



**Universidad Nacional  
Federico Villarreal**

**Vicerrectorado de  
INVESTIGACIÓN**

**ESCUELA UNIVERSITARIA DE POSGRADO**

**“REDES NEURONALES PARA PREDICCIÓN DE  
CONTAMINACIÓN DEL AIRE EN CARABAYLLO-LIMA”**

**TESIS PARA OPTAR POR EL GRADO ACADÉMICO DE:**

**MAESTRO EN INGENIERÍA DE SISTEMAS**

**AUTOR:**

**JACINTO HERRERA RAUL TRINIDAD**

**ASESOR:**

**DRA. ÁNGELES LAZO ANA MARÍA**

**JURADO:**

**DR. SOTO SOTO LUIS**

**DR. MAYHUASCA GUERRA JORGE VICTOR**

**MG. CARRILLO BALCEDA JESUS ELIAS**

**LIMA – PERU**

**2019**

## **DEDICATORIA**

"A mi madre Teodora, a mi  
esposa Nery y a mis hijos Raúl,  
Fatimah y Renzo, por ser fuente  
de constante inspiración".

Raúl Jacinto

## **AGRADECIMIENTO**

En primer lugar debo agradecer a mi núcleo familiar, quienes entendieron la dedicación de horas y días enteros dedicados al presente trabajo de tesis.

Luego quisiera agradecer a la Dra. Ana María Ángeles Lazo por su apoyo constante, su guía, sus consejos y recomendaciones a través del desarrollo de la tesis. También mi agradecimiento a la Escuela de Posgrado, a la Universidad Nacional Federico Villarreal y al cuerpo docente en pleno de la Maestría en Ingeniería de Sistemas por las valiosas enseñanzas transmitidas a mi persona en muchas horas de clases y talleres de formación académica.

Asimismo, mi agradecimiento a la Municipalidad de Carabayllo que me ha permitido gestionar el medio ambiente del entorno urbano, habiendo obtenido valiosa experiencia profesional.

## RESUMEN

En el presente trabajo se desarrolla un modelo de pronóstico capaz de predecir (con múltiples etapas adelantadas) el comportamiento de las variables de contaminación ambiental de material particulado  $PM_{2.5}$  mediante el uso de los datos de concentraciones de contaminantes de material particulado ( $PM_{2.5}$  y  $PM_{10}$ ) y químicos (CO,  $SO_2$ , NO) producidos en la estación automatizada de calidad de aire de Carabayllo.

El modelo ha sido entrenado, con datos reales de la estación automatizada de calidad de aire del distrito de Carabayllo en el intervalo de 2 años, sobre tres diferentes algoritmos de retropropagación y dos modelos de neuronas en una única capa oculta para hallar parámetros de un modelo óptimo de predicción. Experimentalmente se probaron 6 modelos de redes sobre un rango de número de neuronas. La red optimizada fue aplicada sobre un grupo de 72 datos de prueba obteniendo resultados del modelo con un error porcentual medio de -0.1089% lo cual indica un pronóstico preciso para el caso de estudio.

Un aporte de esta investigación es la demostración de que el modelo basado en redes neuronales artificiales es capaz de pronosticar variables de contaminación ambiental de material particulado con buena precisión y en una forma sencilla. Asimismo, el modelo se puede adaptar tanto para pronosticar otros contaminantes del aire (químicos o material particulado) como para datos generados en otras estaciones automatizadas de calidad de aire.

**Palabras Clave:** Redes neuronales artificiales; Retropropagación; Perceptrón multicapa; Series de tiempo; Contaminación del aire; Material particulado; Carabayllo.

## ABSTRACT

This research work develops a model of forecasting able to predict (multistep ahead) the behaviour of air pollution through values of particulate material  $PM_{2.5}$  based upon the use of data of air pollution as particulate material ( $PM_{2.5}$  and  $PM_{10}$ ) and chemical compounds ( $NO_2$ ,  $CO$ ,  $SO_2$ ) measured in the automated air quality station of Carabayllo.

The model has been trained, with actual data obtained from automated air quality station of Carabayllo during two years, on the basis of three backpropagation algorithms and two models of neuron in a alone hidden layer to get parameters of a optimal model of prediction. In a experimental way were tested 6 neural models over a range of neurons. The optimized network was applied over a group of 72 test data getting results of the model with an mean percentage error of -0.1089% which is pointing a accurate forecast for this research.

This research contributes to prove that one model based upon artificial neural network is able to forecast air pollution from particulate material with high accuracy and in an easy way. Likewise, the model can be suited to forecast other air pollution elements (chemical or Particulate material) as for data generated in other automated air quality stations.

**Keywords:** Artificial neural network; Backpropagation; Multilayer perceptron; Time series; Air pollution; Particulate material; Carabayllo.

## TABLA DE CONTENIDOS

AGRADECIMIENTO .....	iii
RESUMEN .....	iv
ABSTRACT .....	v
TABLA DE CONTENIDOS .....	vi
LISTA DE TABLAS .....	viii
LISTA DE FIGURAS .....	ix
<b>INTRODUCCION</b> .....	1
<b>CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA</b> .....	4
1.1    Antecedentes .....	4
1.2    Planteamiento del Problema .....	20
1.2.1    Descripción .....	20
1.2.2    El Problema .....	22
1.3    Objetivos .....	25
1.4    Justificación y Delimitación .....	25
1.4.1    Justificación .....	25
1.4.2    Delimitación .....	26
1.5    Alcances y Limitaciones .....	29
1.5.1    Alcances .....	29
1.5.2    Limitaciones .....	30
1.6    Definición de Variables .....	34
<b>CAPÍTULO II: MARCO TEÓRICO</b> .....	36
2.1    Teorías Generales .....	36
2.1.1    Inteligencia Artificial .....	36
2.1.2    Revisión Histórica de las Redes Neuronales .....	38
2.2    Bases Teóricas .....	41
2.2.1    Contaminación Ambiental .....	41
2.2.2    Series de Tiempo .....	47
2.2.3    Redes Neuronales Artificiales .....	52
2.3    Marco Conceptual .....	84
2.4    Hipótesis .....	87
2.4.1    Hipótesis General $H_0$ .....	87
2.4.2    Hipótesis Específicas $H_i$ .....	88
2.4.3    Operacionalización de Variables .....	88

<b>CAPÍTULO III: MÉTODO</b> .....	91
3.1 Tipo de Investigación.....	91
3.2 Diseño de Investigación.....	93
3.3 Estrategia de prueba de hipótesis.....	97
3.3.1 Error Cuadrático Medio.....	98
3.3.2 Medición del Error de Pronóstico.....	99
3.4 Variables .....	102
3.5 Población .....	103
3.6 Muestra .....	104
3.7 Técnicas de investigación .....	105
3.7.1 Instrumentos de Recolección de Datos .....	105
3.7.2 Procesamiento y Análisis de Datos .....	108
<b>CAPÍTULO IV: PRESENTACION DE RESULTADOS</b> .....	119
4.1 Contratación de Hipótesis .....	119
4.2 Análisis e Interpretación.....	139
<b>CAPITULO V: DISCUSION</b> .....	141
5.1 Discusión .....	141
5.2 Conclusiones.....	142
5.3 Recomendaciones .....	144
5.4 Referencias Bibliográficas.....	146

## LISTA DE TABLAS

<i>Tabla 1: Medidas de error de predicción.....</i>	<i>10</i>
<i>Tabla 2: Matriz de Consistencia del Proyecto .....</i>	<i>24</i>
<i>Tabla 3: Ejemplos de series de tiempo.....</i>	<i>48</i>
<i>Tabla 4: Ejemplo de formato original de datos de la Estación de Calidad de Aire ....</i>	<i>108</i>
<i>Tabla 5: Patrones de entrada (Input).....</i>	<i>114</i>
<i>Tabla 6: Patrones de entrada (Input).....</i>	<i>114</i>
<i>Tabla 7: Diseño de las simulaciones para obtener parámetros óptimos de red.....</i>	<i>121</i>
<i>Tabla 8: Entrenamiento con Levenberg-Marquardt (Neuronas [1x1]) .....</i>	<i>123</i>
<i>Tabla 9: Entrenamiento con Levenberg-Marquardt (Neuronas [1x2]) .....</i>	<i>124</i>
<i>Tabla 10: Entrenamiento con Regulación Bayesiana (Neuronas [1x1]) .....</i>	<i>125</i>
<i>Tabla 11: Entrenamiento con Regulación Bayesiana (Neuronas [1x2]) .....</i>	<i>126</i>
<i>Tabla 12: Entrenamiento con Gradiente de Escala Conjugada (Neuronas [1x1]) .....</i>	<i>127</i>
<i>Tabla 13: Entrenamiento con Gradiente de Escala Conjugada (Neuronas [1x2]) .....</i>	<i>128</i>
<i>Tabla 14: Menores valores de Error Medio Cuadrático .....</i>	<i>129</i>
<i>Tabla 15: Parámetros del Modelo Óptimo de RNA .....</i>	<i>129</i>
<i>Tabla 16: Tabla de valores óptimos de simulación para cada modelo. ....</i>	<i>130</i>
<i>Tabla 17: Estándares de Calidad Ambiental para Aire .....</i>	<i>137</i>
<i>Tabla 18: Comparativo de Valores Pronosticados con ECA.....</i>	<i>138</i>

## LISTA DE FIGURAS

<i>Figura 1: Alcances del proyecto de investigación</i> .....	30
<i>Figura 2: Limitaciones del proyecto de investigación</i> .....	32
<i>Figura 3: Enfoque sistémico del proyecto de investigación</i> .....	34
<i>Figura 4: Neurona biológica</i> .....	53
<i>Figura 5: Representación de una neuronal artificial</i> .....	54
<i>Figura 6: Esquema de una red con tres capas totalmente interconectadas</i> .....	55
<i>Figura 7: Arquitectura de un perceptrón con dos entradas y una salida</i> .....	61
<i>Figura 8: Estructura del Perceptrón Multicapa</i> .....	65
<i>Figura 9: Unidad neuronal elemental no lineal</i> .....	67
<i>Figura 10: Superficie de error para localizar un mínimo</i> .....	71
<i>Figura 11: Modelo No Lineal Auto Regresivo con entrada Externa</i> .....	78
<i>Figura 12: Esquema conceptual</i> .....	86
<i>Figura 13: Esquema de la relación funcional</i> .....	88
<i>Figura 14: Tipos de estudio según orientación</i> .....	92
<i>Figura 15: Esquema de clasificación de tipos de investigación basado en criterios</i> ....	93
<i>Figura 16: Esquema de un Sensor de partículas TEOM</i> .....	107
<i>Figura 17: Arquitectura del modelo de red neuronal de predicción</i> .....	109
<i>Figura 18: Datos originales de PM<sub>2.5</sub> – Enero a Abril de 2018</i> .....	110
<i>Figura 19: Datos válidos de PM<sub>2.5</sub> – Enero a Abril de 2018</i> .....	111
<i>Figura 20: Transformación Lineal (820 datos válidos de PM<sub>2.5</sub>)</i> .....	113
<i>Figura 21: Influencia del número de capas en las RNA</i> .....	116
<i>Figura 22: Entrenamiento con Levenberg-Marquardt (Neuronas [1x1])</i> .....	123
<i>Figura 23: Entrenamiento con Levenberg-Marquardt (Neuronas [1x2])</i> .....	124
<i>Figura 24: Entrenamiento con Regulación Bayesiana (Neuronas [1x1])</i> .....	125
<i>Figura 25: Entrenamiento con Regulación Bayesiana (Neuronas [1x2])</i> .....	126
<i>Figura 26: Entrenamiento con Gradiente de Escala Conjugada (Neuronas [1x1])</i> ...	127
<i>Figura 27: Entrenamiento con Gradiente de Escala Conjugada (Neuronas [1x2])</i> ...	128
<i>Figura 28: Herramienta de entrenamiento de la red</i> .....	131
<i>Figura 29: Modelo de Red Neuronal de Lazo Abierto</i> .....	131
<i>Figura 30: Modelo de Red Neuronal de Lazo Cerrado</i> .....	132
<i>Figura 31: Variación del MSE con el número de iteraciones</i> .....	132
<i>Figura 32: Estado de Entrenamiento de la Red Neuronal</i> .....	133
<i>Figura 33: Histograma de Errores</i> .....	133
<i>Figura 34: Resultados de Entrenamiento, Validación y Prueba</i> .....	134
<i>Figura 35: Respuesta de Salida para la Serie de Tiempo</i> .....	134
<i>Figura 36: Autocorrelación del Error de Simulación</i> .....	135
<i>Figura 37: Simulación con Parámetros Óptimos</i> .....	135
<i>Figura 38: Comparación entre resultados de simulación</i> .....	138

## INTRODUCCION

La presente tesis es un trabajo de investigación cuyo objeto principal es evaluar la factibilidad del uso de técnicas de inteligencia artificial, en particular las redes neuronales artificiales del tipo perceptrón multicapa, para pronosticar valores de contaminantes ambientales del tipo material particulado denominado PM<sub>2.5</sub> en Carabayllo-Lima-Lima. Para este objeto se tomó como muestra al conjunto de datos producidos por la estación de calidad de aire de Carabayllo durante el periodo comprendido desde Enero de 2017 hasta Abril de 2018.

Los modelos de pronóstico fueron realizados mediante redes neuronales del tipo perceptrón multicapa y se evaluaron mediante el indicador denominado error cuadrático medio ( $MSE=0.001009$ ) con lo cual se obtuvieron los parámetros óptimos del modelo de predicción. Se realizó el pronóstico mediante este modelo optimizado de redes neuronales artificiales del tipo perceptrón multicapa y esta predicción se evaluó mediante el indicador denominado error porcentual medio ( $MPE=-0.1089\%$ ).

Con estos indicadores se puede comprobar que las redes neuronales del tipo perceptrón multicapa pueden ser utilizadas para pronosticar con precisión las concentraciones futuras de contaminantes aéreos del tipo material particulado (PM<sub>2.5</sub>) en el distrito de Carabayllo. En base a este pronóstico los órganos de gobierno local pueden gestionar las medidas que correspondan para preservar la salud ambiental de la población. Además pueden implementarse modelos semejantes en otras estaciones de calidad de aire de la provincia de Lima, utilizando como modelo base el presente trabajo.

En la primera parte – Capítulo I, denominado “Planteamiento del Problema”, se presentan antecedentes académicos de casos que brinden un conocimiento previo y se describe la situación problemática. Se plantea el objetivo general y se justifica y delimita este, estableciendo alcances y limitaciones del proyecto de investigación. Luego se definen las variables de investigación.

En la segunda parte – Capítulo II, denominado “Marco Teórico”, presenta un marco teórico referencial en base al cual se desarrolló el presente trabajo, se hace una breve revisión de la inteligencia artificial y se desarrolla la evolución histórica de las redes neuronales. Luego se estudia algunos aspectos de la contaminación ambiental del aire, algunas bases teóricas de las series de tiempo y una revisión detallada de las redes neuronales artificiales del tipo perceptrón multicapa, los métodos de retropropagación, algunos tópicos sobre su arquitectura y las aplicaciones en las cuales puede usarse. Además se establece un marco conceptual para el trabajo, y se plantean la hipótesis general, las hipótesis específicas así como se operacionaliza las variables de investigación.

En la tercera parte - Capítulo III, denominado “Método”, revisa algunos aspectos del tipo de investigación a llevarse a cabo y del diseño de la investigación. Además se revisan indicadores estadísticos que son parte de la estrategia de prueba de hipótesis como son el error cuadrático medio y los indicadores de medición del error del pronóstico. Además se presentan las variables de investigación, la población, la muestra y los instrumentos de recolección de datos, así como parte de los algoritmos utilizados para adecuar las variables de investigación al procesamiento de datos mediante redes neuronales.

En la cuarta parte – Capítulo IV, denominada “Presentación de Resultados”, tiene como objeto de establecer la validez de la hipótesis general, para lo cual se realizan las simulaciones numéricas establecidas en el diseño de la investigación para hallar los parámetros óptimos y con estos realizar los pronósticos correspondientes. Con los valores pronosticados se analizan e interpretan estos valores.

En la quinta parte – Capítulo V, denominado “Discusión”, se discuten los resultados en base a los indicadores estadísticos, se elaboran conclusiones del trabajo y las recomendaciones pertinentes. Luego se presenta la bibliografía utilizada.

Finalmente, como anexos, se presentan la ficha técnica de los instrumentos a utilizar, una breve definición de términos y el programa completo basados en el software Matlab R2015a que permitieron realizar las simulaciones numéricas y los pronósticos de la concentración de contaminación aérea por material particulado  $PM_{2.5}$ , así como los cuadros generados por este programa de simulación.

## **CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA**

### **1.1 Antecedentes**

(Gardner & Dorling, 1998) han demostrado que el perceptrón multicapa puede ser entrenado para aproximar virtualmente cualquier función suavizada y mensurable. A diferencia de otras técnicas estadísticas el perceptrón multicapa no hace asunciones previas con respecto a la distribución de datos. Modela funciones altamente no lineales y puede entrenarse para generalizar con precisión cuando se le presenta datos nuevos y no conocidos. Esto lo convierte en una alternativa atractiva para desarrollar modelos numéricos, y también cuando se selecciona entre enfoques estadísticos.

Los perceptrones multicapa tienen la capacidad de aprender a través del entrenamiento, el cual consiste en asociar (relacionar) una serie de vectores de entrada y salida. Durante el entrenamiento el perceptrón multicapa es presentado repetidamente con los datos de entrenamiento y los pesos en la red son ajustados hasta que ocurra la asociación entrada-salida. Los perceptrones multicapa aprenden de manera supervisada.

Durante el entrenamiento la salida del perceptrón multicapa, para un vector dado de entrada, puede no igualar a la salida deseada. Siendo el error la diferencia entre la salida real y la deseada. El entrenamiento usa la magnitud de este error para determinar en qué grado los pesos de la red deben ajustarse de modo que el error integral del perceptrón multicapa se reduzca. Existen muchos algoritmos que pueden utilizarse para entrenar un perceptrón multicapa. Una vez entrenado, con datos de entrenamiento adecuadamente representativos, el perceptrón multicapa puede generalizar nuevos datos no vistos anteriormente.

(Mendoza & García, 2004) en su trabajo denominado “Desarrollo de un modelo de red neuronal artificial para el pronóstico de los niveles de concentración de ozono en el aire ambiente de Monterrey, N.L.” afirman lo siguiente: “En virtud a los avances logrados en los modelos matemáticos que describen la emisión, transporte, formación y desaparición de los contaminantes atmosféricos, se ha avanzado en la comprensión de la dinámica de los contaminantes de la atmósfera. A pesar de esto, mientras más complejos son los modelos, requieren más y mejor información para ser aplicados con la suficiente certidumbre de que los resultados obtenidos tendrán un valor técnico o científico. Estos modelos determinísticos requieren una gran cantidad de información que no siempre puede conseguirse y los datos disponibles no siempre aseguran buenos resultados en la aplicación del modelo además de que la inversión necesaria para obtener la información para obtener resultados confiables puede ser muy alta.

Bajo esta perspectiva, se recurre a métodos que requieren menor cantidad de información, y que pueden llegar a ser útiles para estudiar la contaminación atmosférica en zonas de la ciudad. Estos métodos se basan en técnicas estadísticas de regresión o ajuste de datos mediante técnicas numéricas para establecer las condiciones químicas de la atmósfera”. De esta aseveración se puede afirmar que es posible determinar el comportamiento de algunas variables de contaminación atmosférica solo basándose en datos disponibles bajo mediciones rutinarias y por ende replicar estas determinaciones a diferentes estaciones de medición de calidad de aire en cualquier parte del territorio. Además, con la tecnología actual, se puede establecer un servicio de evaluación del comportamiento de variables de contaminación ambiental, sin que los centros de procesamiento de datos estén ubicados en la misma localidad que la estación de medición de parámetros de calidad atmosférica.

Los autores también analizan un caso de redes neuronales utilizadas para pronosticar niveles de concentración de ozono. En este caso se empleó una red neuronal tipo perceptrón multicapa con retropropagación. Teniendo como variables de entrada seis meteorológicas y tres químicas; con estas variables se obtiene la concentración de ozono máxima para el día de interés. Se utilizaron redes con tres capas: alimentación, oculta y de salida, con seis a doce nodos en la capa oculta, con funciones de transferencia sigmoideas para la capa interna y lineales para la capa de salida. La red se entrenó con el algoritmo de Levenberg-Marquardt. Las redes se entrenaron con datos históricos de 1998 a 2001 y evaluadas con datos del 2002. Se evaluó el rendimiento comparando los valores de ozono estimados respecto a los observados, contabilizando el porcentaje de acierto sobre tres diferentes niveles de calidad de aire: buena, moderada y mala. Todas las redes mostraron buenos desempeños en sus estimaciones sin presentar diferencias significativas, en general los modelos tuvieron eficiencias mayores al 75%.

(Pérez & Reyes, 2006) en su trabajo denominado “*An integrated neural network model for PM<sub>10</sub> forecasting*”, traducido como “Un modelo integrado de redes neuronales para pronóstico de PM<sub>10</sub>”, desarrollaron un modelo integrado de redes neuronales artificiales para pronosticar el máximo de un promedio de 24 horas de concentraciones de PM<sub>10</sub> con un día de adelanto y lo aplicaron en el caso de cinco estaciones de monitoreo en la ciudad de Santiago, Chile.

Consideraron como entradas al modelo las concentraciones medidas hasta las 7pm en las cinco estaciones del día presente y valores pronosticados de variables meteorológicas. Las salidas del modelo son las concentraciones máximas esperadas para el día siguiente en el sitio de las mismas cinco estaciones. La mayor de las concentraciones en los cinco pronósticos define la calidad del aire para el siguiente día, habiendo tres niveles de calidad: bueno, malo y crítico. Los parámetros de los modelos

se ajustaron con datos de los años 2001 y 2002 para pronosticar condiciones del 2003 y datos del 2002 y 2003 para pronosticar valores del 2004. Los valores pronosticados por el modelo neuronal se compararon con los resultados obtenidos de un modelo lineal con las mismas variables de entrada y un modelo de persistencia.

De acuerdo a los resultados, el modelo neuronal parece más preciso, aunque se aprecia que es muy importante una buena elección de variables de entrada; entre estas, los autores consideran importante el conocimiento de los valores presentes y pronosticados de amplitud térmica (diferencia entre la temperatura máxima y mínima) y el pronóstico de velocidad promedio del viento. Asimismo, los autores consideran que el modelo neuronal puede usarse como una herramienta operacional para pronósticos de calidad de aire en Santiago o, con las adaptaciones apropiadas, en cualquier otra gran ciudad.

En su trabajo “A neural network forecast for daily average PM<sub>10</sub> concentrations in Belgium”, traducido como “Un red neuronal de pronóstico para el promedio diario de concentraciones de PM<sub>10</sub> en Bélgica”, (Hooybergs, Mensink, Dumont, Fierens, & Brasseur, 2005) desarrollan una herramienta de redes neuronales para pronosticar el promedio diario de concentraciones de PM<sub>10</sub> en Bélgica con un día de adelanto. Según los autores el pronóstico de PM<sub>10</sub> en el área urbana de Bélgica puede ser planteado como un problema de regresión. Sobre la base de un conjunto conocido de variables de entrada tenemos que producir una variable de salida, que es en promedio una estimación adecuada para el  $\langle \text{PM}_{10} \rangle_{\text{día } 1}$  objetivo. Esto implica el diseño de un modelo que pueda ajustar la relación entre los parámetros de entrada y objetivo sobre la base de datos históricos.

Desde que el espacio de entrada es multidimensional y la relación funcional con el objetivo es desconocida *a priori* y probablemente no lineal, las técnicas tradicionales de regresión paramétrica no son adecuadas, resultando más adecuado un enfoque de redes neuronales. Los autores basaron su investigación en mediciones desde diez sitios de monitoreo durante el período 1997-2001 y también en simulaciones de parámetros meteorológicos. La variable de entrada más importante fue la altura de capa limitante. Un modelo basado en este parámetro actualmente operativo en línea sirve para monitorear el umbral promedio diario de  $100 \mu\text{g}/\text{m}^3$ . Al expandir el modelo con otros parámetros de entrada se pudo incrementar solo ligeramente el rendimiento. Esto hace concluir a los autores que las fluctuaciones día-a-día de concentraciones de  $\text{PM}_{10}$  en áreas urbanas de Bélgica son en gran parte manejadas por las condiciones meteorológicas y en menor extensión manejadas por cambios en las fuentes antropogénicas.

(Salini-Calderon & Perez-Jara, 2006) en su investigación denominada “Estudio de series temporales de contaminación ambiental mediante técnicas de redes neuronales artificiales” diseñaron una red neuronal artificial (RNA) para predecir valores de concentraciones horarias de material particulado fino en la atmósfera. El estudio está basado en los datos de tres años de series de tiempo de  $\text{PM}_{2.5}$  (material particulado de 2,5 micrones de diámetro), obtenidos en una estación de Santiago de Chile, entre los años 1994 y 1996. Para obtener el espaciamiento óptimo de los datos, así como el número de datos hacia atrás necesarios para pronosticar el valor futuro, se aplicaron dos pruebas estándares: *Información Mutua promedio* (AMI) y *Falsos Vecinos más Cercanos* (FNN).

Se encontró que era más conveniente considerar como entrada los datos de  $\text{PM}_{2.5}$  cada seis horas durante un día (cuatro datos), y en base a ellos predecir el dato siguiente. Una vez fijo el número de variables de entrada y elegida la variable a pronosticar, se

diseñó un modelo predictivo basado en redes neuronales artificiales. Se usó el modelo denominado perceptrón multicapa, alimentado hacia adelante y entrenado mediante la técnica de propagación hacia atrás. Se probaron redes sin capa oculta y con una y dos capas ocultas. El mejor modelo resultó ser con una capa oculta, a diferencia de lo obtenido en un trabajo anterior que reportaba que la red sin capa oculta era más eficiente. Los resultados fueron más precisos que los obtenidos con un modelo de persistencia.

(Camilo-Santana, 2006) en su trabajo denominado “Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana” propone evaluar la capacidad de las redes neuronales en la predicción de series temporales. Una aplicación que pronostica valores futuros de la serie de inflación colombiana demostró que las redes neuronales pueden ser más precisas que las metodologías tradicionales como SARIMA de Box-Jenkins y el suavizamiento exponencial.

También afirma que la utilidad de las redes neuronales, en combinación con otras redes o metodologías tradicionales, se mostró efectiva en el proceso de predicción en términos del MSE. Y que las redes neuronales sin la neurona de tendencia, en la mayoría de los casos exhibieron las menores medidas de error en la predicción, en comparación con la metodología utilizada que propone la tendencia como una neurona de entrada. En todos los casos se estimaron y seleccionaron redes que tuvieran el mejor ajuste posible sobre el conjunto de entrenamiento y la menor medida de error de predicción sobre el conjunto de validación con el fin de evitar el problema de subaprendizaje. En la tabla 1 se aprecia una comparación de medidas de error de predicción tanto para modelos tradicionales como para modelos de redes neuronales.

## Medidas de error de predicción.

Modelo	<i>MSE</i>		<i>MAE</i>	
	<i>h</i> = 1	<i>h</i> = 6	<i>h</i> = 1	<i>h</i> = 6
<i>HW</i>	0.037	0.035	0.183	0.178
<i>SARIMA</i>	0.033	0.036	0.152	0.158
<i>ANN</i> (17, 5, 1)	<b>0.014</b>	0.023	<b>0.096</b>	0.125
<i>ANN</i> (15, 4, 1)	0.018	<b>0.018</b>	0.123	0.124
<i>ANN</i> (17, 2, 2, 1)	0.022	0.019	0.120	<b>0.108</b>

*Tabla 1: Medidas de error de predicción*

Adaptada de “Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana”.

(Ordieres, E.P., R.S., & Salazar, 2005) en su trabajo titulado “Neural network prediction model for fine particulate matter (PM<sub>2.5</sub>) on the US-Mexico border in El Paso (Texas) and Ciudad Juarez (Chihuahua)” traducido como “Modelo de redes neuronales para predicción de material particulado fino (PM<sub>2.5</sub>) en la frontera EEUU-México en El Paso (Texas) y Ciudad Juarez (Chihuahua)” utiliza redes neuronales para predecir las concentraciones de material particulado PM<sub>2.5</sub>. El modelo obtenido será utilizado como una herramienta de control, con el propósito de prevenir situaciones peligrosas que se puedan presentar. Se compararon tres diferentes topologías de redes neuronales para identificar sus fortalezas y debilidades: Perceptrón Multicapa (MLP), Función de Base Radial (RBF) y Perceptrón Multicapa Cuadrado (SMLP). Además, se construyeron dos modelos clásicos, (un modelo de persistencia y una regresión lineal), para así comparar los resultados con los provenientes de los modelos de redes neuronales.

Los resultados logrados con las redes neuronales mostraron que este enfoque no solo sobrepasó a los modelos clásicos sino también mostró valores similares entre las diferentes topologías. Entre los modelos de redes neuronales se tiene que el perceptrón multicapa (MLP) proporciona predicciones aceptables, a pesar de las dificultades

condiciones ambientales de la localidad (aun cuando se consideró los datos de  $PM_{2.5}$  las muestras muestran picos de inmisiones marcados estacionalmente principalmente debido a la presencia de tormentas de arena en esta área). El modelo de perceptrón multicapa cuadrático (SMLP) muestra un comportamiento muy similar aun cuando se necesitan algunas neuronas más en la capa oculta para obtener el mismo error del caso anterior; este hecho tiene una influencia directa en el tamaño de muestra necesario para el correcto entrenamiento.

(Aktan & Bayraktar, 2009) Según estos autores, en su trabajo denominado “*The neural network modeling of suspended particulate matter with autoregressive structure*” traducido como “El modelamiento de material particulado suspendido mediante redes neuronales con estructura autoregresiva”, el conocimiento de las fuentes de contaminación y las emisiones, y sus interacciones con el terreno y la atmósfera son necesarios en el desarrollo de planes de gestión adecuada y estrategias de acción de la contaminación del aire. En este estudio, se investiga la relación entre la concentración total de partículas en suspensión (TSP: Total Suspended Particles, por sus siglas en inglés) y los parámetros meteorológicos, como la velocidad y dirección del viento, la temperatura, la presión del aire, precipitación y humedad relativa. Se investigaron las mediciones de TSP de los últimos dos días, duración de la insolación y la cantidad de sol en la temporada de invierno (noviembre a marzo) en la ciudad de Erzurum entre los años 1990 y 2007.

Los modelos de redes neuronales artificiales (RNA) se construyeron usando una relación auto-regresiva mixta para darse cuenta de la naturaleza estocástica de los niveles de TSP para cada mes de la temporada de invierno. El impacto de la dirección del viento sobre la concentración de TSP se introdujo en el modelo mediante la definición de dos

variables. También se construyeron modelos de regresión lineal para comprobar el rendimiento de las redes neuronales. Los factores más importantes que afectan a la concentración de TSP se tiene que son: el nivel TSP del día anterior ( $TSP_{T-1}$ ), la temperatura esperada ( $t$ ), la velocidad del viento ( $w$ ), la presión de aire ( $p$ ) y la precipitación ( $pc$ ).

(Benvenuto & Marani, 2000) en su trabajo titulado “Neural networks for environmental problems: data quality control and air pollution nowcasting”, es decir “Redes neuronales para problemas ambientales: control de calidad de datos y pronóstico de contaminación del aire”; refieren que las RNA se aplican a los siguientes problemas: i) predicción de término corto y medio para la concentración de contaminantes del aire en áreas urbanas, ii) interpolación y extrapolación de temperatura máxima diaria, iii) reemplazar información distribuida en el tiempo por la distribuida espacialmente (concentración de contaminantes en diferentes sitios de medición). Se comparan datos observados versus pronosticados para probar la eficacia de las RNA en simular procesos ambientales. Los resultados confirman a las RNA como una mejora de los modelos clásicos y muestran la utilidad de las RNA para predicción de series de tiempo.

Los autores muestran que algunos de los problemas difíciles de resolver con los enfoques tradicionales pueden ser manejados con RNA; en particular demuestran que estas redes pueden pronosticar: i) concentraciones de contaminantes de aire en áreas urbanas como funciones de algunas variables químicas y parámetros meteorológicos locales; y ii) interpretar comportamientos espaciales, alcanzando mejores resultados que en el caso donde solo está disponible información temporal. Los resultados confirman la

utilidad de arquitecturas neuronales similares para predecir de contaminantes del aire para áreas urbanas e industriales.

Los autores afirman que las RNA son adecuadas para investigaciones que involucran grandes conjuntos de datos y para problemas en los cuales los valores de entrada y sus correspondientes valores de salida son conocidos pero las relaciones entre las entradas y las salidas son difíciles de relacionar con las técnicas usuales de análisis. Estas condiciones son usualmente encontradas en muchas aplicaciones sobre calidad del aire, debido a las relaciones entre los procesos físicos y químicos en los sistemas ambientales. Las RNA no elimina el trabajo de preparación de los datos: en efecto el conjunto de estos, antes de ser procesados requieren un análisis estadístico cuidadoso para descubrir vínculos entre las variables a fin de evitar patrones redundantes y mejorar la precisión.

El enfoque descrito será un componente para la implementación de una red de monitoreo del aire en el área urbana alrededor de Venecia diseñado para evaluar programas regulatorios para mejorar la calidad ambiental del aire. Las RNA pueden ser utilizadas para controlar la contaminación de la calidad del aire urbano mediante regulación del flujo de tráfico durante condiciones severas del tiempo. Las RNA fueron confirmadas como un enfoque computacional que mejora los modelos clásicos en los cuales la solución se “aprende” desde un conjunto de ejemplos. Es importante notar que las RNA son técnicas de diagnóstico de reconocimiento de patrones, con algunas habilidades predictivas, permitiendo la asimilación de grandes cantidades de datos.

Considere lo siguiente: a) Las RNA pueden proporcionar un soporte para llevar a cabo control de calidad para datos experimentales; b) los resultados pueden mejorar cuando están disponibles datos de áreas geográficamente distribuidas; c) a fin de reemplazar datos no encontrados o evaluar datos aislados (*outliers*) es aconsejable

realizar entrenamiento y pronóstico de las series sobre los mismos datos; d) los pesos proporcionan información útil sobre las variables de pronóstico más importantes y sus vínculos relevantes; e) las RNA son particularmente adecuadas cuando están disponible una gran cantidad de datos (desde que las redes se entrenan mediante ejemplos, a más datos que estas experimentan, mejor es el rendimiento de la red) y cuando no existe una solución simple mediante las técnicas tradicionales.

El trabajo de (Benvenuto & Marani, 2000) muestra que las arquitecturas neuronales son más flexibles que los modelos lineales para reproducir los diferentes comportamientos de los datos. Esto permite plantear las siguientes consideraciones con respecto a las arquitecturas neuronales.

- i) La no linealidad general de las conexiones entrada-salida, incluyendo como un caso especial el límite lineal (el mejor modelo ARMAX puede convertirse en un caso particular de las RNA),
- ii) La gran importancia de las neuronas ocultas como neuronas activas capaces de hacer transformaciones no lineales mediante las funciones de transferencia,
- iii) La complejidad funcional de la arquitectura neuronal está relacionada al número de pesos.

(Tecer, 2007) en su artículo denominado “*Prediction of SO<sub>2</sub> and PM concentrations in a coastal mining area (Zonguldak, Turkey) using an artificial neural network*” es decir, “Predicción de concentraciones de SO<sub>2</sub> y MP en un área minera costera (Zonguldak, Turquía) usando una red neuronal artificial”; propone redes neuronales artificiales para predecir las concentraciones de SO<sub>2</sub> y MP en dos diferentes estaciones en Zonguldak, una ciudad principal en el área minera costera en Turquía. Los modelos establecidos de redes neuronales artificiales involucran parámetros meteorológicos y datos históricos de SO<sub>2</sub>

y material particulado (MP) observados como variables de entrada. Los modelos se basan en la red neuronal de tres capas entrenada por un algoritmo de retropropagación.

Los modelos midieron con precisión la tendencia de las concentraciones de SO<sub>2</sub> y MP. También explicaron la variación diaria de concentración de contaminantes y predijeron con efectividad los valores observados para la ciudad de Zonguldak. Se puede concluir que el conocimiento de la secuencia de valores pasados de concentraciones de contaminación del aire es de alta significancia a fin de predecir sus valores futuros. Las condiciones meteorológicas en el momento de predicción también tienen un efecto importante sobre el modelamiento de contaminación del aire. También se demostró que el modelo RNA proporciona un buen ajuste con los valores medidos de contaminantes del aire. El modelado y monitoreo de calidad del aire sirve para evaluar el éxito de las políticas, proyectos y programas disponibles.

