J. Fao y Q. Yao. *Nonlinear Time Series. Nonparametric and parametric methods*. (Springer Science + Business media, Inc., New York, USA, 2005).
- [34] Box G.E.P. & G.M. Jenkins *Time Series Analysis: Forecasting and Control*. (Holden Day, San Francisco, USA, 1970).
- [35] A.K. Palet y D. Popovic. *Computacional Intelligence in Time Series*. (Springer- Verlag London Limited, USA, 2005).
- [36] B. Pilgram, K. Judd y A. Mess. *Modelling the dynamics of nonlinear time series using canonical variate analysis*. Physica D, **170**, 103-117 (2002).
- [37] M. Small. *Applied nonlinear time series analysis: applications in physics, physiology and finance*. (World Scientific series on Nonlinear Science series A vol. 52, World Scientific Publishing Co., Singapur, 2005).
- [38] H. Kantz y T. Schreiber. *Nonlinear time series análisis*. (Cambridge University Press, 2nd ed., Cambridge, 2004).
- [39] D.M. Walker, y N.B. Tuffillaro, *Phase Space reconstruction using input-output time series data*, Phys. Rev. E, **60**(4), pp. 4008-4013, (1999).
- [40] F. Takens. *Detecting strange attractors in turbulence*; en D.A. Rand, L.S. Young (eds.), *Dynamical Systems and Turbulence*, vol. 898. (Springer Berlin, pp. 365-381, 1981).
- [41] H.D.I. Abarbanel, R. Brown, J.J Sidorowich y L.S. Tsimring. *The analysis of observed chaotic data in physical systems*. Rev. Mod. Phys., **65** (4), pp. 1331-1392 (1993).
- [42] C. Rhodes, y M. Morari, *False-nearest-neighbors algorithm and noise-corrupted time series*, Phys. Rev. E, **55**(5), pp. 6162-6170 (1997).
- [43] J.L. Casti, *Alternate Realities: mathematical models of nature and man*, (John Wiley & Sons, Inc., New York, 1989).
- [44] J. C. Sprott, *How common is Chaos?* Phys. Lett. A, **173**, pp. 21-24 (1993).
- [45] K.T. Alligood, T.D. Sauer, y J.A. Yorke, *Chaos, an Introduction to Dynamical Systems*, (Springer-Verlag, New York, 1996).
- [46] H.G. Schuster, *Deterministic Chaos: An introduction*. (2nd ed., VCH, Weinheim, 1989).
- [47] R.C. Hilborn, *Chaos and nonlinear dynamics, an introductions for scientists and engineers*, 2nd ed., (Oxford University Press, New York, 2000).
- [48] S. Wiggins, *Introduction to applied nonlinear dynamical systems and chaos*, (Springer-Verlag, 2nd ed., New York, 2003).
- [49] W.H. Steeb. *The Nonlinear Workbook*, (World Scientific Publishing Co., 3rd ed., Singapur, 2005).

- [50] I. Prigogine, *Las Leyes del Caos*, (Drakontos Crítica, Grijaldo Mondadori, Barcelona, España, 1997).
- [51] R. V. Solé y S.C. Manrubia, *Orden y Caos en Sistemas Complejos*, (Ediciones UPC, Barcelona, España, 1996).
- [52] J. C. Sprott, *Chaos and Time-Series Analysis*. (Oxford University Press, 2003).
- [53] G.G. Malinietski, *Fundamentos matemáticos de la Sinérgica. Caos, estructuras y simulación por ordenador*. (Editorial URSS, Moscú, 2005).
- [54] E.N. Lorenz, *Irregularity: A fundamental property of the atmosphere*, *Tellus A*, **36**, 98 (1984).
- [55] H.O. Peitgen, H. Jurgens y D. Saupe, *Chaos and Fractal New Frontiers of Science*, 2nd ed., (Springer-Verlag, New York, 2004).
- [56] J. Gao, Y. Cao, W. Tung y J. Hu. *Multiscale análisis of complex time series: integration of chaos and random fractal theory, and beyond*. (Wiley Interscience, John Wiley & Sons, Inc., New Jersey, USA, 2007).
- [57] R.M. May, *Simple mathematical models with very complicated dynamics*, *Nature*, **261**, pp. 459-467 (1976).
