UNIVERSIDAD DE SANTIAGO DE CHILE
FACULTAD DE CIENCIAS
DEPARTAMENTO DE FÍSICA

PRONÓSTICO DE CONCENTRACIONES DE MATERIAL PARTICULADO FINO $PM_{2.5}$ EN LA ATMOSFERA DE SANTIAGO.

SEBASTIÁN RODRIGO ORELLANA REYES

Profesor guía: Dr. Patricio Pérez

Trabajo de titulación presentado a la Facultad de Ciencia,
en cumplimiento parcial de los requisitos exigidos para optar
al título de Ingeniero Físico

Santiago - Chile
2011

PRONÓSTICO DE CONCENTRACIONES DE MATERIAL PARTICULADO FINO $PM_{2.5}$ EN LA ATMOSFERA DE SANTIAGO.

SEBASTIÁN RODRIGO ORELLANA REYES

Este trabajo de titilación 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 Perez Jara ……

Dra. Marina Stepanova SSA. ……

Dr. Ernesto Gramsh Labra ……

Dr. Carlos Balocchi Carreño ……

……
Director Departamento de Física
Mg. Bernardo Carrasco Puentes

PRONÓSTICO DE CONCENTRACIONES DE MATERIAL PARTICULADO FINO $PM_{2.5}$ EN LA ATMOSFERA DE SANTIAGO.

SEBASTIÁN RODRIGO ORELLANA REYES

Trabajo de titulación presentado a la Facultad de Ciencia, en cumplimiento parcial de los requerimientos exigidos para optar al titulo de Ingeniero Físico

Santiago - Chile
2011

AGRADECIMIENTOS

Durante el periodo de la universidad algo que nos marca de gran manera son las amistades que creamos en ella, por esto debo agradecer a mis amigos Francisco Correa, Guillermo Maturana, Beatriz Ramírez, Pedro Álvarez, Beatriz Gallardo, Catalina Torres, Ángela Jiménez y a los muchos que no menciono, por su amistad durante nuestro paso por el Departamento de Física, esta amistad ayudó a formar la persona que soy actualmente.

A todos los profesores que tuve en el trascurso de mi carrera, cada uno de los cuales ayudó en mi formación profesional, agradezco de forma especial a Patricio Perez, por darme la oportunidad de realizar esta tesis y su guía durante el desarrollo de ésta.

Agradezco a mi familia por su incondicional apoyo y muy especialmente a mi novia Beatriz.

PROLOGO

Problema: “Contaminación Ambiental”

¿Por que es problema?.....

¿Que lo produce?.....

¿Que variables intervienen?.....

Para poder dar solución a un problema lo primero es conocerlo a cabalidad.

Una forma de enfrentar el problema es ser capaz de predecirlo, con el fin de tomar las medidas necesarias para disminuirlo o eliminarlo.

TABLA DE CONTENIDOS

ÍNDICE DE TABLAS

CAPITULO 1

CAPITULO 2

CAPITULO 3

CAPITULO 4

CAPITULO 5

ÍNDICE DE FIGURAS

CAPITULO 1

CAPITULO 2

CAPITULO 4

RESUMEN

El material particulado en suspensión en el aire, de tamaño menor a 2.5 micrones (PM2.5), también denominado material particulado fino ha cobrado mucho interés el ultimo tiempo debido a sus nocivos efectos para la salud.

Mientras se pueda predecir la contaminación atmosférica con exactitud, el gobierno será capaz de tomar las medidas adecuadas para preveer estas altas concentraciones.

Este trabajo se enfoca en el desarrollo de nuevos métodos de predicción de concentraciones de material particulado 2.5 ($PM_{2.5}$), que a lo largo aportarían a disminuir niveles de contaminación en la capital.

Actualmente la Universidad de Santiago de Chile tiene en operación un modelo de pronóstico de concentraciones de PM2.5 basado en redes neuronales construido en base a los resultados de la reciente tesis de doctorado de Giovanni Salini [9]. Esta tesis trata de un estudio orientado a mejorar dicho modelo. Estudiando y desarrollando variaciones de este mismo modelo se pueden mejorar los pronósticos.

En primer lugar se estudió la eficiencia del modelo neuronal para distintos números de neuronas en la capa oculta. En segundo lugar se estudió y definió el número de pasos necesarios para entrenar la red neuronal.

Los desarrollos que se realizaron fueron los siguientes:

1. Combinar el método de red neuronal y el método de cluster (conglomerado).
2. Cambiar el pronóstico neuronal basado en el promedio móvil de 24 hrs. a un modelo basado en promedios de 24 horas fijo.
3. Se estudió también el método de predicción utilizado por el gobierno y se introdujo en el modelo neuronal las variables del método Cassmassi.

Los resultados se compararon utilizando los errores porcentuales y los porcentajes de acierto referidos a las mediciones registradas. Al comparar los resultados obtenidos en esta tesis con los resultados desarrollados en la tesis de Giovanni Salini [9], se encuentran mejoras significativas en el pronostico de concentraciones de PM2.5

El mejor resultado se obtuvo combinando los métodos neuronal y conglomerado, en segundo lugar el método que agrega nuevas variables de de entrada, y por ultimo, el método, basado en pronostico de promedio de 24 horas fijo.

CAPITULO 1. INTRODUCCION

Cuando un método de pronóstico de niveles de contaminación predice que la concentración para un día próximo va a ser muy alta, es posible preveer el problema y tomar medidas, por ejemplo, restricción vehicular e impedir que algunas empresas o diferentes fuentes de emisión operen.

Pero, cuando ocurre un episodio de alta contaminación: ¿Cómo afecta a la salud de las personas?, ¿Cuál es la necesidad de evitar eventos de altas concentraciones de contaminación?, ¿No es mejor evitar concentraciones en promedio altas a largo plazo?, ¿Cuál de estos dos últimos fenómenos afecta más?

La ultima pregunta tiene una respuesta muy clara, evaluando las consecuencias que se observan cuando ocurren episodios de altas concentraciones de contaminación: las personas de grupos vulnerables como ancianos, niños, personas con enfermedades respiratorias, etc., se ven afectadas y se observa un alza en consultas médicas y colapso en los servicios de salud primarios.

Para responder las primeras preguntas que se plantearon, se debería hacer un análisis a largo plazo y mucho más detallado, pero independiente del resultado de éste, lo importante es poder evitar que los eventos de altas concentraciones de contaminantes aumenten, ya sea por encima de la norma o no.

Un pronóstico exacto y anticipado de las concentraciones del material particulado permitirá que el gobierno tome las decisiones acertadas que llevarían a la disminución de la contaminación

1.1. Contaminación Ambiental.

La contaminación ambiental hace bastante tiempo que se transformó en un gran problema que demanda una solución inmediata. Para poder contribuir en la solución es necesario, en primer lugar, entender y conocer el problema.

Contaminación es la presencia en un medio, de uno o más compuestos, o cualquier combinación de ellos que produzca un desequilibrio de éste, un daño a los seres que habitan el medio o degrade la calidad de éste.

Específicamente, la contaminación atmosférica se refiere a contaminantes que se encuentran en el aire, sea en forma gaseosa, liquida o sólida.

En este sentido es válido considerar como contaminación ambiental ejemplos como: el polvo en suspensión, aerosoles, smog de los vehículos, humo generado por incendios, emisiones de industrias, etc.

El fenómeno de la contaminación involucra muchas variables: medio ambientales, industriales, de conducta, etc., que deben ser consideradas.

1. 2. Situación de la Ciudad de Santiago de Chile.

Santiago de Chile, ciudad con más de 6 millones de habitantes, capital de Chile, es una ciudad que por su constante y acelerado desarrollo. Se encuentra ubicada en el centro del país y a una altura aproximada de 520 metros sobre el nivel del mar, se encuentra rodeada, por el este y el oeste, por cadenas montañosas. Estas cadenas montañosas son la Cordillera de los Andes (este) y la Cordillera de la Costa (oeste). Ver figura 1.1

El clima de Santiago es mediterráneo, con estaciones como verano e invierno muy marcadas, con meses de Diciembre a Febrero muy calurosos con temperaturas máximas de 35°C, y meses como mayo y junio muy fríos en donde las temperaturas descienden hasta los 0°C.

Figura 1.1: *Santiago de Chile rodeado por cordilleras.*

Debido a sus condiciones topográficas, es importante introducir el concepto de potencial meteorológico de contaminación atmosférica (PMCA), que es un descriptor específicamente meteorológico. Este índice describe condiciones de ventilación y dispersión de contaminantes específicamente para la cuenca de Santiago de Chile. Aplica solamente para el periodo de Abril a Agosto debido a que éste es el periodo crítico de estudio, análisis y monitoreo de concentraciones de contaminantes, y por lo mismo, es asociado a concentraciones de contaminantes.

En la Tabla 1.1, se describen las características y categorías de este índice (Fuente: CENMA – Universidad de Chile).

Tabla 1.1: Potencial Meteorológico de Contaminación Atmosférica (PMCA).

Potencial Meteorológico de Contaminación Atmosférica (PMCA)	
Categoría de PMCA	Condiciones de ventilación/dispersión de contaminantes
1.-Bajo	Muy Buenas
2.-Regular/Bajo	Buenas
3.-Regular	Regulares
4.-Regular/Alto	Malas a Criticas
5.-Alto	Criticas

El PMCA describe condiciones de ventilación específicamente para la cuenca de Santiago relacionados al fenómeno llamado inversión térmica.

Se puede describir inversión térmica como aquel fenómeno en el que se sitúa una capa de aire calido por encima de las masas de aire de menor temperatura. Estas masas de aire de menor temperatura al tener este “tapón” de aire caliente sobre ellas, limita a las capas que puedan moverse de forma

vertical, impidiendo dispersión de contaminantes verticalmente. Debido a las características topográficas de la ciudad de Santiago de Chile, este fenómeno es un factor determinante en la dispersión de contaminantes.