(Barbes, Neagu, Melnic, Ilie, & Velicu, 2009) en su trabajo titulado “*The use of Artificial Neural Network (ANN) for prediction of some airborne pollutants concentration in urban areas*”, traducido como: “Uso de redes neuronales artificiales (RNA) para predicción de concentración de contaminantes aerotransportados en áreas urbanas”; proponen un modelo para la predicción de la concentración de contaminantes inorgánicos transportados por el aire H<sub>2</sub>S-SO<sub>2</sub>, NO-NO<sub>2</sub>-NO<sub>x</sub>, CO-(CO<sub>2</sub>) y PM<sub>10</sub> (material particulado con un diámetro aerodinámico de 10µm o menos) desde un área de riesgo (dos áreas industriales y un área urbana) de Constanta.

El modelo usa funciones no lineales de aproximación, las redes neuronales artificiales (RNA). Estas son utilizadas para los sistemas de fenómenos para los cuales la linealidad entre las diferentes variables no puede ser determinada o aproximada. La RNA puede simular este fenómeno y valores de la concentración de contaminantes del área

urbana a ser estudiada. La presente RNA se usa para entrenar un número pequeño de variables y un gran número de datos (valores medidos). Para el desarrollo y validación del modelo es necesario tener un sistema de monitoreo adecuado y continuo para los datos del fenómeno químico analizado.

Los autores presentan un modelo que usa RNA para pronosticar la concentración de contaminantes aéreos inorgánicos en dos locaciones industriales y un área urbana de Constanta. El uso de RNA, comenzando desde el análisis y finalizando con la prueba y preguntas, puede considerarse un éxito para el objetivo propuesto. Aun trabajando con diferentes variables y datos que no pueden ser simulados con algoritmos determinísticos los resultados de la investigación pueden ser considerados para seleccionar las RNA para uso futuro para previsión y control de los componentes químicos que contaminan la atmosfera.

La RNA usada es la primera de una investigación más elaborada, pero debido a su éxito, se considera que puede usarse para futuros problemas con más complejidad. A fin de superar los problemas originados por la heterogeneidad del CO y H<sub>2</sub>S, el modelo RNA puede ser fácilmente modificado para brindar mejores resultados (minimizando el error de la red) y también los parámetros de entrenamiento (mientras que el coeficiente de propagación y la tasa de aprendizaje puede ser cambiado ligeramente). El uso de las RNA en este artículo es una forma rápida y eficiente de evaluación de los parámetros ambientales medidos. Así, se recomienda la conexión entre las estaciones locales para control ambiental, a fin de ofrecer datos continuos para un procesamiento inmediato y para rápidas decisiones ecológicas por las agencias competentes.

(Thomas & Jacko, 2007) en su trabajo titulado: *“Model for forecasting expressway fine particulate matter and carbon monoxide concentration: Application of regression and*

*neural network models*” traducido como “Modelo para pronosticar material particulado fino y monóxido de carbono de autopistas: Aplicación de regresión y modelos de redes neuronales” afirman que tanto la concentración de  $PM_{2.5}$  como la de CO dependen de un gran número de factores, incluyendo el tráfico, parámetros meteorológicos, reacciones químicas, calidad regional del aire, y parámetros regionales de transporte de contaminantes.

Los autores también mencionan que los modelos determinísticos son computacionalmente más complicados y requieren datos meteorológicos complejos. En cambio afirman que el enfoque de la regresión lineal multivariada se usa comúnmente como una alternativa al enfoque determinístico. También demuestran el uso exitoso de regresión lineal y los modelos de RNA para pronosticar concentraciones de  $PM_{2.5}$  y CO en la autopista. Los indicadores de calidad estadística sugieren que ambos modelos proporcionan una precisión razonable al predecir concentraciones horarias de  $PM_{2.5}$ . Para  $PM_{2.5}$ , el modelo de regresión rindió ligeramente mejor que el modelo RNA como se evidenció por los bajos valores de Error Medio Absoluto. Esto sugiere la existencia de *no linealidades* entre las variables predictoras usadas en este estudio.

(Braga A. L., y otros, 2001) realizan su trabajo titulado “*Health effects of air pollution exposure on children and adolescents in São Paulo, Brazil*”, traducido como: “Efectos en la salud de la exposición a la contaminación del aire sobre niños y adolescentes en Sao Paulo, Brasil”. Este trabajo, los autores muestran una asociación significativa entre la contaminación del aire y las admisiones hospitalarias respiratorias para niños y adolescentes en Sao Paulo, y este efecto varía por edad y contaminante específico del aire. Los niños de dos años o menos parecen ser los el grupo más susceptible, y los adolescentes (mayores de 13 años) fueron también afectados significativamente por la

contaminación del aire. Análisis posteriores estratificando las admisiones respiratorias por causas específicas podrían ayudar a aclarar el rol específico de los contaminantes del aire en la producción de morbilidad respiratoria en sujetos susceptibles.

En la revisión del trabajo de (Pitarque, Ruiz, & Roy, 2000) denominado “Las redes neuronales como herramientas estadísticas no paramétricas de clasificación” se analiza las relaciones entre redes neuronales y técnicas estadísticas convencionales. Estos afirman que, aún con resultados no concluyentes, en general las redes neuronales han mostrado una capacidad clasificatoria igual o superior que las técnicas estadísticas, con la ventaja de poderse utilizar independientemente del cumplimiento de los supuestos teóricos relativos a estas técnicas (denominadas por ello «técnicas no paramétricas»).

Se presentan una serie de simulaciones y aplicaciones sobre datos reales que apoyan esta idea. Se contrastan redes neurales del tipo perceptrón multicapa contra modelos de regresión logística y análisis discriminante en tareas de clasificación, manipulando los patrones de correlación existentes entre los predictores (o variables de entrada) por un lado, y entre predictores con el criterio (variable de salida) por otro. Los resultados muestran que las redes neurales clasifican mejor que las técnicas estadísticas, incluso en la condición *a priori* más favorable a éstas, es decir, cuando existen altas correlaciones entre los predictores con el criterio pero bajas con relaciones entre los predictores

(Alkasassbeh, Sheta, Faris, & Turabieh, 2013) en su artículo “*Prediction of PM<sub>10</sub> and TSP air pollution parameters using artificial neural network autoregressive, external input models: A case study in Salt, Jordan*”, es decir “Predicción de parámetros de contaminación de aire PM<sub>10</sub> y sólidos suspendidos totales usando modelos de redes

neuronales artificiales autoregresivas con entrada externa: Un estudio de caso en Salt, Jordania”; desarrollaron un modelo de Red Neuronal Artificial (RNA) no paramétrica para predecir material particulado  $PM_{10}$  y partículas suspendidas totales en Salt, Jordania.

El presente estudio utilizó un conjunto de datos colectados en los alrededores de la planta de cemento Al-Fuhais en el período de un año (2006-2007) por ocho estaciones de monitoreo. Los modelos de redes neuronales artificiales propuestas consideraron los parámetros meteorológicos: Temperatura, Humedad Relativa, Velocidad del Viento como entradas. Los autores desarrollaron dos modelos de entrada del tipo redes neuronales artificiales autoregresivas con entrada externa (NARX) para proporcionar modelamiento de alto rendimiento para los parámetros de contaminación del aire denominados  $PM_{10}$  y partículas suspendidas totales. Los resultados experimentales muestran que el modelo NARX proporciona buenos resultados de modelamiento utilizando un número limitado de mediciones.

Según (Palazón-González & García-Guzmán, 2004) en su trabajo denominado: “*Modelado de series climatológicas mediante una red neuronal artificial*” considera que las RNA constituyen una alternativa a la generación de modelos empíricos y han sido utilizadas con éxito en tareas de modelado de sistemas y ajuste de funciones no lineales. Su flexibilidad para el ajuste de datos con gran dispersión numérica, junto con la ventaja que supone no tener que precisar el enunciado de una relación matemática. Como funciones de transferencia, las más empleadas son las funciones sigmoideal e identidad. Aunque puede utilizarse cualquier función, es necesario que ésta sea continua y derivable si se va a emplear un método de optimización que haga uso de derivadas. Los autores afirman que, si las funciones de transferencia admiten sucesivas derivadas no nulas, son necesarios menos parámetros y una única capa oculta para conseguir un mismo error.

## 1.2 Planteamiento del Problema

### 1.2.1 Descripción

Así como en la provincia de Lima Metropolitana, en el Distrito de Carabaylo, también son recurrentes los problemas de salud por contaminación ambiental por material particulado (MP) debido a diversos factores entre los cuales se tiene:

- i. La presencia de muchas calles y avenidas sin asfalto o con capa asfáltica deteriorada, ausencia y/o déficit de áreas verdes en diversas bermas centrales de avenidas.
- ii. La presencia de canteras de explotación de minerales no metálicos en las zonas periféricas del distrito, lo que produce un tránsito constante de camiones con minerales no metálicos procesados por las principales avenidas del distrito.
- iii. La topografía del distrito, hace que en la falda de los cerros (estribaciones o contrafuertes andinos) se acumule los contaminantes que provienen de la flota vehicular de Lima Metropolitana.
- iv. Además de esto, existen contaminantes de otro origen, cuyo estudio va más allá del objetivo del presente trabajo.

De acuerdo a (Sanchez-Ccoyllo, 2008) la São Paulo State Environmental Protection Agency, CETESB (Compañía de Tecnología de Saneamiento Ambiental) afirma que el 97% de todas las emisiones de hidrocarburos son emitidos por vehículos, y el 40% de todas las emisiones de material particulado inhalable ( $PM_{10}$ ) provienen de fuentes móviles, en base a esto puede estimarse que también el material particulado  $PM_{2.5}$  provenga de fuentes móviles.

Para la zona en mención, la contaminación ambiental, entre ellas la originada por material particulado viene a ser un reclamo permanente de la población, es a raíz de esto y en concordancia con sus políticas de desarrollo que el SENAMHI, en trabajo coordinado con la Municipalidad de Carabayllo, instala en el mes de junio de 2014 una Estación automatizada de Calidad de Aire en el distrito de Carabayllo; la cual ha permitido tener una serie de datos continuados sobre ciertos parámetros de calidad de aire como son: i) material particulado  $PM_{2.5}$ , ii) material particulado  $PM_{10}$ , iii) Dióxido de Azufre, iv) Monóxido de Carbono, v) Ozono y; vi) Dióxido de Nitrógeno.

Para un mejor manejo de las políticas ambientales de los diferentes niveles de gobierno, se hace necesario tener una proyección de lo que pueda ocurrir en materia de problemas ambientales, siendo así relevante establecer métodos de pronóstico de variables de contaminación ambiental para tomar decisiones proactivas ante los problemas de contaminación mencionados. En este punto debemos señalar que, si bien el presente estudio se hace para evaluar la predicción de variables de contaminación ambiental en la zona de estudio (Carabayllo-Lima-Lima), el enfoque metodológico puede aplicarse para predecir las variables de contaminación ambiental en cualquier estación automatizada de calidad de aire de Lima Metropolitana y por extensión del país. Siempre que exista una provisión suficiente de datos que permita hacer uso de la técnica de redes neuronales propuesta.

Para la proyección de datos existen diversas técnicas, entre las cuales las redes neuronales artificiales han demostrado ser capaces de realizar una proyección más eficiente de datos futuros. Al tener datos monitoreados en forma continua, es posible obtener una serie de tiempo en la cual las variables de contaminación ambiental por

material particulado  $PM_{10}$  y  $PM_{2.5}$  y contaminantes químicos ( $SO_2$ ,  $CO$ ,  $O_3$  y  $NO_2$ ) varían en función al tiempo transcurrido. El trabajo se realizará considerando las variaciones estacionales que pudieran ocurrir y ofrece como resultado valores futuros cercanos de las variables de contaminación ambiental que permitan establecer políticas preventivas para remediar situaciones no deseadas.

El presente estudio se enfoca a la investigación de una zona geográfica específica, ubicada a la altura del km 16 a 22 de la Carretera Lima-Canta, La investigación pretende demostrar que es posible hacer un pronóstico adecuado para las emisiones de material particulado mediante el uso de la técnica denominada redes neuronales artificiales. Bajo la premisa de que es posible hacer una predicción de los parámetros de calidad del aire para la zona de estudio, se propone diseñar y probar un modelo matemático basado en redes neuronales artificiales que sea capaz de predecir, con un margen razonable de error, las proyecciones de las variables de contaminación ambiental. Siendo el objetivo establecer los parámetros del modelo de red neuronal artificial (RNA) del tipo perceptrón multicapa más adecuado que permita realizar este pronóstico y por ende ayudar en forma proactiva a tomar las decisiones más adecuadas para la mitigación de este problema ambiental y con ello los problemas de salud que genera.

## **1.2.2 El Problema**

### **Problema Principal**

¿Cómo puede el distrito de Carabayllo obtener en el futuro cercano indicadores de contaminación ambiental por material particulado que permitan respuestas precisas y oportunas mediante técnicas de inteligencia artificial para que el gobierno local pueda

tomar decisiones adecuadas que permitan remediar los episodios de contaminación que superen los estándares de calidad ambiental?

### **Problemas Específicos**

- ¿Cómo se puede obtener un modelo óptimo de redes neuronales artificiales tipo perceptrón multicapa?
- ¿Cómo se puede utilizar un modelo óptimo de redes neuronales de tipo perceptrón multicapa que permita realizar un pronóstico adecuado del  $PM_{2.5}$  para un período futuro cercano?
- ¿Cómo pueden los entes competentes del gobierno local tomar decisiones adecuadas para mitigar en forma proactiva los casos de contaminación ambiental conociendo los valores futuros cercanos de contaminación ambiental por material particulado  $PM_{2.5}$ .

Problema		Objetivo		Hipótesis	
Principal	Específico	Principal	Específico	Principal	Específico
¿Cómo puede el distrito de Carabayllo obtener en el futuro cercano indicadores de contaminación ambiental por material particulado que permitan respuestas precisas y oportunas mediante técnicas de inteligencia artificial para que el gobierno local pueda tomar decisiones adecuadas que permitan remediar los episodios de contaminación que superen los estándares de calidad ambiental?	¿Cómo se puede obtener un modelo óptimo de redes neuronales artificiales tipo perceptrón multicapa?	Obtener en el futuro cercano indicadores de contaminación ambiental mediante redes neuronales para predicción de la variable denominada Material Particulado PM <sub>2.5</sub> en base a series de tiempo de material particulado y contaminantes químicos generadas por la Estación de Calidad de Aire de Carabayllo, para que el gobierno local pueda tomar decisiones adecuadas que permitan remediar los episodios de contaminación que superen los estándares de calidad ambiental.	Determinar los parámetros óptimos del modelo de redes neuronales artificiales tipo perceptrón multicapa que permitan obtener indicadores confiables de contaminación ambiental.	¿Es posible predecir valores futuros cercanos de material particulado PM <sub>2.5</sub> (Y) mediante la aplicación de redes neuronales artificiales del tipo perceptrón multicapa (X) para la zona geográfica definida en Carabayllo-Lima-Lima?	¿Es posible determinar el modelo óptimo de redes neuronales artificiales del tipo perceptrón multicapa?
	¿Cómo se puede utilizar un modelo óptimo de redes neuronales de tipo perceptrón multicapa (X) que permita realizar un pronóstico adecuado del PM <sub>2.5</sub> para un período futuro cercano?		Obtener un pronóstico adecuado de contaminación ambiental por material particulado PM <sub>2.5</sub> mediante una red neuronal del tipo perceptrón multicapa con sus parámetros optimizados.		¿Es posible utilizar el modelo óptimo de redes neuronales artificiales del tipo perceptrón multicapa (X) para realizar una predicción adecuada del valor de material particulado PM <sub>2.5</sub> (Y) para periodos futuros cercanos?
	¿Cómo pueden los entes competentes del gobierno local tomar decisiones adecuadas para mitigar en forma proactiva los casos de contaminación ambiental conociendo los valores futuros cercanos de contaminación ambiental por material particulado PM <sub>2.5</sub> ?		Comparar los valores pronosticados de PM <sub>2.5</sub> con los Estándares de Calidad Ambiental y emitir las recomendaciones técnicas para que los entes competentes del gobierno local tomen las decisiones adecuadas para mitigar problemas de contaminación ambiental.		¿Es posible comparar los valores pronosticados de PM <sub>2.5</sub> (Y) con los estándares de calidad ambiental para evaluar la contaminación ambiental para la zona geográfica definida en Carabayllo-Lima-Lima?

Tabla 2: Matriz de Consistencia del Proyecto

### **1.3 Objetivos**

#### **Objetivo General**

Obtener en el futuro cercano indicadores de contaminación ambiental mediante redes neuronales para predicción de la variable denominada Material Particulado  $PM_{2.5}$  en base a series de tiempo de material particulado y contaminantes químicos generadas por la Estación de Calidad de Aire de Carabaylo, para que el gobierno local pueda tomar decisiones adecuadas que permitan remediar los episodios de contaminación que superen los estándares de calidad ambiental.

#### **Objetivos Específicos**

- Determinar los parámetros óptimos del modelo de redes neuronales artificiales tipo perceptrón multicapa que permitan obtener indicadores confiables de contaminación ambiental.
- Obtener un pronóstico adecuado de contaminación ambiental por material particulado  $PM_{2.5}$  mediante una red neuronal del tipo perceptrón multicapa con sus parámetros optimizados.
- Comparar los valores pronosticados de  $PM_{2.5}$  con los Estándares de Calidad Ambiental y emitir las recomendaciones técnicas para que los entes competentes del gobierno local tomen las decisiones adecuadas para mitigar problemas de contaminación ambiental.

### **1.4 Justificación y Delimitación**

#### **1.4.1 Justificación**

Esta investigación se justifica debido a que la contaminación por partículas ocasiona daño sobre la salud de la población expuesta. Desde que es imposible evitar la ocurrencia de este tipo de contaminación se requiere enfocar los esfuerzos en gestionar en forma proactiva el manejo de las situaciones ocasionadas. Así, en base al conocimiento e información generada por el presente trabajo pueden formularse políticas, proyectos y programas que contribuyan a la protección y salvaguarda de la salud, en especial de la población vulnerable, ante las ocurrencias de contaminación ambiental por material particulado de hasta 2.5 micras de diámetro que exceda los estándares de calidad ambiental nacionales. Asimismo el modelo estudiado, con ligeras variaciones, puede ser utilizado para determinar pronósticos de todas las variables que se producen en la Estación de Calidad de Aire del distrito de Carabayllo. También es posible aplicar el modelo desarrollado para pronosticar datos en cualquiera de las 10 estaciones automatizadas de calidad de aire del SENAMHI.

#### **1.4.2 Delimitación**

Se observa que existen problemas en el campo de la contaminación ambiental y se desea iniciar una investigación al respecto. Se conoce que el tema es muy amplio y el estudio de todos los aspectos demandaría un esfuerzo muy grande, ya que existen muchos temas relacionados a esa problemática. Además, las soluciones deben darse de inmediato. Se decide así enfocar el estudio en un solo aspecto de la situación y para ello se inicia un proceso de análisis del problema consultando fuentes bibliográficas, informes de investigaciones pasadas sobre el tema que aspira estudiar, consulta a expertos, con el objeto de recabar información para aclarar y precisar el problema, es decir, especificar con exactitud el campo de estudio, abarcando solo un aspecto relacionado con el problema de contaminación.

Esta consulta indica que los aspectos más resaltantes relacionados con el tema de la contaminación ambiental son: a) contaminación de las aguas; b) contaminación del aire y c) contaminación del suelo. Al tener estos tres casos resaltantes del tema de contaminación ambiental, se observa dos cosas: la primera, que son temas todavía muy amplios para estudiarlos todos a la vez; segundo, que no se posee el tiempo ni los recursos para ello. Por tanto, se decide tomar solo el aspecto de la contaminación del aire, donde aún se pueden apreciar dos grandes aspectos: contaminación por partículas y por contaminantes químicos, campos que aun cuando están relacionados, involucran un mayor esfuerzo para su estudio conjunto. De estos campos, se ha elegido estudiar el tema de contaminación por partículas, sin desmerecer la importancia de los otros temas, con el objeto de tener conocimiento pleno de los efectos de la contaminación ambiental.

Así, el problema de investigación se basa en el desarrollo de un modelo matemático para predicción de la variable de contaminación ambiental denominada  $PM_{2.5}$  en base a redes neuronales artificiales del tipo perceptrón multicapa que manipulen series de tiempo de este contaminante en conjunción con la evolución en el tiempo de otros contaminantes tanto de material particulado como químicos y de ser necesario otras variables que posibiliten minimizar el error de predicción. En base a este desarrollo el problema de investigación debe concluir en recomendaciones para que quienes ejercen la administración de la ciudad tomen las decisiones correspondientes con la mejor información posible.

En las líneas posteriores se realiza la delimitación del problema de investigación en el tiempo y espacio, en cuanto a los conceptos a utilizar y asimismo en referencia a los organismos estatales que tienen relevancia para la investigación realizada.

### **Delimitación Espacial**

El estudio de caso se llevará a cabo en el Distrito de Carabayllo, ubicado al norte de la provincia de Lima Metropolitana en la zona denominada Lima Norte a la altura de los km 16 a 22 de la Carretera Lima-Canta (Av. Túpac Amaru). El distrito a estudiar tiene una extensión total del 346 km<sup>2</sup> y un área urbana de 70 km<sup>2</sup>, en términos específicos el estudio se llevará a cabo en la Estación Meteorológica y de Calidad de Aire ubicada en la Urb. Santa Isabel a la altura del km 16 de la Carretera Lima-Canta (Coordenadas UTM 18L 278497.05 E y 8683455.00 S, a 182 msnm).

### **Delimitación Temporal**

Es una investigación de actualidad, es decir la evaluación se realizará el año 2018, con una evaluación preliminar de aproximadamente 2,888 muestras para el intervalo de tiempo desde Enero de 2018 hasta Abril de 2018.

### **Delimitación Conceptual**

Se trabajará el marco teórico con referencias a las redes neuronales artificiales y sus aplicaciones para predicción en temas ambientales, los efectos sobre la salud de los contaminantes ambientales denominados material particulado y contaminantes químicos, los métodos analíticos para determinarlos, así como las técnicas de muestreo de material particulado que ocurren en la estación meteorológica.

### **Delimitación Funcional**

Esta investigación será realizada con el concurso de la Municipalidad de Carabayllo y el Servicio Nacional de Meteorología e Hidrografía del Perú (SENAMHI), quienes proporcionarán los datos necesarios para diseñar el modelo de redes neuronales

artificiales del tipo perceptrón multicapa y a quienes se les informará los resultados obtenidos para la toma de decisiones correspondientes.

## **1.5 Alcances y Limitaciones**

### **1.5.1 Alcances**

Dentro del campo de conocimiento de la contaminación ambiental, con un enfoque muy simplificado para fines del presente trabajo, tenemos el estudio de la contaminación del aire, del suelo y del agua, si consideramos este universo de contaminación ambiental, seleccionamos solamente la contaminación del aire. Dentro de este campo consideramos la división entre contaminación por material particulado y por contaminantes químicos. Seleccionamos como nuevo universo de datos la contaminación ocasionada por material particulado. A su vez, esta contaminación por material particulado enfoca temas como el estudio de partículas de hasta 10 micras, partículas de hasta 2.5 micras y, según tendencias más recientes, partículas de hasta 0.1 micras. Para el caso de estudio hemos seleccionado la contaminación por partículas de hasta 2.5 micras.

A partir de este punto y considerando este tipo de contaminación como un nuevo universo, hacemos un quiebre en la consecución del alcance del proyecto de investigación y consideramos los conjuntos de técnicas de predicción, divididos en métodos matemáticos y modelos físicos. Al considerar el estudio solamente de los métodos matemáticos un nuevo universo, tenemos los métodos denominados deterministas (descripción de los procesos físico-químicos de la atmosfera) y los métodos estadísticos; estos a su vez pueden ser divididos en dos subconjuntos basados en redes neuronales artificiales y el conjunto de otros métodos tales como el histograma, el kernel, el falso vecino más cercano y los árboles de clasificación y regresión. En base a esta clasificación,

que se aprecia esquematizada en la figura siguiente, es que restringimos el alcance a la predicción de la variable de contaminación ambiental denominada material particulado de hasta 2.5 micras de diámetro mediante la aplicación de una técnica matemática del tipo estadística (regresión o ajuste de datos) denominada redes neuronales artificiales.

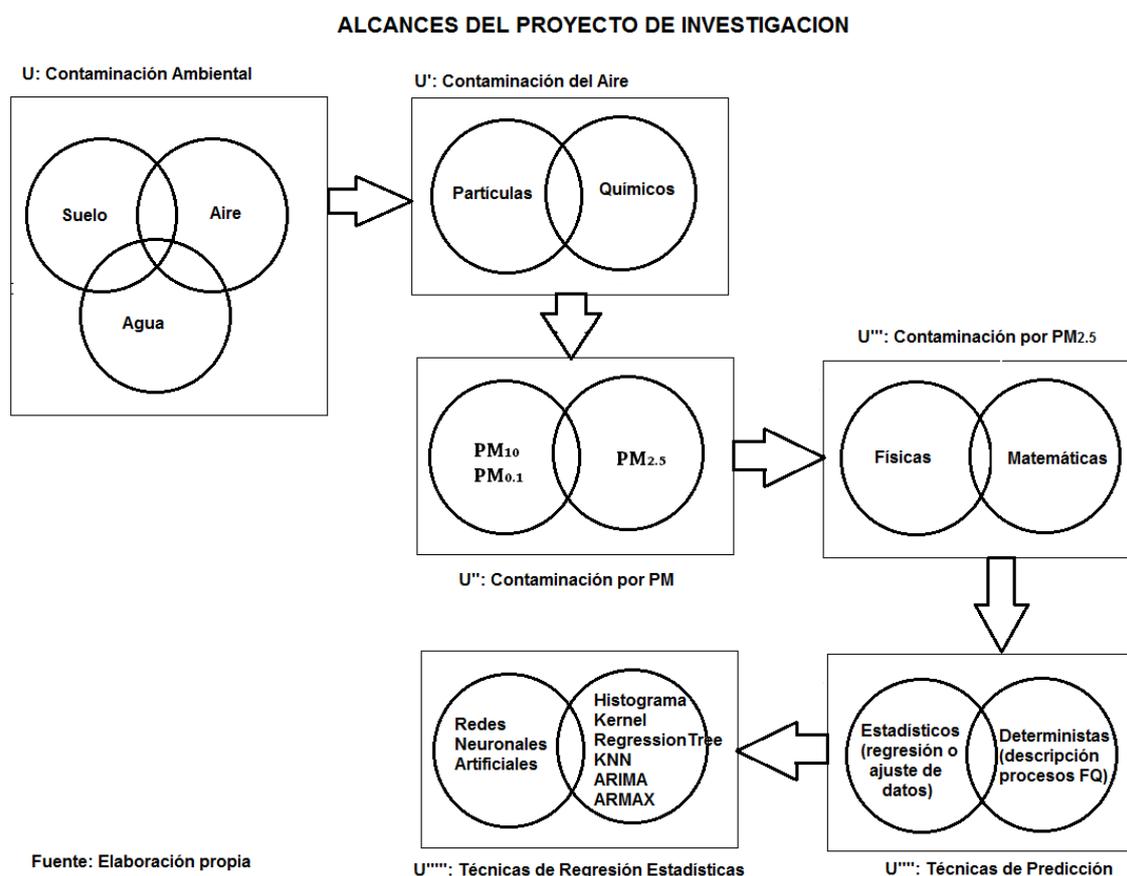


Figura 1: Alcances del proyecto de investigación

### 1.5.2 Limitaciones

Entre las limitaciones, esquematizadas en la figura siguiente, que se presentan para realizar el proyecto de investigación denominado: “Redes neuronales artificiales para predicción de contaminación del aire en Carabayllo-Lima” se tienen las siguientes:

- La presente investigación en el tiempo sólo comprende los datos de operación de la estación de calidad de aire de Carabayllo que corresponden a dos años calendario.

- Aun cuando la estación de calidad de aire de Carabayllo genera datos de contaminación ambiental por material particulado y contaminantes químicos, el presente trabajo solo se limitará al tema de contaminación por material particulado, sin dejar de mencionar la relevancia que también posee los contaminantes químicos, los cuales merecen un estudio aparte por su complejidad.
- La investigación solo utiliza las técnicas no paramétricas como son las redes neuronales artificiales, sin embargo se conoce y acepta que las técnicas estadísticas paramétricas también pueden ser fuente de conocimiento para el tema en discusión
- El presupuesto de investigación solamente proviene de fuentes propias del investigador, siendo esto un limitante para la adquisición de más recursos técnicos, bibliográficos y otros para mejorar en enfoque teórico del tema en estudio.
- La carencia de recursos para investigar a tiempo completo, implica que solo se dedican pocas horas a la semana para realizar esta investigación.
- Existe una carencia de material bibliográfico disponible a nivel comercial, lo cual obliga a trabajar con los escasos libros existentes y restringe la utilización de libros más actualizados en el tema de estudio. Una mención aparte debe hacerse para indicar que gran parte de este trabajo es en base a artículos de investigación descargados desde la página web de google académico (<https://scholar.google.com>)

## LIMITACIONES DEL PROYECTO DE INVESTIGACION

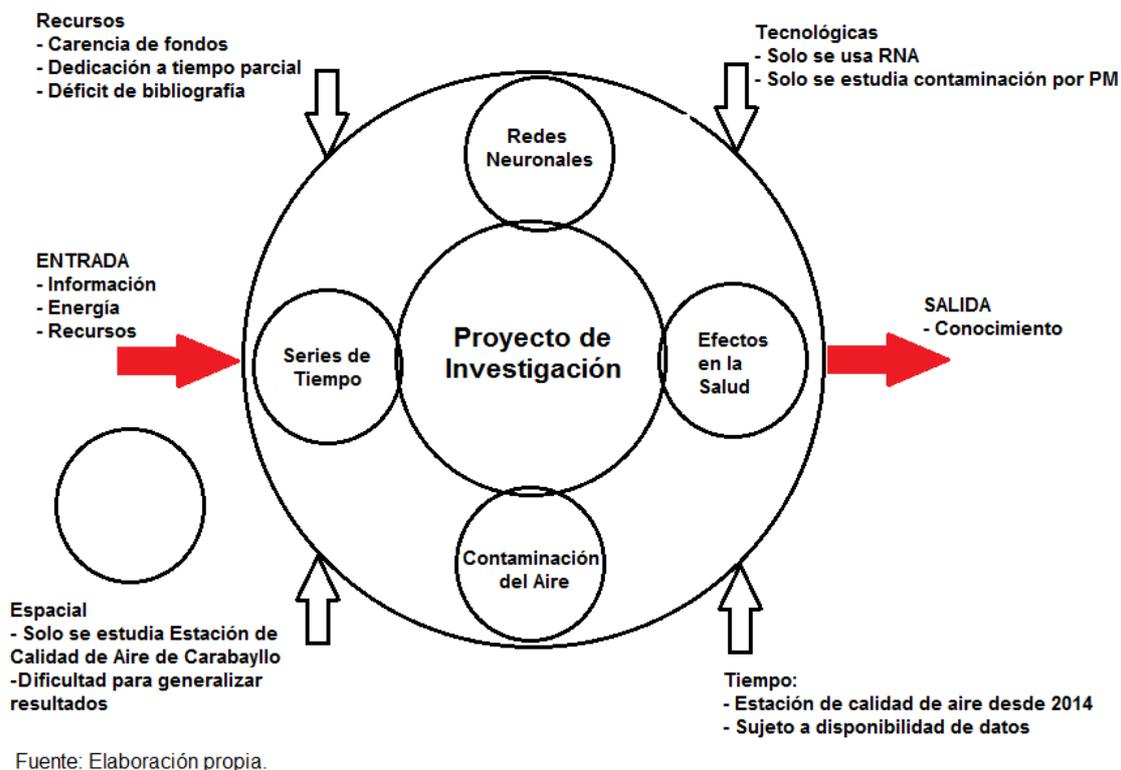


Figura 2: Limitaciones del proyecto de investigación

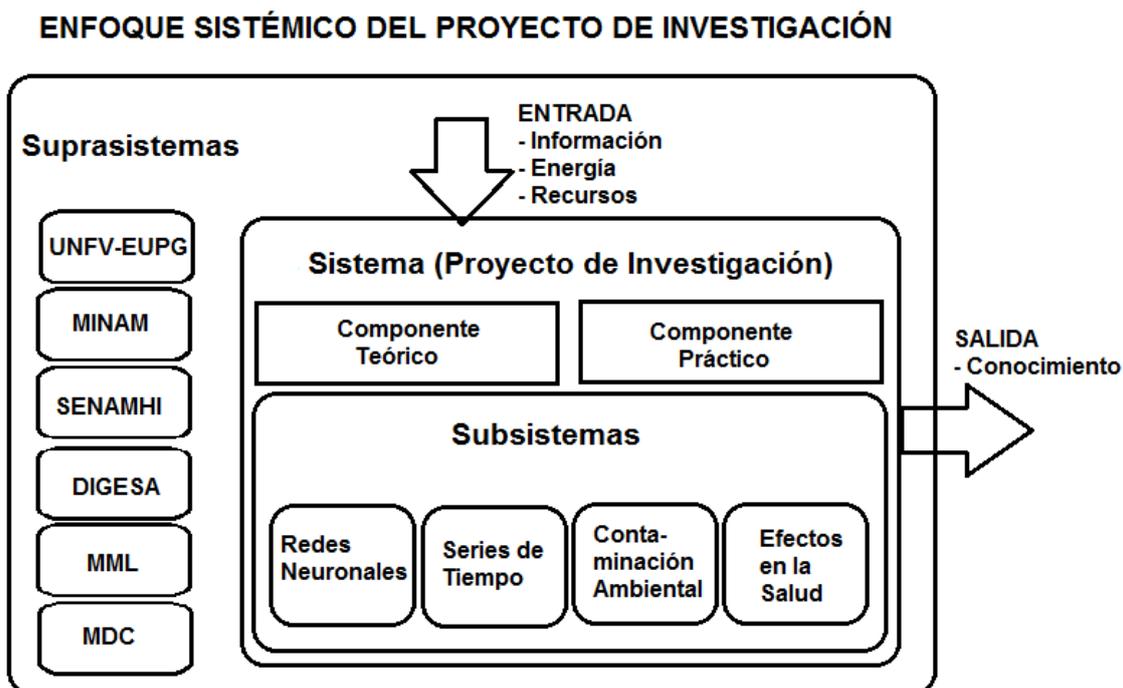
### El Enfoque de Sistemas Blandos

Es claro que el proyecto de investigación no actúa como un objeto aislado, o en términos de análisis de sistemas blandos, es un sistema cerrado. Más bien el proyecto se comporta como un sistema abierto que recibe entradas de información, energía, recursos humanos y materiales y produce una salida que podemos definir como conocimiento, utilizado para tomar decisiones relevantes con respecto a los problemas de contaminación ambiental que justifican su realización.

Si consideramos como un sistema base al proyecto de investigación, tenemos que considerar que existen suprasistemas, definidos como sistemas que incluyen a los sistemas de menor jerarquía desde el punto de vista de pertenencia. Entre estos tenemos

la Escuela Universitaria de Post-Grado de la Universidad Nacional Federico Villarreal, el Ministerio del Ambiente, el Servicio Nacional de Meteorología e Hidrología del Perú, la Dirección General de Salud Ambiental, la Municipalidad Metropolitana de Lima y la Municipalidad de Carabaylo.

Estos suprasistemas se caracterizan por cuanto incluyen al sistema denominado proyecto de investigación, dado que este sistema depende de estos, sea para el abastecimiento de datos así como también para la emisión o regulación de estándares de calidad ambiental e incluso para tomar las decisiones adecuadas bajo la presunción de que se haya generado el conocimiento de las variables de contaminación ambiental, objeto de la investigación. Para el sistema denominado proyecto de investigación, se han definido varios subsistemas que formarán el cuerpo de conocimientos, tales como el estudio de las series de tiempo, las redes neuronales, la contaminación por partículas y los efectos en la salud. En la figura siguiente se puede apreciar la jerarquía de sistemas implementado para este caso de investigación, resaltando el funcionamiento del sistema y su principal producto que es la generación de conocimiento, básica para que los tomadores de decisiones elaboren las políticas ambientales que se correspondan con los valores obtenidos.



Fuente: Elaboración propia. Basado en el análisis de los sistemas blandos

*Figura 3: Enfoque sistémico del proyecto de investigación*

## 1.6 Definición de Variables

### Variable Independiente (X):

Modelo de Predicción basado en Red Neuronal Artificial tipo Perceptrón Multicapa (MLP) modelo no lineal autoregresivo con entrada externa (NARX).

### Indicadores

#### Basados en fenómenos físicos

- Número de variables de entrada (1 a N)
- Número de variables de salida (1 a N)

#### Basados en el modelo matemático

- Número de capas (1 a N)
- Número de retrasos (1 a N)

- Tipo de funciones de activación (logsig, tansig, purelin, etc.)
- Pesos (Reales mayores que cero)
- Tipo de predicción (etapa única o multietapa)

### **Variable Dependiente (Y) -**

Resultados obtenidos mediante la aplicación del modelo de predicción (valores futuros cercanos)

- Concentración de Material Particulado  $PM_{2.5}$  (ppm)

### **Prueba Estadística**

- Error de correlación
- Error Mínimo Cuadrático (MSE o Mean Square Error, por sus siglas en inglés).
- Error Porcentual Medio (MPE o Mean Percentage Error, por sus siglas en inglés).