- [58] G.M. Zaslavsky, *Chaotic dynamics and the origin of statistical laws*, *Phys. Today*, **8**, (1999).
- [59] J. Holzfuss, y W. Lauterborn, *Liapunov exponents from a time series of acoustic chaos*, *Phys. Rev. A*, **39**, pp. 2146-2152 (1989).
- [60] D. Vassiliadis, A.S. Sharma, y K. Papadopoulos, *Lyapunov exponent of magnetospheric activity from AL time series*, *Geophys. Res. Letters*, **18**(8), pp. 1643-1646 (1991).
- [61] A.M. López Jiménez, C. Camacho Martínez vara de Rey y A.R. García Torres, *Effect of parameter calculation in direct estimation of the Lyapunov exponent in short time series*, *Discrete Dynamics in Nature and Society*, **7**(1), pp. 41-52 (2002).
- [62] M. Sato, y Y. Sawada, *Measurements of the Lyapunov spectrum from a chaotic time series*, *Phys. Rev. Letters*, **55**(10), pp. 1082-1085 (1985).
- [63] T. Tanaka, K. Aihara, y M. Taki, *Lyapunov exponents of random time series*, *Phys. Rev. E*, **54**(2), pp. 2122-2124, (1996).
- [64] J.P. Eckmann, S.O. Oliffson Kamphorst, D. Ruelle y S. Ciliberto, *Liapunov exponents from time series*, *Phys. Rev. A*, **34**, pp. 4971-4979 (1986).
- [65] Y.A. Kravtsov (ed.), *Limits of Predictability*. (Springer-Verlag, Berlin, Heidelberg, 1993).
- [66] J. Liu, T. Zhang y S. Yu. *Chaotic phenomenon and the maximum predictable time scales of observation series of urban hourly water consumption*. *Journal of Zhejiang University Science*, **5** (9) pp. 1053-1059 (2004).
- [67] R.S. Huang, *Chaos and its applications* (Wuan university Press, pp. 115-135, Wuhan, 2000).
- [68] A. Wolf, J.B. Swift, H.L. Swinney y J.A. Vastano. *Determining Lyapunov exponent from time series*. *Physica D*, **16**, 285-317 (1985).

- [69] M.T. Rosenstein, J.J. Collins y C.J. De Luca. *A practical method for calculating largest Lyapunov exponents from small data sets*, *Physica D*, **65**, pp. 117-134 (1993).
- [70] A. Komori *et al*, *Correlation dimension and largest Lyapunov exponent for broadband edge turbulence in the compact helical system*, *Phys. Rev. Letters*, **73**(5), pp. 660-663 (1994).
- [71] K.E. Chlouverakis y J.C. Sprott, *A comparison of correlation and Lyapunov dimensions*. *Physica D*, **200**, pp. 156-164 (2005).
- [72] P. Grassberg, P. y I. Procaccia, *Characterization of strange attractors*, *Phys. Rev. Letters*, **50**(3), pp. 346-349 (1983).
- [73] M. Small. *Detecting determinism in time series: the method of surrogate data*, *IEEE transactions on circuit and systems-I: Theory and applications*, **50**(5) pp. 663-672, (2003).
- [74] J.C. Principe, A. Rathie, y J.M. Kuo, *Prediction of chaotic time series with neural networks and the issue of dynamic modeling*, *Inter. Journal of Bifurcation and Chaos*, **2**(4) pp. 989-996 (1992).
- [75] E.R. Kandel & J-H- Schwartz. *Principles of neural science*. (2nd Edition, Elsevier Science Pub. CO, Inc., New York, 1985).
- [76] S.W.Kuffler & J.G. Nicholls, *De la neurona al cerebro*. (1ra edición, Editorial Reverté, Barcelona, España, 1982).
- [77] J.A. Freeman y D.M. Skapura, *Redes Neuronales, Algoritmos, aplicaciones y técnicas de programación*. (Addison-Wesley/Diaz de Santos, Wilmington, Delaware, USA, 1993).
- [78] B. Martín del Río y A. Sanz. *Redes Neuronales y Sistemas Borrosos*. (Editorial RA-MA, Madrid, 2001).