En la década de los ´80, la contaminación atmosférica se transformó en un grave problema. Dentro de algunas ciudades de Latinoamérica, la situación de la contaminación atmosférica es parecida o incluso peor. Según un estudio hecho por la consultora norteamericana William M. Mercer, las ciudades con mayor contaminación atmosférica son: Ciudad de México, Sao Paulo (Brasil), en menor manera, se encuentra Santiago de Chile.

Debido a esta crisis, el gobierno empezó a tomar medidas. En Santiago se realizan mediciones de concentraciones de contaminantes desde finales de los años 80, específicamente en 1989 se realizó el primer inventario de emisiones de la ciudad de Santiago de Chile. El último inventario realizado fue el año 2005, y establece los porcentajes de aporte de cada fuente de contaminación.
A continuación, se muestran los resultados más representativos respecto al último inventario de emisiones realizado el año 2005:

- Industria (24,6%)
- Vehículos livianos (18,2%)
- Camiones (14,5%)
- Sector residencial (11%)
- Buses (8%)

En comparación al inventario de emisiones anterior (2000), la mayor alza fue producida por el sector residencial, que pasó de un 5% el año 2000 a un 11% el 2005. La mayor disminución de porcentajes respecto al inventario anterior fueron los buses, que pasaron de un 22% a un 8%.

Cabe destacar que más del 57% de las emisiones contaminantes en Santiago son responsabilidad de la industria, vehículos livianos y camiones [15].

En la ciudad de Santiago, se implementó en el año 1988, una red de monitoreo de contaminantes atmosféricos, (RED MACAM), esta red, consta de estaciones de monitoreo de concentraciones de contaminantes que operan de manera permanente.

En la figura 1.2 se muestra la ubicación geográfica de las estaciones analizadas en este estudio.

***Figura 1.2**: Localización de la red de monitoreo MACAM en la ciudad de Santiago.*

1. 3. Material Particulado.

Dentro de la contaminación atmosférica se define como material particulado o también llamado “Partículas Totales Suspendidas” a la acumulación de partículas sólidas o liquidas generadas a partir de una actividad natural o antropogénica. Cuando uno observa una erupción volcánica, todo el “humo” de la erupción que se observa corresponde a material particulado. Bajo el mismo ejemplo, la masa de aire brumosa sobre la ciudad de Santiago en episodios de altas concentraciones, corresponde a material particulado.

Los contaminantes en partículas son de una amplia variedad de tamaños, formas y composiciones químicas que producen daño a la salud en mayor o menor grado.

Una consecuencia negativa del material particulado en el ambiente son los efectos nocivos sobre la salud. Ya que las partículas son de tamaño reducido, estas pueden entrar por las vias respiratorias y afectar el cuerpo.

Por esta última razón se estableció un tamaño para partículas respirables, PM10, que corresponde a la sigla de material particulado de tamaño menor a 10 micrones, este material por su tamaño es capas de penetrar las vías respiratorias hasta los pulmones. Sin embargo existe una clasificación mas detallada que corresponde a PM2.5, material particulado de tamaño menor a 2.5 micrones, este ultimo denominado *Material Particulado Fino*.

El material particulado fino ha tomado mayor importancia en el último tiempo, debido a los efectos muy nocivos para la salud, ya que es capaz de penetrar con mayor facilidad las vías respiratorias y depositarse en los pulmones y alvéolos.

Actualmente existe una normativa vigente que rige en Santiago de Chile respecto a las concentraciones de PM10, en relación a PM2.5 existe una normativa que comenzara a regir el 1 de diciembre del 2012. En el año 2009 se realizó la presentación de la Resolución exenta N° 4624 [1], que establece el anteproyecto de normas primarias de calidad ambiental, para material particulado fino PM2.5, la cual describe concentraciones límites diarias y anuales.

Actualmente existe una normativa vigente que rige en Santiago de Chile respecto a las concentraciones de PM10, en relación a PM2.5 existe una normativa que comenzara a regir el 1 de diciembre del 2012. En el año 2009 se realizó la presentación de la Resolución exenta N° 4624 [1], que establece el anteproyecto de normas primarias de calidad ambiental, para material particulado fino PM2.5, la cual describe concentraciones límites diarias y anuales.

En la tabla 1.2 se muestran las normas de PM10 y PM2.5 para algunos países, incluido Chile.

Tabla 1.2: Normas de calidad del aire para PM10 y PM2.5.

	PM10		PM2.5	
	máximo promedio 24 hrs	promedio anual	máximo promedio 24 hrs	promedio anual
	$\mu g/m^3$	$\mu g/m^3$	$\mu g/m^3$	$\mu g/m^3$
OMS	50	20	25	10
EPA-USA	150		35	15
Australia			25	8
México	120	50	65	15
Argentina	150	50	35	15
Chile	150	50	50*	25*
EEA	50	40		25**

* se aplicará a partir del 1 de enero del 2012
** el año 2015 se espera reducirlo a 20
EEA European Environment Agency (Agencia europea de medio ambiente)

Esta nueva norma para la ciudad de Santiago [1] y de acuerdo a su última publicación en el diario oficial el día 9 de mayo del 2011, clasifica de la siguiente forma las concentraciones diarias:

Normal : Concentraciones hasta 50 [µg/m^3].

Regular : Concentraciones mayores a 50 [µg/m^3] hasta 79 [µg/m^3].

Alerta : Concentraciones mayores a 80 [µg/m^3] hasta 109 [µg/m^3].

Preemergencia : Concentraciones mayores a 110 [µg/m^3] hasta 169 [µg/m^3].

Emergencia : Concentraciones mayores a 170 [µg/m^3].

Actualmente en la ciudad Santiago de Chile, se realizan 2 pronósticos simultáneos e independientes de concentraciones de PM10. Uno realizado por la Universidad de Santiago de Chile, a cargo del Profesor Dr. Patricio Pérez; y otro realizado por la Dirección Meteorológica de Chile (utilizado oficialmente por el Gobierno): el modelo Cassmassi. Respecto a PM2.5 existe un modelo en operación desarrollado en la Universidad de Santiago de Chile.

En Santiago el monitoreo de contaminantes es realizado mediante la RED MACAM, este monitoreo comenzó a realizarse en el año1987 en las comunas de Santiago Centro, Las Condes e Independencia. A medida que pasó el tiempo se fueron agregando mas estaciones de monitoreo. En la actualidad la red de monitoreo cuenta con 11 estaciones de monitoreo capaces de medir distintos contaminantes como: PM10, PM2.5, NO, NO2, O3, NOX, entre otros.

El método en el que se realizan las mediciones de material particulado, ya sea PM10 o PM2.5 es mediante un equipo llamado Microbalanza gravimétrica TEOM (Tappered Element Oscillating Microbalance). El funcionamiento de este equipo es el siguiente:

El equipo toma aire directamente del medio ambiente a un flujo constante.

Este flujo de aire se hace pasar a través de un filtro (este filtro depende de que material particulado se desee medir).

El material particulado que se desea medir es acumulado en este filtro.

la microbalanza mide el peso del filtro periódicamente, mediante esto es capas de obtener concentraciones [µg/m^3], en donde: [µg] corresponde a la unidad del peso correspondiente al material acumulado en el filtro y [m^3] corresponde al volumen total de aire que se hizo pasar por el filtro.

Con el fin de revisar la situación histórica de la ciudad de Santiago de Chile, se presentarán a continuación, comparaciones respecto a concentraciones en distintos periodos de PM10 y PM2.5.

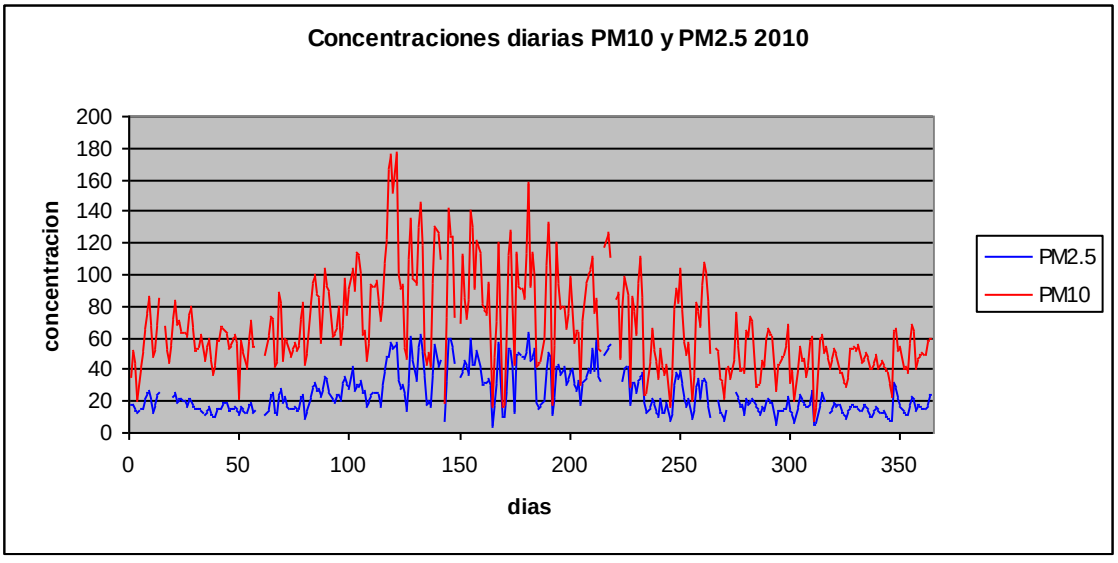

Figura 1.3: *Distribución diaria del año 2010 de PM10 y PM2.5.*

La figura 1.3 muestra la distribución diaria encontrada en el año 2010 en la ciudad de Santiago, específicamente en la estación de monitoreo de la comuna

de Santiago ubicada en el parque O´Higgins. Se puede observar que las concentraciones mayores se encuentran acotadas en los meses de abril a agosto, en la figura 1.3 corresponden al periodo de días 90-260.