## CAPÍTULO II: MARCO TEÓRICO

### 2.1 Teorías Generales

#### 2.1.1 Inteligencia Artificial

La investigación en inteligencia artificial genera el término “red neuronal”, mediante el cual se intenta comprender y modelar el comportamiento del cerebro humano. (Baughman & Liu, 1995) enuncian: *“la inteligencia artificial es una parte de la ciencia de la computación que se relaciona con el diseño de sistemas inteligentes de computación, es decir, sistemas que exhiban características que asociamos con la inteligencia en el comportamiento humano”*. Esta definición implica que la inteligencia artificial resuelva problemas que involucren o necesiten inteligencia humana.

De otro lado, (Baughman & Liu, 1995) también definen: *“inteligencia artificial es la rama de la ciencia de la computación que se ocupa de métodos de solución de problemas simbólicos y no algorítmicos”*. Esta segunda definición implica que la inteligencia artificial no usa un algoritmo, secuencia de pasos, para resolver un problema y además el procesamiento simbólico trata con símbolos no numéricos y nombres.

Las tres principales tecnologías basadas en inteligencia artificial que poseen una creciente importancia tecnológica y comercial son los sistemas expertos, las redes neuronales y los sistemas de lógica difusa (Baughman & Liu, 1995). Para el presente trabajo solo se consideró las bases teóricas para redes neuronales artificiales.

#### **Los sistemas expertos, las redes neuronales y el procesamiento subsimbólico**

Un sistema experto (sistema basado en el conocimiento), es un programa de computadora que usa conocimiento profundo y de alta calidad para resolver problemas complejos y

avanzados que usualmente requieren humanos expertos. Los sistemas expertos operan simbólicamente, en una escala macroscópica, procesando símbolos y nombres no numéricos. Estos requieren conocimiento de las relaciones y no le es relevante como se desarrollan estas relaciones.

De otro lado, las redes neuronales usan procesamiento subsimbólico, caracterizado por interacciones microscópicas que eventualmente se manifiestan como comportamiento inteligente macroscópico y simbólico. En base a esto (Baughman & Liu, 1995) definen a las redes neuronales de la siguiente manera: *“una red neuronal es un sistema de computación compuesto de un número de nodos (unidades de procesamiento) simples y altamente conectados, los cuales procesan información mediante su respuesta de estado dinámico a entradas externas”*.

El objetivo de una red neuronal es asociar (relacionar) un conjunto de patrones de entrada a su correspondiente conjunto de patrones de salida. La red ejecuta esta asociación mediante aprendizaje desde una serie de ejemplos pasados definiendo conjuntos de correspondencia de entrada y salida para un sistema dado. La red entonces aplica lo que ha aprendido a un nuevo patrón de entrada para predecir la salida apropiada.

Para el modelamiento de computadoras, las redes neuronales utilizan el mismo tipo de estructura del comportamiento inteligente. Se ha propuesto una red altamente interconectada de neuronas (nodos o unidades de proceso) para este propósito. Mediante una computadora, se ingresa información a la red de neuronas artificiales. Estas neuronas interactúan matemáticamente en forma desconocida por el usuario. Eventualmente, basado en la entrada, la red produce una salida que asocia el patrón macroscópico esperado entrada-salida. El procesamiento subsimbólico, microscópico que ocurre en la red se manifiesta a sí mismo como un comportamiento macroscópico y simbólico.

### 2.1.2 Revisión Histórica de las Redes Neuronales

Esta revisión se basa en el trabajo de (Serrano, Soria, & Martin, 2010).

**Lashey** (1950) afirma que “el proceso de aprendizaje es un proceso distribuido y no local a una determinada área del cerebro”. Hebb, un estudiante de Lashey, basada en su maestro determina una de las reglas de aprendizaje más usadas en la regla del conexionismo denominada aprendizaje hebbiano. Las contribuciones de Hebb fueron publicadas en su libro *The Organization of the Behavior*, es decir *La Organización del Comportamiento*.

**McCulloch y Pitts** (1943) fijan las características de trabajo de la neurona denominada con su nombre, esta neurona es un dispositivo binario (salida 0 ó 1) con un umbral de funcionamiento por debajo del cual está inactiva y puede recibir entradas excitatorias o inhibitorias con acción absoluta: si existe alguna de estas entradas la neurona permanece inactiva. Si no existe ninguna entrada inhibidora se determina la resultante de las entradas excitadoras y si esta es mayor que el umbral, la salida es 1 y si no, la salida es 0.

**Rochester, Holland, Haibt y Duda** (1956) verifican mediante simulaciones una teoría neuronal basada en el postulado de Hebb. Este trabajo, de carácter práctico, implicó hacer suposiciones no halladas en el postulado de Hebb, entre estas acotaron el valor de las sinapsis, las cuales con el modelo hebbiano podían crecer sin límite.

**Selfridge y Rosenblatt** (1958) plantean implementaciones físicas de sistemas conexionistas. Selfridge plantea el modelo conocido como *Pandemonium*, este sistema consta de una serie de capas compuestas conocidas como “demonios”, cada una de las diferentes capas de este sistema se reparten las diferentes tareas a realizar. Rosenblatt presenta una nueva aproximación al problema de reconocimiento de patrones mediante

la introducción del perceptrón. Rosenblatt planteó un dispositivo que realizara tareas que interesaran a los psicólogos.

**Widrow y Hoff (1960)** presentan su modelo ADALINE, por Adaptive Linear Neuron (o Adaptive Linear Element). El sistema de Widrow estaba regido por un sencillo algoritmo de aprendizaje denominado LMS (Least Mean Square). Con este trabajo se propone un sistema adaptativo que puede aprender de forma más precisa y rápida que los perceptrones existentes. El trabajo de Widrow permitió el desarrollo de un área del procesamiento digital de señales denominado “control adaptativo”.

**Minsky y Papert (1969)** presentan su trabajo denominado “Perceptrones”, donde pusieron de manifiesto que los perceptrones solo podían resolver problemas linealmente separables, asimismo plantearon la inutilidad de la extensión de los perceptrones (a sistemas multicapa). El trabajo de estos autores paralizó durante 10 años el desarrollo de la teoría de redes neuronales.

**Kohonen y Anderson (1977)** proponen el modelo de memoria asociativa en forma simultánea, en este modelo artificial planteado la neurona es un sistema lineal que usa como regla de aprendizaje la regla de Hebb modificada, denominado asociador lineal.

**Grossberg (1980)** establece un nuevo principio de auto-organización desarrollando la red neuronal denominada ART o “Adaptive Resonance Theory” traducido como Teoría de Resonancia Adaptativa. Además planteó diferentes modelos neuronales de utilidad práctica.

**Hopfield (1982)** desarrolla la idea de una función de energía para comprender la dinámica de una red neuronal recurrente con uniones sinápticas simétricas. En este primer trabajo solo permite salidas bipolares (0 ó 1). Posteriormente, amplía la función de energía permitiendo la salida continua de las neuronas.

En 1982 **Kohonen** publica su trabajo sobre mapas auto-organizativos que se ordenan sobre reglas simples, el aprendizaje sobre este modelo no requiere ser supervisado.

**Fukushima, Miyake e Ito** presentan la red neuronal *neocognitron*, dispositivo capaz de ser aplicado con éxito en problemas de reconocimiento de patrones. Barto, Sutton y Anderson estudian el aprendizaje reforzado y su aplicación en control, aquí se plantea un nuevo tipo de aprendizaje donde no es necesario un conocimiento total del error cometido por la red, sino tan sólo el signo de este.

En 1983, **Kirkpatrick, Gellat y Vecchi** describen un procedimiento de optimización, posteriormente usado por **Ackley, Hinton y Sejnowski** en el desarrollo de un algoritmo de aprendizaje estocástico. Con esto se busca evitar la obtención de un mínimo local en el proceso de aprendizaje de una red, el procedimiento de optimización derivado de la física estadística aun cuando tiene baja velocidad de convergencia.

En 1986, **Rumelhart, Hinton y Williams** desarrollan el algoritmo de aprendizaje de retropropagación (*backpropagation*) para redes neuronales multicapa mostrando la potencia del método desarrollado. Broomhead y Lowe y Poggio y Girosi trabajan sobre el diseño de redes en capas usando Funciones de Base Radial (Radial Basis Functions, RBF), las máquinas de vectores soporte, el desarrollo de la unión entre elementos neuronales y difusos. Elman, Jordan, Robinson, Fallside, Williams y Zisper trabajan sobre la predicción de series temporales y sobre la consecución de redes recurrentes. Wan realiza una generalización de las redes Time Delay Neural Network (redes neuronales con retraso en el tiempo, orientadas a la utilización sobre series temporales), en su trabajo los filtros sinápticos eran filtros digitales del tipo FIR (Finite Impulse Response). Una extensión de esta idea sería considerar estos pesos con filtros del tipo IIR (Infinite Impulse Response).

## 2.2 Bases Teóricas

Con el objetivo de implementar el programa de predicción de variables de contaminación ambiental, se desarrolló un modelo informático capaz de realizar una predicción multietapa adelantada (*multistep ahead*), basado en redes neuronales del tipo perceptrón multicapa con la arquitectura no lineal autoregresiva con entrada externa (NARX). En este capítulo se hará una revisión breve de la contaminación ambiental por partículas y sus efectos, las series de tiempo y las redes neuronales artificiales tipo perceptrón multicapa con énfasis en las capacidades predictivas.

### 2.2.1 Contaminación Ambiental

En la actualidad, el país en general y la ciudad de Lima Metropolitana en particular, sufren los embates de los diversos tipos de contaminación ambiental, que van desde la presencia de residuos sólidos en diversos distritos de la Capital, desmontes y escombros en las orillas del Río Chillón en los distritos de Carabayllo y Comas, así como en las orillas del mar en los distritos de Magdalena y San Miguel. La presencia de una alta carga de contaminantes que se arrojan a los ríos Rímac y Chillón que convierten a estos en ríos con alto grado de contaminación, sin ningún tipo de vida en su cauce.

En medio de estos temas de contaminación y a consecuencia del desarrollo urbano e industrial, del parque vehicular y de las características geográficas mismas de la ciudad metropolitana entre otras causas antrópicas y también naturales, se produce la contaminación del aire por contaminantes químicos y contaminantes debido a material particulado (PM, por su denominación común en inglés *Particulate Material*). En el presente trabajo, se aborda tan solo este tipo de contaminación debida a partículas, la cual

tiene sus propias características sin desmerecer el estudio que también debe ser realizado para los otros tipos de contaminación.

### **Contaminación del Aire por Material Particulado**

La contaminación por material particulado (PM) es una mezcla de partículas sólidas y líquidas suspendidas en el aire que varían en número, tamaño, forma, área superficial, composición química, solubilidad y origen. La distribución de tamaño de las partículas suspendidas totales (TSP: *Total Suspended Particles*, por sus siglas en inglés) en el aire ambiente es trimodal, incluyendo partículas gruesas, finas y ultrafinas. El muestreo selectivo en tamaño de material particulado se refiere a coleccionar partículas debajo, sobre o dentro de un rango específico de tamaño aerodinámico usualmente seleccionado para tener una relevancia especial para la inhalación y fuentes de deposición o toxicidad.

Dado que los muestreadores no diferencian con precisión el tamaño de partícula, usualmente se define como tamaño de partícula relativo al 50% de punto de corte a un diámetro aerodinámico específico (tal como 2.5 o 10 micras) y una pendiente de la curva de efectividad de la muestra. Las partículas gruesas se derivan primariamente de la suspensión de polvo, suelo, u otros materiales de la corteza de caminos, granjas, minas, tormentas de viento, volcanes y otros. Las partículas gruesas también incluyen sales marinas, polen, moho, esporas y otras partes de las plantas. Las partículas gruesas a menudo se indican por concentración de masa de partículas mayores al punto de corte de 2.5 micras. (Pope & Dockery, 2006)

Las partículas ultrafinas se definen como partículas con un diámetro aerodinámico de 0.1 micras. El aire ambiente en entornos urbanos e industriales está constantemente

recibiendo emisiones frescas de partículas ultrafinas de fuentes relacionadas a combustión, tales como gases exhaustos de vehículos y reacciones fotoquímicas atmosféricas. Estas partículas ultrafinas primarias, sin embargo, tienen una vida muy corta (minutos a horas) y crecen rápidamente (mediante coagulación y/o condensación) para formar agregados complejos de mayor tamaño, pero típicamente permanecen como parte de PM<sub>2.5</sub>. Existe recientemente mayor interés en partículas ultrafinas, debido a que es más probable que las partículas ultrafinas poco solubles se trasladen desde los pulmones a la sangre y otras partes del cuerpo; que las partículas de mayor dimensión. (Pope & Dockery, 2006).

### **Exposición del hombre al material particulado**

(Strauss & Mainwaring, 2011) en su libro denominado “Contaminación del aire: causas, efectos y soluciones” enuncian que “el hombre, por vivir en comunidades desarrolladas y por ende contaminadas, expone solo parte de su piel a la atmósfera pero cada día inhala aproximadamente 7,500 litros de aire de modo tal que sus pulmones y sistema respiratorio general están en contacto con ella y tienen la potencialidad de retener cualquier sustancia nociva que pueda estar contenida en ese aire. Se han visto consideraciones de incidentes de contaminantes de aire que la nariz, la garganta y sistema bronquial son los que con más frecuencia se ven afectados. Esto también sucede con otros contaminantes; aun cuando el pulmón no sea el órgano objetivo, si provee la ruta a través de la cual el contaminante penetra en el sistema”. De esta lectura se concluye que es de extrema importancia para el ser humano la calidad del aire que inhala, siendo esto más relevante cuando consideramos los grupos humanos vulnerables, tal como son los niños, ancianos y en general aquellas personas que sufren problemas de salud como el asma o similares.

## **Efectos de la contaminación por material particulado**

(Pope & Dockery, 2006) en su texto denominado “Health Effects of Fine Particulate Air Pollution: Lines that Connect” traducido como Efectos sobre la salud de contaminación de aire por partículas finas: Líneas que conectan, afirman que “se ha avanzado en la evaluación de los efectos en la salud del material particulado (PM) a diferentes escalas de tiempo de exposición y en la exploración de la forma de función concentración-respuesta. Existe también evidencia de efectos en la salud cardiovascular relacionada al material particulado (PM) y creciente conocimiento con referencia a mecanismos patofisiológicos generales que vinculan la exposición a PM con morbilidad y mortalidad cardiopulmonar.

La investigación evidencia que la exposición a la polución del aire en partículas finas tiene efectos adversos en la salud cardiopulmonar. Aun cuando esta investigación ha sido motivada por políticas de salud ambiental pública, estos resultados tienen importancia en los campos científicos, médicos y de salud pública que son más amplias que los debates sobre los estándares legales de calidad de aire”. Asimismo, los autores también afirman que “se han observado los efectos biológicos del PM, y se proponen varios mecanismos generales de vías de ingreso, incluyendo: (1) la exposición de largo plazo (origina progresión más rápida de enfermedad pulmonar obstructiva crónica), y la exposición aguda (exacerba las enfermedades pulmonares existentes); (2) estrés oxidativo pulmonar y sistémico, inflamación, aterosclerosis, y enfermedades cardiovasculares relacionadas; (3) cambios adversos en la función cardíaca autonómica; (4) alteraciones vasculares; (5) translocación sistémica de PM y efectos protrombóticos; (6) modificación de defensas e inmunidad; y (7) daño pulmonar inducido por PM, disminución en la función pulmonar, dificultad respiratoria, e hipoxemia. Aun cuando ninguna de estas formas ha sido plenamente explorada, esta evidencia proporciona varias formas

mecánicas interdependientes hipotéticas que pueden vincular a la exposición a PM con la morbilidad y mortalidad cardiopulmonar”.

Los párrafos anteriores que vinculan a la salud con la exposición al material particulado (PM) a diferentes escalas de tiempo de exposición y la certeza de que la salud cardiopulmonar se ve afectada, en especial de las personas más vulnerables, es una justificación suficiente para emprender el desafío de predecir, en una forma alterna a los esquemas de predicción de las entidades estatales correspondientes, el comportamiento de la contaminación ambiental por material particulado y mediante esto, elaborar las herramientas de gestión con las cuales los gobiernos locales y metropolitano deben tomar las decisiones adecuadas para salvaguardar la salud de los ciudadanos.

(Wark & Warner, 2007) afirman que las partículas pueden tener efectos tóxicos de una o más de las tres maneras siguientes:

- a. La partícula puede ser intrínsecamente tóxica debido a sus características inherentes físicas o químicas.
- b. La partícula puede interferir con uno o más de los mecanismos que despejan usualmente el aparato respiratorio.
- c. La partícula puede actuar como un conductor a una sustancia tóxica absorbida.

Es extremadamente difícil obtener una relación directa entre la exposición a varias concentraciones de partículas y los efectos resultantes sobre la salud del hombre, siendo importante la cantidad de tiempo de exposición.

## **Mediciones de la Contaminación del Aire**

(De-Nevers, 1998) indica que existen dos clases de mediciones de la contaminación del aire; mediciones del ambiente (concentraciones de contaminante en el aire que el público respira, o monitoreo del ambiente) y mediciones en la fuente (concentraciones o índice de emisiones, o ambas cosas, de las fuentes de contaminación o pruebas de las fuentes). Para controlar las concentraciones de contaminantes, se deben regular el tiempo, el lugar y la cantidad de sus emisiones. Por tanto, se deben medir los índices de emisiones de diversas fuentes de contaminantes del aire.

### **Medición del Ambiente en Lima Metropolitana.**

#### **Monitoreo Manual**

El SENAMHI cuenta con cuarenta puntos de muestreo de contaminantes sólidos sedimentables (CSS) instalados en la Zona Metropolitana de Lima - Callao que recopilan información desde el año 2002 hasta la fecha. Estos valores obtenidos son comparados con el límite referencial establecido por la Organización Mundial de la Salud de  $5t/km^2/mes$ . Se evalúan los valores promedio, máximos y mínimos y se explica su interrelación con las condiciones meteorológicas observadas. Con la información histórica recopilada se realizó un diagnóstico de la caracterización por contaminantes sólidos sedimentables, donde se identificaron las zonas más críticas de la ciudad de Lima Metropolitana – Callao. El estudio sirvió como línea de base para el diseño de la red automática de calidad de aire del SENAMHI. (Sánchez-Ccoyllo, y otros, 2011).

### **Monitoreo Automático**

La red de monitoreo de la calidad del aire del SENAMHI en Lima Metropolitana consta de cinco estaciones dotados con equipos automáticos que monitorean de una manera constante los contaminantes: material particulado ( $PM_{10}$ ), dióxido de azufre ( $SO_2$ ), dióxido de nitrógeno ( $NO_2$ ) y ozono superficial ( $O_3$ ). Se encuentran ubicadas en los distritos de San Borja, Ate, Jesús María, Santa Anita y Villa María del Triunfo, tal como se muestra en el cuadro 5 y mapa 1. Las estaciones de Santa Anita y Villa María del Triunfo cuentan con una estación meteorológica que mide presión, humedad relativa, radiación ultravioleta y temperatura del aire (Sánchez-Ccoyllo, y otros, 2011).

#### **2.2.2 Series de Tiempo**

Estas líneas se basan en el trabajo de (Arellano, 2016) y de (Isasi-Viñuela & Galván-León, 2004); con ligeras adaptaciones para vincular la teoría de series de tiempo a la predicción.

##### **a. Generalidades**

La planificación racional exige prever los sucesos del futuro que probablemente vayan a ocurrir. La previsión, a su vez, se suele basar en lo que ha ocurrido en el pasado. Se tiene pues un nuevo tipo de inferencia estadística que se hace acerca del futuro de alguna variable o compuesto de variables basándose en sucesos pasados. La técnica más importante para hacer inferencias sobre el futuro con base en lo ocurrido en el pasado, es el *análisis de series de tiempo*. Las aplicaciones se dan en áreas como: economía, física, geofísica, química, electricidad, en demografía, en marketing, en telecomunicaciones, en transporte, etc.

<i>Series De Tiempo</i>	<i>Ejemplos</i>
1. Series económicas:	- Precios de un artículo - Tasas de desempleo - Tasa de inflación - Índice de precios, etc.
2. Series Físicas:	- Meteorología - Cantidad de agua caída - Temperatura máxima diaria - Velocidad del viento (energía eólica) - Energía solar, etc.
3. Geofísica:	- Series sismológicas
4. Series demográficas:	- Tasas de crecimiento de la población - Tasa de natalidad, mortalidad - Resultados de censos poblacionales
5. Series de marketing:	- Series de demanda, gastos, ofertas
6. Series de telecomunicación:	- Análisis de señales
7. Series de transporte:	- Series de tráfico

*Tabla 3: Ejemplos de series de tiempo*

Uno de los problemas que intenta resolver las series de tiempo es el de predicción. Esto es dado una serie  $\{x(t_1), \dots, x(t_n)\}$  el objetivo es describir el comportamiento de la serie, investigar el mecanismo generador de la serie temporal, buscar posibles patrones temporales que permitan sobrepasar la incertidumbre del futuro.

Se estudiará el diseño de un modelo para explicar la estructura y prever la evolución de una variable que observamos a lo largo del tiempo. La variables de interés puede ser macroeconómica (índice de precios al consumo, demanda de electricidad, series de exportaciones o importaciones, etc.), microeconómica (ventas de una empresa, existencias en un almacén, gastos en publicidad de un sector) o física (velocidad del viento en una central eólica, temperatura en un proceso, caudal de un río, concentración en la atmósfera de un agente contaminante), entre otras. Para esta investigación específica se ha revisado la bibliografía referida a predicciones de series de tiempo, en especial del multietapa adelantada.

## b. Definición de series de tiempo

En diversas áreas del conocimiento las observaciones de interés se obtienen en instantes sucesivos y equidistantes de tiempo, por ejemplo, cada hora, durante 24 horas, mensuales, trimestrales, semestrales o bien registradas por algún equipo en forma continua.

Se denomina *Serie de Tiempo* a un conjunto de mediciones de cierto fenómeno o experimento registradas secuencialmente en el tiempo. Estas observaciones serán denotadas por  $\{x(t_1), x(t_2), \dots, x(t_n)\} = \{x(t) : t \in T \subseteq \mathbb{R}\}$  con  $x(t_i)$  el valor de la variable  $x$  en el instante  $t_i$ . Si  $T = \mathbb{Z}$  se dice que la serie de tiempo es discreta y si  $T = \mathbb{R}$  se dice que la serie de tiempo es continua. Cuando  $t_{i+1} - t_i = k$  para todo  $i = 1, \dots, n-1$ , se dice que la serie es equiespaciada, en caso contrario será no equiespaciada.

En adelante se trabajará con series de tiempo discreta, equiespaciadas en cuyo caso asumiremos y sin pérdida de generalidad que:  $\{x(t_1), x(t_2), \dots, x(t_n)\} = \{x(1), x(2), \dots, x(n)\}$ .

## c. Análisis de Series de Tiempo

El primer paso en el análisis de series de tiempo, consiste en graficar la serie. Esto nos permite detectar las componentes esenciales de la serie.

El gráfico de la serie permitirá:

*i. Detectar Outliers:* se refiere a puntos de la serie que se escapan de lo normal. Un outlier es una observación de la serie que corresponde a un comportamiento anormal del fenómeno (sin incidencias futuras) o a un error de medición. Se debe determinar desde fuera si un punto dado es outlier o no. Si se concluye que lo es, se debe omitir o reemplazar por otro valor antes de analizar la serie.

ii. **Permite detectar tendencia:** la tendencia representa el comportamiento predominante de la serie. Esta puede ser definida vagamente como el cambio de la media a lo largo de un periodo (ver figura 1.2).

iii. **Variación estacional:** la variación estacional representa un movimiento periódico de la serie de tiempo. La duración de la unidad del periodo es generalmente menor que un año. Puede ser un trimestre, un mes o un día, etc. Matemáticamente, podemos decir que la serie representa variación estacional si existe un número  $s$  tal que  $x(t) = x(t + k \cdot s)$ .

Todos estos fenómenos presentan un comportamiento estacional (anual, semanal, etc.)

iv. **Variaciones irregulares (componente aleatorio):** los movimientos irregulares (al azar) representan todos los tipos de movimientos de una serie de tiempo que no sea tendencia, variaciones estacionales y fluctuaciones cíclicas.

#### d. **Predicción de Series de Tiempo**

La predicción sobre series temporales consiste en, a partir de una muestra de la serie de tiempo, conocer los valores en el futuro, es decir, su comportamiento a lo largo del tiempo. La idea es estimar  $X(t)$  en un instante  $n + k$  posterior al último dato observado en  $t=n$ ,  $k=1, 2, 3, \dots$ . Pueden distinguirse dos casos:

##### i. **Predicción en un paso de tiempo**

Consiste en predecir el valor de la serie en el instante de tiempo inmediatamente siguiente al instante actual  $t$ , a partir de las muestras disponibles hasta dicho instante de tiempo. Es decir, predecir el valor de  $x(t+1)$  utilizando un cierto número de muestras anteriores  $x(t)$ ,  $x(t-1)$ ,  $x(t-2)$ ,  $\dots$ , número que depende de la serie temporal.

## ii. Predicción en múltiples pasos de tiempo

Consiste en predecir el comportamiento de la serie, no sólo en el instante inmediato siguiente al instante  $t$  actual, sino en un futuro más lejano, concretamente en el intervalo de predicción  $[t+1, t+h+1]$ , siendo  $h$  un número natural que representa el horizonte de predicción. Es decir, consiste en predecir los valores de la serie  $x(t+1)$ ,  $x(t+2)$ , ..., y  $x(t+h+1)$  a partir de la información disponible en el instante de tiempo actual  $t$ .

## e. Justificación

Cuando no se dispone de las ecuaciones diferenciales que describan el comportamiento de una serie de tiempo, se hace necesario recurrir a métodos aproximativos para construir modelos que permitan resolver el problema de predicción. La dificultad de predicción de las series de tiempo depende del comportamiento dinámico de la serie, si bien existen modelos periódicos y estables capaces de ser explicados mediante modelos basados en técnicas clásicas y lineales, es relevante en esta reseña las series de tiempo cuyo comportamiento puede ser descrito por modelos no lineales autoregresivos con entrada externa (NARX). Estos modelos se caracterizan porque recogen el comportamiento temporal de la serie expresando el valor de la serie en el instante de tiempo  $t+1$  como una función no lineal de  $r+1$  valores de la serie temporal en instantes anteriores de tiempo, es decir:

$$x(t+1) = F[x(t), x(t-1), \dots, x(t-r)] + \varepsilon(t)$$

donde  $t$  es la variable discreta tiempo,  $\varepsilon(t)$  es un error residual que se asume ruido blanco Gaussiano, y  $F$  es una función no lineal desconocida y que, por tanto, debe ser estimada o aproximada a partir de un conjunto de datos observados de la serie temporal,  $x(t)_{t=0..N}$

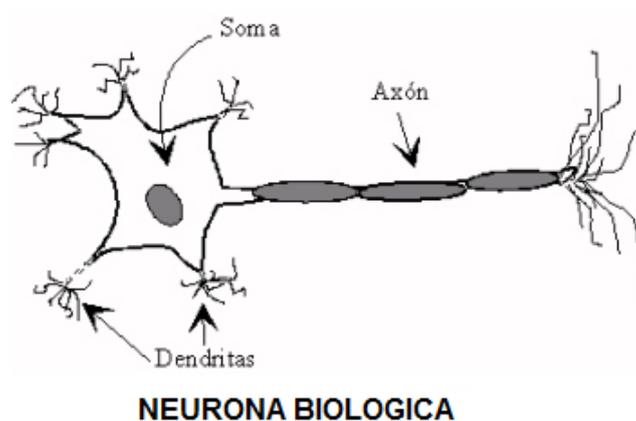
Por tanto, la construcción de modelos no lineales autoregresivos (NAR) involucra la determinación de la función **F**, a partir del conjunto de muestras disponibles, mediante técnicas de aproximación, entre las cuales se encuentran las redes de neuronas artificiales.

### **2.2.3 Redes Neuronales Artificiales**

#### **Redes Neuronales Biológicas**

Las redes neuronales artificiales (RNA) están relacionadas en forma remota a sus contrapartes biológicas, existen características de la función cerebral que han inspirado el desarrollo de las redes neuronales artificiales. El cerebro consiste en un gran número (aproximadamente  $10^{11}$ ) de elementos altamente conectados (aproximadamente 104 conexiones por elemento) denominadas neuronas.

Estas neuronas tienen tres componentes principales: las dendritas, el cuerpo de la célula y el axón. Las dendritas son como redes receptoras de señales nerviosas que portan señales eléctricas al cuerpo de la célula. La célula efectivamente suma y señala un umbral para estas señales de entrada. El axón es una única fibra que porta la señal desde el cuerpo de la célula hacia otras neuronas. El punto de contacto entre el axón de una célula y una dendrita de otra es denominado una sinapsis. Es el arreglo de neuronas y la fuerza de las sinapsis individuales, determinada por un complejo proceso químico que establece la función de la red neuronal. La figura siguiente muestra un esquema simplificado de una neurona biológica.



*Figura 4: Neurona biológica.*

Fuente: <http://avellano.usal.es/~lalonso/RNA/>

### **Redes Neuronales Artificiales**

Las redes neuronales artificiales no se aproximan a la complejidad del cerebro. Sin embargo, hay dos semejanzas claves entre las redes neuronales artificiales y biológicas. Primero, los bloques de construcción de ambas redes son simples dispositivos computacionales (aun cuando las redes neuronales artificiales son mucho más simples que las biológicas) que están altamente interconectados. Segundo, las conexiones entre neuronas determinan la función de la red (Hagan & Demuth, 2016).

**La Neurona Artificial:** Es un elemento que posee un estado interno, denominado nivel de activación, y recibe señales que le permiten, en su caso, cambiar de estado.

Se denomina  $S$  al conjunto de estados posibles de la neurona, pudiendo ser números naturales  $S = \{0, 1, 2, \dots\}$  o un intervalo continuo de valores, por ejemplo  $S = [0, 1]$ . Las neuronas poseen una función que les permite cambiar de nivel de activación a partir de las señales que reciben, esta se denomina *función de activación*. Las señales que recibe cada neurona pueden provenir del exterior o de otras neuronas a las cuales está conectada.

El nivel de activación de una neurona depende de las entradas recibidas y de los valores sinápticos, pero no de anteriores valores de estados de activación. Para calcular el estado de activación se ha de calcular la entrada total ( $\sum_i$ ) a la neurona  $i$ . Este valor se calcula como la suma de las entradas ponderadas.

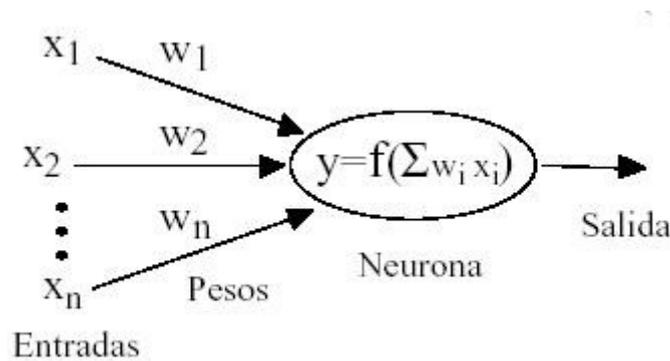


Figura 5: Representación de una neuronal artificial

Donde las entradas se pueden representar por los vectores

$$X = (x_1, x_2, \dots, x_N) \text{ y } W = (w_1, w_2, \dots, w_N)$$

La sumatoria que corresponde al cuerpo de la neurona, suma todas las entradas ponderadas algebraicamente, produciendo una salida denominada  $\sum_i$ .

$$\text{donde } \sum_i = x_1 w_1 + x_2 w_2 + \dots + x_N w_N$$

En forma vectorial se define así:  $\sum_i = \mathbf{X}^T \cdot \mathbf{W}$

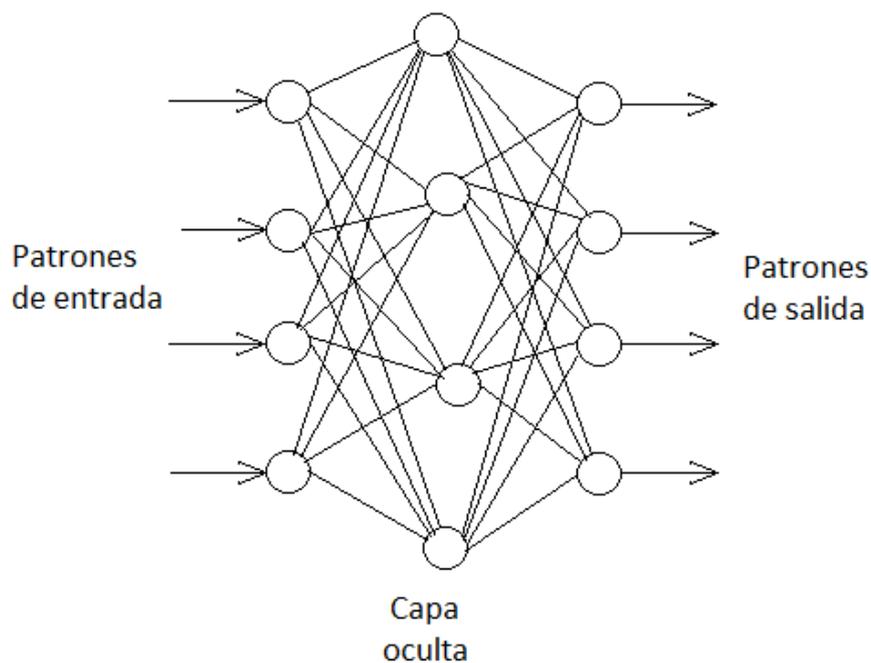
Por tanto la salida de la unidad neuronal se define como la aplicación de una función de activación no lineal a  $\sum_i$ , siendo  $y = F(\sum_i) = F(\mathbf{X}^T \cdot \mathbf{W})$ .

Desde el punto de vista del desarrollo de código de programación, el proceso de solución de una neurona artificial se reduce al requerido para una multiplicación de matrices como unidad elemental, la aplicación de una función de activación (no lineal) y luego un

escalamiento bottom-up para recorrer todas las unidades neuronales que existen en un determinado modelo.

### Estructura básica de la red neuronal artificial

Al modo en el cual las neuronas artificiales se conectan entre si se denomina patrón de conectividad o arquitectura de la red. La estructura básica de interconexión entre neuronas artificiales es la de red multicapa, mostrada en la figura siguiente. Esta es una estructura típica de implementación de la técnica denominada retropropagación. El primer nivel lo constituyen las neuronas de entrada, estas unidades reciben los valores de unos patrones representados como vectores que sirven de entrada a la red. Luego hay una serie de capas intermedias (ocultas) cuyas unidades responden a rasgos particulares que pueden aparecer en los patrones de entrada, puede haber una a más capas ocultas. El último nivel es de salida, estas sirven como salida de toda la red.



*Figura 6: Esquema de una red con tres capas totalmente interconectadas*

Adaptado de: Redes de neuronas artificiales, un enfoque práctico.

Bajo esta perspectiva, se puede definir una *red neuronal artificial (RNA)* como un grafo cuyos nodos están constituidos por neuronas (unidades de proceso) idénticas y que propagan información a través de los arcos. En este grafo se distinguen tres tipos de nodos: los de entrada, los de salida y los ocultos.

**Definición formal de una red neuronal** (Martín-del-Brío & Sanz-Molina, 2006).

Una **red neuronal** es un grafo dirigido, con las siguientes propiedades:

1. A cada nodo  $i$  se le asocia una variable de estado  $x_i$
2. A cada conexión  $(i, j)$  de los nodos  $i$  y  $j$  se asocia un peso  $w_{ij}$  que pertenece a los reales ( $\mathbb{R}$ ).
3. A cada nodo  $i$  se asocia un umbral  $\Theta_i$
4. Para cada nodo  $i$  se define una función  $f_i(x_j, w_{ij}, \Theta_i)$ , que depende de los pesos de sus conexiones, del umbral y de los estados de los nodos  $j$  a él conectados. Esta función proporciona el nuevo estado del nodo.

La red funciona de la siguiente manera: para cada vector de entrada, este se introduce a la red copiando cada vector en la célula de entrada correspondiente. Cada célula de la red, una vez recibida la totalidad de sus entradas, las procesa y genera una salida que es propagada a través de las conexiones entre células, llegando como entrada a la célula destino. Una vez que la entrada ha sido completamente propagada por toda la red, se producirá un vector de salida, cuyos componentes son cada uno de los valores de salida de las células de salida.