- [79] J.R. Hiler y V.J. Martínez. *Redes Neuronales Artificiales. Fundamentos, modelos y aplicaciones*. (1ra edición, Addison- Wesley Iberoamericana S.A. y RA-MA, Wilmington, Delaware, USA, 1995).
- [80] G.A. Salini, *Almacenamiento de Información en Redes Neuronales Acopladas*, tesis de Magister, Departamento de Física, Universidad de Santiago de Chile, Santiago, 2004.
- [81] W.S Mc Culloch & W.A. Pitts, *A logical calculus of the ideas inmanente in nervous activity*, *Bull. Math. Biophys.*, **5**, pp. 115-133 (1943).
- [82] D.O. Hebb. *The organization of behavior: a neurophysiological study* (John Wiley, New York, NY, 1949)
- [83] F. Rosenblatt. *Principles of Neurodynamics*. (Spartan Books, New York, 1962).
- [84] D. Rumelhart, J. Mc Clelland & PDP Research Group, *Parallel distributed processing*. Vol. 1,2. (MIT Press, Cambridge, Mass. 1988).
- [85] I. Guyon, *Neural Networks and applications tutorial*, *Physics Reports*, **20**, pp. 215-259 (1991).
- [86] J.J. Hopfield. *Neural Networks as physical Systems with emergent computacional habilities*. *Proc. Natl. Acad. Sci. USA*, **79**, pp. 2554 (1982).
- [87] M.L. Minsky & S. Pappert, *Perceptrons* (MIT Press, Cambridge, Mass. 1969).
- [88] D.J. Amit, H. Gutfreund y H. Sompolinsky, *Spin-Glass models of neural Networks*, *Phys. Rev. A*, **32**, pp. 100-1018, (1985).

- [89] D. R. Hush y B.G. Horne, *Progress in supervised neural Networks*, IEEE Signal Processing Magazine, January, pp. 8-39 (1998).
- [90] T. Kohonen, *Self-Organized formation of topologically correct feature maps*, Bio. Cybern., **43**, pp. 59-69 (1982).
- [91] H. Ritter, T. Martínez y K. Schulten, *Neural Computation and Self-Organizing Maps*, (Addison-Wesley Pub. Co., Ma, USA, 1992).
- [92] B. Kosko, *Neural Networks and Fuzzy Systems: a dynamical systems approach to machine intelligence*. (Prentice Hall Inc, New jersey, USA, 1992)
- [93] I. Matus y P. Pérez, *Neural networks with state-dependent synapses*, Phys. Rev. A **41**, pp. 7013-7019 (1990).
- [94] P. Pérez y G.A. Salini. *Storage of words in a neural network*. Phys. Lett. A, **181**, pp. 61 (1993).
- [95] P.J. Werbos. *Beyond Regression: new tools for prediction and analysis in the Behavioral Sciences*. PHD thesis, Harvard University (1974).
- [96] B. Parker, *Learning Logic*. Technical Report TR-47, Center for computation research in economics management science. (MIT Press, Cambridge, MA, 1985).
- [97] W.L Martínez y A.R. Martínez. *Exploratory data analysis with Matlab*. (Chapman & Hall/CRC Press UK, London, 2005).
- [98] D. Maravall Gómez-Allende. *Reconocimiento de formas y visión artificial. Capítulo 6: Algoritmos de agrupación de clases (clustering)*. (Addison-Wesley Iberoamericana y RA MA Wilmington, Delaware, USA, 1994).
- [99] D. Vlachogiannis y A. Sfetsos. *Time series forecasting of hourly PM10 values: model intercomparison and the development of localized linear approaches*, en Proceedings of Air Pollution XIV: Fourteenth International Conference on Modelling, Monitoring and Management of Air Pollution, pp. 85-94, (edited by J.W.S. Longhurst, University of the West of England, UK and C.A. Brebbia, Wessex institute of Technology, UK., WIT Press, 2006).
- [100] J. Dudhia, G.A. Grell, Y. Guo, K. Manning, W. Wang, y J. Chiszar, *PSU/NCAR Mesoscale Modelling System Tutorial Class Notes and User's Guide: MM5 modelling system version 3, Mesoscale and Microscale meteorology Division*, (NCAR, Boulder, Colorado, USA, 2000).
- [101] Grell, G., J. Dudhia, and D. Stauffer, *A description of the fifth generation Penn State/NCAR mesoscale model (MM5)*. (NCAR Technical Note, CAR/TN 398+ STR, 1994).