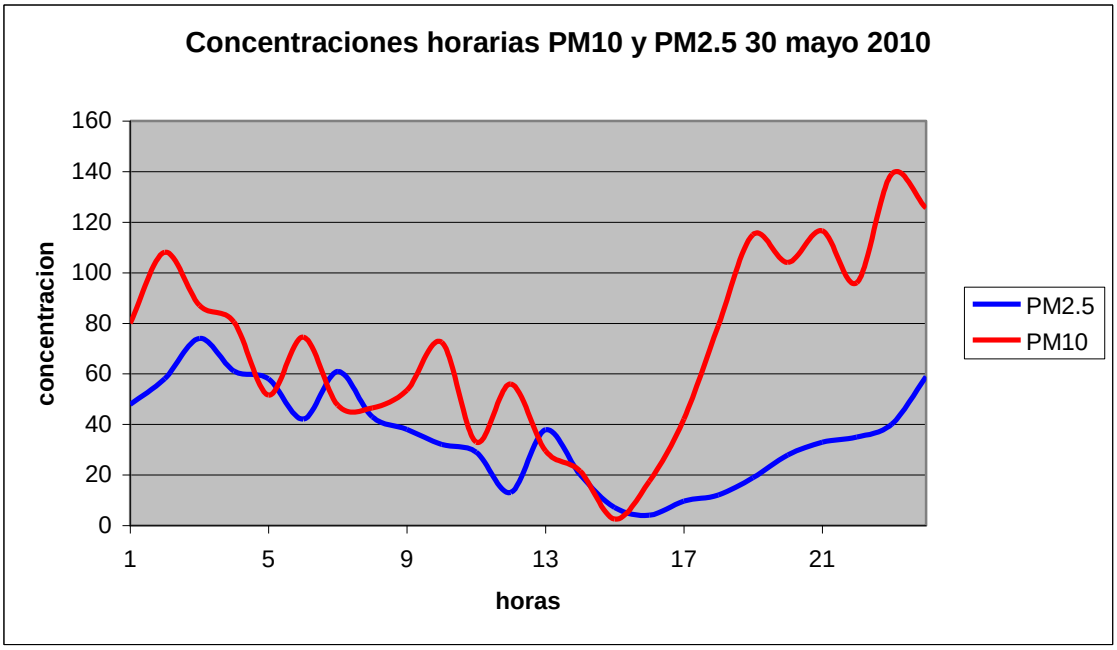

***Figura 1.4**: Distribución horaria del día 30 de mayo 2010 para PM10 y PM2.5.*

La figura 1.4, muestra la distribución horaria durante un día completo. En ella se puede apreciar que durante el día, las concentraciones varían bastante. También se aprecia que las mayores concentraciones horarias se presentan a horas cercanas a la media noche.

De las figuras 1.3 y 1.4, se puede establecer a simple vista, que las concentraciones de PM2.5 corresponden aproximadamente al 50% de las concentraciones de PM10. Hay que tener presente que PM2.5 está contenido dentro de las mediciones de concentraciones de PM10.

Enfocándonos en concentraciones de PM2.5 a continuación se mostrara el historial de eventos críticos (alertas y preemergencias), por año, en el periodo comprendido entre los años 2001-2007.

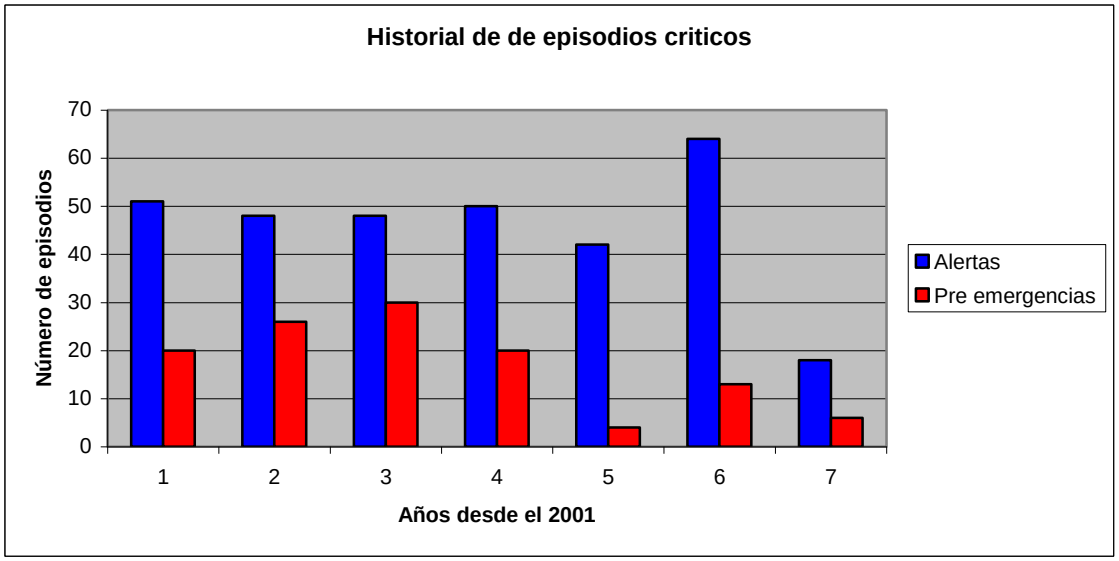

***Figura 1.5**: Historial de eventos críticos.*

Observando la figura 1.5 se pueden hacer los siguientes comentarios:

- El número de eventos correspondiente a Alertas es siempre mayor al de preemergencias.
- A medida que pasan los años el número de eventos preemergencias en general presenta un descenso.

1.4. Modelos de Predicción.

Algunos de los métodos de predicción y/o pronóstico de concentraciones de contaminantes que se utilizaran para el desarrollo de este trabajo serán algunos de los que se utilizan actualmente:

- Redes neuronales.
- Método de Cluster (Conglomerado).
- Modelo Cassmassi.

Los primeros 2 métodos, Redes neuronales y Conglomerado, han sido desarrollado en la Universidad de Santiago de Chile [9]
El modelo Cassmassi es el utilizado por el gobierno de Chile para predecir eventos de contaminación en la ciudad de Santiago.

1.4.1. Método de Redes Neuronales.

Las Redes Neuronales Artificiales representan una técnica de modelación matemática, que imita el proceso de aprendizaje que ocurre en el sistema nervioso de nuestro cerebro.
Las redes neuronales son modelos que intentan reproducir el comportamiento del cerebro, mediante un algoritmo computacional que simula la interacción que ocurre cuando las neuronas realizan sinapsis.

Se denomina Red a la disposición que se realiza de un grupo de neuronas. Estas podría organizarse en 3 grupos, entradas, oculta y salida (Ver figura 1.6), puntos importantes a describir.

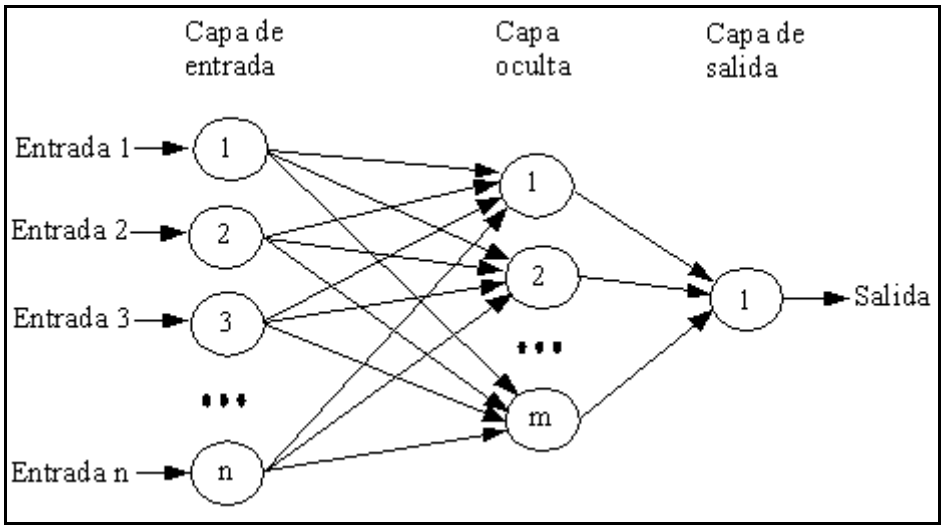

***Figura 1.6**: Esquema de la estructura de red neuronal*

Neuronas de entrada, consiste en las variables de entrada de información o datos a la red neuronal.

Neuronas de salida, estas corresponden a las salida o resultados que entrega la red neuronal, en un modelo de pronostico, estas serian los resultados pronosticados.

Neuronas de las capas ocultas. Corresponden a las neuronas que realizan las interacciones entre las neuronas de entrada y salida, estas neuronas componen el procesamiento de una red neuronal.

Existen varios algoritmos que permiten ir corrigiendo el error de pronóstico; uno de los más usados y que se utilizara en estos desarrollos es el denominado

retro-propagación (backpropagation), que consiste en propagar el error hacia atrás durante el entrenamiento de la red, desde la capa de salida hacia la de entrada, permitiendo así la adaptación de las constantes con el fin de reducir dicho error.

La importancia de la red backpropagation consiste en su capacidad de auto adaptar las constantes o “pesos” de las neuronas de las capas intermedias para aprender la relación que existe entre un conjunto de patrones dados y sus salidas correspondientes.

Cuando se crea una red neuronal se deben considerar las siguientes características:

- Arquitectura de la red, esto es, disposición y número de neuronas.
- Tipos de conexión
- Tipos de dinámica o algoritmo de actualización de la red.
- Regla de aprendizaje
- Tipo de función de transferencia.