El esquema de funcionamiento de una RNA por capas semejante a la figura anterior, puede describirse mediante la siguiente ecuación:

$$S = F [ F (X.W_1).W_2 ]$$

donde  $W_1$  y  $W_2$  son los pesos de la primera y segunda capa respectivamente,  $F$  es la función de activación idéntica en todas las neuronas,  $X$  es el vector de entrada a la red y  $S$  es el vector de salida que la red produce.

Se puede demostrar que si la función de activación es lineal, el introducir más capas en la red es irrelevante; existirá siempre una red con una sola capa equivalente a cualquier otra con un número de capas arbitrario, y su cálculo es trivial. Dado que las redes de una sola capa tienen limitaciones, de esto se desprende que las redes multicapa utilizan funciones de activación no lineales que potencian a la red para solucionar problemas en forma genérica.

**Aprendizaje:** es la determinación de los valores precisos de los pesos para todas las conexiones, de modo que la red quede capacitada para la resolución eficiente de un problema. El proceso de aprendizaje consiste en ir introduciendo los ejemplos del conjunto de aprendizaje, y modificar los pesos de las conexiones siguiendo un determinado esquema de aprendizaje. Esto se repite hasta que se cumpla cierto criterio de convergencia.

#### **Criterios de Convergencia:**

- **Número fijo de ciclos:** se decide con anterioridad cuantas veces será presentado el conjunto de aprendizaje, y una vez superado ese número se detiene el proceso y se acepta la red resultante.
- **Error por debajo de una cantidad pre-establecida:** se define una función error y un valor aceptable para dicho error. Se detiene el aprendizaje de la red cuando se produce un valor de error por debajo del valor aceptable pre-fijado. Para el caso

que la red no pueda obtener un valor por debajo del pre-fijado se debe tener un criterio adicional de parada.

- **Modificación de pesos irrelevante:** en modelos donde las conexiones se modifican cada vez con menor intensidad. Si en el proceso de aprendizaje se llega a un momento donde ya no se producen variaciones significativas de los valores de los pesos, se dice que la red ha convergido y se detiene el proceso de aprendizaje.

### **Tipos de Aprendizaje**

- **Algoritmos de aprendizaje supervisado.** En este caso los datos del conjunto de aprendizaje tienen dos atributos: los datos y cierta información relativa a la solución del problema. El esquema de aprendizaje supervisado utiliza esta información para modificar las conexiones, mediante la comparación entre la salida producida y la esperada, y en consecuencia actúa modificando apropiadamente los valores de los pesos. Dado un miembro del conjunto de entrenamiento como entrada de la red, podríamos saber cuál debería ser la salida y como difiere de lo obtenido en la realidad (Anderson, 2007).
- **Algoritmos de aprendizaje no supervisado.** Los datos del conjunto de aprendizaje solo tienen información de los ejemplos, y no hay nada que permita guiar en el proceso de aprendizaje. La red modificará los valores de los pesos a partir de información interna. Cuando se utiliza aprendizaje no supervisado, la red trata de determinar características de los datos del conjunto de entrenamiento: rasgos significativos o redundancias. Estos modelos son también conocidos como sistemas auto organizados, debido a que la red se ajusta únicamente de los valores

recibidos como entrada. Los algoritmos no supervisados son difíciles de construir y utilizar y usualmente hacen supuestos sobre lo que se trata de hacer.

- **Algoritmos de aprendizaje por refuerzo.** Es una variante del aprendizaje supervisado en el que no se dispone de información concreta del error cometido por la red para cada ejemplo de aprendizaje, sino que simplemente se determina si la salida producida para dicho patrón es o no adecuada.

### **Conjuntos de Entrenamiento y Validación**

Para poder determinar si la red produce salidas adecuadas, se divide el conjunto de entrenamiento en dos conjuntos denominados de entrenamiento y de validación. El conjunto de entrenamiento se utiliza para aprender los valores de los pesos. El conjunto de validación es utilizado para medir el error. De este modo, y con datos que no han sido parte del aprendizaje, se mide la eficacia de la red para resolver un problema. Si el error sobre el conjunto de validación es pequeño, entonces se garantiza la capacidad de generalización de la red.

### **Características de los conjuntos de entrenamiento y validación**

- El conjunto de validación debe ser independiente del de aprendizaje. No puede haber ningún tipo de sesgo en el proceso de selección de los datos de validación.
- El conjunto de validación debe cumplir las propiedades de un conjunto de entrenamiento.

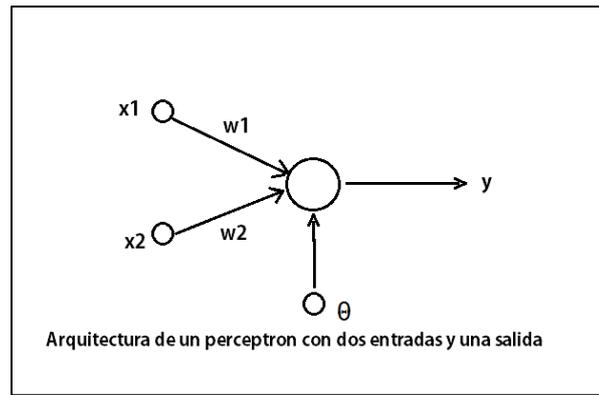
Asimismo, el conjunto de validación puede utilizarse durante el aprendizaje para guiarlo en conjunción con el entrenamiento. El procedimiento es:

1. Asignar a los pesos valores aleatorios
2. Introducir todos los ejemplos del conjunto de entrenamiento, modificando los pesos de acuerdo con el esquema de aprendizaje supervisado elegido.
3. Introducir todos los ejemplos del conjunto de validación. Obtener el error producido al predecir dichos ejemplos.
4. Si el error calculado en el paso anterior está por encima de cierto valor umbral, ir a (2)
5. Acabar el proceso de aprendizaje y dar como salida la red obtenida.

### **El Perceptrón.**

Es un sistema capaz de realizar tareas de clasificación de forma automática a partir de un conjunto de ejemplos de clases diferentes, para determinar las ecuaciones de las superficies que hacían de frontera entre dichas clases. El sistema se basó sobre la información de ejemplos (patrones de entrenamiento) existentes de diferentes clases. Estos son los que aportan la información necesaria para que el sistema construya las superficies discriminantes y también actúa como un discriminador para ejemplos nuevos desconocidos. Al final del proceso, el sistema es capaz de determinar la clase de pertenencia para cualquier ejemplo nuevo.

La arquitectura de la red consta de una estructura monocapa, en la que hay un conjunto de células de entrada (una o varias según las condiciones del problema) y una o varias células de salida. Cada una de las células de entrada tiene conexiones con todas las células de salida, y estas conexiones son las que determinan la superficie de discriminación del problema.



*Figura 7: Arquitectura de un perceptrón con dos entradas y una salida*

La solución del perceptrón se presenta en la siguiente ecuación:

$$Y = F(\sum w_i x_i + \Theta) \quad (\text{Ecuación I})$$

donde  $x_i$  son las entradas,  $Y$  la salida,  $w_i$  los pesos y  $\Theta$  el umbral.

Tal que

$$F(s) = \begin{cases} 1 & \text{si } s > 0 \\ -1 & \text{en caso contrario} \end{cases}$$

La función de salida  $F$  es binaria y al ser un discriminante de clases, una salida binaria puede ser fácilmente traducible a una clasificación en dos categorías, de la siguiente manera:

- Si la red produce salida 1, la entrada pertenece a la categoría A
- Si la red produce salida -1, la entrada pertenece a la categoría B

Para el caso de dos dimensiones, la ecuación anterior se transforma en

$$w_1x_1 + w_2x_2 + Q = 0$$

Dado que esta ecuación define una recta, en caso de ser solución al problema discriminará entre las clases existentes en los datos de entrenamiento. Si las clases están separadas, se puede demostrar que dicha recta discriminante existe. El perceptrón, en su proceso de aprendizaje determina la ecuación de esa recta a partir de los datos de entrenamiento.

### **Definición formal del perceptrón**

Dados dos conjuntos de puntos en  $\mathbb{R}^N$ :  $A = (a_1, \dots, a_{na})$  y  $B = (b_1, \dots, b_{nb})$ , obtener el hiperplano:  $w_1x_1 + \dots + w_nx_n + \Theta = 0$ , de tal forma que:

$$\forall \mathbf{a} \in A : w_1a_1 + \dots + w_na_n + Q > 0$$

$$\forall \mathbf{b} \in B : w_1b_1 + \dots + w_nb_n + Q < 0$$

Este proceso constituye el aprendizaje del perceptrón y se realiza mediante un proceso iterativo en el que paulatinamente se van modificando los valores de los pesos de las conexiones, hasta encontrar los valores que determinan las ecuaciones discriminantes. En función de la red neuronal esto es equivalente a encontrar los valores de las conexiones se hagan:

$$\forall \mathbf{a} \in \mathbf{A} : y(\mathbf{a}) = 1$$

$$\forall \mathbf{b} \in \mathbf{B} : y(\mathbf{b}) = -1$$

### **El Perceptrón Multicapa**

Nótese que este resumen de teoría se basa en los aportes de (Isasi-Viñuela & Galván-León, 2004), y han sido enriquecidos y actualizados con la lectura de autores como

(Baughman & Liu, 1995) ; (Martín-del-Brío & Sanz-Molina, 2006), (Serrano, Soria, & Martín, 2010) y otros también referenciados.

El perceptrón multicapa o red multicapa con conexiones hacia adelante es una generalización del perceptrón simple y se consolidó como resultado de las limitaciones de dicha topología con referencia al problema de la separabilidad no lineal. Se había demostrado que la combinación de varios perceptrones, lo cual implicaba la inclusión de una o más capas ocultas, podría convertirse en una solución adecuada para tratar problemas no lineales. Sin embargo, esto solucionaba el problema de adaptar los pesos de la capa de entrada a la capa oculta, desde que la regla de aprendizaje del perceptrón no se aplica en este caso. Luego se presentó una manera de retropropagar los errores medidos en la salida de la red hacia las neuronas ocultas, creando la regla delta generalizada, la cual es una generalización de la regla delta para funciones de activación no lineales y redes multicapa.

En el teorema de aproximación universal se demuestra que el perceptrón multicapa es un aproximador universal, dado que cualquier función continua sobre un compacto de  $\mathbb{R}^N$  puede aproximarse con un perceptrón multicapa, con al menos una capa oculta de neuronas. Así, puede considerarse al perceptrón multicapa como una nueva clase de funciones - tal como polinomios, las funciones trigonométricas, etc. - para aproximar o interpolar relaciones no lineales entre datos de entrada y salida.

Dentro del campo de redes neuronales, el perceptrón multicapa es una de las arquitecturas más utilizadas en la resolución de problemas tanto debido a su capacidad como aproximador universal como su fácil aplicación y utilización. Este tipo de redes han sido aplicadas con éxito a la resolución de problemas en una gran variedad de diferentes áreas, tales como: reconocimiento del habla, reconocimiento de caracteres ópticos,

reconocimiento de caracteres escritos, control de procesos, modelización de sistemas dinámicos, conducción de vehículos, diagnósticos médicos, predicción de series temporales, etc.

### **Ventajas y Desventajas:**

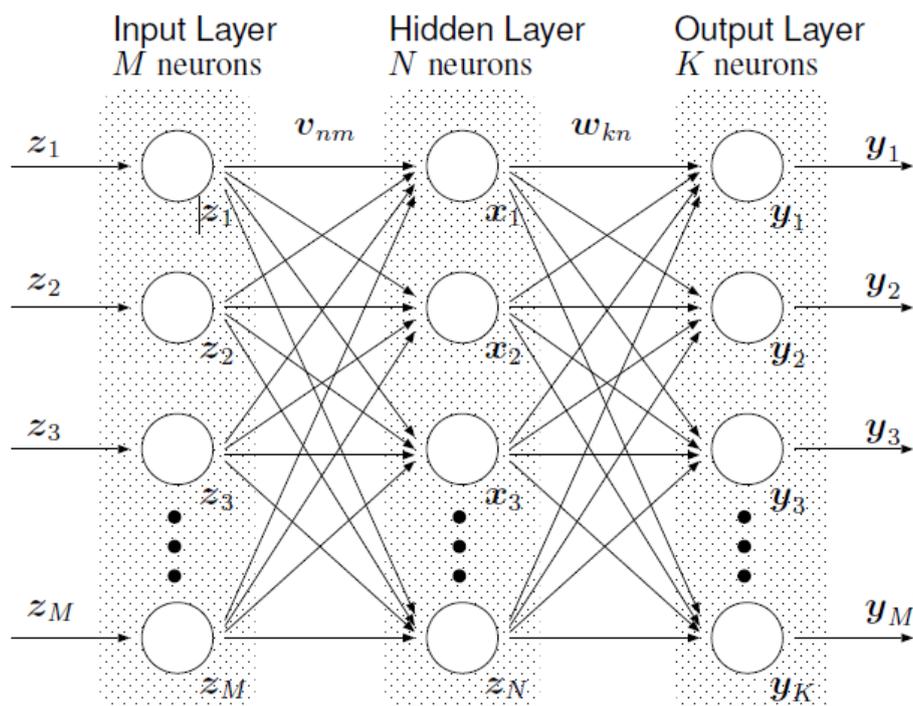
El perceptrón multicapa tiene habilidad para aprender a partir de un conjunto de ejemplos, aproximar relaciones no lineales, filtrar ruido en los datos, etc. Esto hace que sea un modelo adecuado para enfocar problemas reales.

De otro lado, el perceptrón multicapa tiene también una serie de limitaciones, como el largo proceso de aprendizaje para problemas complejos que dependen de un gran número de variables; dificultad en codificar problemas reales mediante valores numéricos, dificultad para realizar un análisis teórico de la red debido a la presencia de componentes no lineales y a la alta conectividad.

### **Arquitectura del Perceptrón Multicapa**

El perceptrón multicapa (MultiLayer Perceptrón, MLP por sus siglas en inglés) es una red formada por una capa de entrada, al menos una capa oculta y una capa de salida.

Esta arquitectura se caracteriza porque tiene sus neuronas agrupadas en capas de diferentes niveles. Cada una de las capas está formada por un conjunto de neuronas y se distinguen tres tipos de capas diferentes: la capa de entrada, las capas ocultas y la capa de salida. En la figura siguiente se observa esta arquitectura.



Fuente: Isokawa et al. "Quaternionic Multilayer Perceptron with Local Analyticity"

Figura 8: Estructura del Perceptrón Multicapa

Fuente: (Isokawa, Kusakabe, Matsui, & Peper, 2003)

La figura muestra una estructura formada por neuronas que propagan la señal hacia la salida. Las conexiones entre neuronas se denominan pesos. Estos pesos son optimizados por el algoritmo de aprendizaje.

Las neuronas de la capa de entrada se encargan únicamente de recibir las señales o patrones externos y propagan dichas señales a todas las neuronas de la siguiente capa. La última capa actúa como salida de la red, proporcionando al exterior la respuesta de la red para cada uno de los patrones de entrada. Las neuronas de las capas ocultas (*hidden layer*) realizan un procesamiento no lineal de los patrones o señales recibidas.

Las conexiones del perceptrón multicapa siempre están dirigidas hacia adelante, esto significa que las neuronas de una capa se conectan con las neuronas de la siguiente capa,

siendo por esto denominadas “redes alimentadas hacia adelante” (redes feedforward). Las conexiones entre las neuronas llevan asociado un número real, denominado peso de la conexión. Todas las neuronas llevan también asociado un umbral, para el caso del perceptrón multicapa se trata como una conexión más con entrada constante e igual a 1.

### **Tipos de conectividad**

- **Conectividad total:** todas las neuronas de una capa están conectadas a todas las neuronas de la siguiente capa. Es decir, las neuronas de la capa de entrada están conectadas a todas las neuronas de la primera capa oculta, las neuronas de la primera capa oculta se conectan a las neuronas de la siguiente capa, etc. Se dice entonces que la red está totalmente conectada.
  
- **Conectividad parcial:**
  - Redes con conexiones de todas o ciertas neuronas de una determinada capa a neuronas de capas posteriores, aunque no inmediatamente posteriores.
  - Redes en las que ciertas neuronas de ciertas capas no están conectadas a neuronas en la siguiente capa, es decir el peso de la conexión es constante e igual a cero.

Al enfocar problemas con el perceptrón multicapa, usualmente se parte de una arquitectura con conectividad total, es decir, todas las neuronas de una capa están conectadas a todas las neuronas de la siguiente capa. No es posible demostrar que si se utilizan arquitecturas en las que se eliminan conexiones o se añaden conexiones de una capa a capas no inmediatamente posteriores, se puedan obtener mejores resultados. Sin embargo, si la naturaleza del problema lo requiere, se pueden diseñar redes basadas en perceptrón multicapa con características de conectividad parcial.

### Características:

- Es una estructura altamente no lineal
- Presenta tolerancia a fallos
- El sistema es capaz de establecer una asociación (relación) entre dos conjuntos de datos
- Existe la posibilidad de realizar una implementación hardware.

### Funciones de Activación para el Perceptrón Multicapa

La propagación se realiza de modo tal que cada neurona hace una combinación lineal de las señales precedentes de las neuronas de la capa anterior siendo los coeficientes de esta combinación los pesos sinápticos. Luego se aplica una función de activación no lineal. Estas neuronas, conocidas como neuronas no lineales, aseguran la alta no linealidad en el MLP. En la siguiente figura se muestra una neurona no lineal:

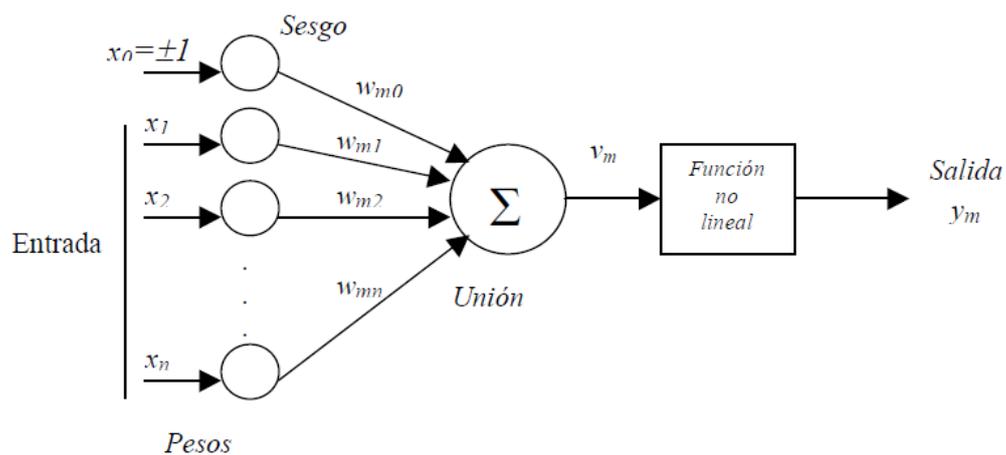


Figura 9: Unidad neuronal elemental no lineal

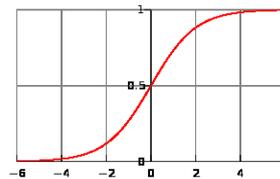
Adaptado de: Serrano y Colaboradores

La función no lineal que se aplica a la salida de la neurona se conoce como función de activación y debe cumplir únicamente dos condiciones: ser continua y diferenciable. Para el perceptrón multicapa las funciones más utilizadas son:

- **Sigmoide:** Su expresión se muestra en la ecuación siguiente y toma valores entre 0 y 1 para una variación de la variable independiente  $x$  entre  $+\infty$  y  $-\infty$

$$f_1(x) = \frac{1}{1 + e^{-x}}$$

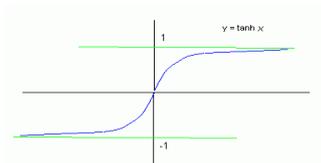
Sigmoide



- **Tangente hiperbólica:** esta función toma valores entre -1 y +1 para una variación de la variable independiente  $x$  entre  $+\infty$  y  $-\infty$ . En este caso, con -1 se codifica la mínima actividad de la neurona y con +1 la máxima. La expresión se muestra así:

$$f_2(x) = \frac{1 - e^{-x}}{1 + e^{-x}}$$

Tangente hiperbolica



Ambas funciones son crecientes con dos niveles de saturación: el máximo que presenta salida 1, y el mínimo con salida 0 para la función sigmoide y salida -1 para la tangente hiperbólica, como se puede apreciar en los gráficos correspondientes a cada función.

Es usual que la función de activación del perceptrón multicapa sea común a todas las neuronas de la red y es elegida por el diseñador, basándose únicamente en los valores de activación que se desee que alcancen las neuronas. Ambas funciones están

relacionadas mediante la expresión:  $f_2(x) = 2 f_1(x) - 1$ , por lo que la utilización de una u otra se elige únicamente en función del recorrido que interese.

Debe notarse que de acuerdo con la naturaleza del problema, las neuronas de salida se distinguen del resto de las neuronas de la red, utilizando otro tipo de función de activación. En este caso, las más usadas son la función identidad y la función escalón.

Se tiene un gran número de posibilidades para disponer estos elementos básicos, el número de neuronas que forman las capas de entrada y salida está determinado por el problema, mientras que el número de capas ocultas y de neuronas en cada una de ellas no está fijado por el problema ni por ninguna regla teórica por lo que quien diseña la red decide la arquitectura en función a la aplicación que se hace de la red y tras realizar un barrido en los valores de estos dos factores.

La alternativa de implementar una red con muchas capas ocultas y una gran cantidad de neuronas en ellas parece lógica, sin embargo esto aumenta la carga computacional lo que implica mayor dificultad de implementación en tiempo real y un incremento del tiempo de aprendizaje por parte de la red. Además la red pierde con esta solución su capacidad de generalización, al aumentar el número de neuronas en la capa oculta, aumenta el número de pesos sinápticos, por lo que la red está caracterizada por más parámetros; esto permite un mejor modelamiento de los patrones utilizados pero se pierde capacidad de generalización ya que un patrón no usado en el modelamiento tendrá más dificultades a la hora de ajustarse a un modelo con un gran número de parámetros.

### **Algoritmo de Retropropagación**

La regla o algoritmo de aprendizaje es el mecanismo mediante el cual se van adaptando y modificando todos los parámetros de la red. En el caso del perceptrón multicapa se trata

de un algoritmo de aprendizaje supervisado, es decir, la modificación de los parámetros se realiza para que la salida de la red sea lo más próxima posible a la salida proporcionada por el supervisor o salida deseada. Por tanto, para cada patrón de entrada a la red es necesario disponer de un patrón de salida deseada.

Dado que el objetivo es que la salida de la red sea lo más próxima posible a la salida deseada, el aprendizaje de la red se formula como un problema de minimización del siguiente modo:

$$\text{Min}_W E$$

siendo  $W$  el conjunto de parámetros de la red – pesos y umbrales- y  $E$  una función error que evalúa la diferencia entre las salidas de la red y las salidas deseadas. En la mayor parte de los casos, la función error se define como:

$$E = \frac{1}{N} \sum_{n=1}^N e(n)$$

donde  $N$  es el número de patrones o muestras y  $e(n)$  es el error cometido por la red para el patrón  $n$ , dado por:

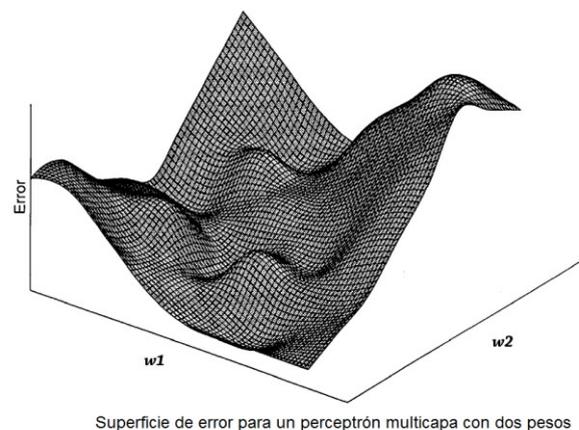
$$e(n) = \frac{1}{2} \sum_{i=1}^{nc} (s_i(n) - y_i(n))^2$$

Siendo  $Y(n) = (y_1(n), y_2(n), \dots, y_{nc}(n))$  y  $S(n) = (s_1(n), s_2(n), \dots, s_{nc}(n))$  los vectores de salidas de la red y salidas deseadas para el patrón  $n$ , respectivamente.

De este modo, si  $W^*$  es un mínimo de la función error  $E$ , en dicho punto el error es próximo a cero, lo cual implica que la salida de la red es próxima a la salida deseada, alcanzando así la meta de la regla de aprendizaje.

Por tanto, el aprendizaje del perceptrón multicapa es equivalente a encontrar un mínimo de la función error. La presencia de funciones de activación no lineales hace que la respuesta de la red sea no lineal respecto a los parámetros ajustables, por lo que el problema de minimización es un problema no lineal y, como consecuencia, tienen que utilizarse técnicas de optimización no lineales para su resolución. Dichas técnicas están, generalmente, basadas en una adaptación de los parámetros siguiendo una cierta dirección de búsqueda. En el contexto de redes neuronales, y en particular para el perceptrón multicapa, la dirección de búsqueda más comúnmente usada es la dirección negativa del gradiente de la función  $E$  –método de descenso del gradiente–, pues conforme al cálculo de varias variables, esta es la dirección en la que función decrece.

Puede apreciarse en la siguiente figura la complejidad del cálculo de un perceptrón multicapa de solo dos pesos ( $y = F(x_1 w_1 + x_2 w_2)$ ). Se hace necesario entonces el desarrollo de métodos optimizados de búsqueda de los mínimos de la función error, en especial cuando hay una gran cantidad de pesos que deben ser ajustados.



*Figura 10: Superficie de error para localizar un mínimo*

## Métodos Optimizados para localización de Mínimos

No obstante, se han desarrollado también métodos de búsqueda aleatoria para localizar el mínimo de dicha función y métodos basados en técnicas evolutivas en los que la búsqueda está guiada por una función de adecuación. Los siguientes aspectos de la optimización de funciones pueden ser considerados según (Cueva, Oliva-Navarro, Diaz-Cortés, & Osuna-Enciso, 2010).

- Un problema de optimización puede ser formulado como un proceso donde se desea encontrar el valor óptimo  $X^*$  que minimiza o maximiza la función objetivo  $f(X)$ , tal que: Minimizar  $f(x)$ ,  $x=(x_1, \dots, x_n) \in \mathbb{R}$  y  $x \in X$
- Debido a que  $f(x)$  puede ser no lineal, se plantean procesos iterativos para la exploración eficiente del espacio de búsqueda.
- Existen dos grupos de algoritmos para resolver esta clase de problemas: las técnicas clásicas (uso del gradiente de la función  $f(x)$ ) y los métodos evolutivos (no necesitan información de la derivada)
- Si bien, en el presente trabajo se utilizan métodos clásicos para resolver los problemas de optimización (minimización de la función Error  $\epsilon$ ).

Cabe tener presente los diferentes métodos evolutivos para mejorar en posteriores trabajos las soluciones planteadas. Entre estos métodos evolutivos tenemos:

- Búsqueda aleatoria
- Templado simulado
- algoritmos genéticos
- estrategias evolutivas
- optimización por enjambre de partículas

- evolución diferencial
- búsqueda armónica
- sistemas inmunológicos artificiales
- optimización basada en electromagnetismo
- colonia artificial de abejas

### **Diseño de la Arquitectura del Perceptrón Multicapa**

Para resolver un problema utilizando el perceptrón multicapa, uno de los primeros pasos a realizar es el diseño de la arquitectura de la red. Este diseño implica la determinación de la función de activación a emplear, el número de neuronas y el número de capas de la red. La elección de la función de activación se hace en base al recorrido deseado, y elegir una u otra no influye en la capacidad de la red para resolver el problema.

Con respecto al número de neuronas y capas, algunas de estos parámetros vienen dados por el problema y otros deben ser elegidos por el diseñador. Así, por ejemplo, tanto el número de neuronas en la capa de entrada, como en número de neuronas en la capa de salida vienen dados por las variables que definen el problema.

En algunas aplicaciones prácticas, no hay lugar a duda sobre el número de entradas y salidas. Sin embargo, existen problemas en los que el número de variables de entrada relevantes para el problema no se conoce con exactitud. En estos casos se dispone de una gran número de variables, algunas de las cuales podrían no aportar información relevante a la red y su utilización podría complicar el aprendizaje pues implicaría arquitecturas de gran tamaño y con alta conectividad. En estas situaciones es conveniente realizar un análisis previo de las variables de entrada más relevantes al problema y descartar aquellas que no aportan información a la red. Este análisis puede llegar a ser una tarea complicada

y requerir técnicas avanzadas, como técnicas basadas en análisis de correlación, análisis de componentes principales, análisis de sensibilidad de redes de neuronas, y técnicas basadas en algoritmos genéticos, entre otras.

El número de capas ocultas y el número de neuronas en estas capas deben ser elegidos por el diseñador. No existe un método o regla que determine el número óptimo de neuronas ocultas para resolver un problema dado. En la mayor parte de las aplicaciones prácticas, estos parámetros se determinan por ensayo y error.

Partiendo de una arquitectura ya entrenada, se realizan cambios aumentando y disminuyendo el número de neuronas ocultas y el número de capas hasta conseguir una arquitectura adecuada para el problema a resolver, que pudiera no ser la óptima, pero que proporciona una solución.

Si bien el número de neuronas ocultas puede influir en el comportamiento de la red –capacidad de generalización de la red-, es necesario indicar que en el caso del perceptrón multicapa, generalmente, el número de neuronas ocultas no es parámetro significativo, pues dado un problema, pueden existir una gran cantidad de arquitecturas capaces de resolver de manera adecuada dicho problema. Además, añadir o eliminar una neurona oculta no influye, de manera significativa, en la capacidad de la red.

En la actualidad existen líneas de investigación abiertas centradas en la determinación automática del número óptimo de neuronas ocultas, así como de capas ocultas, para cada problema en particular. En el caso del perceptrón multicapa, la mayor parte de estos trabajos se basan en la utilización de técnicas evolutivas, las cuales realizan una búsqueda en el espacio de las arquitecturas de redes guiada por la optimización del rendimiento de la red.

## **Aspectos Prácticos de las Redes Neuronales Artificiales**

### **Aplicabilidad**

(Martín-del-Brío & Sanz-Molina, 2006), en su libro denominado Redes neuronales y sistemas borrosos, establecen las condiciones para aplicar redes neuronales artificiales con el objetivo de evaluar la contaminación atmosférica. Enuncian que para plantear un problema y resolverlo mediante redes neuronales artificiales debe tener ciertas características de aplicabilidad, las cuales son:

#### **a. Características que debe cumplir el problema:**

- No se dispone de un conjunto de reglas sistemáticas que describan complementemente el problema.
- Se dispone de muchos ejemplos o casos históricos, condición indispensable para aplicar la técnica de RNA.
- Los datos procedentes del problema son imprecisos o incluyen ruido.
- El problema es de elevada dimensionalidad.
- Si los métodos de redes neuronales proporcionan una alternativa mucho más rápida y sencilla de desarrollar que otras técnicas convencionales.
- Si las condiciones son cambiantes, en este caso se puede hacer uso de la capacidad de las RNA para adaptarse a esos cambios re-entrenando el sistema con nuevos ejemplos.

#### **b. Características que hacen desaconsejable el empleo de RNA.**

- Existe un algoritmo que resuelve con total eficacia el problema. Es el caso de los problemas puramente algorítmicos o numéricos (multiplicación de números, matrices, FFT, etc.)

- No se dispone de un número elevado de casos (ejemplos) para entrenar la red neuronal.
- Tareas críticas o potencialmente peligrosas, cuya resolución deba ser siempre perfectamente predecible y explicable. A veces no resulta fácil interpretar la operación de la red neuronal o predecir con total fidelidad el resultado que pueda proporcionar en todos los casos posibles.

## **Tipos de Redes Neuronales Artificiales**

### **Redes Estáticas**

En estas redes las salidas pueden ser calculadas directamente de las entradas a través de las conexiones hacia adelante. En esta clase podemos encontrar las redes: Perceptrón Multicapa, Funciones de Base Radial, Red Neuronal Probabilística (PNN), etc. Todas ellas tienen como característica común el no poseer memoria, es decir, sólo son capaces de transformar un conjunto de entradas en un conjunto de salidas, de tal manera que una vez establecidos todos los parámetros de la red las salidas únicamente dependen de las entradas. En general la relación deseada de entradas y salidas se determina en este caso externamente mediante alguna forma de ajuste de los parámetros del sistema supervisado. Este tipo de redes se han empleado con éxito en muchos problemas de clasificación, como funciones lógicas, así como en el campo de la aproximación funcional.

### **Redes Dinámicas**

En estas redes las salidas no solo dependen de las entradas actuales a la red, sino también de las entradas, salidas o estados actuales o previos de la red. Estas a diferencia de las

anteriores permiten establecer una relación entre salidas y entradas y/o salidas y entradas previas. Esto añade cierta memoria a estas redes. Matemáticamente, esta memoria se traduce en la aparición de ecuaciones diferenciales o ecuaciones en diferencia formando parte del modelo de las mismas. Como ejemplos de este tipo de redes encontramos la red de Hopfield, la red de retardos en el tiempo (Time Delay Neural Network), la red de tiempo discreto (Time Discrete Neural Network), etc. Las redes neuronales dinámicas se han revelado útiles en problemas de modelización de la dinámica directa e inversa de sistemas complejos, tales como robots, cohetes, naves espaciales, etc., así como en la modelización de circuitos secuenciales y en la conversión de texto a voz.

### **Redes Neuronales NAR**

Las redes neuronales tipo no lineal autoregresivas (NAR) son redes dinámicas recurrentes, con conexiones de retroalimentación que incluyen varias capas de la red. En este tipo de red solo está involucrada una serie de tiempo. Los valores futuros de una serie de tiempo  $y(t)$  se predicen solo en base a valores pasados de la misma serie. Esta forma de predicción es denominada no lineal autoregresiva (NAR) y se expresa de la siguiente manera:

$$y(t) = F(y(t-1), y(t-2), \dots, y(t-n_y))$$

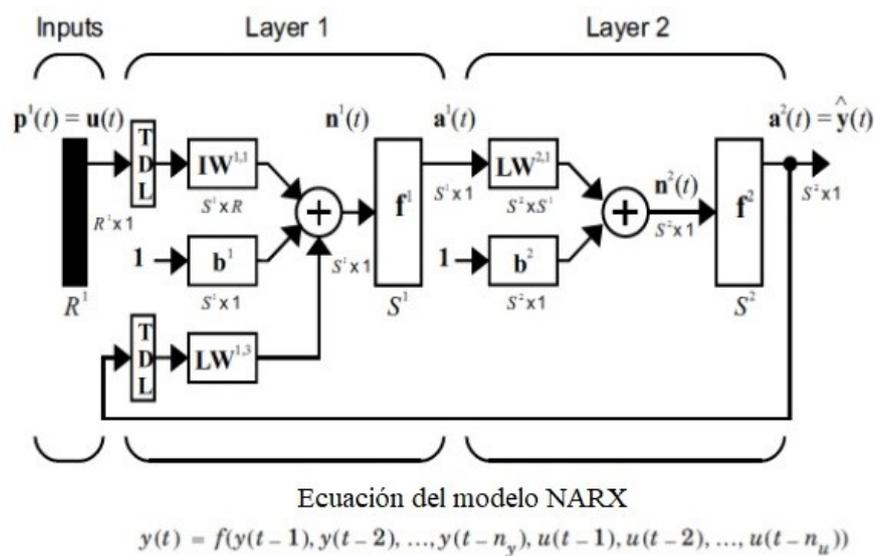
### **Redes Neuronales NARX:**

Las redes neuronales tipo no lineales autoregresivas con entradas exógenas (NARX) es una red dinámica recurrente, con conexiones de retroalimentación que incluyen varias capas de la red. El modelo NARX se basa en el modelo línea ARX, que es comúnmente usado en el modelamiento de series de tiempo.

La ecuación que define el modelo NARX es:

$$y(t) = F [y(t-1), y(t-2), \dots, y(t-n_y), u(t-1), u(t-2), \dots, u(t-n_u)]$$

donde el valor próximo de la señal de salida dependiente  $y(t)$  es regresada sobre valores previos de la señal de salida y valores previos de una señal de entrada independiente (exógena). Se puede implementar un modelo NARX usando una red neuronal feedforward para aproximar la función  $F$ . El diagrama siguiente muestra la red resultante, donde una red feedforward de dos capas se usa para la aproximación. Esta implementación también permite un modelo ARX, donde la entrada y salida pueden ser multidimensionales.