- [102] V. Kotrini y K. Lagouvardos, *MM5 fine scale simulations for the support of Athens 2004 Olympic Games: evaluation of the first year of operational activities*, (MM5 Workshops, División 3, 2002).
- [103] D.W. Byun y J.K.S. Ching (Eds.). *Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling system*, (US EPA Report No. EPA/600/R-99/030, Office of research and Development, Washington, DC., 1999).
- [104] T.L. Otte, et al. *Linking the ETA model with the community multiscale air quality (CMAQ) modelling system to build a national air quality forecasting system*. Weather and Forecasting. American Meteorological Society, **20**(3), pp. 367-385 (2005).

- [105] O. Jorba, T. Loridan, P. Jiménez-Guerrero, C. Pérez, y J.M. Baldasano, *Development of a linkage between the Advanced Research Weather and Forecasting model and the CHIMERE Chemistry-transport Model*, (Environmental Modelling and Software, 2008).
- [106] Z.I. Janjic, J.P. Gerrity Jr. y S. Nickovic. *An alternative approach to nonhydrostatic modelling*. Monthly Weather Review, **129**, pp. 1164-118 (2001).
- [107] X. Tie, S. Madronich, G. Li, Z. Ying, R. Zhang, A.R. García, J. Lee-Taylor y Y. Liu. *Characterization of chemical oxidants in Mexico city: a regional chemical dynamical model (WRF-Model) study*. Atmospheric Environment, **41**, pp. 1989–2208, (2007) Elsevier.
- [108] L.A. Cifuentes, J. Vega, K. Köpfer y L.B. Lave. *Effect of the fine fraction of particulate matter versus the coarse mass and other pollutants on daily mortality in Santiago, Chile*. Journal of Air and Waste Management Association **50**, pp. 1287–1298 (2000).
- [109] J. Schwartz, L.M. Neas, 2000. *Fine particles are more strongly associated than coarse particles with acute respiratory health effects in school children*. Epidemiology **11**, pp. 6–10 (2000).
- [110] R.J. Klemm, R.M. Mason, C.M. Heilig, L.M. Neas y D.W. Dockery, *Is daily mortality associated specifically with fine particles? Data reconstruction and replication of analyses*. Journal of Air and Waste Management Association **50**, pp. 1215–1222 (2000).
- [111] R.D. Brook, J.R. Brook, B. Urch, R. Vincent, S. Rajagopalan y F. Silverman. *Inhalation of fine particle air pollution and ozone cause acute vasoconstriction in healthy adults*. Circulation **105**, pp. 1534–1536 (2002).
- [112] C.A. Pope III, R.T. Burnet, M.J. Thun, E. E. Calle, D. Krewski, K. Ito y G.D. Thurston. *Lung cancer, cardiopulmonary mortality and long term exposure to fine particulate air pollution*. Journal of the American Medical Association **287**, pp. 1132–1141 (2002).
- [113] P. Pérez y J. Reyes, *Prediction of maximum of 24-h average of PM10 concentrations 30 h in advance in Santiago, Chile*, Atmospheric Environment, **36**, pp. 4555-4561 (2002).
- [114] P. Pérez y J. Reyes, *Prediction of particulate air pollution using neural techniques*, Neural Computation & Application, **10**, pp. 165-171, (2001).
- [115] J. Reyes y P. Pérez, *Predictibilidad del material particulado PM2.5 en Santiago de Chile utilizando técnicas de modelación de sistemas dinámicos y redes neuronales*, Contribuciones Científicas y Tecnológicas, ACB, N° 121, pp. 107-108, diciembre (1998).
- [116] P. Pérez y P. Rodríguez, *A model for PM_{2.5} forecasting in Santiago, Chile*. en Proceedings of Air Pollution XIV: Fourteenth International Conference on Modelling, Monitoring and Management of Air Pollution, pp. 95-99, (edited by J.W.S. Longhurst, University of the West of England, UK and C.A. Brebbia, Wessex institute of Technology, UK., WIT Press 2006).
- [117] P.J. Silva, E.L. Vawdrey, M. Crbett, M. Erupe. *Fine particle concentrations and composition during winter time inversions in Logan, Utah, USA*. Atmospheric Environment **41**, pp. 5410–5422 (2007).