Para nuestro caso la arquitectura de la red se mantuvo relativamente fija, a excepción del capitulo 5 en el que se agregaron 2 variables de entrada y cambio el numero entradas en la red. La disposición básica usada para la red neuronal consistió en 15 neuronas de entrada, 12 neuronas en la capa oculta y 4 neuronas de salida.

La regla de aprendizaje utilizada fue Norm-Cum-Delta que corresponde a una regla de aprendizaje que acumula los cambios y actualizaciones de los pesos al final del total de datos entrenados, de esta forma la taza de aprendizaje de la red es independiente del numero de datos a entrenar.

La función de transferencia para todas las conexiones de neuronas es la misma y corresponde al tipo sigmoidal. (ver figura 1.7).

$$f(t) = \frac{1}{1+e^{-t}} \tag{1.1}$$

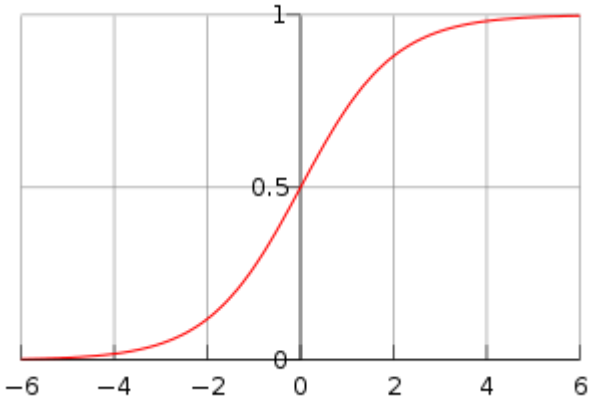

Figura 1.7: *grafico de la función Sigmoidal.*

Una red neuronal por si sola no puede resolver ningún problema. La forma de utilizar la red neuronal es entrenándola de forma que la misma red establece los pesos de cada neurona creando (matemáticamente), las relaciones internas entre las capas de neuronas.

Como se dijo anteriormente, el poder de las redes neuronales viene del entrenamiento, por lo que cualquier “problema” del que se pueda obtener algún historial y que pueda ser manejado como variables y “vectorizado”, podrá ser procesado con redes neuronales.

Algunos ejemplos de aplicaciones de redes neuronales son las siguientes:

- Clasificación y reconocimiento de imágenes, señales, patrones de voz
- Pronostico climático.
- Filtrado de señales

Las principales características positivas de las redes neuronales son:

- Aprendizaje
- Robustez
- Autoorganización
- Tolerancia a fallos
- Flexibilidad
- Tiempo real
- Rapidez del calculo

El modelo de pronóstico basado en redes neuronales desarrollado en la Universidad de Santiago de Chile [9], será la base a partir este trabajo. Para

este caso de predicción de concentraciones de PM2.5 se disponen los datos de la siguiente forma, entregados por el SEREMI de Salud [14].

Tabla 1.3: distribución de datos en vector de trabajo para red neuronal.

Estación La Florida			Estación Las Condes			Estación Santiago			Estación Pudahuel						Estación La Florida	Estación Las Condes	Estación Santiago	Estación Pudahuel
		Pro			Pro			Pro			Pro	Δ°c	Δ°c					
18	19	24	18	19	24	18	19	24	18	19	24	hoy	mañ	pmca	Out	Out	Out	Out
26	29	30	25	25	20	28	13	25	19	23	31	18	16	40	31	24	33	39
20	24	30	22	25	23	27	22	32	27	26	39	16	9	40	34	33	31	38
22	15	16	26	27	18	26	19	18	16	16	20	6	17	60	33	21	40	41
17	51	28	23	33	19	25	36	30	23	46	32	17	17	60	40	25	53	63
75	53	40	30	70	22	37	32	53	37	34	63	17	8	40	45	47	44	52

En la tabla 1.3 se pueden describir:

- 18: Concentración horaria de PM2.5 medida a las 18 horas del presente día para cada estación de contaminación.
- 19: Concentración horaria de PM2.5 medida a las 19 horas del presente día para cada estación de contaminación.
- Pro 24: Concentración de promedio móvil de 24 horas medido a las 19 horas del presente día para cada estación de contaminación.
- Δ°C hoy: Amplitud térmica en la ciudad para el presente día
- Δ°C mañ: Amplitud térmica pronosticada para el siguiente día
- PMCA: índice de PMCA pronosticado para el siguiente día.
- OUT: Concentración de máximo del promedio móvil de 24 horas para cada estación de contaminación.

Cada columna de datos representa una variable (ya se de entrada o de salida), cada fila de datos representa un dia específico. Las 15 primeras columnas (en color amarillo) representan los datos de entrada, las primeras 12 columnas se dividen en grupos de 3, esto es: las 3 primeras columnas representan las variables de entrada directas para la primera estación (estación L, Las Condes). las siguientes 3 columnas representan los variables de entrada directas para la segunda estación (estación M, La Florida). Así sucesivamente para las estaciones N (Santiago, Parque O'higgins) y O (Pudahuel), de las tres columnas referidas a cada estación, el primero (18), es la medición de $PM_{2.5}$ a las 18 hrs, la segunda (19), es la medición a las 19 hrs. de concentración de $PM_{2.5,}$ la tercera (PRO 24), es el promedio móvil de las 24 hrs media a las 20 horas del presente día. La columna numero 13 (∆°C hoy), corresponde a la variación de temperatura del día de hoy, la columna 14 (∆°C mañ), corresponde a la variación de temperatura del día de mañana, la columna numero 15 corresponde al PMCA (Potencial meteorológico de contaminación atmosférica), discutido en el punto 1.2 de este Capitulo.

Las últimas 4 columnas (en color verde claro), corresponden a los valores que se midieron en concentración de 24 horas móvil de $PM_{2.5}$ para el día de mañana referidas a cada estación de monitoreo.

El motivo de elegir las variables descritas anteriormente, se explica de la siguiente forma:

- Las concentraciones horarias a las 18 y 19 horas indica la tendencia de la distribución de concentración.
- La concentración de 24 horas medido a las 19 horas indica la acumulación de 24 horas de PM2.5,
- Las amplitudes térmicas se escogieron debido a que tienen relación con el fenómeno inversión térmica.
- El PMCA se escogió por lo específico que es referido a la ciudad de Santiago respecto a dispersión de contaminantes.

Dentro del programa de creación de la red neuronal hay un parámetro llamado EPOCH, este valor lo entrega el programa luego de leer los archivos en donde se encuentran las variables de entrada y salida. La condición de factibilidad para la red neuronal (que sea capaz de realizar cálculos) consiste en que la suma del numero de neuronas en la capa oculta multiplicado por las variables de entrada más el número de neuronas en la capa oculta multiplicado por las variables de salida no puede ser mayor que la EPOCH:

$$V_{in}\, N + V_{out}\, N < \text{EPOCH} \qquad (1.2)$$

En donde:

- V_{in}: numero de variables de entrada (en nuestro caso 15).
- V_{out}: numero de variables de salida (en nuestro caso 4).
- N: numero de neuronas en la capa oculta.

El concepto de EPOCH corresponde al número de días que serán entrenados. El motivo por el cual se debe cumplir la ecuación 1.2 es debido a que el número de ecuaciones (días de entrenamiento), debe ser mayor a las incógnitas (constantes de cada neurona en la capa oculta). Si por algún motivo se elige un número de neuronas que no cumpla la ecuación 1.2, se tendrán más incógnitas que ecuaciones y la red neuronal no será eficiente al realizar los pronósticos.

1.4.2. Método de Conglomerado

El método de Conglomerado, en ingles Cluster, consiste en una técnica basada en separación mediante clases o rangos. La finalidad de este método es determinar cuan dispersos son los datos de acuerdo a un vector de referencia. En pocas palabras, esta técnica consiste en clasificar los datos y realizar mediciones para verificar cuan cerca o lejos de estas clasificaciones se encuentran los datos.

Aplicando estos conceptos al pronóstico de concentraciones de contaminantes y utilizando los mismos datos de entrada y salida de los utilizados en el método neuronal se puede comentar lo siguiente.

La clasificación utilizada es la misma para todos los desarrollos.

Los vectores de referencia se calcularán promediando los datos de entrenamiento para cada una de las clases.

El vector de referencia se obtendrá de los datos que se usan como entrenamiento.

Comúnmente para calcular la distancia de cada vector al vector referencial se utiliza la distancia Euclidiana entre vectores.

$$d_E(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2} \qquad (1.3)$$

El subíndice E determina la clase, los subíndices *i* y *j* se refieren al vector, con la condición $i \neq j$.

La mayor fortaleza de este método es la capacidad de realizar clasificaciones precisas de datos sin la necesidad de tener muchos datos de entrenamiento.

Debido a este motivo este método se utilizará en el capitulo 3 para corregir concentraciones altas, que corresponderán a eventos que pocas veces ocurren.

1.4.3. Método Cassmassi

En 1998 el Servicio de Salud Metropolitano del Medio Ambiente (SESMA), dio inicio al segundo modelo de pronóstico de calidad del aire para Santiago denominado CASSMASSI, que comenzó a operar de manera definitiva en el año 2000.

Este modelo predice concentraciones de PM10, y su uso está restringido al año calendario comprendido entre los meses de Abril – Agosto.

Desarrollado por Joseph Cassmassi, el modelo esta basado en algoritmos de calculo obtenidas mediante estadísticas de regresión múltiple, que determinan las relaciones entre las variables predictivas (de entrada) y la variable que se va a predecir (de salida). A través de estos métodos se obtuvo una ecuación lineal que corresponde a la ecuación de pronóstico de concentración de PM10. Este análisis se realizo para cada estación de monitoreo de la RED MACAM por separado generando una serie de coeficientes especifica para cada estación.