Fuente: [www.mathworks.com](http://www.mathworks.com)

Figura 11: Modelo No Lineal Auto Regresivo con entrada Externa

### Aplicaciones potenciales

(Baughman & Liu, 1995) proporcionan ciertos alcances sobre las aplicaciones potenciales de las redes neuronales artificiales:

- Clasificación – Usa valores de entrada para predecir una categoría de salida. Por ejemplo, dado síntomas y resultados de laboratorio, se determina la enfermedad más probable.
- Predicción – Usa variables de entrada para predecir un valor de salida. Por ejemplo, dado la temperatura, humedad y velocidad del viento, predecir la tasa de evaporación.
- Asociación de Datos – Aprende asociaciones libres de errores o datos ideales, luego clasifica o asocia datos que contienen errores. Por ejemplo aprende cinco patrones ideales y luego reconoce patrones de entrada ruidosos como uno de los cinco patrones.
- Conceptualización de datos – Analiza datos y determina relaciones conceptuales. Por ejemplo, segmenta datos con muchos atributos de modo que se puedan inferir las relaciones de agrupamiento.
- Filtrado de Datos – Suaviza una señal de entrada. Por ejemplo, suaviza una señal de electrocardiograma.
- Optimización – Determina el valor óptimo. Por ejemplo, determina la longitud mínima de viaje para un vendedor.

En general, en aplicaciones de ingeniería donde los problemas son complejos, no lineales y con incertidumbre; una red neuronal bien estructurada y dimensionada puede abordar con éxito un problema de modelamiento complejo. Aún más, debido a que las redes neuronales son herramientas de cómputo no lineales, puede fácilmente aplicarse a problemas de procesamiento no lineal. Además cuando se comparan a modelos empíricos ajustados a curvas, las redes neuronales son relativamente menos sensitivas a ruido e información incompleta, y así pueden manejar el problema de incertidumbre.

## Fortalezas y Debilidades

### Fortalezas

- La información se distribuye sobre un campo de nodos, esto proporciona mayor flexibilidad de la encontrada en el procesamiento simbólico, donde la información se mantiene en una ubicación fija.
- Las RNA tienen la capacidad de aprender, si ocurre un error o una situación nueva que produce resultados imprecisos del sistema, se puede usar una técnica de entrenamiento de error-corrección para corregirlo ajustando la fuerza de las señales emitidas de los nodos hasta que el error de salida desaparezca. En este punto, el sistema ha “aprendido” cómo manejar la nueva entrada. Cuando el sistema encuentre esa situación en el futuro, la red lo modelará en forma adecuada.
- Las RNA permiten indexación extensiva del conocimiento, la indexación del conocimiento es la capacidad de almacenar una gran cantidad de información y accederla con facilidad. Las RNA almacenan conocimiento en dos formas: por las conexiones entre nodos y los factores de peso de estas conexiones.
- Las RNA son más idóneas para procesar datos ruidosos, incompletos o inconsistentes. Ningún nodo dentro de una RNA es directamente responsable de asociar una cierta entrada con una cierta salida. En vez de eso, cada nodo codifica una *microcaracterística* del patrón de entrada-salida. Este concepto implica que cada nodo solo afecta ligeramente el patrón de entrada-salida, así minimizando los efectos de datos incompletos o ruidosos en un nodo dado. Solamente cuando ensamblamos todos los nodos juntos en una única red coordinada, estas microcaracterísticas asocian el patrón macroscópico de entrada-salida.
- Las RNA asemejan el proceso de aprendizaje humano, la mayoría de aprendizaje y solución de problemas ocurre por ensayo-error. Del mismo modo las RNA pueden

ser entrenadas para ajustar iterativamente la fuerza de las conexiones entre los nodos. Después de numerosos ajustes iterativos, la red puede predecir adecuadamente las relaciones causa-efecto.

- Abstracción automatizada, a través del entrenamiento con datos numéricos, las RNA pueden determinar automáticamente lo esencial en las relaciones de entrada-salida. No se necesita un experto en determinado campo del conocimiento para establecer esta relación.
- Potencial para uso online, las RNA pueden tomar mucho tiempo para ser entrenadas, pero una vez entrenadas, estas pueden calcular resultados para una entrada determinada muy rápidamente; teniendo el potencial para ser usado en línea en un sistema de control.

### **Debilidades**

Es necesario comprender las fortalezas y debilidades de las redes neuronales a través de experiencias prácticas. En base a esto, será más factible determinar si una red neuronal es adecuada para una aplicación determinada. Debe entenderse que las redes neuronales son simplemente otra herramienta de modelamiento por computadora que ofrece algunas fortalezas distintas sobre algunas técnicas tradicionales de modelamiento. Algunas de sus debilidades son:

- i. **Tiempos extensos de entrenamiento** – El entrenamiento puede tomar un largo tiempo lo suficiente para hacer impráctica a la red. Los problemas más simples requieren al menos 1,000 etapas de tiempo para entrenar a la red, y los problemas complejos requieren hasta 75,000 etapas. Se debe considerar que con las

computadoras más potentes, los tiempos de entrenamiento pueden dejar de ser un problema.

ii. **Gran cantidad de datos de entrenamiento** – Si se tiene pocos datos de entrada-salida sobre un problema o proceso, se puede reconsiderar el uso de redes neuronales, desde que ellas se basan principalmente en gran cantidad de datos. En consecuencia, las redes neuronales están mejor adecuadas para problemas con una gran cantidad de datos históricos, o aquellos que permiten entrenar la red con un simulador separado.

iii. **No se garantiza resultados óptimos** – La mayoría de técnicas de entrenamiento son capaces de “entrenar” la red, pero eso no garantiza que la red opere adecuadamente. El entrenamiento puede “sesgar” la red, haciéndola precisa en algunas regiones operativas, pero imprecisa en otras. Además, la red puede quedarse en forma inadvertida en un mínimo local durante el entrenamiento.

iv. **No se garantiza 100% de confiabilidad** – Mientras que este concepto se aplica para todas las aplicaciones de computación, este punto es particularmente verdad para redes neurales con datos limitados de entrenamiento.

## **Redes Neuronales Artificiales vs Modelos Empíricos**

### **Datos Ruidosos o Incompletos**

- Las RNA tienen una mejor capacidad de filtrado debido al concepto de microcaracterística.
- Debido a que cada nodo solo codifica una microcaracterística del patrón de entrada-salida, esto afecta al patrón de entrada-salida solo ligeramente.

- Las RNA son masivamente paralelas, de modo que cada nodo opera independiente de los otros. Podemos ver cada nodo como un procesador y ver que estos operan en paralelo.
- La red no depende mucho de un nodo en particular, a diferencia de un modelo empírico que depende de una variable independiente.
- Debido a este paralelismo, las RNA tienen una mejor capacidad de filtrado
- Rinden mejor que los modelos empíricos con datos ruidosos o incompletos.

### **Adaptación**

- Las RNA son más adaptativas que los modelos empíricos.
- Las RNA tienen algoritmos de entrenamiento específicos, donde se ajustan los factores de peso entre nodos hasta que se logre el patrón deseado entrada-salida.
- Si las condiciones cambian tal que el rendimiento de la red es inadecuado, podemos entrenar bajo estas nuevas condiciones para corregir su rendimiento.
- Se puede diseñar la red para que actualice periódicamente su rendimiento entrada-salida, proporcionando un modelo continuo, en línea y autocorregido.
- Los modelos empíricos típicos no tienen esta capacidad.

### **Número de Variables**

- Las RNA son sistemas reales MIMO ‘múltiple-entrada-múltiple-salida’ (por su descripción en inglés Multiple-Input-Multiple-Output).
- La mayoría de modelos empíricos asocian (relacionan) uno, o a lo más dos o tres variables dependientes.
- Las RNA pueden asociar (relacionan) muchas variables independiente con muchas variables dependientes como se necesite.

- Las RNA rinden mejor en reconocimiento de patrones que los sistemas tradicionales de modelamiento empírico.

### 2.3 Marco Conceptual

#### **Definición:**

Bosquejo de la idea o conjunto de ideas sobre los elementos que se visualizan relacionados en torno a nuestro objeto de estudio, que sirve de base para estructurar el fundamento teórico de la investigación.

Representa un conjunto de ideas que, conjugadas con las perspectivas teóricas, la información previa que se tenga al respecto y nuestra experiencia con el fenómeno a estudiar, nos brinda una idea más clara del objeto de estudio a abordar.

Constituye un entramado o red donde las ideas, los conceptos y ciertos argumentos se presentan de manera lógica.

Una delimitación que representa un conjunto de ideas, conceptos y argumentos entrelazados de manera lógica.

#### **Importancia:**

Es útil para visualizar en forma sencilla un fenómeno de la realidad.

En base a esto se deducen algunas conclusiones respecto a la posible relación entre *variables*, conclusiones que por no haber sido verificadas aún son planteadas en forma de *hipótesis*.

Clarifica el pensamiento respecto a la investigación que se desea realizar, así como de la especificación del fenómeno a estudiar (objeto de estudio) y de los elementos directamente con él relacionados.

Brinda las bases para elegir el aparato teórico que fungirá como base para nuestra investigación.

### **Estructura del Marco Conceptual:**

El aspecto estructural del marco conceptual, contemplan dos apartados: *El esquema conceptual* y *la definición conceptual* de los términos que lo integran.

Un *esquema conceptual* puede ser entendido como una representación gráfica o a través de símbolos, de un conjunto de conceptos que corresponden a una realidad, y cuya interrelación puede mostrarse en forma hipotética o teórica.

La *definición conceptual* consiste en delimitar los conceptos establecidos en el esquema conceptual para tener una idea clara de ellos. Con esto se da a entender a los lectores y a la comunidad científica el significado otorgado a estos conceptos.

En el gráfico mostrado a continuación apreciamos un Esquema Conceptual, como expresión de la realidad acotada existente para el tema de predicción de contaminación ambiental por material particulado.

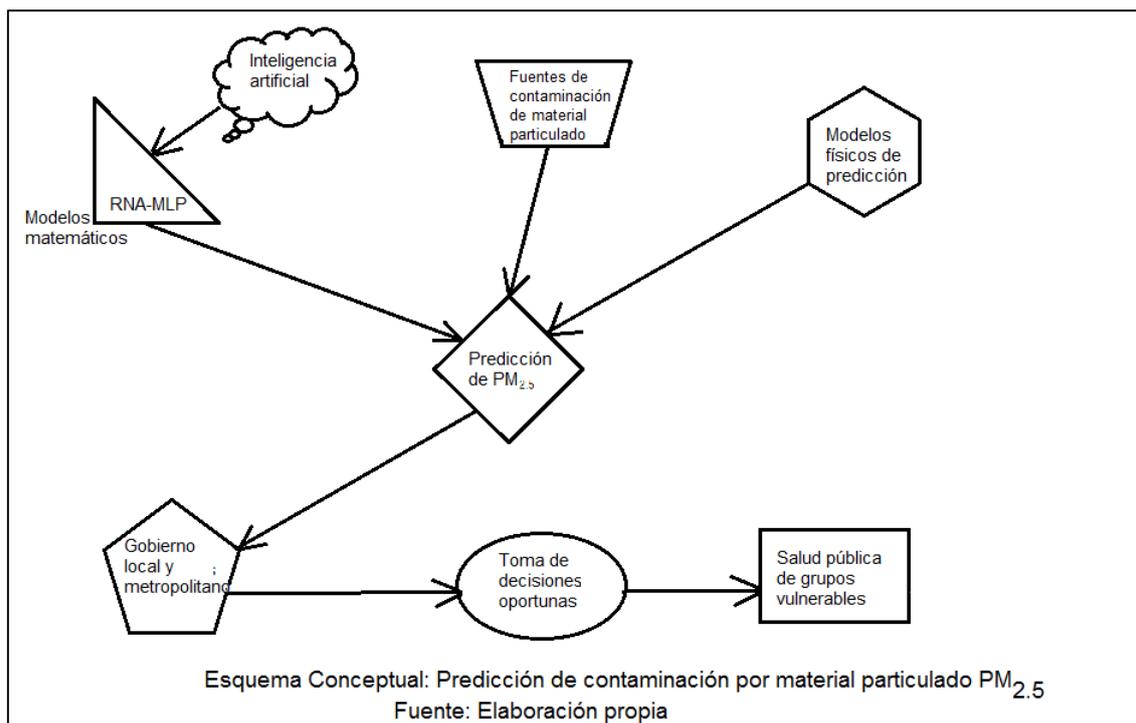


Figura 12: Esquema conceptual

**Definiciones Conceptuales** para los objetos del esquema conceptual anterior:

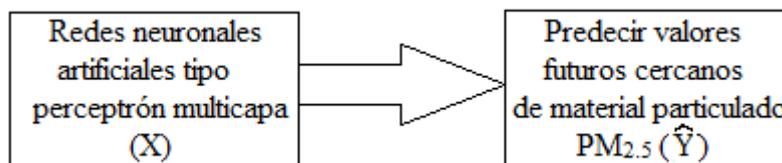
- Inteligencia artificial: una parte de la ciencia de la computación que se relaciona con el diseño de sistemas inteligentes de computación, es decir, sistemas que exhiban características que asociamos con la inteligencia en el comportamiento humano
- Redes neuronales artificiales: sistema de computación compuesto de un número de nodos (unidades de procesamiento) simples y altamente conectados, los cuales procesan información mediante su respuesta de estado dinámico a entradas externas.
- Modelos físicos de predicción: son aquellos que requieren de estudios del modelo del fluido o en túneles aerodinámicos de viento, son muy complejos.

- Fuentes de contaminación por material particulado: son las fuentes que originan la contaminación, pueden ser antropogénicas o naturales, están gobernadas por las leyes de la meteorología
- Predicción de contaminación por material particulado PM<sub>2.5</sub>: para este caso particular se refiere a la predicción de la contaminación de material particulado mediante la aplicación de los modelos matemáticos denominados redes neuronales artificiales
- Gobierno local y metropolitano: son las instancias de gobierno que en forma efectiva pueden tomar decisiones político-administrativas que amortigüen los efectos de la contaminación ambiental en los casos considerados extremos
- Toma de decisiones: ciencia de ejecutar los hechos adecuados para resolver diferentes tipos de problemas, basados en información oportuna y de calidad
- Salud pública de grupos vulnerables: se refiere a la salud de los grupos denominados vulnerables, en especial niños y ancianos. En este caso de estudio se refiere al daño que pueden causar episodios de niveles altos de contaminación por material particulado.

## 2.4 Hipótesis

### 2.4.1 Hipótesis General H<sub>0</sub>

¿Es posible predecir valores futuros cercanos de material particulado PM<sub>2.5</sub> (Y) mediante la aplicación de redes neuronales artificiales del tipo perceptrón multicapa (X) para la zona geográfica definida en Carabayllo-Lima-Lima?



*Figura 13: Esquema de la relación funcional*

#### 2.4.2 Hipótesis Específicas H<sub>i</sub>

- ¿Es posible determinar el modelo óptimo de redes neuronales artificiales del tipo perceptrón multicapa?
- ¿Es posible utilizar el modelo óptimo de redes neuronales artificiales del tipo perceptrón multicapa para realizar una predicción adecuada del valor de material particulado PM<sub>2.5</sub> para periodos futuros cercanos?
- ¿Es posible comparar los valores pronosticados de PM<sub>2.5</sub> con los estándares de calidad ambiental para evaluar la contaminación ambiental para la zona geográfica definida en Carabayllo-Lima-Lima?

#### 2.4.3 Operacionalización de Variables

Con respecto a la variable independiente, se realizarán experimentos modificando los parámetros del modelo de Red Neuronal Artificial: número de capas ocultas, número de neuronas en capa oculta, tipo de vector en las neuronas y tipo de algoritmo de entrenamiento, con el objetivo de determinar los parámetros óptimos de la arquitectura de la red de modo que se pueda realizar un pronóstico adecuado.

En cuanto a la variable dependiente, se obtendrá inicialmente los pronósticos de cada uno de los experimentos diseñados con el programa Matlab, utilizando el criterio del error mínimo cuadrático (MSE). Luego, en una segunda etapa se utilizará los parámetros del modelo óptimo para realizar una simulación del programa Matlab que permitirá obtener las predicciones de 72 datos horarios, se realizará la transformación lineal inversa y se hará la comparación con los estándares de calidad ambiental para facilitar la toma de decisiones por parte del gobierno local, el detalle se muestra en el capítulo 4.

### **Variable Independiente (X)**

Modelo de Predicción basado en Red Neuronal Artificial tipo Perceptrón Multicapa.

### **Dimensiones**

#### **A. Modelo Físico**

##### Indicadores

- Número de variables de entrada (1 a N)
- Número de variables de salida (1 a N)

#### **B. Modelo Matemático**

##### Indicadores

- Número de capas (1 a N)
- Número de retrasos (1 a N)
- Tipo de funciones de activación (sigmoideal, tangente hiperbólica, etc.)
- Pesos (Reales mayores que cero)
- Tipo de predicción (etapa única o multietapa)

**Variable Dependiente (Y)**

Resultados obtenidos mediante la aplicación del modelo de predicción (valores futuros cercanos de concentración de Material Particulado PM<sub>2.5</sub> en ppm)

**Indicadores**

- Error de correlación
- Error Mínimo Cuadrático (MSE o Mean Square Error, por sus siglas en inglés).
- Error Porcentual Medio (MPE o Mean Percentage Error, por sus siglas en inglés).

## CAPÍTULO III: MÉTODO

### 3.1 Tipo de Investigación

Concepto: Se denomina así a los diferentes procedimientos, que de acuerdo a su finalidad, alcance y pertinencia, se determinan como los más apropiados para enfocar un problema de investigación y alcanzar los objetivos propuestos. El tipo de estudio puede definirse como una clasificación genérica de la investigación. De acuerdo a (Morales-Islas, 2016) los tipos de investigación se clasifican de acuerdo a determinados criterios, siendo estos los siguientes:

**Según tendencia**, existen los tipos de investigación denominados:

- Cualitativa. Muestra información sobre el pensamiento de las personas, resumen los sentimientos de la muestra, se usa para grupos pequeños, la muestra participa en el proceso de investigación, se pretende transformar la realidad en referencia a la variable de estudio.
- Cuantitativa. Describe, analiza o experimenta con la variable de estudio; pretende dar un aporte científico sobre un hecho; es real, tangible, observable y medible, reproducible, generalizable, predecible; se toma el mayor número posible de individuos a investigar.

**Según orientación**, existen los tipos denominados experimento y observación. En la observación se tienen varios subcasos entre los que destacan los longitudinales y comparación de causa-efecto.

Los criterios que definen el tipo de estudio (Mendez-Ramirez, Namihira-Guerrero, Moreno-Altamirano, & Sosa-de-Martinez, 2017) son: i) período en que se capta la información; ii) evolución del fenómeno estudiado; iii) comparación de poblaciones

muestrales; y, iv) interferencia del investigador en el estudio. La siguiente figura muestra una clasificación de los tipos de investigación mediante este criterio.

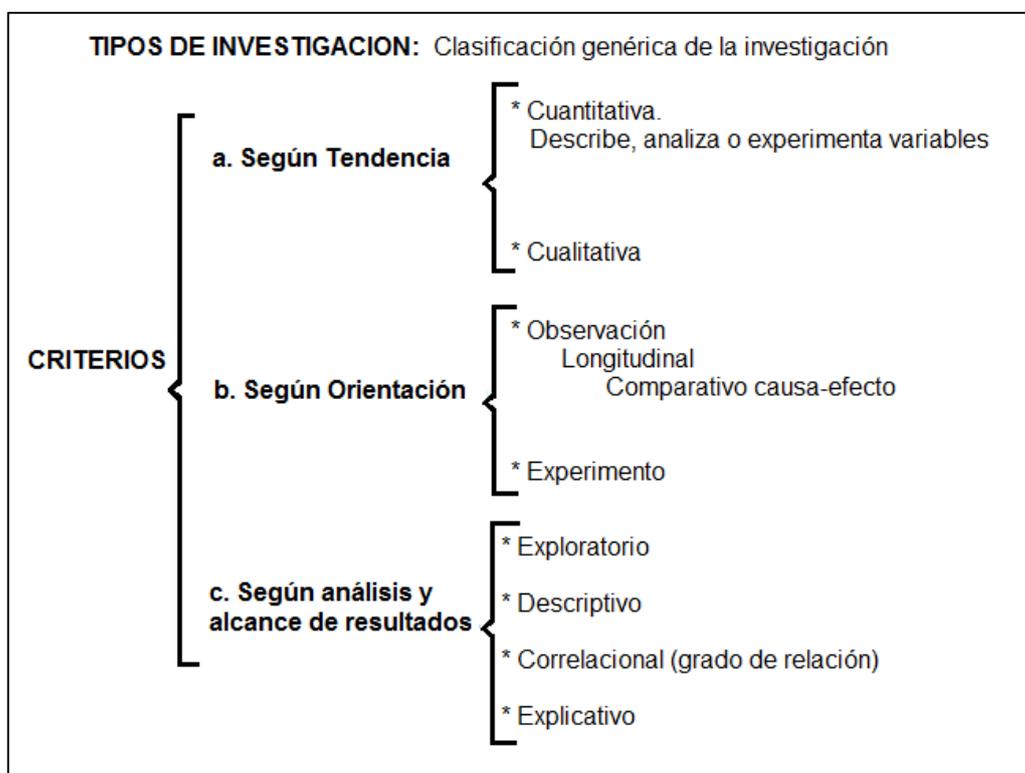
<b>Observacion</b>	Transversal	Retrospectivo o Prospectivo	Descriptivo	Estudio descriptivo	
			Comparativo	Estudio comparativo	
	Longitudinal	Retrospectivo		Descriptivo	Estudio de revisión de casos
				Comparativo de efecto a causa	Estudios de casos y controles
				<b>Comparativo de causa a efecto</b>	<b>Estudio no concurrente o de perspectiva histórica</b>
		Prospectivo		Descriptivo	Estudio de una cohorte
				Comparativo	Estudio concurrente o de varias cohortes
Experimento	Prospectivo	Longitudinal	Comparativo	Experimental	

*Figura 14: Tipos de estudio según orientación.*

#### **Según Análisis y Alcance de resultados**

- Estudios Exploratorios: también conocido como estudio piloto, son aquellos que se investigan por primera vez o son estudios muy poco investigados. También se emplean para identificar una problemática.
- Estudios Descriptivos: Describen los hechos como son observados.
- Estudios Correlacionales: Estudian las relaciones entre variables dependientes e independientes, o sea se estudia la correlación entre dos variables.
- Estudios Explicativos: Este tipo de estudio busca el porqué de los hechos, estableciendo relaciones de causa-efecto. (Hernandez-Sampieri, Fernandez-Collado, & Baptista-Lucio, 2003) establecen estos cuatro tipos de investigación, basándose en la estrategia de investigación que se emplea, ya que el diseño, los datos que se recolectan, la manera de obtenerlos, el muestreo y otros componentes del proceso de investigación son distintos.

La figura siguiente muestra un esquema de clasificación de los tipos de investigación en base a determinados criterios.



*Figura 15: Esquema de clasificación de tipos de investigación basado en criterios*

### 3.2 Diseño de Investigación

Según la web (explorable.com, 2017) el diseño de investigación es el plan o estrategia para responder a la pregunta de investigación. Este diseño señala el camino a seguir para alcanzar los objetivos del estudio y llegar a demostrar o rechazar la hipótesis. Sirve asimismo para el control de las variables y evita la interferencia de las variables extrañas.

El diseño constituye la estructura de cualquier trabajo científico. Brinda dirección y sistematiza la investigación. Los diferentes tipos de diseños de investigación tienen diferentes ventajas y desventajas.

## **Formulación del Diseño de Investigación**

Con el objeto de recabar la información necesaria para absolver las preguntas de investigación, se debe seleccionar un diseño de investigación. Es decir, la manera práctica y precisa que el investigador adopta para cumplir con los objetivos del estudio. Desde que el diseño de investigación indica los pasos a seguir para el logro de dichos objetivos, es necesario que antes de la selección de un diseño en particular se tenga claridad en los objetivos de investigación. En el diseño se selecciona la manera de cómo conseguir respuesta a las interrogantes o hipótesis planteadas dependen de la investigación. Es debido a esto que existen diferentes tipos de diseños de investigación, de los cuales se debe elegir el adecuado para llevar a cabo la investigación.

**Tipos de Diseño de Investigación:** El éxito de los resultados de la investigación dependen de la elección adecuada del diseño de investigación. Líneas abajo se tienen los diferentes tipos de investigación en forma resumida.

Cada diseño posee sus propias características, la eficacia de este diseño depende de que se ajuste lo mejor posible a la investigación. Los diseños experimentales son propios de la investigación cuantitativa, mientras que los no experimentales se aplican en ambos enfoques (cualitativo o cuantitativo). Así, existen dos diseños principales: los experimentales (de laboratorio) y los no experimentales basados en la temporalización de la investigación.

### **Investigación Experimental (de laboratorio)**

Se ocupa de la orientación dirigida a los cambios y desarrollos, tanto en la esfera de las ciencias naturales como de las sociales. El control adecuado es el factor esencia del método utilizado. La ley de variable única debe cumplirse en toda situación experimental.

Esta investigación se presenta mediante la manipulación de una variable no comprobada, en condiciones controladas con rigurosidad, con el objeto de describir de que modo y porque motivo se produce una situación o acontecimiento particular. Este diseño de investigación tiene las siguientes etapas:

- Presencia de un problema para el cual será realizada una revisión bibliográfica
- Identificación y definición del problema
- Definición de hipótesis y variables y la operacionalización de estas
- Diseño del plan experimental
- Prueba de confiabilidad de los datos
- Realización del experimento
- Tratamiento de los datos

### **Investigación no Experimental según la temporalización**

**Método Transversal**: Es el diseño de investigación que recolecta datos en un solo momento y en un tiempo único. El propósito de este método es describir variables y analizar su incidencia e interrelación en un momento dado.

- Diseños transversales descriptivos: son aquellos que tienen como objetivo indagar la incidencia y los valores en que se manifiestan una o más variables.
- Diseños transversales correlacionales: se encargan de describir las relaciones entre dos o más variables en un momento determinado.
- Diseños transversales correlacionales/causales: son aquellos en los cuales las causas y efectos ya ocurrieron en la realidad (estaban dados y manifestados) y el investigador los observa y reporta

**Método Longitudinal**: Es el diseño de investigación que recolecta datos a través del tiempo en puntos o periodos especificados, para hacer inferencias respecto al cambio, sus determinantes y consecuencias. Leído en (<http://metodologia02.blogspot.pe>, 2017).

- Diseños longitudinales de tendencia (trend): son aquellos que analizan cambios a través del tiempo (en las variables o sus relaciones), dentro de alguna población en general.
- Diseños longitudinales de evolución de grupo o cohorte: son estudios que examinan cambios a través del tiempo en subpoblaciones o grupos específicos. Atención a las cohortes o grupos de individuos vinculados de alguna manera, generalmente grupos por edad.
- Diseños longitudinales panel: son similares a las dos clases de diseños antes señalados, solo que el mismo grupo de sujetos es medido en todos los tiempos y momentos.

Del análisis de los diseños de investigación presentados, se concluye que el trabajo a realizar se ubica dentro de la denominada Investigación No Experimental, siendo un Diseño longitudinal es que se va a considerar.

Consideramos los siguientes componentes para el diseño de investigación:

- Definir la información requerida
- Planear las fases exploratoria, descriptiva o causal de la investigación
- Especificar los procedimientos de medida y graduación
- Construir y probar un cuestionario o equivalente a una forma apropiada para la recolección de datos.

- Especificar el proceso de muestreo y el tamaño de la muestra
- Desarrollar el plan de análisis de datos.

### **Diseño de investigación para el presente estudio**

El desarrollo del presente estudio consta de una secuencia de cuatro etapas:

- i. Levantamiento y preparación de los datos. Los datos de campo se obtuvieron mediante el requerimiento de estos datos (en formato de hoja de cálculo) generada en la Estación de Calidad de Aire de Carabayllo. El pre-procesamiento de los datos se realizó mediante el software Matlab.
- ii. Recopilación de la información, incluyó la recolección de libros, tesis, artículos tanto en medio físico como virtual; así como consulta a páginas web de instituciones vinculadas al tema en estudio. Esta información se analizó, se clasificó, se catalogó y se formó el respaldo bibliográfico para la tesis en curso.
- iii. Construcción del modelo predictivo de multietapa adelantada (*multistep ahead*), basada en las redes neuronales artificiales denominadas perceptrón multicapa con la técnica denominada no lineal auto regresiva con entrada exógena (NARX) como propuesta de trabajo.
- iv. Validación del modelo de predicción, mediante comparación del modelo predictivo con datos históricos reales en la estación de estudio y estimación del margen de error.

### **3.3 Estrategia de prueba de hipótesis**

Para probar la hipótesis debemos obtener los resultados de las predicciones deseadas y compararlas con los conjuntos de validación, utilizando algunos de los

parámetros de comparación de errores. Esto implica en una fase inicial formular la herramienta matemática de inteligencia artificial denominada redes neuronales artificiales tipo perceptrón multicapa que permita ejecutar el modelo predictivo de multietapa adelantada deseado con el mínimo error posible. Para probar la hipótesis se utilizará la herramienta estadística denominada error cuadrático medio.

### 3.3.1 Error Cuadrático Medio

La definición de Error Cuadrático Medio puede describir a un estimador o a un predictor:

- **Estimador** (función matemática que asocia una muestra de datos a un parámetro de la población desde la cual se toma la muestra)
- **Predictor** (función matemática que asocia entradas arbitrarias a una muestra de valores de alguna variable aleatoria).

El error cuadrático medio (MSE o Mean Square Error por su denominación en inglés) de un **estimador** (o procedimiento para estimar una cantidad no observada) mide el promedio de los cuadrados de los errores –diferencia entre el estimador y lo que es estimado. El MSE es una medida de la calidad de un estimador, es siempre no negativo y los valores cercanos a cero son mejores. Este enfoque penaliza los errores mayores de pronósticos, ya que eleva cada uno al cuadrado.

El MSE evalúa la calidad de un **predictor** (una función que asocia entradas arbitrarias con una muestra de valores de alguna variable aleatoria. Considerado como un predictor se puede definir el MSE del siguiente modo:

Si  $\hat{Y}$  es un vector de  $n$  predicciones, e  $Y$  es el vector de valores observados correspondientes a las entradas a la función que genera las predicciones, entonces el MSE del predictor puede ser estimado por:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

El Error Cuadrático Medio (MSE) permite comparar los diferentes modelos de pronóstico y seleccionar al que mejor predice los valores, también sirven para monitorear el desempeño de un modelo: cuando se incrementa de repente, significa que el modelo ya no es adecuado.

### 3.3.2 Medición del Error de Pronóstico

Para un pronóstico particular el error ( $e_t$ ) es:  $e_t = y_t - \hat{y}_t$

Donde

$y_t$ : valor real de la cantidad que se está pronosticando

$\hat{y}_t$ : valor pronosticado

El uso del estimador Error Cuadrático Medio (Mean Square Error, MSE) posee un nivel de incertidumbre asociado a su estimación. Mediante los procedimientos de medición del error del pronóstico se puede medir que tan adecuado puede ser un modelo o una técnica para pronosticar.

La decisión para utilizar una técnica de pronóstico en particular, depende de si la técnica producirá errores de predicción que se valoren como suficientemente

pequeños. Los siguientes indicadores evalúan el nivel de precisión de cada método o técnica:

### **Error Porcentual Absoluto Medio (MAPE)**

El Error Porcentual Absoluto Medio (Mean Absolute Percentage Error o MAPE por sus siglas en inglés) es una medida de la precisión de un método de pronóstico y se expresa en porcentaje. Se calcula encontrando el error absoluto en cada período, dividiendo éste entre el valor real observado, para ese período y después promediando estos errores absolutos de porcentaje. Su fórmula es:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

Donde  $y_t$  es el valor real de la cantidad que se está pronosticando,  $\hat{y}_t$  es el valor pronosticado y  $n$  es el número de veces que la variable es pronosticada.

El MAPE indica que tan grandes son los errores de pronóstico comparados con los valores reales de la serie. También se utiliza para comparar la precisión de la misma u otra técnica sobre dos series diferentes.

Tiene la desventaja de no poder usarse si existen valores cero. Asimismo, para pronósticos demasiado bajos el porcentaje de error no puede exceder el 100%; pero para pronósticos demasiado altos no hay límite superior para el porcentaje de error. Cuando se usa MAPE para comparar la precisión de los métodos de predicción este es sesgado en el sentido que se seleccionará en forma sistemática al método con pronósticos demasiado bajos.

### **Error Porcentual Medio (MPE)**

El Porcentaje Medio de Error (Mean Percentage Error o MPE por sus siglas en inglés), se utiliza para determinar si un método de pronóstico está sesgado (pronóstico consistentemente alto o bajo). El MPE se calcula encontrando el error en cada periodo, dividiendo esto entre el valor real de ese periodo y promediando después estos porcentajes de error.

Debido a que se utilizan los valores reales en vez de los absolutos en la fórmula, los errores positivos y negativos pueden compensarse entre ellos; como resultado de esto la fórmula puede ser usada como una medida del sesgo en los pronósticos. Una desventaja de esta medición ocurre cuando alguno de los valores es cero. Su fórmula es:

$$MPE = \frac{100\%}{n} \sum_{t=1}^n \frac{y_t - \hat{y}_t}{y_t}$$

Donde  $y_t$  es el valor real de la cantidad que se está pronosticando,  $\hat{y}_t$  es el valor pronosticado y  $n$  es el número de veces que la variable es pronosticada.

Si un pronóstico no está sesgado (propiedad deseable de un estimador), la ecuación del MPE producirá un porcentaje cercano a cero. Si el resultado es un porcentaje negativo grande, el método de pronóstico está sobrestimado de manera

consistente<sup>1</sup>. Si el resultado es un porcentaje positivo grande, el método de pronóstico está subestimado de forma consistente.

Estos conceptos fueron leídos en las siguientes páginas web (Varios, 2018).

### 3.4 Variables

#### Variable Independiente (X)

Modelo de Predicción basado en Red Neuronal Artificial tipo Perceptrón Multicapa.

#### Indicadores

##### Basados en fenómenos físicos

- Número de variables de entrada (1 a N)
- Número de variables de salida (1 a N)

##### Basados en el modelo matemático

- Número de capas (1 a N)
- Número de retrasos (1 a N)
- Tipo de funciones de activación (sigmoideal, tangente hiperbólica, etc.)
- Pesos (Reales mayores que cero)
- Tipo de predicción (etapa única o multietapa)

---

<sup>1</sup> Se dice que un estimador es **consistente** cuando éste converge a su *valor verdadero* cuando el número de datos de la muestra tiende a infinito.

## **Variable Dependiente (Y)**

Resultados obtenidos mediante la aplicación del modelo de predicción (valores futuros cercanos de concentración de Material Particulado PM<sub>2.5</sub> en ppm)

## **Indicadores**

- Error de Correlación
- Error Mínimo Cuadrático (MSE o Mean Square Error, por sus siglas en inglés).
- Error Porcentual Medio (MPE o Mean Percentage Error, por sus siglas en inglés).

### **3.5 Población**

La población de estudio está constituida por las diez (10) Estaciones de Calidad del Aire<sup>2</sup> instaladas por el Servicio Nacional de Meteorología, Hidrología y Navegación (SENAMHI) en la ciudad de Lima Metropolitana. Se ha incluido en esta población solo a las estaciones que generan información que permite gestionar la calidad de aire en la provincia de Lima. Se han considerado solo las estaciones tipo automatizadas debido a que generan un flujo continuo de datos, los cuales se consideran adecuados para el presente trabajo.

---

<sup>2</sup> Estaciones de calidad del aire instaladas por el SENAMHI, localizadas en los distritos de Ate, San Borja, Jesús María, Santa Anita y Villa María del Triunfo, Carabaylo, Puente Piedra, San Martín de Porres, Huachipa y San Juan de Lurigancho.

### 3.6 Muestra

Para el presente trabajo se consideran muestras a cada estación de calidad del aire ubicadas en una zona geográfica específica (provincia de Lima) y de una tipología específica (estación automatizada). Esto tiene como objetivo tener una población homogénea.

Para esta investigación se ha considerado a una de las estaciones de calidad de aire que cumplen las condiciones de estar dentro de la provincia de Lima y son de la clase automatizada: Estación de Calidad de Aire de Carabayllo, instalada el año 2014 por el SENAMHI en el distrito de Carabayllo y mediante un convenio de colaboración con el gobierno local del distrito.

Una vez seleccionada la muestra y a través de la Gerencia de Servicios a la Ciudad y Medio Ambiente se solicitó al SENAMHI los datos de calidad del aire referidos a los años 2017 y 2018, los cuales fueron proporcionados en formato de hoja de cálculo.

Dadas las características especiales que reviste el procesamiento de datos con la técnica de redes neuronales artificiales, se utilizarán la totalidad de los datos horarios recabados (es decir se utilizará el íntegro de los datos), exceptuando los registros que poseen campos vacíos.