- [118] J.B. Ordieres, E.P. Vergara, R.S. Capuz y R.E. Salazar, *Neural network prediction model for fine particulate matter (PM_{2.5}) on the US-Mexico border in El Paso (Texas) and Ciudad Juarez (Chihuahua)*, Environmental Modelling & Software **20**, pp. 547-559, (2005).
- [119] S. Thomas y R.B. Jacko. *Model for forecasting expressway fine particulate matter and carbon monoxide concentrations: application of regression and neural network models*. Journal of Air and Waste Management Association **57**, pp. 480–488 (2007).
- [120] G.A. Salini y P. Pérez. *Estudio de series temporales de contaminación ambiental mediante técnicas de redes neuronales artificiales*, EMFIMIN V Encuentro de modelos físicos y matemáticos en ingeniería, USACH, Santiago, Chile, modalidad presentación oral, pp. 25, 10 al 11 de Noviembre de 2005.
- [121] C-K Lee y S-C Lin, *Chaos in air pollutant concentration (APC) Time series*, Aerosol and Air Quality research, **8**,(4) pp. 381-391 (2008).
- [122] M. Lanfredi y M. Macchiato, *Searching for low dimensionality in air pollution time series*, Europhysics letters, **40**(6), pp. 589-594 (1997).
- [123] V.N. Khokhlov, A.V. Glushkov, N.S. Loboda y Y.Y. Bunyakova *Short-range forecast of atmospheric pollutants using non-linear prediction method*. Atmospheric Environment, **42**, pp. 7284–7292, (2008) Elsevier.
- [124] A.K. Gautam, A.B. Chelani, V.K. Jain, y S. Devotta, *A new scheme to predict chaotic time series of air pollutant concentrations using artificial neural network and nearest neighbour searching*, Atmospheric Environment, **42**, pp. 4409-4417 (2008).
- [125] G.A. Salini y P. Pérez. *Fine particulate matter forecasting in Santiago, Chile*. Proceeding of 26th International Symposium on Forecasting, pp. 122,123, oral presentation, Santander, Spain, June 10-14 (2006).
- [126] G.A. Salini y P. Pérez. *Estudio de series temporales de contaminación ambiental mediante técnicas de redes neuronales artificiales*, Revista Chilena de Ingeniería (SCielo), **14**, N° 3, pp. 284-290, 2006.
- [127] G.A. Salini *Pronosticando series temporales no lineales varios pasos en adelante mediante técnicas no lineales y redes neuronales artificiales* EMFIMIN VI Encuentro de modelos físicos y matemáticos en ingeniería, USACH, Santiago, Chile, modalidad presentación oral, pp. 71-72, 15 al 16 de Noviembre de 2007.
- [128] G.A. Salini y P. Pérez. *Using three methods to forecast PM_{2.5} in a large city*. Primer Encuentro de Radio-Astronomía y Meteorología en Valparaíso, con presentación en forma oral. Hotel Gala Viña del Mar, 25 de Septiembre de 2008.
- [129] G.A. Salini y P. Pérez *PM_{2.5} forecasting one day in advance in Santiago, Chile*. Proceedings XVI Simposio Chileno de Física, UTFSM, Valparaíso, pp. 240, 12-14 Nov, 2008.
- [130] P. Pérez & G.A. Salini. *PM_{2.5} forecasting in a large city: Comparison of three methods*, Atmospheric Environment, **42**, pp. 8219–8224, (2008) Elsevier.
- [131] J. Varon, P. Marik, *Carbon monoxide poisoning* Disponible en: <http://www.ispub.com/journals/OJEICM/vOL1n2/CO.htm>

- [132] K.Wark, C.F.Warner and W. T. Davis, *Air Pollution its origin and control*, (Addison-Wesley, 3rd ed., 1997), Chap. IV.
- [133] M. Bady et al., *Comparative study of concentration and distribution of CO and NO in an urban area: Gaussian plume model and CFD analysis*, Air Pollution XIV, WIT Press, **86**, pp. 55-64 (2006).
- [134] P. Pérez, R. Palacios y A. Castillo, *Carbon monoxide concentration forecasting in Santiago*, Air & Waste Managa. Assoc., **54**, pp. 908-913, (2004).