Por ejemplo, en el caso de la estación de Pudahuel:

$$Y_0 = 39{,}4\ V_0 + 0{,}33\ C_0 + 2{,}06\ T_0 + 0{,}21\ D_0 - 21{,}7 \qquad (1.4)$$

En donde:

- Y_0: Máximo del promedio móvil de 24 horas de PM10 esperado para el día siguiente en [µg/m^3].
- V_0: Estabilidad atmosférica pronosticada para el día siguiente, la cual es una variable discreta que puede variar entre 1 y 5 (PMCA).
- C_0: Promedio de 24 horas de PM10 medido a las 10:00 AM del día presente en la estación Pudahuel en [µg/m^3].
- T_0: Temperatura en grados °C del nivel 925 hPa medido a las 12 UTC del día presente con el radio sondeo de Rocas de Santo Domingo (80 Km. al oeste de Santiago).
- D_0: Cambio en las ultimas 24 horas para la altura del nivel 500 hPa nivel medido a las 12 UTC del día presente con el radio sondeo de Rocas de Santo Domingo.

Analizando las variables de esta ecuación, se observa que se tienen variables (C_0), relacionadas directamente a concentraciones de PM10, las demás variables involucradas son netamente variables atmosféricas.

1.5. Métodos de Análisis de Datos

Para poder realizar una comparación entre los resultados encontrados en la tesis se Giovanni Salini [9] y nuestros resultados, primero se clasificaron lo

datos de la misma forma, de acuerdo a las concentraciones. La clasificación que se utilizó para los desarrollos posteriores, fue la misma clasificación utilizada en el modelo desarrollado por la Universidad de Santiago de Chile [9]. Para los cálculos, y posterior análisis se dejará establecida como clasificación de concentraciones la siguiente distribución:

Clase A (BUENA), concentraciones hasta 35 [µg/m^3], inclusive.
Clase B (REGULAR), concentraciones entre a 35 [µg/m^3] y 65 [µg/m^3].
Clase C (ALERTA), concentraciones entre a 65 [µg/m^3] y 100 [µg/m^3].
Clase D (PREEMERGENCIA), concentraciones mayores a 100 [µg/m^3].

Esta clasificación, se estableció cuando se desarrolló el pronóstico basado en redes neuronales realizado por Giovanni Salini [9], y se basó en la norma para PM2.5 en USA. Actualmente esta norma establece como norma diaria 35 µg/m^3, la norma anterior establecía 65 µg/m^3. Basados en estos 2 valores se establecieron las 2 primeras clases. Las 2 siguientes clasificaciones se establecieron eligiendo como critico el valor 100 µg/m^3, que corresponde a 4 veces el valor establecido por la Organización Mundial de la Salud (OMS).

Cuando se realiza un pronostico de alguna concentración de contaminantes es necesario hacer una validación con el fin de analizar cuan eficiente es el método en pronosticar dichas concentraciones.

Para realizar esta validación se utilizarán 2 métodos: uno basado en porcentajes de error del valor pronosticado respecto al valor medido, otro se basa en un análisis de porcentajes de aciertos específicos para cada clase.

1.5.1. Porcentaje de Error

Un método para comparar los resultados obtenidos de los pronósticos de concentraciones de los distintos desarrollos consistió en comparar el porcentaje de error de los resultados basados en las concentraciones medidas.

La ecuación utilizada para realizar el cálculo de porcentajes de error esta dada por:

$$PE = \frac{1}{2}\sum_{1}^{n}\left|\frac{y_{tp} - y_{ta}}{y_{ta}}\right| \times 100 \qquad (1.5)$$

En donde:

- y_{tp} : Es el valor pronosticado
- y_{ta} : El valor medido.
- n : Es el número total de los casos.

Este método de comparación resulta válido para comparar el modelo de redes neuronales debido a que este modelo entrega un valor preciso de concentración de PM2.5.

1.5.2. Porcentajes de Acierto

Cuando se desea realizar comparaciones de eficiencia de un método de pronóstico referidos a alguna clase en particular lo más válido es contabilizar directamente el acierto para cada día.

Cuando se utiliza como valor de eficiencia, el porcentaje de error, existe un tema que se deja sin desarrollo, cuando se obtiene un porcentaje especifico de error para un método y para una estación en particular (o un valor promedio), este porcentaje de error es referido únicamente al error aritmético respecto al valor medido en un periodo en particular, este método no permite hacer comentarios respecto a las clases en las que se encuentra la concentración pronosticada.

Debido a esto y con el fin de analizar los aciertos en cada clase, y en específico las clases que corresponden a eventos críticos (clases C y D), se utilizara y compararan los aciertos de cada método referidos específicamente a cada clase. De esta forma es posible hacer un análisis respecto a la efectividad en predecir de cada método.

$$\%_{acierto} = 100 \frac{N_{aciertos}}{N_{totales}} \tag{1.6}$$

La ecuación 1.6 muestra la formula para el cálculo de porcentaje de aciertos, en donde:

$N_{aciertos}$: corresponde al número total de aciertos de una clasificación específica.

$N_{totales}$: corresponde al número total de eventos de una clasificación específica.

1.6. Propuestas a Desarrollar

Para lograr desarrollar un modelo de pronóstico de contaminantes es importante considerar las variables involucradas en el fenómeno que son variadas: atmosféricas, topográficas, sociales (comportamientos, cultura), etc. Un tema importante es ser capaz de encontrar las variables más “relevantes”. Mientras mas relevantes son las variables, mas preciso puede llegar a responder el modelo de pronóstico.

El conocer la relación o importancia de cada variable involucrada es lo mas importante en cualquier problema, si nos enfocamos en encontrar la mayor cantidad de variables se podría llegar a un punto en donde seria imposible cuantificar o interpretar tal cantidad de entradas, en este sentido la eficiencia de los modelos de pronostico se pierde cuando se le agregan variables en exceso.

En primer lugar y como base para los desarrollos posteriores, se realizará un estudio para determinar el número de neuronas más eficiente (que entregue los mejores resultados de % de error). Consistirá en que la red neuronal se creará para un mismo caso con distintos número de neuronas en la capa oculta. en paralelo y para cada número de neuronas en la capa oculta se analizará el número optimo de pasos para entrenar cada red. De esta forma se encontrara no solo el número óptimo de neuronas a utilizar para los posteriores desarrollos con red neuronal, también se obtendrá el número optimo de pasos para realizar los entrenamientos de la red neuronal.

Como segundo desarrollo se realizará una combinación del método neuronal y el método de cluster. Utilizando los resultados de los cálculos de pronósticos mediante red neuronal y mediante cluster, se realizará una separación inicial por clases, para luego en cada clase utilizar un método en específico.

Un tercer desarrollo propuesto será la implementación de un método de pronóstico de concentraciones mediante redes neuronales pero basadas en pronóstico de concentraciones de 24 horas fijo.

Un último desarrollo que se realizará será la adición de 2 variables nuevas al método de redes neuronales, que se utilizan en el método de pronóstico de Cassmassi, y están relacionadas directamente con el fenómeno de inversión térmica.

CAPITULO 2. PRONOSTICO DE CONCENTRACIONES DE $PM_{2.5}$

El objetivo principal de este capitulo es encontrar los parámetros mas eficientes en la creación de la red neuronal en función del número de neuronas en la capa oculta y en función del numero de pasos de entrenamiento.

2.1. Número Óptimo de Neuronas y Pasos de Entrenamiento

Una variable importante en el método de redes neuronales es el número de neuronas en la capa oculta.

El procedimiento consistió en considerar un conjunto de datos especifico, que en nuestro caso fue el periodo 2001, 2002, 2003. De este consunto total de datos se escogió un subconjunto de datos que se utilizarán para testear el pronostico, este subconjunto se escogió del año 2002 y corresponde a los meses de mayo, junio y julio. Se debe dejar claro que estos datos se excluyen del total de datos con el que se entrenara la red.

El enfoque en este desarrollo fue encontrar el número de pasos necesarios para que el pronóstico sea más preciso.

En un comienzo del desarrollo al crear la red se dispusieron 2 neuronas en la capa oculta, luego se procedió a entrenar la red con los datos del 2001 al 2003 (excluyendo el subconjunto), el numero de veces entrenamientos que se realizo para todos los números de neuronas fue de 20000, 40000, 60000, 80000 y

100000, teniendo la red entrenada se procedió a realizar un pronóstico, para luego comparar las concentraciones medidas y calculadas para el subconjunto mayo, junio y julio del año 2002.

La referencia para evaluar el resultado fue el porcentaje de error de cada pronóstico para cada una de las estaciones de monitoreo.

En primer lugar se concluyo que los mejores resultados referidos a porcentajes de error se obtuvo al realizar mas de 80000 pasos de entrenamiento para todos los números de neuronas en la capa oculta, por esto, el numero optimo que se utilizara para los cálculos posteriores, será de 100000 pasos.
Los resultados obtenidos a porcentajes de error para los distintos números de neuronas en la capa oculta, se muestran en la tabla 2.1.

Tabla 2.1: Porcentajes de error en función de el numero de neuronas en la capa oculta de la red neuronal

	TEST 2001- 2002 - 2003 PRUEBA 2002 mayo junio julio			
neuronas capa oculta	**Estación La florida**	**Estación Las Condes**	**Estación Santiago**	**Estación Pudahuel**
2	16	20	18	24
3	11	16	17	23
4	14	16	18	19
5	11	14	14	16
6	10	14	11	21
7	10	15	11	17
8	10	15	11	16
9	12	15	14	16
10	12	15	15	17
11	10	15	10	15
12	10	14	9	13
13	10	14	10	14
14	10	14	10	12
15	10	15	11	14
16	10	14	11	13

A continuación en la figura (2.1) se muestra un grafico de los resultados mostrados en la tabla (2.1).