Son 2,888 registros equivalentes a información de cada hora con datos de los contaminantes que han sido registrados en la estación de calidad de aire de Carabayllo, estos datos fueron registrados entre el mes de enero de 2018 al mes de abril de 2018. Cabe destacar que durante el año 2017 no hubo registros del año 2017 para el contaminante ambiental Material Particulado  $PM_{2.5}$ . Estos contaminantes son los siguientes:

- Material particulado,  $PM_{10}$
- Material particulado,  $PM_{2.5}$
- Dióxido de azufre,  $SO_2$
- Dióxido de nitrógeno,  $NO_2$
- Ozono,  $O_3$
- Monóxido de carbono,  $CO$

Cabe destacar que a lo largo de la operación de la estación de calidad de aire, se han tenido varias ocasiones en las que, por diversos motivos, se detuvo el registro de los datos, por ende existe una cantidad del total de registros, tanto en los datos de material particulado  $PM_{2.5}$  como en otros datos, que están sin valor. Para estos casos se utilizará el software Matlab R2015a para eliminar estos campos vacíos, esto con el objeto de que cada fila de datos considerada tenga un valor en cada uno de los contaminantes, de modo que se pueda hacer el procesamiento posterior sin la dificultad que representa la manipulación de datos vacíos en alguna columna. Después de este procesamiento se obtuvo una hoja de cálculo con datos distribuidos en cinco columnas por 820 filas que involucran datos completos para cada celda.

### **3.7 Técnicas de investigación**

#### **3.7.1 Instrumentos de Recolección de Datos**

Dentro de los tipos de medición de contaminación del aire, los datos recolectados corresponden al tipo de mediciones en ambiente (concentraciones de contaminantes a la que está expuesto el público) o monitoreo del ambiente. El

SENAMHI realiza el monitoreo del ambiente tanto en forma manual como automatizada con fines de evaluar valores promedio, máximos y mínimos, comparar con los límites referenciales establecidos por la Organización Mundial de la Salud y explicar su relación con las condiciones meteorológicas observadas. Dentro de Lima Metropolitana el SENAMHI posee una red de monitoreo de calidad de aire con estaciones dotadas de equipos automáticos que monitorean de manera constante los contaminantes denominados: material particulado,  $PM_{10}$ ; material particulado,  $PM_{2.5}$ ; dióxido de azufre,  $SO_2$ ; dióxido de nitrógeno,  $NO_2$ ; ozono,  $O_3$  y monóxido de carbono,  $CO$ . Entre las estaciones automatizadas establecidas en Lima Metropolitana se tiene como muestra a la ubicada en el distrito de Carabaylo, de la cual se han obtenido los datos utilizados en el presente proyecto de investigación.

En la estación de Carabaylo se han instalado los siguientes equipos:

#### Analizadores de contaminantes químicos

- Analizador de Ozono ( $O_3$ )
- Analizador de Óxido de Nitrógeno (NO)
- Analizador de Dióxido de Azufre ( $SO_2$ )
- Analizador de Monóxido de Carbono (CO)
- 

#### Analizadores de contaminantes particulados

- Analizador de  $PM_{10}$
- Analizador de  $PM_{2.5}$

Para la recolección de datos referidos a partículas se cuenta con la tecnología automatizada basada en monitoreo continuo de partículas en el ambiente. Los equipos de análisis automático continuo toman mediciones continuas directas de masa de las partículas mediante el uso de un instrumento denominado micro balanza de oscilación de elementos cónicos (*tapered element oscillating microbalance*, TEOM por sus siglas en inglés). Este tipo de monitores se usan para medir partículas aerotransportadas con excelente precisión de corto término.

**Mecanismo de funcionamiento:** Microbalanza oscilatoria de elementos cónicos, la muestra de aire pasa a través de un filtro que es parte de un sistema que vibra a su resonancia característica. El material particulado colectado sobre el filtro aumenta la masa vibrante y por lo tanto decrece la frecuencia de oscilación en forma proporcional. La concentración del material particulado es calculado a partir de una calibración que relaciona la frecuencia de vibración y la cantidad de material particulado, teniendo en cuenta el volumen de muestra de aire. Leído en (<https://www.thermofisher.com>, 2017).

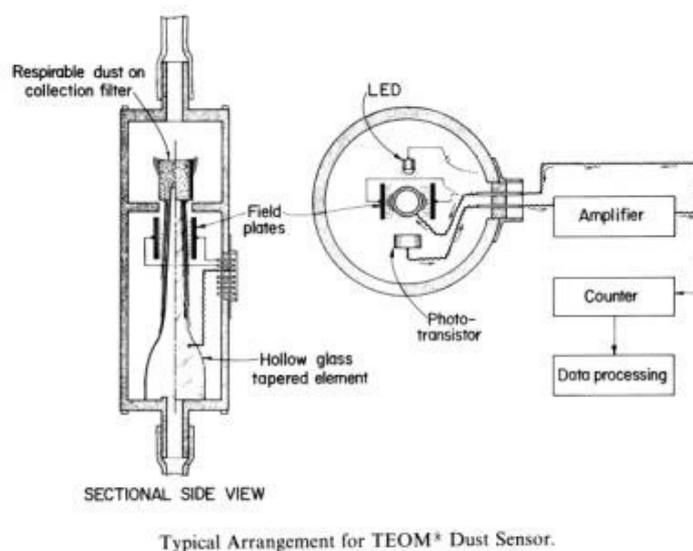


Figura 16: Esquema de un Sensor de partículas TEOM

En el siguiente cuadro se muestran el formato original de datos de calidad de aire emitidos por la estación automatizada de calidad de aire de Carabayllo, y brindados por el Servicio Nacional de Meteorología e Hidrología hacia la Municipalidad de Carabayllo en el marco del convenio de cooperación vinculado a la transmisión de información de calidad de aire para los fines de mejora ambiental persigue la entidad edil.

USUARIO: MUNICIPALIDAD DE CARABAYLLO							
ESTACION	CODIGO	DEPARTAMENTO	PROVINCIA	DISTRITO	LONGITUD	LATITUD	ALTITUD
CARABAYLLO	111286	LIMA	LIMA	CARABAYLLO	77° 02' "W"	11° 54' "S"	190 m s.n.m
PARÁMETRO: MATERIAL PARTICULADO MENOR A 2.5 MICRAS, PM2.5 ( $\mu\text{g}/\text{m}^3$ )							
22/02/2018 17:00	16.45						
22/02/2018 18:00	27.03						
22/02/2018 19:00	32.32						
22/02/2018 20:00	37.61						
22/02/2018 21:00	49.77						
22/02/2018 22:00	42.64						
22/02/2018 23:00	4.47						
23/02/2018 00:00	25.28						
23/02/2018 01:00	44.95						
23/02/2018 02:00	28.96						
23/02/2018 03:00	12.97						
▶	PM10_1h	<b>PM2.5_1h</b>	NO2	CO	<b>SO2</b>	O3	⊕

Tabla 4: Ejemplo de formato original de datos de la Estación de Calidad de Aire

### 3.7.2 Procesamiento y Análisis de Datos

Considerando el marco teórico, se eligió la herramienta de inteligencia artificial denominada red neuronal artificial tipo perceptrón multicapa bajo el modelo no lineal autoregresivo con entrada exógena. Esto debido a su capacidad de predecir el

comportamiento de las series de tiempo aplicadas al caso de variables de contaminación ambiental del aire clasificadas como material particulado.

### Arquitectura del modelo predictivo de red neuronal artificial

En la figura siguiente se presenta la arquitectura de red neuronal artificial basada en un perceptrón multicapa con retroalimentación del tipo NARX (no lineal autoregresivo con entrada exógena). Se presentan además los patrones de entrada, las capas, las neuronas pertenecientes a estas y las salidas.

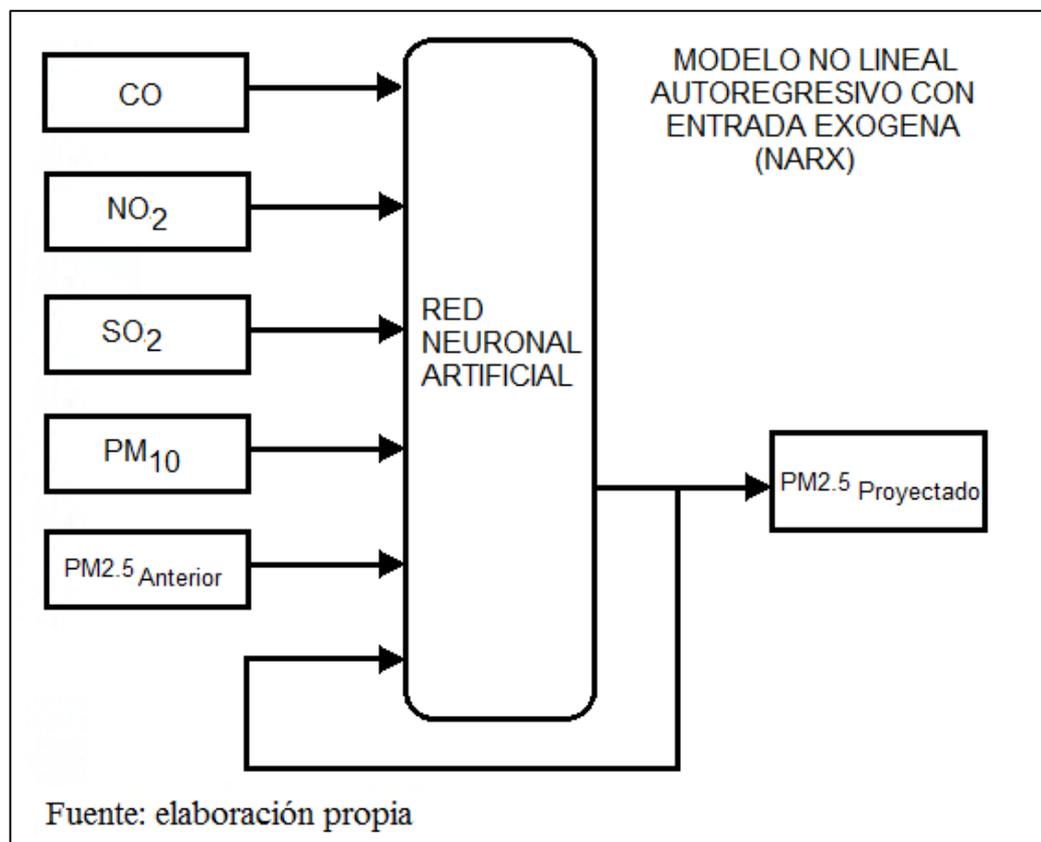


Figura 17: Arquitectura del modelo de red neuronal de predicción

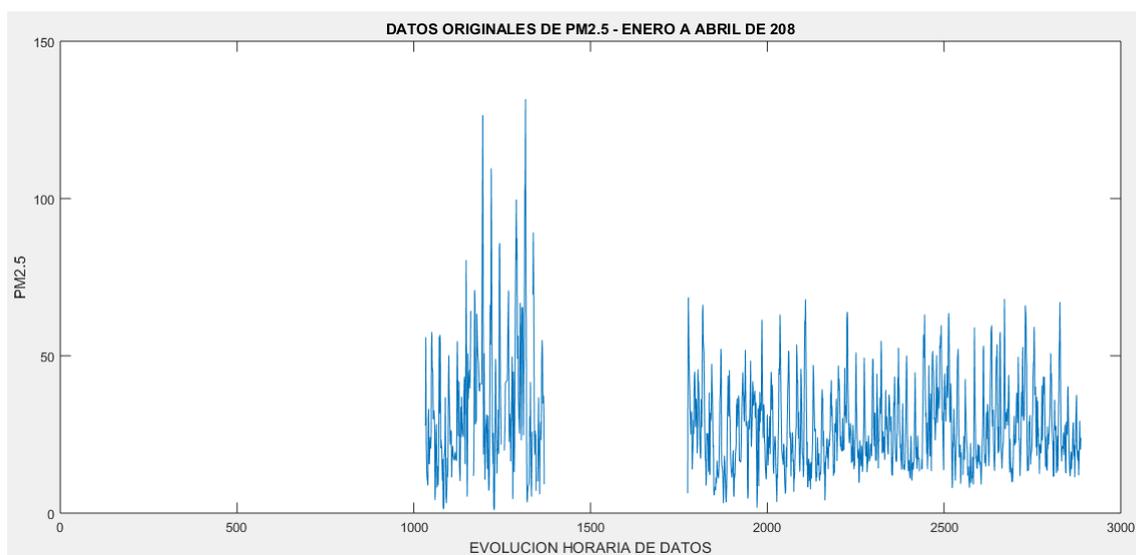
## Procesamiento inicial de datos de entrada

### Importación de Datos.

Para el trabajo inicial con el modelo NARX (no lineal autoregresivo con entrada externa) es necesario importar cinco columnas de datos. Mediante estos comandos se obtiene un vector de 2,888 filas por 5 columnas, este incluye elementos vacíos producto de omisiones de registro en la estación automatizada, por lo que es necesario eliminar los registros sin valores. Para esto deben trasladarse los datos numéricos recibidos en formato de hoja de cálculo Excel hacia el formato especial de datos utilizado por Matlab. El siguiente ejemplo de código Matlab permite este proceso para la variable  $PM_{10}$  y se repite un procedimiento semejante para cada uno de los contaminantes.

```
% comando para importar datos de hoja de cálculo
uiimport

%% Importar datos
VarName1 =xlsread('D:\matlab_lenovo\MUNI-CARABAYLLO-
29JUN18.xls','PM10_1h','B1:B2888');
```



*Figura 18: Datos originales de  $PM_{2.5}$  – Enero a Abril de 2018*

## Eliminación de registros sin valores

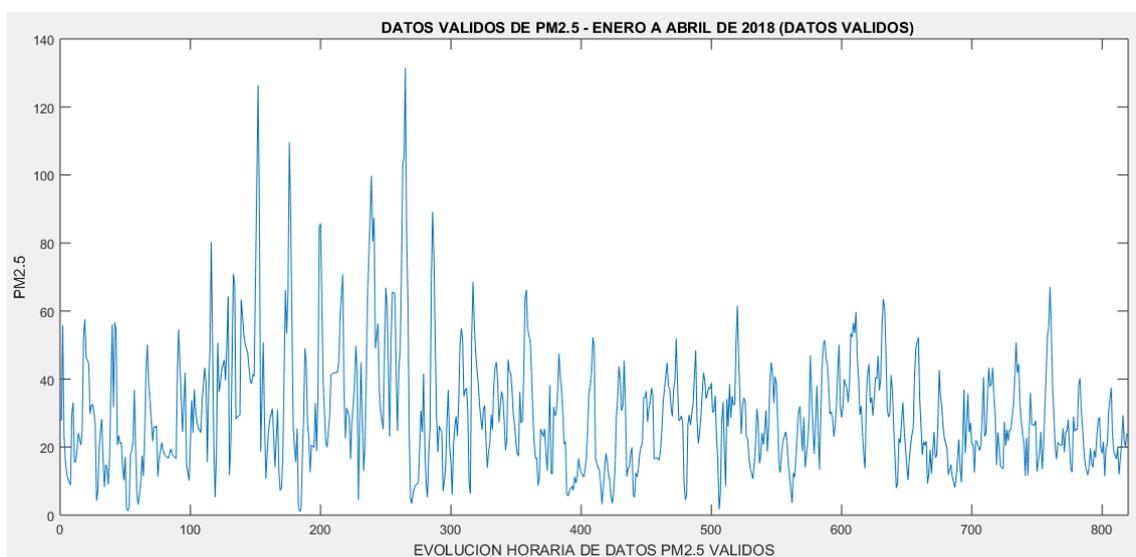
Mediante los siguientes comandos Matlab eliminamos los registros que contienen un campo sin valor (NaN, not a number). Esto permitirá asegurar un procesamiento de datos sin incurrir en distorsiones que se producen al procesar campos vacíos.

```
% H es un vector que contiene los valores importados
% Eliminación de Filas que tengan NaN en cada columna

H(~any(~isnan(H), 2), :)=[];

% Se obtiene la matriz 820x5 double
```

La aplicación del comando *isnan* permite la eliminación de los datos NaN, para realizar el procesamiento de datos solamente sobre valores positivos, eliminando los campos que contienen los datos NaN. Se obtiene de esta manera un vector de 820 filas por 5 columnas de datos válidos para el procesamiento posterior. Es decir no hubo más datos eliminados, y se trabajará con un vector de valores positivos de  $PM_{2.5}$ .



*Figura 19: Datos válidos de  $PM_{2.5}$  – Enero a Abril de 2018*

### Normalización de las entradas (Transformación Lineal)

Se debe considerar que las variables de entrada pueden tener diferencias de valores de varios órdenes de magnitud de forma que el aprendizaje de la red se verá influenciado por estas diferencias, ya que el incremento de pesos de una neurona es proporcional a su entrada. Este problema puede evitarse asignando valores parecidos a las variables de entrada mediante la normalización de sus valores. Para el presente proyecto, se plantea la utilización de la normalización que transforma los datos al rango  $[-1 +1]$  mediante el siguiente segmento de programa Matlab que realiza la transformación lineal para  $PM_{10}$  y semejantes para otros contaminantes:

```
% Transformación Lineal para  $PM_{10}$ 

% Selección de la columna 1 ( $PM_{10}$ )
T1=H_1(:,1);

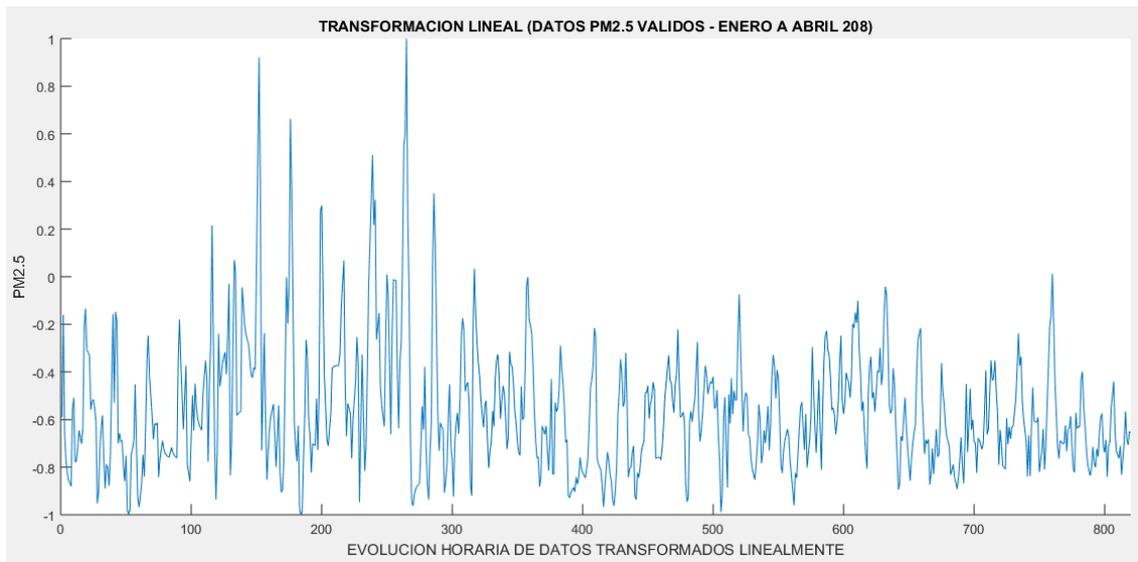
% Máximo y mínimo de los datos de entrada
max1 = max(T1);min1 = min(T1);

% Definimos valores máximo y mínimo de datos de salida
min2 = -1;max2 = 1;

% Creamos  $PM_{10\_B}[]$  array ordinario con datos transformados
for i= 1:length(T1)
     $PM_{10\_B}(i)=((T1(i)-min1)/(max1-min1))*(max2-min2)+min2;$ 
end

% Transposición de  $PM_{10\_B}$ 
 $PM_{10\_B}=PM_{10\_B}';$ 
```

En base a los datos transformados, se elaboró el siguiente gráfico que representa la evolución del comportamiento de los valores diarios del material particulado  $PM_{2.5}$  dentro del rango  $[-1 +1]$ . Con este conjunto de datos se procedió a hacer las evaluaciones de parámetros para el diseño de la red neuronal artificial.



*Figura 20: Transformación Lineal (820 datos válidos de PM<sub>2.5</sub>)*

### **Patrones de entrada (input de la red)**

Los factores que influyen directamente en la predicción de material particulado PM<sub>2.5</sub> son: valores anteriores de PM<sub>2.5</sub>, valores de otros contaminantes químicos, porcentaje de humedad relativa, temperatura, contaminantes químicos, entre otros. Debe considerarse que para el caso de estudio aún no se cuenta con información de datos meteorológicos. Desde que se ha observado que uno de los valores que influencia fuertemente son los propios valores anteriores del material particulado PM<sub>2.5</sub>, así como los de otros contaminantes por material particulado PM<sub>10</sub> y contaminantes químicos y debido a la naturaleza del proyecto de investigación, se considera solo como patrones de entrada para la red neuronal artificial tipo perceptrón multicapa a los valores anteriores de material particulado PM<sub>2.5</sub> y de otros contaminantes químicos presentados en la tabla siguiente:

<b>Patrones de entrada</b>	<b>Unidad</b>	<b>Tipo</b>	<b>Rango</b>
PM <sub>2.5</sub>	ppm	Numérico (Real)	> 0
PM <sub>10</sub>	ppm	Numérico (Real)	> 0
SO <sub>2</sub>	ppm	Numérico (Real)	> 0
NO <sub>2</sub>	ppm	Numérico (Real)	> 0
CO	ppm	Numérico (Real)	> 0
PM <sub>2.5</sub> anterior	ppm	Numérico (Real)	> 0

*Tabla 5: Patrones de entrada (Input)*

Estos elementos de los patrones de entrada se han ordenado como un arreglo de vectores (inicialmente cinco columnas por 820 filas) para alimentar a la red y utilizar las capacidades de aproximador universal de funciones de estas para estimar la función de salida de acuerdo a lo siguiente:

#### **Patrones de salida (output de la red)**

En base a los patrones descritos en la Tabla 5 y las muestras de datos históricos de la concentración de material particulado PM<sub>2.5</sub>, se procedió a evaluar la influencia de los patrones de entrada sobre el comportamiento de la concentración de material particulado PM<sub>2.5</sub>. Con esto se obtiene un arreglo de vectores de dimensión 1xN como patrón de salida, lo cual representan los valores estimados de material particulado PM<sub>2.5</sub>.

<b>Patrones de salida</b>	<b>Unidad</b>	<b>Tipo</b>	<b>Rango</b>
PM <sub>2.5</sub>	ppm	Numérico (Real)	> 0

*Tabla 6: Patrones de entrada (Input)*

### **Determinación del número de capas ocultas de la red**

No existe una metodología específica para determinar el número de capas ocultas (*hidden layer*) en una red neuronal artificial. Sin embargo, cuando el número de capas es mayor a uno, la capacidad de discriminación en el espacio  $\mathbf{R}^N$  se hace más potente, aun cuando en el caso de  $N=3$ , esta capacidad no necesariamente aumenta con más capas ocultas. De este hecho se puede deducir que las capas ocultas variarían entre un mínimo de una y un máximo de tres capas ocultas. La figura 20 presenta los espacios de discriminación que se pueden lograr con diversos números de capas ocultas.

Para el presente caso y con base en investigaciones realizadas por (Baughman & Liu, 1995) se indica que *“se encontró en forma consistente que adicionar una segunda capa oculta mejora significativamente la capacidad de predicción de la red, sin algún efecto perjudicial en la generalización del conjunto de datos de prueba”*. Igualmente se afirma en el texto (pp. 102) que *“agregar una tercera capa oculta, produce capacidades de predicción similares a las de dos capas, pero requiere tiempos mayores de entrenamiento debido a estructuras más complejas”*. De esta manera la recomendación que realiza el autor es el uso de una configuración de dos capas ocultas en relación 30:15 para la arquitectura inicial de la mayoría de las redes neuronales. En base a esto, se formuló la arquitectura inicial con dos capas ocultas y con un número de neuronas en relación de 2:1 entre la primera y segunda capa. Sin embargo, debido a los extensos tiempos de entrenamiento se decidió por utilizar una sola capa oculta donde las neuronas tuvieran la forma de vectores [1x1] y [1x2].

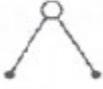
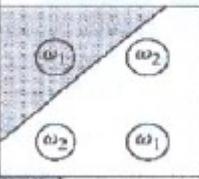
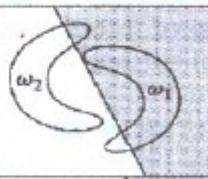
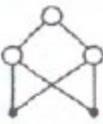
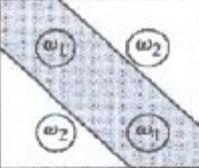
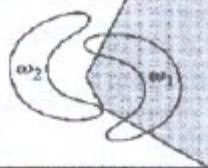
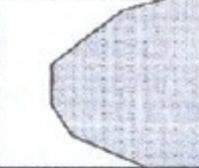
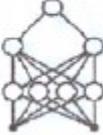
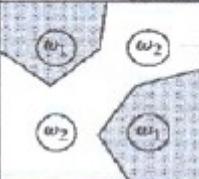
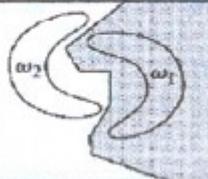
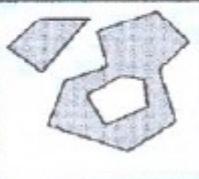
Network structure	Type of decision region	Solution to exclusive-OR problem	Classes with meshed regions	Most general decision surface shapes
Single layer 	Single hyperplane			
Two layers 	Open or closed convex regions			
Three layers 	Arbitrary (complexity limited by the number of nodes)			

Figura 21: Influencia del número de capas en las RNA

### Determinación del número de neuronas por capa

Para iniciar el cálculo se utilizó un múltiplo de 2 en la única capa oculta, se varió el tamaño de los vectores utilizados en las neuronas de la capa oculta utilizando [1x1] y [1x2]. Se han introducido retrasos (delays) en el vector de entrada y en el de realimentación para los valores de concentración de  $PM_{2.5}$  estimada, desde que el valor actual de concentración está directamente relacionado con los N últimos valores registrados.

Para determinar el número óptimo de neuronas en las capas ocultas se ha realizado un proceso de ensayo-error, siendo el objetivo diseñar el número adecuado de neuronas en ambas capas ocultas para que estas pudieran aprender las características de las potenciales asociaciones (relaciones) entre los datos muestrales.

Las funciones de activación (funciones de transferencia) utilizadas para la red neuronal artificial fueron: *tansig* y *purelin*. El siguiente código permite esa selección:

```
net.layers{1}.transferFcn = 'logsig';  
net.layers{2}.transferFcn = 'purelin';
```

Los retrasos en el caso del modelo propuesto de red neuronal artificial hacen que los datos de ingreso no solo dependan de variables externas, sino también de sí mismo en periodos anteriores. La colocación de retrasos a los valores de concentración de  $PM_{2.5}$ , tanto a la concentración estimada como a la de los periodos anteriores, desde que el valor actual de concentración de  $PM_{2.5}$  está directamente relacionado con los últimos N valores registrados.

Inicialmente se tomaron dos retardos para la formulación de la red neuronal, aun cuando este valor varió de acuerdo a los ensayos y simulaciones realizadas, utilizando también cuatro retardos, tanto en la capa de input como en la de feedback.

### **Funciones de activación**

Desde que se conoce que las funciones de activación tienen usos determinados, tal como se puede apreciar:

- Sigmoide y tangente hiperbólica: redes de predicción y redes de pronóstico de procesos que modelan sistemas dependientes del tiempo.
- Función de base radial: redes de clasificación.

Luego, y como estamos en un caso de predicción seleccionamos la función de activación más adecuada como es la tangente hiperbólica (*tansig*, en Matlab). Aplicamos

esta función para la única oculta. Para la función de activación de salida en cambio se utilizó la función lineal (*purelin*, en Matlab).

Una vez definido el tipo de red neuronal artificial a ser utilizado (red neuronal artificial tipo perceptrón multicapa con retrasos), así como sus parámetros y conjuntos de datos de entrada y de salida, se procedió a realizar el diseño de la arquitectura de la RNA. Considerando las conexiones de este tipo, los retrasos y la retroalimentación convierten a esta red en un modelo dinámico.

En base a los conceptos enunciados y utilizando el programa fuente Matlab versión R2015a diseñado para la predicción de multietapa adelantada (Anexo 01), se hicieron las pruebas detalladas en el capítulo IV.

## CAPÍTULO IV: PRESENTACION DE RESULTADOS

### 4.1 Contrastación de Hipótesis

Del capítulo II extraemos las hipótesis planteadas:

#### **Hipótesis General $H_0$**

¿Es posible predecir valores futuros cercanos de material particulado  $PM_{2.5}$  (Y) mediante la aplicación de redes neuronales artificiales del tipo perceptrón multicapa (X) para la zona geográfica definida en Carabayllo-Lima-Lima?

#### **Hipótesis Específicas $H_i$**

- ¿Es posible determinar el modelo óptimo de redes neuronales artificiales del tipo perceptrón multicapa?
- ¿Es posible utilizar el modelo óptimo de redes neuronales artificiales del tipo perceptrón multicapa (X) para realizar una predicción adecuada del valor de material particulado  $PM_{2.5}$  (Y) para periodos futuros cercanos?
- ¿Es posible comparar los valores pronosticados de  $PM_{2.5}$  (Y) con los estándares de calidad ambiental para evaluar la contaminación ambiental para la zona geográfica definida en Carabayllo-Lima-Lima?

A partir de este planteamiento debemos contrastar la hipótesis planteada de modo que pueda verificarse su cumplimiento o nulidad

### **Análisis de la Hipótesis Específica (H<sub>1</sub>)**

Del análisis de la primera hipótesis específica (H<sub>1</sub>) se tiene que se ha podido determinar el modelo óptimo de redes neuronales artificiales tipo perceptrón multicapa mediante una serie de simulaciones sobre la versión R2015a del programa Matlab que permite predecir, con un error cuadrático medio (MSE) mínimo de 0.001009, el conjunto de valores futuros cercanos.

De acuerdo a las simulaciones realizadas se ha logrado el primer objetivo específico del presente trabajo de tesis como es obtener un modelo óptimo de redes neuronales artificiales del tipo perceptrón multicapa para predecir la concentración de material particulado PM<sub>2.5</sub> mediante el uso de parámetros de calidad de aire incluyendo la misma concentración de material particulado PM<sub>2.5</sub>. Este modelo de redes neuronales artificiales del tipo perceptrón multicapa utiliza la técnica denominada “No Lineal Autoregresiva con entrada externa (NARX)”.

El conjunto de datos de entrada y salida se colectó en formato de hoja de cálculo desde la estación meteorológica y de calidad de aire del Servicio Nacional de Meteorología, Hidrología y Navegación ubicado en Carabayllo por un periodo de 2 años (2017-2018). Mediante el software de procesamiento matemático Matlab los conjuntos de datos de input y de target fueron divididos aleatoriamente en dos subconjuntos de datos: entrenamiento y prueba.

Se diseñaron y desarrollaron tres diferentes modelos de redes neuronales artificiales del tipo perceptrón multicapa en la versión R2015a de Matlab, mediante la utilización de los

siguientes parámetros de entrada: material particulado  $PM_{10}$ , material particulado  $PM_{2.5}$ , dióxido de azufre  $SO_2$ , dióxido de nitrógeno  $NO_2$ , monóxido de carbono  $CO$ ; y un solo parámetro de salida: material particulado  $PM_{2.5}$

Los modelos de redes neuronales artificiales del tipo perceptrón multicapa se entrenaron utilizando los siguientes tres algoritmos de retropropagación:

- Levenberg-Marquardt (LM)
- Regulación Bayesiana (BR)
- Scaled Conjugate Gradient (SCG)

Asimismo, se entrenaron siguiendo dos estructuras de datos para las neuronas de la capa oculta: los vectores tipo  $[1 \times 1]$  y  $[1 \times 2]$ .

Se diseñaron los siguientes seis grupos de experimentos, cada uno de los cuales se incluyó un barrido de ocho simulaciones para obtener el menor valor del criterio de evaluación seleccionado. Estas evaluaciones fueron ejecutadas mediante un programa en Matlab versión R2015a:

<b>Algoritmo de Entrenamiento</b>	<b>Capas Ocultas</b>	<b>Neuronas <math>[1 \times 1]</math></b>	<b>Neuronas <math>[1 \times 2]</math></b>
Levenberg-Marquardt	1	2 hasta 16	$[2 \ 1]$ hasta $[16 \ 8]$
Regulación Bayesiana	1	2 hasta 16	$[2 \ 1]$ hasta $[16 \ 8]$
Scaled Conjugate Gradient	1	2 hasta 16	$[2 \ 1]$ hasta $[16 \ 8]$

*Tabla 7: Diseño de las simulaciones para obtener parámetros óptimos de red*

El criterio de evaluación seleccionado está dado por el Error Cuadrático Medio (MSE), calculado mediante la siguiente fórmula:

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2$$

$y_t$ : Valor de la concentración de  $PM_{2.5}$  en el intervalo horario respectivo.

$\hat{y}_t$ : Valor pronosticado de  $y_t$

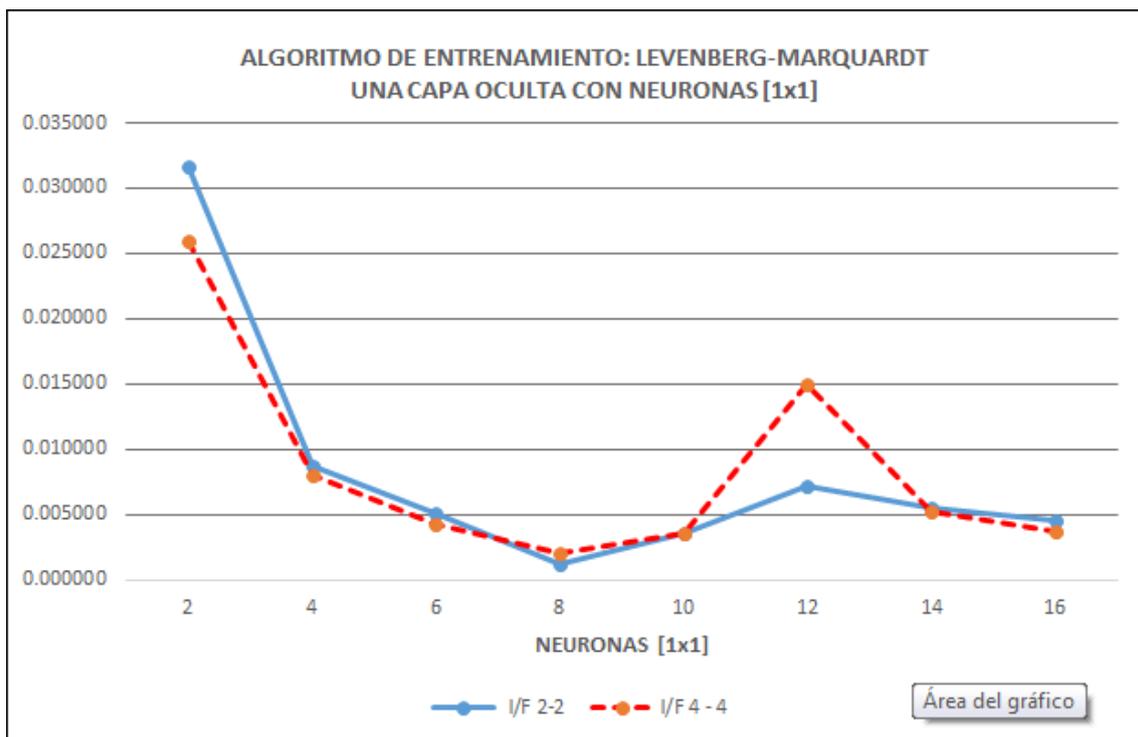
$e_t = y_t - \hat{y}_t$ : Residuo o error de pronóstico

Las tablas y figuras siguientes nos muestran los resultados de las simulaciones de la red neuronal tipo perceptrón multicapa bajo el entrenamiento con los diferentes algoritmos de retropropagación, esto detalla el procedimiento de obtención de los parámetros del modelo óptimo de redes neuronales artificiales del tipo perceptrón multicapa.