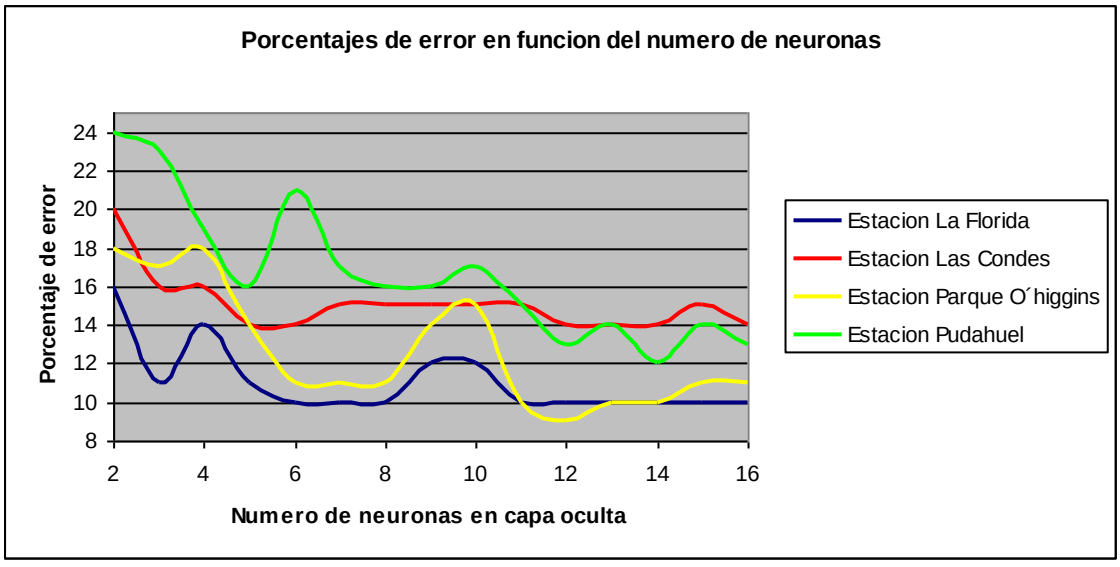

Figura 2.1: *Porcentaje de error en función de número de neuronas en la capa oculta.*

Para obtener un resultado más referencial, se calculó el error promedio para las 4 estaciones, este resultado se muestra en la tabla 2.2.

Tabla 2.2: Porcentajes de error promedio en función del número de neuronas en la capa oculta.

neuronas	% error
2	19,5
3	16,75
4	16,75
5	13,75
6	14
7	13,25
8	13
9	14,25
10	14,75
11	12,5
12	11,5
13	12
14	11,5
15	12,5
16	12

Para analizar los porcentajes de error promedio se graficara los datos expuestos en la tabla 2.2.

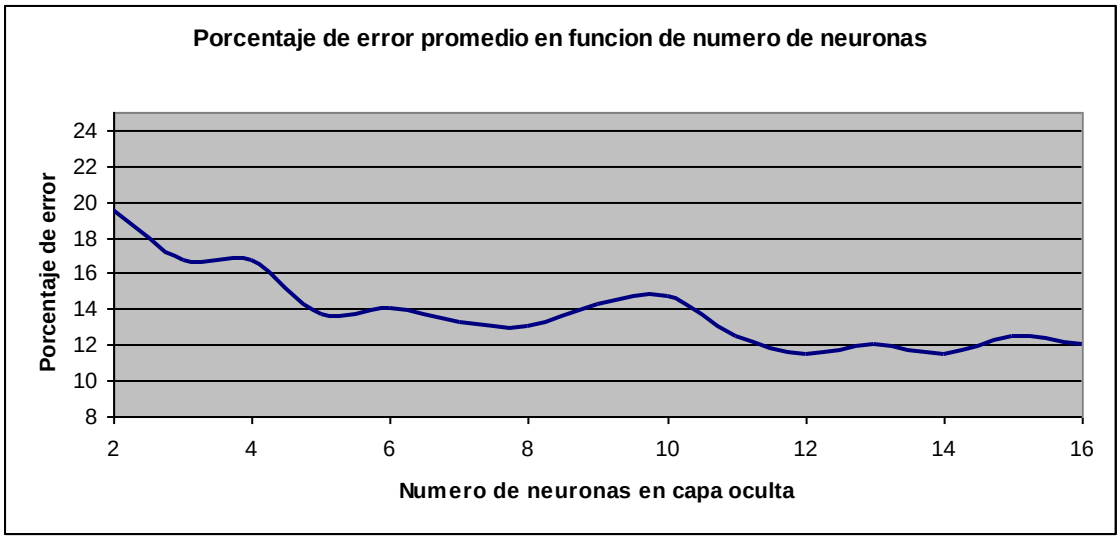

Figura 2.2: *Porcentaje de error promedio en función de número de neuronas en la capa oculta.*

En la tabla 2.2 se destacan en amarillo los menores porcentajes de error obtenidos. El numero optimo de neuronas para obtener mejor porcentaje de error (menor), ocurre al tener 12 y 14 neuronas.

Para elegir el numero optimo de neuronas en primer lugar se analizara el numero limite de neuronas para esta red, el numero de neuronas limite para este caso es de 17 neuronas. Al tener el número límite de 17 neuronas se elegirá como numero óptimo a 12 neuronas. Se elimina como opción valida las 14 neuronas debido a que se encuentra cercana a las 17 neuronas que corresponde a las neuronas limites.

A continuación se muestra una tabla con los resultados de porcentajes de error obtenidos para la tabla completa de datos correspondientes a los años 2004, 2005, 2006 y 2007. Se realizará una comparación respecto a los datos

obtenidos en el trabajo de Tesis [9], en esta referencia se tienen resultados de porcentajes de error basados en cálculos realizados con 20 neuronas en la capa oculta.

Tabla 2.3: Porcentajes de error utilizando 12 neuronas.

	12 neuronas				
Año	Estación La Florida	Estación Las Condes	Estación Santiago	Estación Pudahuel	Promedio
2004	15%	22%	17%	24%	19,5%
2005	14%	18%	18%	23%	18,25%
2006	18%	24%	21%	24%	21,75%
2007	17%	22%	17%	25%	20,25%

Con el fin de comparar nuestros resultados, a continuación se muestran los resultados realizados en el trabajo de Giovanni Salini [9],

Tabla 2.4: Porcentajes de error utilizando 20 neuronas [9].

	20 neuronas				
Año	Estación La Florida	Estación Las Condes	Estación Santiago	Estación Pudahuel	Promedio
2004	15%	14%	23%	31%	20,75%
2005	16%	18%	20%	32%	21,5%
2006	17%	17%	25%	31%	22,5%
2007	18%	18%	20%	26%	20,5%

Al observar ambas tablas, 2.3 y 2.4, se aprecia la mejora en los porcentajes de error al utilizar 12 neuronas en comparación con 20 neuronas.

A continuación se mostrarán los porcentajes de acierto respecto a las clases definidas como criticas, clases C y D. la clase C corresponde a eventos de “Alertas Ambientales”, la clase D corresponde a eventos de “Pre emergencia ambiental”

Tabla 2.5: Porcentajes de aciertos para eventos clase C.

Año	Eventos totales clase C	% Aciertos
2004	52	81
2005	58	78
2006	65	58
2007	55	60

Tabla 2.6: Porcentajes de aciertos para eventos críticos (C y D).

Año	Eventos totales clase C	% Aciertos
2004	72	94
2005	63	86
2006	76	64
2007	78	67

A continuación se presentan los resultados del trabajo de Giovanni Salini [9] pronósticos basados en cálculos de redes neuronales con 20 neuronas en la capa oculta.

Tabla 2.7: Porcentajes de aciertos para eventos clase C [9].

Año	Eventos totales clase C	% Aciertos
2004	52	67
2005	58	40
2006	65	52
2007	55	71

Tabla 2.8: Porcentajes de aciertos para eventos críticos (C y D) [9].

Año	Eventos totales clase C	% Aciertos
2004	72	62
2005	63	38
2006	76	51
2007	78	61

Comparando las tablas 2.5 y 2.7 se aprecian mejoras de porcentajes de aciertos referidos a eventos de Alerta ambiental.

Para casos con mayor cantidad de eventos (clases B y C), se obtuvieron mejores resultados al utilizar 12 neuronas, TABLA 2.5.

CAPITULO 3. IMPLEMENTACION DE MÉTODO COMBINANDO RED NEURONAL Y CLUSTER

3.1. Presentación y Justificación de la Propuesta

Con el objetivo de mejorar el pronóstico de concentraciones se buscará combinar ambos métodos de predicción: el método neuronal y el método de Cluster. Ambos métodos tienen buenos porcentajes de aciertos y bajos porcentajes de error pero al compararlos, el método de cluster es mas preciso al pronosticar clases menos frecuentes, que corresponden a clases A y D, por su parte el método de redes neuronales es mas eficiente en pronosticar concentraciones que se encuentran en las clases B y C, que corresponden a los eventos de mayor ocurrencia.

3.2. Consideración

Hay que tener en cuenta que los cálculos de red neuronal son directamente en concentraciones, a diferencia del método de cluster que arroja solamente la clase en que se encontrara la predicción.

Basados en esto último, la metodología consiste en hacer correr ambos programas para cada vector de datos y comparar los resultados aplicando los siguientes criterios:

- Cuando ambos métodos pronostican el resultado en la misma clase, se mantendrá el resultado obtenido por el método neuronal.
- Cuando los resultados se encuentren en las clases intermedias pero los resultados de ambos métodos no coinciden se priorizara el método neuronal.
- Cuando los resultados se encuentren en las clases extremas y los resultados de ambos métodos no coincidan, se priorizara el resultado obtenido por el método cluster, ajustando el resultado numérico de pronostico de la red neuronal de manera que pertenezca a la clase que fue predicha por el método de cluster.