**ALGORITMO DE ENTRENAMIENTO: LEVENBERG-MARQUARDT  
UNA CAPA OCULTA CON [1x1] NEURONAS**

		PARÁMETROS				RESULTADOS	
CAPAS OCULTAS	NEURONAS POR CAPA	RETRASOS				MSE	
		INPUT	FEEDBACK	INPUT	FEEDBACK	I/F 2-2	I/F 4-4
<b>1</b>	2	2	2	4	4	0.031630	0.026010
	4	2	2	4	4	0.008702	0.008081
	6	2	2	4	4	0.005124	0.004254
	8	2	2	4	4	0.001280	0.002095
	10	2	2	4	4	0.003593	0.003526
	12	2	2	4	4	0.007218	0.014970
	14	2	2	4	4	0.005579	0.005292
	16	2	2	4	4	0.004560	0.003722
<b>MÍNIMO</b>						<b>0.001280</b>	

*Tabla 8: Entrenamiento con Levenberg-Marquardt (Neuronas [1x1])*



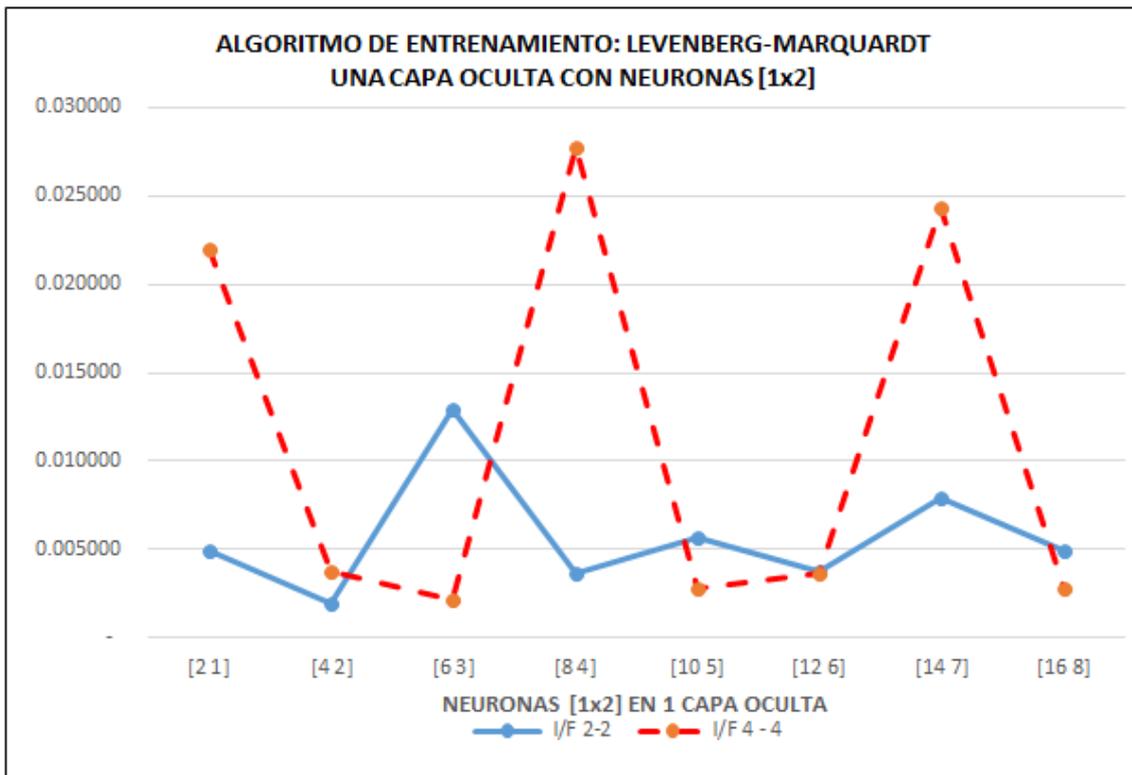
Fuente: Elaboración propia

*Figura 22: Entrenamiento con Levenberg-Marquardt (Neuronas [1x1])*

**ALGORITMO DE ENTRENAMIENTO: LEVENBERG-MARQUARDT  
UNA CAPA OCULTA CON [1x2] NEURONAS**

		EXPERIMENTOS				RESULTADOS	
CAPAS OCULTAS	NEURONAS POR CAPA	RETRASOS				MSE	
		INPUT	FEEDBACK	INPUT	FEEDBACK	I/F 2-2	I/F 4-4
<b>1</b>	[2 1]	2	2	4	4	0.004915	0.021960
	[4 2]	2	2	4	4	0.001924	0.003806
	[6 3]	2	2	4	4	0.012940	0.002122
	[8 4]	2	2	4	4	0.003694	0.027760
	[10 5]	2	2	4	4	0.005728	0.002837
	[12 6]	2	2	4	4	0.003748	0.003664
	[14 7]	2	2	4	4	0.007901	0.024330
	[16 8]	2	2	4	4	0.004895	0.002787
<b>MÍNIMO</b>						<b>0.001924</b>	

*Tabla 9: Entrenamiento con Levenberg-Marquardt (Neuronas [1x2])*



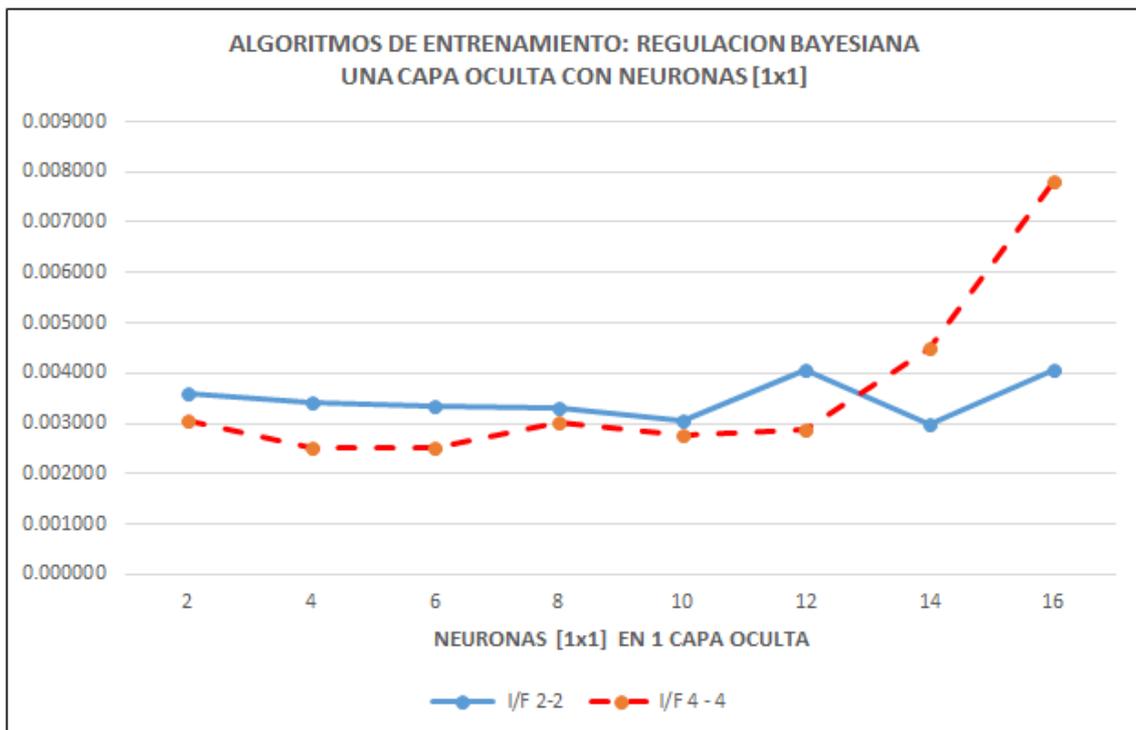
Fuente: Elaboración propia

*Figura 23: Entrenamiento con Levenberg-Marquardt (Neuronas [1x2])*

**ALGORITMO DE ENTRENAMIENTO: REGULACION BAYESIANA  
UNA CAPA OCULTA CON NEURONAS [1x1]**

CAPAS OCULTAS	NEURONAS POR CAPA	PARÁMETROS				RESULTADOS	
		RETRASOS				MSE	
		INPUT	FEEDBACK	INPUT	FEEDBACK	I/F 2-2	I/F 4-4
<b>1</b>	2	2	2	4	4	0.003600	0.003069
	4	2	2	4	4	0.003403	0.002512
	6	2	2	4	4	0.003349	0.002513
	8	2	2	4	4	0.003300	0.003002
	10	2	2	4	4	0.003042	0.002758
	12	2	2	4	4	0.004047	0.002876
	14	2	2	4	4	0.002970	0.004490
	16	2	2	4	4	0.004074	0.007796
<b>MÍNIMO</b>						<b>0.002512</b>	

*Tabla 10: Entrenamiento con Regulación Bayesiana (Neuronas [1x1])*

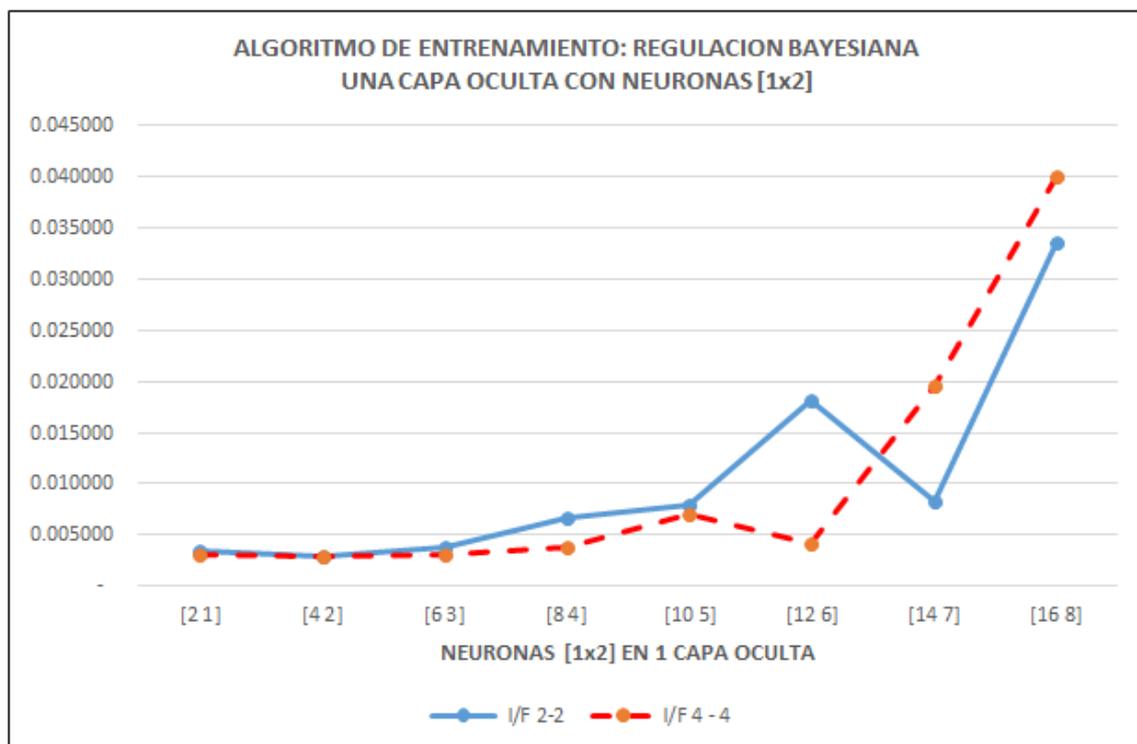


*Figura 24: Entrenamiento con Regulación Bayesiana (Neuronas [1x1])*

**ALGORITMO DE ENTRENAMIENTO: REGULACION BAYESIANA  
UNA CAPA OCULTA CON NEURONAS [1x2]**

CAPAS OCULTAS	NEURONAS POR CAPA	EXPERIMENTOS				RESULTADOS	
		RETRASOS				MSE	
		INPUT	FEEDBACK	INPUT	FEEDBACK	I/F 2-2	I/F 4-4
<b>1</b>	[2 1]	2	2	4	4	0.003535	0.003097
	[4 2]	2	2	4	4	0.002892	0.002837
	[6 3]	2	2	4	4	0.003770	0.003034
	[8 4]	2	2	4	4	0.006623	0.003823
	[10 5]	2	2	4	4	0.007987	0.007099
	[12 6]	2	2	4	4	0.018070	0.004132
	[14 7]	2	2	4	4	0.008249	0.019560
	[16 8]	2	2	4	4	0.033580	0.039970
<b>MÍNIMO</b>						<b>0.002837</b>	

*Tabla 11: Entrenamiento con Regulación Bayesiana (Neuronas [1x2])*

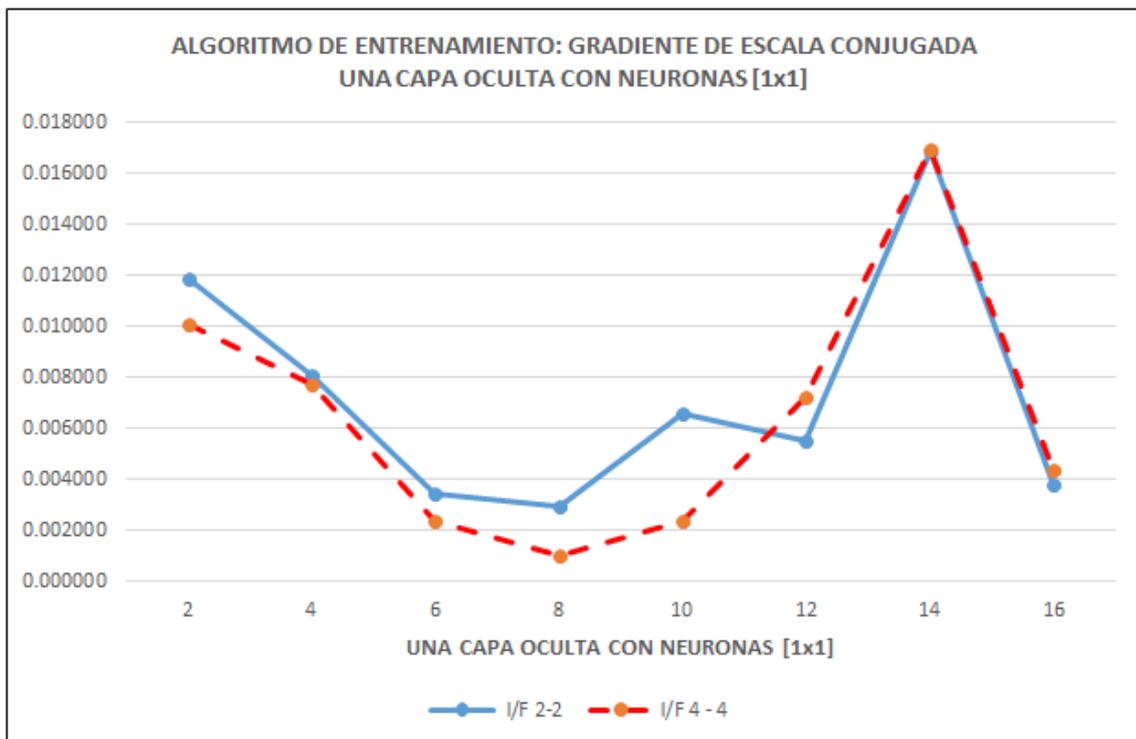


*Figura 25: Entrenamiento con Regulación Bayesiana (Neuronas [1x2])*

**ALGORITMO DE ENTRENAMIENTO: GRADIENTE DE ESCALA CONJUGADA  
UNA CAPA OCULTA CON NEURONAS [1x1]**

		EXPERIMENTOS				RESULTADOS	
CAPAS OCULTAS	NEURONAS POR CAPA	RETRASOS				MSE	
		INPUT	FEEDBACK	INPUT	FEEDBACK	I/F 2-2	I/F 4-4
1	2	2	2	4	4	0.011860	0.010060
	4	2	2	4	4	0.008087	0.007753
	6	2	2	4	4	0.003437	0.002406
	8	2	2	4	4	0.002953	0.001009
	10	2	2	4	4	0.006578	0.002372
	12	2	2	4	4	0.005487	0.007211
	14	2	2	4	4	0.016860	0.016910
	16	2	2	4	4	0.003836	0.004402
<b>MÍNIMO</b>						<b>0.001009</b>	

*Tabla 12: Entrenamiento con Gradiente de Escala Conjugada (Neuronas [1x1])*

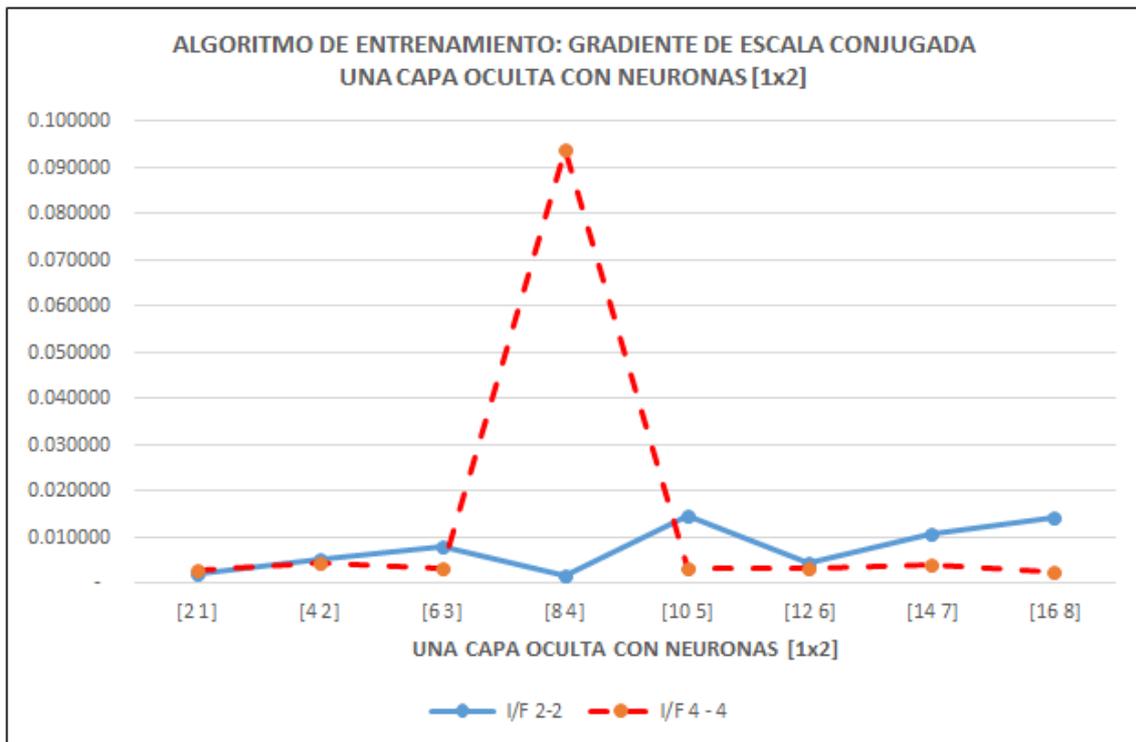


*Figura 26: Entrenamiento con Gradiente de Escala Conjugada (Neuronas [1x1])*

**ALGORITMO DE ENTRENAMIENTO: GRADIENTE DE ESCALA CONJUGADA  
UNA CAPA OCULTA CON NEURONAS [1x2]**

		EXPERIMENTOS				RESULTADOS	
CAPAS OCULTAS	NEURONAS POR CAPA	RETRASOS				MSE	
		INPUT	FEEDBACK	INPUT	FEEDBACK	I/F 2-2	I/F 4-4
1	[2 1]	2	2	4	4	0.002139	0.002693
	[4 2]	2	2	4	4	0.005218	0.004339
	[6 3]	2	2	4	4	0.008107	0.003285
	[8 4]	2	2	4	4	0.001693	0.093860
	[10 5]	2	2	4	4	0.014740	0.003339
	[12 6]	2	2	4	4	0.004250	0.003078
	[14 7]	2	2	4	4	0.010810	0.004099
	[16 8]	2	2	4	4	0.014040	0.002327
<b>MÍNIMO</b>						<b>0.001693</b>	

*Tabla 13: Entrenamiento con Gradiente de Escala Conjugada (Neuronas [1x2])*



*Figura 27: Entrenamiento con Gradiente de Escala Conjugada (Neuronas [1x2])*

En base a los valores de error medio cuadrático<sup>3</sup> (MSE) de las simulaciones basadas en la red neuronal artificial tipo perceptrón multicapa con variación del algoritmo de entrenamiento expresado en las tablas anteriores, seleccionamos los mínimos de estos valores y se presenta el resumen de estos en la tabla siguiente.

<b>MENORES VALORES DE ERROR MEDIO CUADRATICO (MSE) PARA LOS DIFERENTES ESCENARIOS</b>			
<b>CAPAS OCULTAS</b>	<b>ALGORITMOS DE ENTRENAMIENTO</b>	<b>LEVENBERG-MARQUARDT</b>	<b>REGULACION BAYESIANA</b>
		<b>SCALE CONJUGATED GRADIENT</b>	
<b>UNA CAPA OCULTA [1x1]</b>		0.001280	0.002512
<b>UNA CAPA OCULTA [1x2]</b>		0.001924	0.002837
<b>MSE MÍNIMO</b>		<b>0.001009</b>	

Fuente: Elaboración propia.

*Tabla 14: Menores valores de Error Medio Cuadrático*

En base a este valor mínimo de error medio cuadrático (MSE mínimo), identificamos los parámetros que corresponden a este y obtenemos la siguiente tabla 14.

<b>PARÁMETROS DEL MODELO ÓPTIMO</b>		
<b>ALGORITMO DE ENTRENAMIENTO</b>		<b>GRADIENTE DE ESCALA CONJUGADA</b>
<b>RETRASOS</b>	<b>INPUT</b>	4
	<b>FEEDBACK</b>	4
<b>CAPAS OCULTAS</b>		1
<b>TIPO DE NEURONAS</b>		[1x1]
<b>NEURONAS EN CAPA OCULTA</b>		8
<b>MSE MÍNIMO</b>		<b>0.001009</b>

Fuente: Elaboración propia.

*Tabla 15: Parámetros del Modelo Óptimo de RNA*

<sup>3</sup> MSE: función de que mide el rendimiento de la red de acuerdo a la media de los errores cuadráticos. En Matlab se calcula mediante: `Perf = mse(red_neural, target, output)`

### Análisis de la Hipótesis Específica (H<sub>2</sub>)

Después de la obtención de los parámetros del modelo óptimo, se procede a ejecutar la simulación de la red neuronal artificial tipo perceptrón multicapa bajo el modelo no lineal autoregresivo utilizando el mismo algoritmo, pero solamente con los parámetros óptimos. Como resultado de esta simulación se obtienen los valores pronosticados consistentes en 72 datos horarios pronosticados (equivalentes a 3 días) y se muestran los parámetros detallados del proceso de pronóstico originados por la versión R2015a del software Matlab.

La evaluación de estos parámetros permitirá además apreciar la factibilidad de escalar el modelo para un uso intensivo y para ser aplicado en entornos geográficos similares. Este pronóstico permite afirmar que se ha cumplido la hipótesis específica 2, habiéndose obtenido una predicción adecuada de los valores del parámetro de contaminación ambiental denominado material particulado PM<sub>2.5</sub>. La tabla siguiente muestra un resumen de los valores mínimos y sus correspondientes parámetros de evaluación.

MODELO	ALGORITMO DE ENTRENAMIENTO	CAPAS	TIPO DE NEURONA	MSE	PARAMETROS	
					RETRASOS	NEURONAS
1	LEVENBERG-MARQUARDT	1	[1x1]	0.001280	2-2	8
2	LEVENBERG-MARQUARDT	1	[1x2]	0.001924	2-2	[4 2]
3	REGULACION BAYESIANA	1	[1x1]	0.002512	4-4	4
4	REGULACION BAYESIANA	1	[1x2]	0.002837	4-4	[4 2]
5	<b>GRADIENTE DE ESCALA CONJUGADA</b>	<b>1</b>	<b>[1x1]</b>	<b>0.001009</b>	<b>4-4</b>	<b>8</b>
6	GRADIENTE DE ESCALA CONJUGADA	1	[1x2]	0.001693	2-2	[8 4]