La clase A no sera considerada puesto que es una clase de baja ocurrencia y de mas bajas concentraciones, en ese sentido son mas prioritarias las clases en las cuales las concentraciones son mas altas.

A continuación se muestra el número de eventos sucedidos en cada clase y para cada año, que corresponde a los meses de Abril hasta Agosto.

Tabla 3.1: Numero de episodios por año para cada clase.

	Clases			
Año	A	B	C	D
2004	19	49	52	20
2005	13	55	58	5
2006	13	60	65	13
2007	15	58	56	22

3.3. Comparación Método Cluster y Red Neuronal

Al revisar la tabla se observa que los episodios que más ocurren son los comprendidos en las clases B y C,

A continuación se muestran los resultados de los porcentajes de acierto de cada método:

Tabla 3.2: Porcentajes de acierto para el método de red neuronal y cluster.

	red			
	A	B	C	D
2004	42%	88%	81%	15%
2005	54%	84%	78%	0%
2006	38%	82%	58%	0%
2007	53%	90%	59%	0%

cluster			
A	B	C	D
84%	55%	62%	70%
77%	62%	62%	20%
85%	62%	40%	62%
93%	59%	50%	77%

Observando ambas tablas y fijándonos en los porcentajes mas altos para cada caso podemos percatarnos que los mayores porcentajes de acierto del método de redes neuronales es para los casos en los que se tienen mayor cantidad de datos (eventos), por su parte el método de Cluster parece ser mas efectivo al pronosticar casos con menor cantidad de eventos, estas clases serian clases A y D.

Haciendo una comparación con los cálculos desarrollados por Giovanni Salini [9] se puede concluir en primer lugar que para episodios de de clases B y C, se obtiene un mejor porcentaje de aciertos. La diferencia es debida al uso de menos neuronas en la capa oculta, en este caso se utilizo 12 neuronas a diferencia de las 20 neuronas utilizadas en sus desarrollos.

3.4. Resultados

A continuación se presentan los resultados combinados, esto es porcentajes de acierto calculados mediante el método de Cluster para clases A y D, y resultados de aciertos calculados mediante el método de red neuronal para las clases C y D.

Tabla 3.3: Porcentajes de acierto para método combinado.

	A	B	C	D
2004	84%	82%	67%	70%
2005	77%	75%	40%	20%
2006	85%	70%	52%	62%
2007	93%	72%	71%	77%

Estos resultados de acierto resultan ser bastante altos respecto a los modelos actuales utilizados en Santiago.

CAPITULO 4. IMPLEMENTACION DE MODELO NEURONAL USANDO EL PROMEDIO DE 24 HORAS FIJO

4.1. Propuesta.

Todos los modelos de pronóstico actuales utilizados en la ciudad de Santiago de Chile, utilizan como parámetro de entrada concentraciones basadas en promedio móviles de 24 horas, consideradas como mas representativas.

Esto no permite visualizar las concentraciones más altas y que consiste el mayor problema a la salud, como en caso de las concentraciones medidas cada hora.

4.2. Consideraciones

Para poder trabajar los datos primero hicieron coincidir los datos con fechas, de acuerdo al historial de datos de la página del Sistema de Información Nacional de Calidad del Aire (SINCA del Ministerio del Medio Ambiente) [12].

La única diferencia en el vector de datos original es que en este nuevo caso, las variables de salida serán las concentraciones de 24 horas fijas. Esto es la concentración promedio del día.

Hay que tener en cuenta que al realizar esta modificación, se esta pronosticando otro fenómeno.

Para realizar la comparación entre el modelo actual (24 horas móvil) y esta nueva implementación (24 horas fijo), se utilizará como herramienta el porcentaje de error en la concentración que se predice versus el porcentaje de error en la concentración medida.

A continuación se mostrara un grafico de un día en particular con episodio crítico en el que se aprecian las concentraciones horarias, 24 horas móvil y 24 horas fijas.

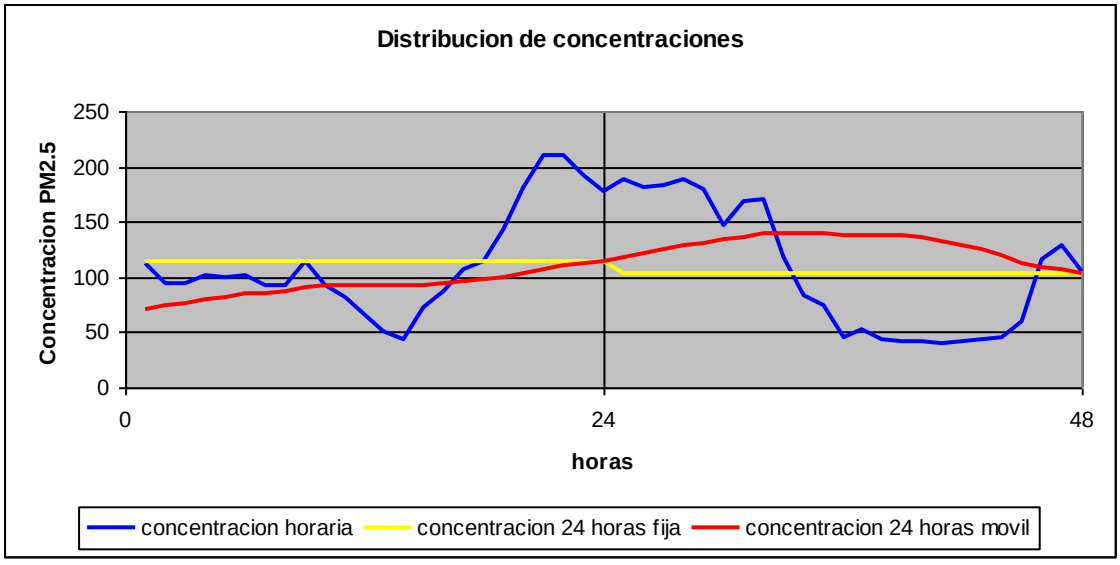

***Figura 4.1**: Porcentajes de acierto para método combinado.*

Observando y analizando la figura 4.1 se verifica lo descrito anteriormente, al cambiar a concentraciones de 24 horas fija, se esta pronosticando un fenómeno completamente distinto, por lo tanto en cierta forma no podrían ser comparables del todo sus resultados.

4.3. Desarrollo

Basándonos en la propuesta descrita anteriormente se procedió a cambiar la estructura de los datos, puesto que el vector de datos contenía en las variables de salida concentraciones basadas en muestras de promedios móviles, todas estas concentraciones se cambiaron a concentraciones en 24 horas fijas.

Una vez realizado este cambio se procedió a utilizar el modelo de redes neuronales para pronosticar las concentraciones y verificar alguna variación en los resultados.

4.4. Resultados

A continuación se expondrán los resultados obtenidos de acuerdo a los porcentajes de error en las concentraciones que se predicen.

La primera tabla (4.1), consiste en los resultados obtenidos de acuerdo al método que se utiliza actualmente (24 horas móvil).

Tabla 4.1: porcentajes de error en predicción con 24 horas móvil.

Año	Est. La Florida	Est. Las Condes	Est. Santiago	Est. Pudahuel	Cantidad de datos
2004	14	14	16	24	401 (126)
2005	14	16	18	19	388 (111)
2006	16	16	21	25	364 (148)
2007	16	17	17	23	385 (53)

A continuación y con el fin de comparar resultados, se mostrara la tabla de resultados referida al método que se desarrolla en este capitulo (24 horas fijo).

Tabla 4.2: porcentajes de error en predicción con 24 horas fijo.

Año	Est. La Florida	Est. Las Condes	Est. Santiago	Est. Pudahuel	Cantidad de datos
2004	14	23	15	21	401 (126)
2005	10	17	11	15	388 (111)
2006	15	24	13	19	364 (148)
2007	4	8	4	6	385 (53)

En ambas tablas (4.1 y 4.2) se la columna “cantidad de datos” se muestra la cantidad de datos que se usaron de entrenamiento y en paréntesis se muestra la cantidad de datos que se pronosticaron.

Analizando ambas tablas (4.1 y 4.2), no es posible establecer claramente una mejoría de los porcentajes de error. Debido a que algunos porcentajes mejoran y otros empeoran, no se puede establecer una mejoría clara al utilizar el método de 24 horas fijo.

Con el fin de analizar mas profundamente este método de implementación de 24 horas fijo, se realizara una comparación respecto al porcentaje de aciertos referidos a los eventos críticos (clases C y D).

En primer lugar se comparara el número de eventos críticos ocurridos para el modelo de 24 horas móvil y para el modelo de 24 horas fijo.

Tabla 4.3: Numero de eventos críticos modelo 24 horas móvil.

	clase C	clase D
2004	52	20
2005	58	5
2006	65	13
2007	56	22

Tabla 4.4: Numero de eventos críticos modelo 24 horas fijo.

	clase C	clase D
2004	37	11
2005	28	2
2006	45	4
2007	15	3

La diferencia encontrada en el numero de eventos es esperable, debido a que el modelo de 24 horas fijo, se basa en el promedio de las 24 horas de un día completo, desde las 1 hasta las 24 horas, al ser un promedio diario, es menos representativo para eventos de mayores concentraciones. Esto puede verse claramente en la figura 4.1.

A continuación se mostraran los resultados de aciertos de ambos métodos para las clases críticas.

Tabla 4.5: porcentaje de acierto casos críticos modelo 24 horas móvil.

	clase C	clase D
2004	81%	15%
2005	78%	0%
2006	58%	0%
2007	59%	0%

Tabla 4.6: porcentajes de acierto casos críticos modelo 24 horas fijo.

	clase C	clase D
2004	70%	0%
2005	68%	0%
2006	58%	0%
2007	53%	0%

Haciendo una comparación de los porcentajes de aciertos para las clases criticas, es claro que al utilizar un modelo de 24 horas fijo, se obtienen resultados menos asertivos.

CAPITULO 5. NUEVA VARIABLE

5.1. Propuesta

Debido a las características geográficas de la cuenca de Santiago, las variables meteorológicas ($\Delta T°$, PMCA), son muy importantes para poder predecir las concentraciones de contaminantes. Debido a esta importancia se buscara involucrar una nueva variable que esté relacionada con la meteorología, que ayude a mejorar el pronóstico.

Basándonos en el modelo de predicción utilizado por el gobierno (Modelo Cassmasi), nos centraremos en estudiar su ecuación de predicción de concentraciones para la estación de monitoreo de Pudahuel para encontrar nuevas variables que puedan ser un aporte en nuestros desarrollos. Cabe destacar que el modelo Cassmassi predice concentraciones de PM10, a diferencia de nuestro modelo, pero de igual forma se utiliza como referencia para la adición de nuevas variables. Estas variables están relacionadas con la Altura de la Capa de Inversión Térmica que es monitoreada en Santo Domingo (Quinta Región), [10].

El motivo por el cual se escogió esta estación de monitoreo es por que es la estación en la que se perciben mayores concentraciones de PM2.5, que constituyen la mayor amenaza para la salud de la población.

5.2. Variable Nueva

Con el objetivo de mejorar el pronostico, un punto que se desarrollara será el de incluir nueva(s) variables de entrada. Para encontrar una nueva variable se observara el método de pronostico Cassmassi y nos basaremos en la ecuación de pronostico de la estación de Pudahuel.

$$Y_0 = 39{,}4\ V_0 + 0{,}33\ C_0 + 2{,}06\ T_0 + 0{,}21\ D_0 - 21{,}7 \quad \text{Ecuación 5.1}$$

En donde:

- Y_0: Máximo del promedio móvil de 24 horas de PM10 esperado para el día siguiente en [µg/m^3].
- V_0: Estabilidad atmosférica pronosticada para el día siguiente, la cual es una variable discreta que puede variar entre 1 y 5 (PMCA).
- C_0: Promedio de 24 horas de PM10 medido a las 10:00 AM del día presente en la estación Pudahuel en [µg/m^3].
- T_0: Temperatura en grados °C del nivel 925 hPa medido a las 12 UTC del día presente con el radio sondeo de Rocas de Santo Domingo (80 Km. al oeste de Santiago).
- D_0: Cambio en las ultimas 24 horas para la altura del nivel 500 hPa nivel medido a las 12 UTC del día presente con el radio sondeo de Rocas de Santo Domingo.

Considerando esta ecuación como válida (con algunas correcciones debido a sus resultados no exactos en un 100%), se consideraran las variables que intervienen.

Analizando las variables de la ecuación 5.1, se observa que las variables V_0 y C_0 ya se utilizan (considerando en nuestro caso a C_0 una medición de PM2.5), las variables que esta ecuación (5.1) que en nuestro modelo no están involucradas son: T_0 y D_0.

Ambas variables están relacionadas con el fenómeno de inversión térmica, que corresponde a un fenómeno atmosférico importante cuando se refiere a ventilación y dispersión de contaminantes.

5.3. Desarrollo

Para agregar como variable a la temperatura del nivel 925 hPa y el cambio en las ultimas 24 horas de la altura del nivel 500 hPa, primero se procedió a agregar fechas a los datos que se tienen, al igual que en el capitulo 4, esto se realizo haciendo coincidir los datos que se tienen con los datos del Sistema de Información Nacional de Calidad del Aire (SINCA del Ministerio del Medio Ambiente).

Teniendo la tabla de datos con 2 variables adicionales, se modifica la creación de la red neuronal. En este nuevo caso se tendrán 17 variables de entrada y se mantendrán las 4 variables de salida.

5.4. Resultados

A continuación se muestra la tabla de datos que muestra eventos para los años 2004, 2005, 2006 y 2007.

Tabla 5.1: numero de eventos.

	numero de eventos			
	A + B	C	D	C + D
2004	61	48	20	68
2005	62	37	4	41
2006	62	54	13	67
2007	28	18	4	85

A continuación se presenta la tabla de datos que muestra los porcentajes de acierto para las distintas clases interesadas.

Tabla 5.2: Porcentajes de acierto.

	15 entradas			
	A + B	C	D	C + D
2004	92%	77%	15%	85%
2005	97%	81%	0%	83%
2006	90%	54%	0%	61%
2007	96%	61%	0%	68%

	17 entradas			
	A + B	C	D	C + D
2004	92%	83%	25%	90%
2005	97%	84%	25%	85%
2006	90%	52%	0%	60%
2007	96%	61%	0%	68%

Los porcentajes de error calculados para ambos se muestran en la tabla 5.3.

Tabla 5.3: Porcentajes de error.

	15 entradas			
	Estación La Florida	Estación Las Condes	Estación Santiago	Estación Pudahuel
2004	14%	20%	15%	22%
2005	11%	15%	13%	19%
2006	17%	22%	20%	23%
2007	5%	6%	5%	8%

	17 entradas			
	Estación La Florida	Estación Las Condes	Estación Santiago	Estación Pudahuel
2004	14%	20%	18%	22%
2005	12%	15%	14%	19%
2006	17%	22%	21%	23%
2007	5%	6%	5%	8%

Al analizar la tabla 5.3 no se observan mejoras en los porcentajes de error al utilizar 2 variables extras, de hecho analizando la estación Santiago, se observa que los porcentajes de error empeoran al utilizar 2 variables extras.

Analizando ahora la tabla 5.2 se observan que los porcentajes de aciertos para las clases A y B no hay diferencias al utilizar las nuevas variables.

Las únicas mejoras encontradas se encuentran en las clases correspondientes a los episodios críticos, estos son, clase C, clase D y clases C y D sumadas.

CAPITULO 6. CONCLUSIONES

Los resultados obtenidos al combinar ambos métodos, (Red Neuronal y Método de Cluster), son los que tienen mayores porcentajes de aciertos. Puesto que su desarrollo consiste en eligir los mejores porcentajes de aciertos para pronosticar cada clase, se espera de antemano una mejora en los pronósticos También se pudo comprobar que al cambiar el número de neuronas en la capa oculta se podía mejorar los resultados obtenidos anteriormente [9].

Al realizar la implementación de la variable de salida como un promedio de 24 horas fijas, si bien no se puede hacer una comparación estricta respecto al método tradicional de 24 horas móvil, este estudio puede servir como referencia para la creación de modelos basados en la normativa de otros países. Esta implementación no resultó muy favorable respecto a las tasas de acierto, ya que el promedio de 24 horas fijo usado corresponde al ultimo promedio móvil de 24 horas de la metodología usada anteriormente.. Este promedio tiene en general un error relativamente mayor debido a la lejanía (en tiempo) con respecto a los datos de entrada.

La introducción dentro del modelo neuronal, de las variables relacionadas con la altura de la capa de inversión térmica, usadas en modelo Cassmassi de PM10 no mostró disminución en el porcentaje de error. Sin embargo si nos centramos en los resultados referidos a los porcentajes de aciertos, se encuentra una pequeña mejoría. Esta mejoría fue solo encontrada para los

episodios críticos, pues para los episodios normales, los porcentajes de acierto se mantuvieron iguales.

Una importante conclusión general es que los actuales modelos de pronóstico de PM2,5 pueden ser mejorados por medio de una exploración adecuada de variables relevantes y algoritmos alternativos.

CAPITULO 7. REFERENCIAS

[1] Resolución exenta N° 4624, Santiago 10 de Agosto 2009 que establece Anteproyecto de norma primaria de calidad ambiental para material particulado respirable MP2.5**.** Ministerio del medio ambiente, http://www.sinia.cl/1292/w3-article-47699.html

[2] Patricio Perez; Giovanni Salini, *$PM_{2.5}$ Forecasting in Santiago, Chile: Comparison of three methods,* Atmospheric Environment 42, pp 8219-8224, 2008.

[3] Patricio Perez; Jorge Reyes, *An integrated neural network model for PM10 forecasting*, Atmospheric Environment 40, 2845-2851, 2006

[4] Patricio Perez; Alex Trier, *Prediction of NO and NO_2 concentrations near a street with heavy traffic in Santiago, Chile*, Atmospheric Environment, 35, pp 1783-1789, (2001).

[5] Patricio Perez; Alex Trier; Jorge Reyes, *Prediction of $PM_{2.5}$ concentrations several hours in advance using neural networks in Santiago*, *Chile*, Atmospheric Environment 34, pp 1189-1196, 2000

[6] Raúl G. E. Morales, *Contaminación Atmosférica Urbana, Episodios críticos de contaminación ambiental en la ciudad de Santiago*, Primera Edición, Editorial Universitaria, pp. 55-80, 2006.

[7] OMS 2006, parámetros de calidad del aire, normas para material particulado.

[8] Gobierno de Chile CONAMA-Región Metropolitana, Actualización del inventario de emisiones de contaminantes atmosféricos en la región metropolitana 2005. www.sinia.cl

[9] G.A. Salini, *Desarrollo de un modelo para pronosticar concentraciones extremas de PM2.5 en Santiago*, Tesis de Doctorado, Departamento de Física, Universidad de Santiago de Chile, Santiago, 2009.

[10] Sounding Wyoming, http://weather.uwyo.edu/upperair/sounding.html

[11] CENMA, http://aire.cenma.cl/

[12] SINCA, http://sinca.mma.gob.cl/index.php/region/index/id/M

[13] 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.

[14] SEREMI, http://www.seremisaludrm.cl/sitio/pag/aire/indexjs3aireindices-prueba.asp

[15] SINIA, Sistema Nacional de Información Ambiental, *Actualización del inventario de emisiones de contaminantes atmosféricos en la región metropolitana 2005*. http://www.sinia.cl/