*Tabla 16: Tabla de valores óptimos de simulación para cada modelo.*

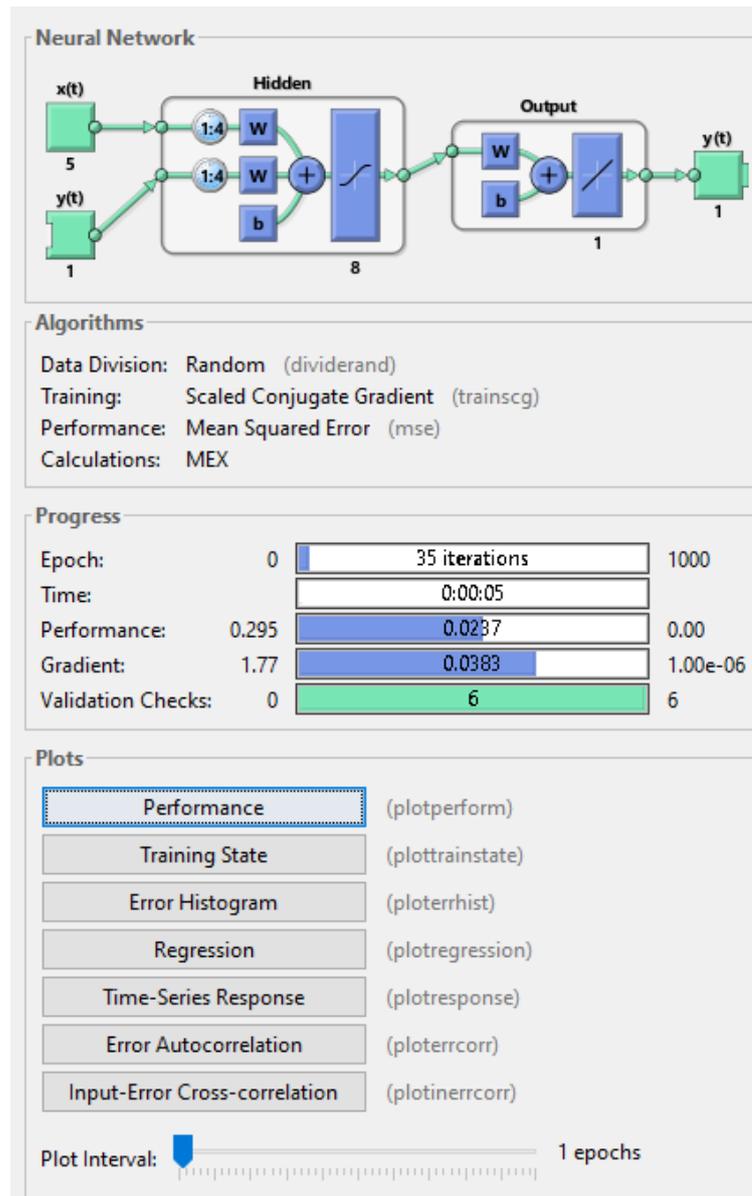


Figura 28: Herramienta de entrenamiento de la red

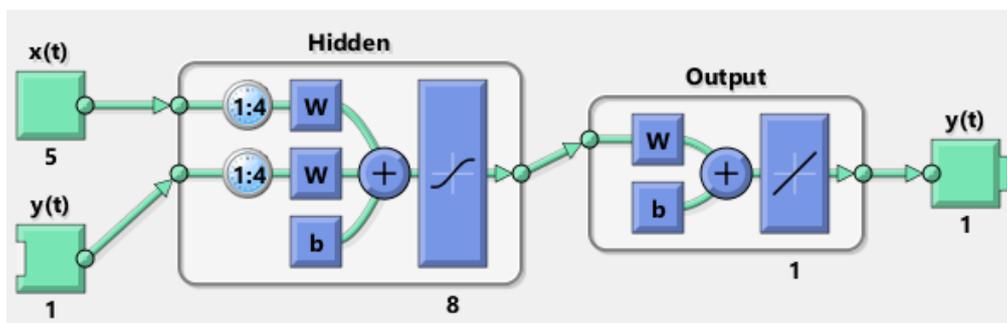


Figura 29: Modelo de Red Neuronal de Lazo Abierto

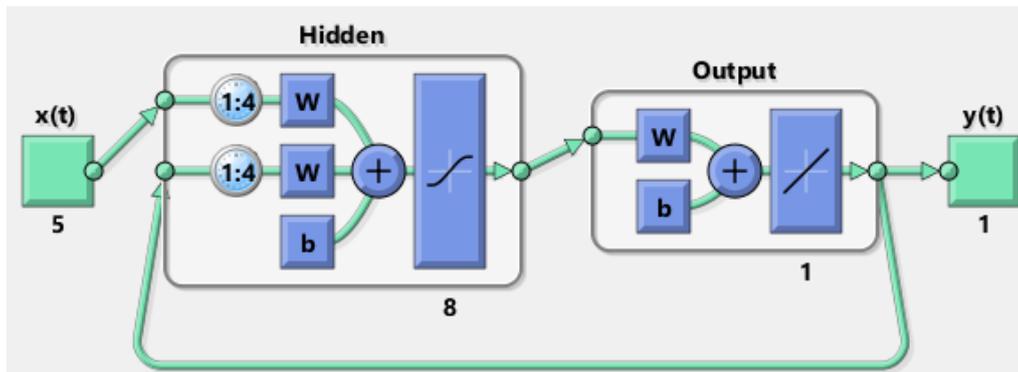


Figura 30: Modelo de Red Neuronal de Lazo Cerrado

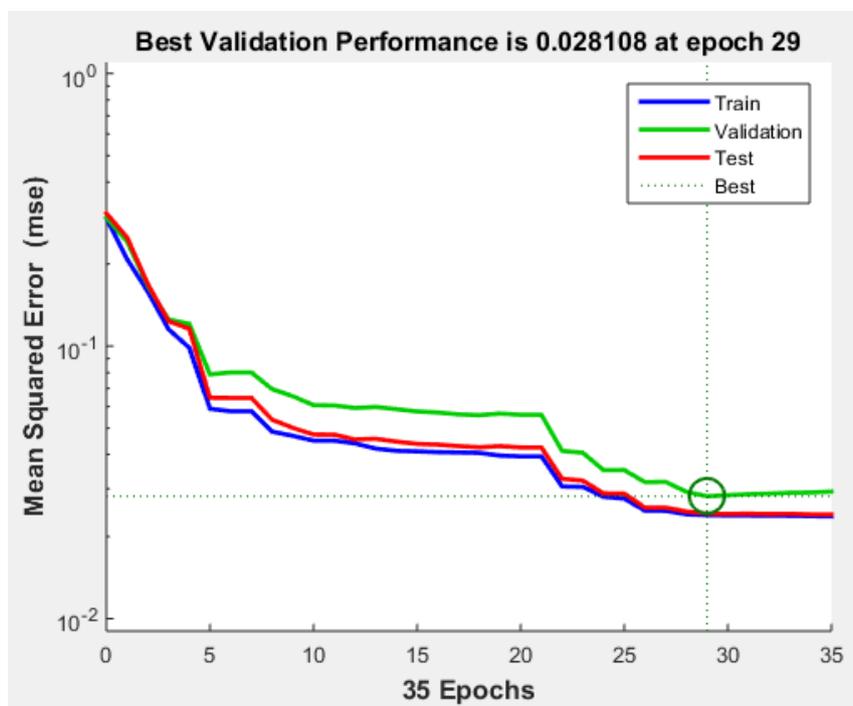


Figura 31: Variación del MSE con el número de iteraciones

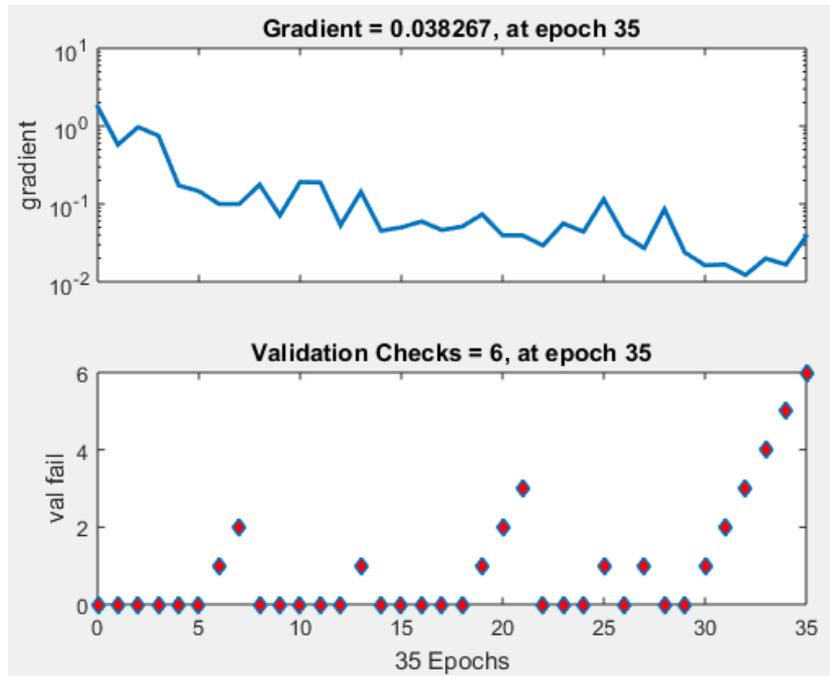


Figura 32: Estado de Entrenamiento de la Red Neuronal

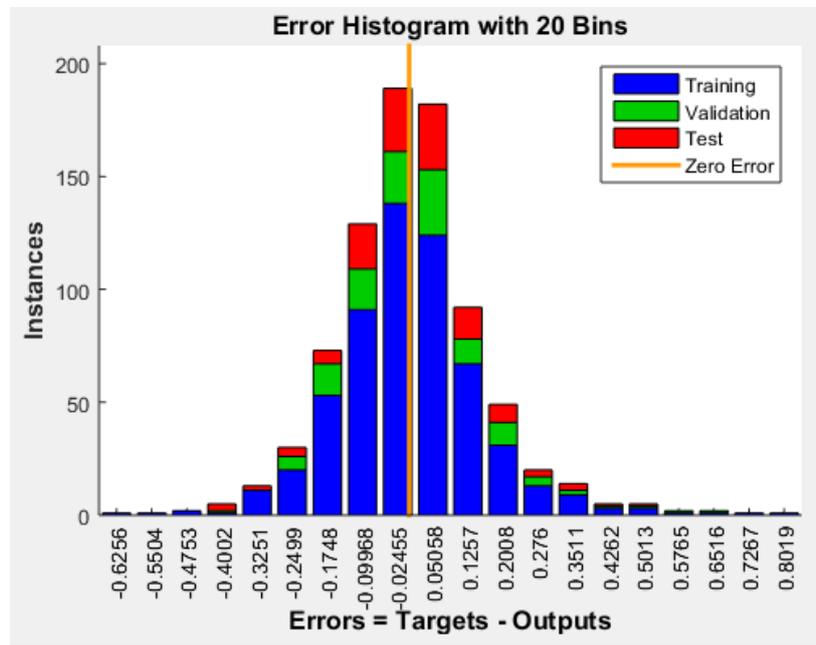


Figura 33: Histograma de Errores

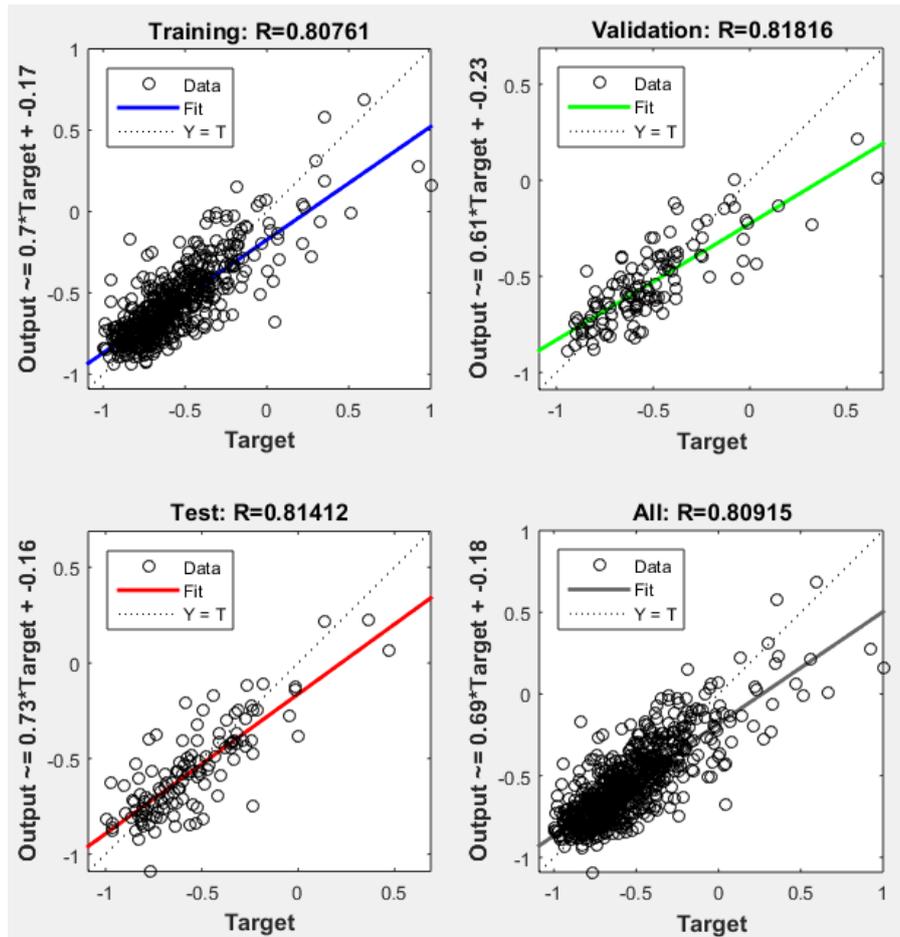


Figura 34: Resultados de Entrenamiento, Validación y Prueba

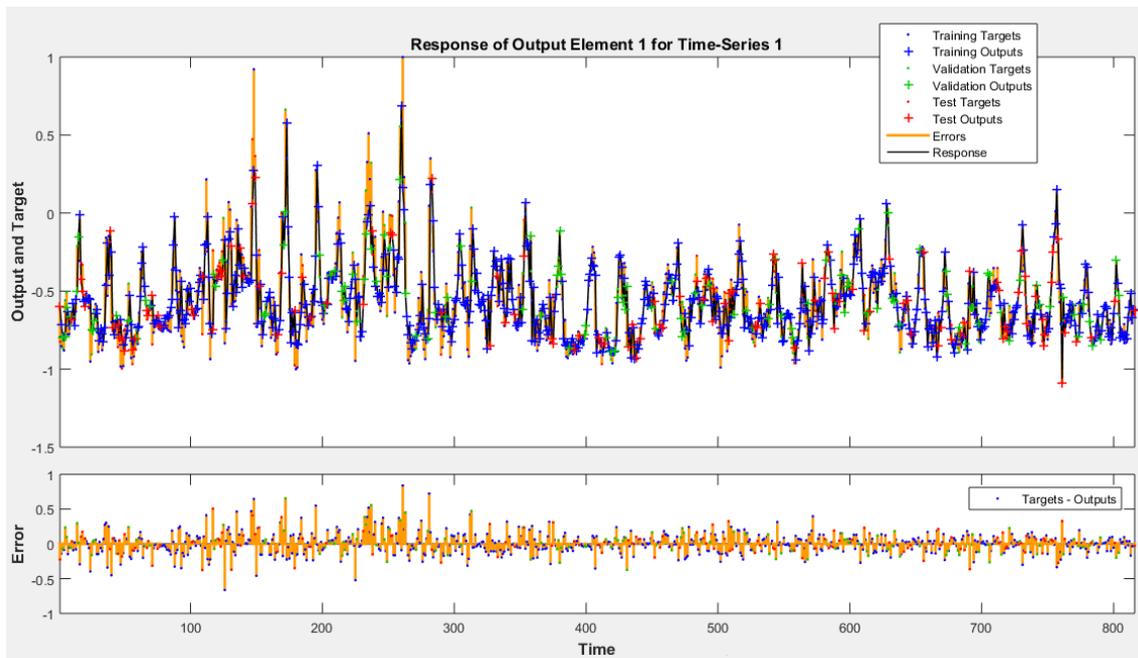


Figura 35: Respuesta de Salida para la Serie de Tiempo

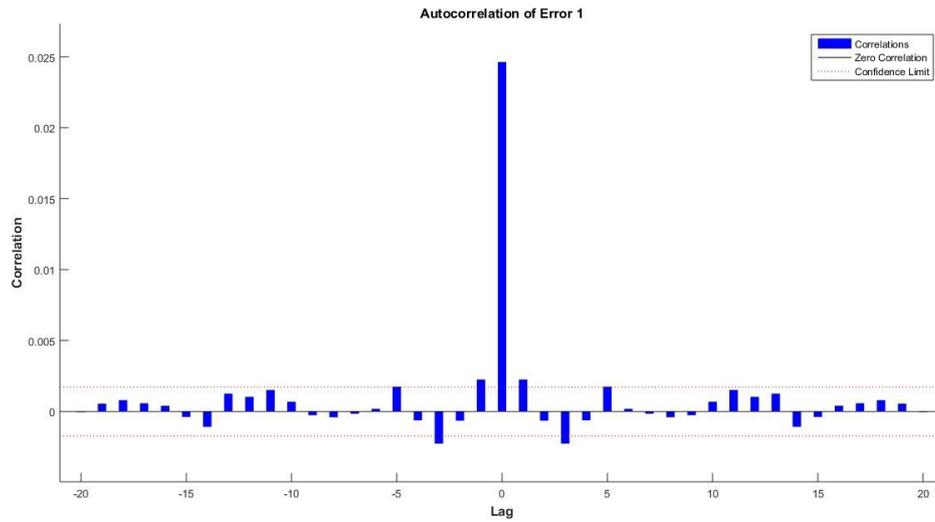


Figura 36: Autocorrelación del Error de Simulación

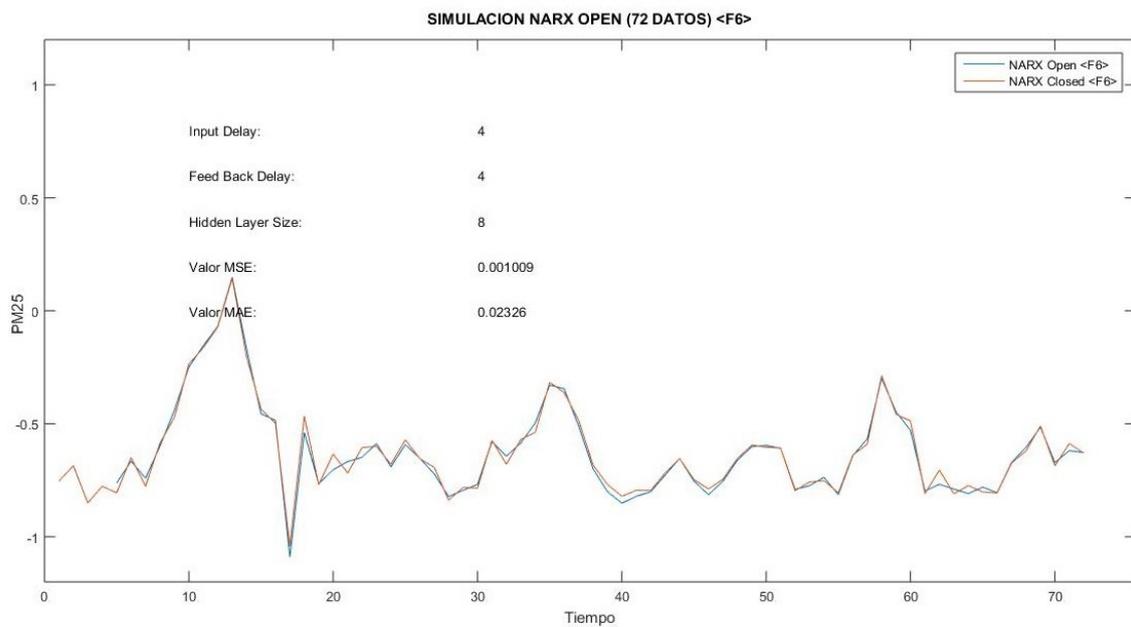


Figura 37: Simulación con Parámetros Óptimos

### Análisis de la Hipótesis Específica (H<sub>3</sub>)

Del análisis de los valores obtenidos de la simulación utilizando los parámetros óptimos se tiene que una primera etapa para evaluar esta hipótesis específica requiere de realizar

una inversión de la transformación lineal de los datos obtenidos para poder realizar la comparación con estándares de calidad ambiental establecidos para el estado peruano.

### **Inversión de la Transformación Lineal**

Mediante las operaciones de transformación lineal, se convirtió todos los conjuntos de valores al rango de diseño del proyecto de investigación que es desde -1 hasta +1. Luego de ejecutar las corridas experimentales mediante los programas en Matlab versión R2015a se obtienen los datos pronosticados mediante la simulación con los parámetros del modelo óptimo (con el MSE más cercano a cero) en el rango de -1 hasta +1.

Se hace necesario entonces invertir la transformación lineal para obtener valores reales de concentraciones de material particulado  $PM_{2.5}$  para el horizonte de predicción establecido. Estos valores serán utilizados asimismo para verificar la tercera hipótesis específica ( $H_3$ ) donde se requieren valores normales de los datos de concentración de material particulado  $PM_{2.5}$ .

### **Estándares de Calidad Ambiental**

Mediante el Decreto Supremo N° 003-2017-MINAM “*Aprueban Estándares de Calidad Ambiental (ECA) para Aire y establecen Disposiciones Complementarias*” publicado en el diario El Peruano el 7 de junio de 2017, se han establecido estándares para la calidad ambiental del aire.

Estándares de Calidad Ambiental para Aire

Parámetros	Período	Valor [µg/m <sup>3</sup> ]	Criterios de evaluación	Método de análisis <sup>[1]</sup>
Benceno (C <sub>6</sub> H <sub>6</sub> )	Anual	2	Media aritmética anual	Cromatografía de gases
Dióxido de Azufre (SO <sub>2</sub> )	24 horas	250	NE más de 7 veces al año	Fluorescencia ultravioleta (Método automático)
Dióxido de Nitrógeno (NO <sub>2</sub> )	1 hora	200	NE más de 24 veces al año	Quimioluminiscencia (Método automático)
	Anual	100	Media aritmética anual	
Material Particulado con diámetro menor a 2,5 micras (PM <sub>2,5</sub> )	24 horas	50	NE más de 7 veces al año	Separación inercial/filtración (Gravimetría)
	Anual	25	Media aritmética anual	
Material Particulado con diámetro menor a 10 micras (PM <sub>10</sub> )	24 horas	100	NE más de 7 veces al año	Separación inercial/filtración (Gravimetría)
	Anual	50	Media aritmética anual	
Mercurio Gaseoso Total (Hg) <sup>[2]</sup>	24 horas	2	No exceder	Espectrometría de absorción atómica de vapor frío (CVAAS) o Espectrometría de fluorescencia atómica de vapor frío (CVAFS) o Espectrometría de absorción atómica Zeeman.  (Métodos automáticos)
Monóxido de Carbono (CO)	1 hora	30000	NE más de 1 vez al año	Infrarrojo no dispersivo (NDIR) (Método automático)
	8 horas	10000	Media aritmética móvil	
Ozono (O <sub>3</sub> )	8 horas	100	Máxima media diaria NE más de 24 veces al año	Fotometría de absorción ultravioleta (Método automático)
Plomo (Pb) en PM <sub>10</sub>	Mensual	1,5	NE más de 4 veces al año	Método para PM <sub>10</sub> (Espectrofotometría de absorción atómica)
	Anual	0,5	Media aritmética de los valores mensuales	
Sulfuro de Hidrógeno (H <sub>2</sub> S)	24 horas	150	Media aritmética	Fluorescencia ultravioleta (Método automático)

NE: No Exceder.

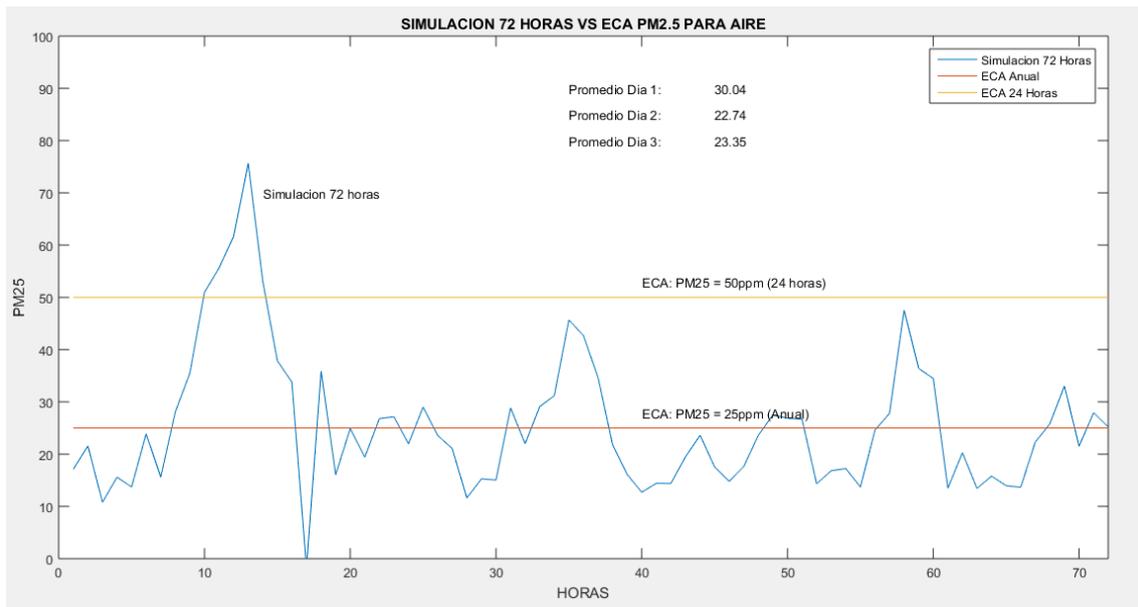
<sup>[1]</sup> o método equivalente aprobado.

<sup>[2]</sup> El estándar de calidad ambiental para Mercurio Gaseoso Total entrará en vigencia al día siguiente de la publicación del Protocolo Nacional de Monitoreo de la Calidad Ambiental del Aire, de conformidad con lo establecido en la Séptima Disposición Complementaria Final del presente Decreto Supremo.

Fuente: Decreto Supremo N° 003-2017-MINAM

Tabla 17: Estándares de Calidad Ambiental para Aire

Desde que los datos de la simulación con el método óptimo de retropropagación y con los parámetros óptimos tienen como resultado final valores que representan la contaminación del aire por material particulado PM<sub>2,5</sub>, se ve mediante el gráfico siguiente que es posible comparar los estándares de calidad ambiental (ECA) establecidos por el estado peruano y realizar el análisis comparativo para inferir situaciones que puedan estar por encima de los estándares establecidos. En base a este análisis de se puede evaluar la incidencia de contaminación atmosférica en la zona geográfica determinada por el distrito de Carabayllo, cumpliéndose así la tercera hipótesis específica.



*Figura 38: Comparación entre resultados de simulación y estándares de calidad del aire (D.S. 003-2017-MINAM)*

Además de la comparación que se aprecia en la figura 40, se elaboró la siguiente tabla con el objeto de ver en cuanto los valores pronosticados sobrepasan los estándares de calidad ambiental, las frecuencias y otros valores estadísticos que puedan ser de utilidad para la toma de decisiones.

PRONOSTICO		MATERIAL PARTICULADO PM <sub>2.5</sub>		
DIA	HORAS	>= 25	>= 50	PROMEDIO
1	01:24	12	5	30.0362
2	25:48	7	0	22.7434
3	49:72	11	0	23.3476
<b>TOTAL</b>	<b>01:72</b>	<b>30</b>	<b>5</b>	<b>25.3757</b>

*Tabla 18: Comparativo de Valores Pronosticados con ECA*

### **Análisis de la Hipótesis General $H_0$**

Luego de evaluar las tres hipótesis específicas planteadas, se puede afirmar con respecto a la Hipótesis General  $H_0$ , expresada como:

*¿Es posible predecir valores futuros cercanos de material particulado  $PM_{2.5}$  (Y) mediante la aplicación de redes neuronales artificiales del tipo perceptrón multicapa (X) para la zona geográfica definida en Carabayllo-Lima-Lima?*

Que se verifica el cumplimiento de la hipótesis general  $H_0$ , con lo cual se puede afirmar que es posible utilizar la técnica de inteligencia artificial denominada redes neuronales artificiales del tipo perceptrón multicapa para predecir valores futuros cercanos de material particulado para una zona geográfica definida.

### **4.2 Análisis e Interpretación**

En primer lugar se obtuvo los parámetros óptimos del modelo de redes neuronales del tipo perceptrón multicapa para la predicción de la evolución de la concentración de material particulado  $PM_{2.5}$ . En un segundo paso se utilizó estos parámetros óptimos para obtener los valores de las concentraciones de material particulado  $PM_{2.5}$  en el intervalo de 72 horas del futuro cercano. Después de obtener estos resultados y previa transformación lineal inversa, se han comparado estos valores con los estándares de calidad ambiental (ECA) para material particulado  $PM_{2.5}$ . En base a este procedimiento se puede enunciar lo siguiente:

- Con un buen grado de aproximación (error cuadrático medio o MSE de 0.001009), ha sido posible pronosticar valores futuros cercanos que abarcan datos referidos a 72 horas o 3 días, los cuales junto con la transformación lineal inversa y la posterior comparación con las normas de calidad de aire del estado peruano puede ser utilizadas para diseñar políticas de gobierno local del distrito de Carabayllo que permitan mejorar los episodios de calidad de aire que superen los máximos valores establecidos en los estándares de calidad del aire.
- Los experimentos fueron llevados a cabo mediante el diseño de una red neuronal artificial del tipo perceptrón multicapa con una sola capa oculta. Se analizaron tres técnicas de entrenamiento mediante retro propagación, realizando dos análisis para cada algoritmo que consistieron en variar el número de neuronas en la capa oculta utilizando vectores de dimensiones  $[1 \times 1]$  y  $[1 \times 2]$ .
- Mediante corridas previas de la simulación de la red neuronal artificial del tipo perceptrón multicapa y mediante el modelo no lineal autoregresivo con entrada externa, se determinó que el número de neuronas en la única capa oculta varíe desde 2 hasta 16 en el caso del vector  $[1 \times 1]$  y desde  $[2 \ 1]$  hasta  $[16 \ 8]$  para el caso de los vectores  $[1 \times 2]$ .
- Para los rangos establecidos de simulación se obtuvieron tiempos razonables de entrenamiento, además los mejores resultados siempre se obtienen para casos de un número neuronas del tipo múltiplo de 2 y en especial del formato  $2^N$  en la única capa oculta del experimento.
- Para el experimento se han considerado determinados parámetros de entrada entre los cuales se incluye en forma recursiva al parámetro material particulado  $PM_{2.5}$ . La inclusión recursiva del parámetro a pronosticar implica el uso del modelo no lineal autoregresivo con entrada externa.

## CAPITULO V: DISCUSION

### 5.1 Discusión

- Se verifica que los modelos matemáticos basados en redes neuronales artificiales pueden utilizarse como herramientas de predicción para concentraciones de material particulado  $PM_{2.5}$  en áreas urbanas. Un modelo estadístico<sup>4</sup> basado en redes neuronales artificiales, entrenado con datos específicos en tiempo y ubicación geográfica, puede utilizarse con menos recursos que el uso de un modelo determinístico<sup>5</sup>.
- Mediante el pronóstico de las variables de contaminación del aire por material particulado  $PM_{2.5}$  se pueden evaluar los resultados de proyectos de mejora de la calidad ambiental. Los resultados de este modelo pueden promover las capacidades de los gobiernos locales para identificar las metas y objetivos concernientes a establecer políticas y diseñar proyectos de mitigación ambiental de la calidad del aire.
- Las redes neuronales artificiales están restringidas a un periodo de tiempo y una determinada ubicación, desde que estas siempre requieren ser entrenadas con datos locales. En principio los modelos de redes neuronales artificiales no son adecuados para predecir distribuciones de concentración espacial en grandes áreas. Sin embargo, el uso de las redes neuronales en forma simultánea en diversas

---

4 Modelo estadístico: permite una diagnosis de la calidad del aire mediante extrapolación de las medidas obtenidas en una determinada zona geográfica. Ecuación entre calidad del aire y variables de emisión.

5 Modelo determinístico: requiere información sobre fuentes y cantidades, composición química y procesos físico-químicos atmosféricos de los elementos contaminantes. Son complejos y de alto costo

zonas geográficas puede ser útil para establecer un mapa de probables eventos de contaminación en un área geográfica extensa.

- Se conoce que la formación de las partículas atmosféricas, tanto primarias (origen antropogénico: procesamiento de minerales, combustión, incendios) como secundarias (reacciones químicas, condensación o coagulación) está influenciada por las concentraciones de otros contaminantes atmosféricos y ciertas condiciones meteorológicas como porcentaje de humedad y radiación solares. Debe destacarse que en este caso no se tuvo disponibilidad de datos meteorológicos, sin embargo queda la propuesta de realizar este procesamiento numérico utilizando como datos de entrada al modelo a los parámetros meteorológicos.

## 5.2 Conclusiones

- El presente trabajo plantea la construcción de un modelo matemático, basado en redes neuronales artificiales, para pronosticar valores de contaminación ambiental por material particulado PM2.5 en el distrito de Carabayllo; esta propuesta demuestra un modelo innovador de pronóstico **basado solamente en datos**, los cuales pueden ser obtenidos en grandes cantidades desde las estaciones automatizadas de calidad de aire. De esta manera se propone un modelo que pueda realizar predicciones en el futuro cercano con un nivel aceptable de desempeño.
- Los resultados obtenidos demuestran que los modelos basados en redes neuronales artificiales tipo perceptrón multicapa pueden ser aplicado al pronóstico de series de tiempo de calidad ambiental de calidad del aire debido a su capacidad

de aprendizaje del tipo supervisado. Debe definirse para este objetivo los parámetros óptimos que permitan obtener resultados adecuados en el pronóstico.

- De las simulaciones realizadas para los experimentos propuestos se observó que los resultados más precisos se obtuvieron mediante los modelos de redes neuronales del tipo perceptrón multicapa con una sola capa oculta conteniendo un número de neuronas múltiplo de 2 y en especial en los números del tipo  $2^N$ .
- Los mejores resultados se obtuvieron con los parámetros del modelo 5 (algoritmo de entrenamiento Gradiente de Escala Conjugado, con 8 neuronas del tipo vector [1x1] en la única capa oculta), seguido de los modelos 1 y 6 respectivamente. Estos modelos son los que poseen los menores errores cuadráticos medios. En general todos los modelos planteados poseen un nivel de error bajo, lo cual lleva a concluir que los modelos basados en redes neuronales tipo perceptrón multicapa consiguen un buen nivel de aproximación para el pronóstico de valores de contaminación ambiental por materia particulado.
- El modelo 5 con MSE más bajo (MSE=0.001009) presenta además un error porcentual medio bastante bajo (MPE=-0.1089%) lo cual señala que la técnica no presenta mucho sesgo, desde que el valor porcentual MPE es cercano a cero.
- En el presente estudio se han considerado 6 modelos de RNA y se han evaluado los errores medio cuadráticos para el rango de neuronas en capa oculta considerados, en general los modelos que tuvieron las neuronas tipo vector [1x1] tuvieron un mejor rendimiento que los del tipo [1x2], todo esto en una única capa

oculta. Se destaca de esto que los pronósticos para predicción de calidad del aire poder ser un referente para tomar decisiones adecuadas y mejorar la calidad ambiental en la zona de influencia.

- El modelo RNA ha sido aplicado al distrito de Carabayllo, mediante los datos de la estación de calidad del aire del mencionado distrito. En principio, su aplicación y uso debe limitarse a esta zona geográfica. Sin embargo, la metodología propuesta puede utilizarse en otros distritos que tengan estación de calidad del aire, previo entrenamiento de la red para su uso adecuado.
  
- Se deben obtener los parámetros óptimos para los modelos predictivos RNA de contaminación ambiental por material particulado, debido a que para cada área geográfica evaluada las condiciones topográficas y meteorológicas son variables. En el caso de una ciudad como Lima Metropolitana, por su extensión, se tiene que el área de estudio circunscrita al distrito de Carabayllo configura un microclima tal que se puede considerar un área geográfica aislada para realizar en presente estudio.
  
- Se concluye que para conseguir una mejor precisión en las estimaciones de calidad de aire, deben considerarse otros factores, tales como datos meteorológicos, estacionales y la posibilidad de tener datos para un mayor rango de años.

### **5.3 Recomendaciones**

- Se han utilizado los datos de dos años completos, lo cual es un conjunto de datos lo suficientemente grande para obtener resultados de cierta confiabilidad, se estima que existe oportunidad de obtener buenos resultados utilizando un conjunto

de datos de menor dimensión, de al menos un año, para obtener resultados razonables.

- Se recomienda utilizar la metodología para generar modelos propios en otras ciudades del país donde existan estaciones de calidad del aire. Así se podrá hacer un análisis local de las condiciones químicas que determinan los niveles de concentración de contaminantes atmosféricos en las diferentes zonas del país.
- Debe tenerse en cuenta que existen diferentes técnicas de inteligencia artificial que permiten el tratamiento de datos cuantitativos, por lo que resulta atractivo rediseñar los modelos planteados bajo diferentes técnicas de aprendizaje supervisado, tales como funciones de base radial y algunas variaciones del perceptrón multicapa. Además también se recomienda el uso de otras técnicas de entrenamiento que van más allá de la retropropagación.
- Si bien en el presente trabajo se ha realizado un pronóstico bastante aceptable de las variables de contaminación del aire por material particulado y debido a la simplicidad de su aplicación, se recomienda evaluar su aplicación en diversos sectores de las investigaciones ambientales y verificar si es factible su aplicación.
- Los modelos predictivos de contaminantes de calidad de aire pueden ser optimizados mediante utilización de información meteorológica y topográfica en la zona a estudiar. Si se superan las limitaciones de información sobre estas variables se recomienda evaluar algunos modelos de redes neuronales artificiales que incorporen estas variables como datos de entrada.

## 5.4 Referencias Bibliográficas

- Aktan, M., & Bayraktar, H. (2009). The Neural Network Modeling of Suspended Particulate Matter with Autoregressive Structure. *Atatürk University. Faculty of Engineering*, 32-37.
- Alkasassbeh, M., Sheta, A. F., Faris, H., & Turabieh, H. (2013). Prediction of PM10 and TSP air pollution parameters using artificial neural network autoregressive, external input models: A case study in Salt, Jordan. *Middle-East Journal of Scientific Research*, 999-1009.
- Anderson, J. A. (2007). *Redes Neurales*. México: Alfaomega Grupo Editor.
- Arellano, M. (16 de Diciembre de 2016). *Introducción al Análisis Clásico de Series de Tiempo*. Obtenido de [www.5campus.com/leccion/seriest](http://www.5campus.com/leccion/seriest): [www.5campus.com](http://www.5campus.com)
- Barbes, L., Neagu, C., Melnic, L., Ilie, C., & Velicu, M. (2009). The use of Artificial Neural Network (ANN) for prediction of some airborne pollutants concentration in urban areas. *Rev. Chim (Bucuresti)*.
- Baughman, D., & Liu, Y. (1995). *Redes Neuronales en Bioprociamiento e Ingeniería Química*. Virginia.: Academic Press.
- Benvenuto, F., & Marani, A. (2000). Redes neuronales para problemas ambientales: Control de calidad de datos y pronóstico de contaminación del aire. *Global Nest: The Int. Journal Vol. 2*, 281-292.
- Braga, A. L., Saldiva, P. H., Pereira, L. A., J.C., M. J., Conceicao, G. M., Lin, C. A., Dockery, D. W. (2001). Health effects of air pollution exposure on children and adolescents in São Paulo, Brazil. *Pediatric Pulmonology*, 106-113.
- Braga, A. L., Saldiva, P. H., Pereira, L. A., J.C., M. J., Conceicao, G. M., Lina, C. A., . . . Dockery, D. W. (2001). Health effects of air pollution exposure on children and adolescents in Sao Paulo Brazil. *Pediatric Pulmonology*, 106-113.
- Braga, A. L., Saldiva, P. H., Pereira, L. A., Menezes, J. J., Conceicao, G. M., Lin, C. A., . . . Dockery, D. W. (2001). Health effects of air pollution exposure on children and adolescents in Sao Paulo Brazil. *Pediatric Pulmonology*, 106-113.
- Camilo-Santana, J. (2006). Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana. *Revista Colombiana de Estadística*, 77-92.
- Cueva, E. V., Oliva-Navarro, D. A., Diaz-Cortés, M. A., & Osuna-Enciso, J. V. (2010). *Optimización – Algoritmos programados con MatLab*". Madrid: Editorial Alfaomega.
- De-Nevers, N. (1998). *Ingeniería de Control de la Contaminación del Aire*. México: McGraw Hill Interamericana de Editores S.A.
- explorable.com. (03 de 08 de 2017). <https://explorable.com/es/disenos-de-investigacion>. Obtenido de <https://explorable.com/es/disenos-de-investigacion>

- Gardner, M., & Dorling, S. (1998). Artificial neural networks (the multilayer perceptron) a review of applications in the atmospheric sciences. *Atmospheric Environment*, 2627-2636.
- Hagan, M. T., & Demuth, H. B. (2016). *Neural Network Design*. Texas: eBook.
- Hernandez-Sampieri, R., Fernandez-Collado, C., & Baptista-Lucio, P. (2003). *Metodología de la Investigación*. Buenos Aires: Mc Graw Hill.
- Hooybergs, J., Mensink, C., Dumont, G., Fierens, F., & Brasseur, O. (2005). A neural network forecast for daily average PM10 concentrations in Belgium. *Atmospheric Environment*, 3279-3289.
- <http://metodologia02.blogspot.pe>. (16 de 02 de 2017).  
<http://metodologia02.blogspot.pe/p/operacionalizacion-de-variables.html>.  
Obtenido de <http://metodologia02.blogspot.pe>
- <https://www.thermofisher.com>. (30 de 6 de 2017).  
<https://www.thermofisher.com/order/catalog/product/TEOM1405>. Obtenido de  
<https://www.thermofisher.com/order/catalog/product/TEOM1405>
- Isasi-Viñuela, P., & Galván-León, I. M. (2004). *Redes de neuronas artificiales: un enfoque práctico*. Madrid: Ed. Pearson Educación.
- Isokawa, T., Kusakabe, T., Matsui, N., & Peper, F. (2003). Quaternion Neural Network and Its Application. *Knowledge-Based Intelligent Information and Engineering Systems*, 318-324.
- Martín-del-Brió, B., & Sanz-Molina, A. (2006). *Redes Neuronales y Sistemas Borrosos*. Madrid: Editorial RaMa.
- Mendez-Ramirez, I., Namihira-Guerrero, D., Moreno-Altamirano, L., & Sosa-de-Martinez, C. (15 de 10 de 2017). <https://es.scribd.com/document/314668647/El-Protocolo-de-Investigacion-MENDEZ>. Obtenido de Mendez:  
<https://es.scribd.com/document/314668647/El-Protocolo-de-Investigacion-MENDEZ>
- Mendoza, A., & García, I. (2004). Desarrollo de un Modelo de Red Neuronal Artificial para el Pronóstico de los niveles de concentración de ozono en el aire ambiente de Monterrey, N.L. *III Congreso Internacional sobre Métodos Numéricos en Ingeniería y Ciencias Aplicadas*, 1-10.
- Morales-Islas, M. I. (17 de 02 de 2016).  
[http://cvonline.uaeh.edu.mx/Cursos/Maestria/MTE/seminario\\_de\\_inv/-UNIDAD%203-3/Tipos\\_de\\_estudios.pdf](http://cvonline.uaeh.edu.mx/Cursos/Maestria/MTE/seminario_de_inv/-UNIDAD%203-3/Tipos_de_estudios.pdf). Obtenido de  
<http://cvonline.uaeh.edu.mx>
- Ordieres, J., E.P., V., R.S., C., & Salazar, R. (2005). Neural network prediction model for fine particulate matter (PM2.5) on the US-Mexico border in El Paso (Texas) and Ciudad Juarez (Chihuahua). *Environmental Modelling & Software*, 547-559.
- Palazón-González, J., & García-Guzmán, A. (2004). Modelado de series climatológicas mediante una red neuronal artificial. *Ingeniería del Agua*.

- Pérez, P., & Reyes, J. (2006). An integrated neural network model for PM10 forecasting. *Atmospheric Environment*, 2845-2851.
- Pitarque, A., Ruiz, J. C., & Roy, J. F. (2000). Las redes neuronales como herramientas estadísticas no paramétricas de clasificación. *Univesidad de Valencia*.
- Pope, C., & Dockery, D. W. (2006). Health Effects of Fine Particulate Air Pollution: Lines that Connect. *J. Air & Waste Manage. Assoc.*, 709-742.
- Salini-Calderon, G., & Perez-Jara, P. (2006). Estudio de series temporales de contaminación ambiental mediante técnicas de Redes Neuronales Artificiales. *Ingeniare. Revista chilena de ingeniería*, 284-290.
- Sanchez-Ccoyllo, O. (2008). Vehicular particulate matter emissions in road tunnels in Sao Paulo, Brazil. *Envvironmental Monit. Assess.*
- Sánchez-Ccoyllo, O. R., Ordoñez, C. G., Gavidia, M. E., Silva, J. S., Angulo, J. L., & Blas, D. (2011). *Evaluación de la calidad del aire en Lima Metropolitana*. Lima: SENAMHI.
- Serrano, A. J., Soria, E., & Martin, J. D. (2010). *Redes Neuronales Artificiales*. Cataluña: Escuela Superior Técnica de Engyneria.
- Strauss, W., & Mainwaring, S. (2011). *Contaminación del aire: causas, efectos y soluciones*. México: Editorial Trillas.
- Tecer, L. (2007). Predicción de SO2 and PM Concentrations in a Coastal Mining Area (Zonguldak, Turkey) Using an Artificial Neural Network. *Polish Journal of Environmental Studies*, 633-638.
- Thomas, S., & Jacko, R. B. (2007). Model for forecasting expressway fine particulate matter and carbon monoxide concentration: application of regression and neural network models. *Journal of the Air & Waste Management Association*, 480-488.
- Varios. (06 de 10 de 2018). [https://en.wikipedia.org/wiki/Mean\\_percentage\\_error](https://en.wikipedia.org/wiki/Mean_percentage_error). Obtenido de [https://en.wikipedia.org/wiki/Mean\\_percentage\\_error](https://en.wikipedia.org/wiki/Mean_percentage_error): <http://composi.info/universidad-iberoamericana.html?page=16>
- Wark, K., & Warner, C. F. (2007). *Contaminacion del Aire: Origen y Control*. México: Editorial Limusa S.A.

**ANEXO N° 1****FICHA TECNICA DE LOS INSTRUMENTOS**

En el presente trabajo de investigación de utilizaron los siguientes instrumentos de cálculo y procesamiento de datos:

<b>Hardware</b>	
Tipo de CPU	Laptop
Procesador	Intel ® Core™ i3
Velocidad de Procesador	1.70 GHz
Memoria RAM	8 GB

<b>Software</b>	
Sistema Operativo	Windows 8.1 Pro a 64 bits
Procesador de Textos	MS Word 13
Hoja de Cálculo	MS Excel 13
Simulación y Pronóstico	Matlab R2015a

## ANEXO N° 2

### DEFINICION DE TÉRMINOS

**Inteligencia Artificial:** rama de la ciencia de la computación que trata de métodos simbólicos y no-algorítmicos de solución de problemas.

**Red Neuronal Artificial:** sistema de computación compuesto de un número de nodos simples y altamente interconectados o elementos de procesamiento, los cuales procesan información mediante su respuesta de estado dinámico a entradas externas.

**Perceptrón:** Tipo de red neuronal artificial desarrollado por Frank Rosenblatt, también puede entenderse como perceptrón la neurona artificial y unidad básica de inferencia en forma de discriminador lineal, que constituye este modelo de red neuronal artificial. Esto debido a que el perceptrón puede usarse como neurona dentro de un perceptrón más grande u otro tipo de red neuronal artificial.

**Perceptrón Multicapa:** El perceptrón multicapa con conexiones hacia adelante es una generalización del perceptrón simple con la inclusión de una o más capas ocultas. Se consolidó como resultado de la limitación de dicha topología con referencia al problema de la separabilidad no lineal.

**Retropropagación:** Es un algoritmo de aprendizaje supervisado que se usa para entrenar redes neuronales artificiales. El algoritmo consiste en minimizar un error (comúnmente cuadrático) por medio de descenso de gradiente, la parte esencial del algoritmo es cálculo de las derivadas parciales de dicho error con respecto a los parámetros de la red neuronal.

**Material Particulado:** ( $PM_{10}$ ,  $PM_{2.5}$ ,  $PM_{0.1}$ ) es un contaminante aéreo, mezcla de partículas sólidas y líquidas suspendidas en el aire que varían en número, tamaño, forma, área superficial, composición química, solubilidad y origen.

**Monitoreo Automático:** Efectuado por estaciones de calidad de aire dotadas con equipos automáticos que monitorean de una manera constante los contaminantes: material particulado ( $PM_{2.5}$ ,  $PM_{10}$ ), dióxido de azufre ( $SO_2$ ), dióxido de nitrógeno ( $NO_2$ ) y ozono superficial ( $O_3$ ) entre otros.

**Series de Tiempo:** Conjunto de mediciones de cierto fenómeno o experimento registradas secuencialmente en el tiempo. Estas observaciones serán denotadas por  $\{x(t_1), x(t_2), \dots, x(t_n)\} = \{x(t) : t \in T \subseteq \mathbb{R}\}$  con  $x(t_i)$  el valor de la variable  $x$  en el instante  $t_i$ .

**Predicción de Series de Tiempo:** La predicción sobre series temporales consiste en, a partir de una muestra de la serie de tiempo, conocer los valores en el futuro, es decir, su comportamiento a lo largo del tiempo. La idea es estimar  $X(t)$  en un instante  $n + k$  posterior al último dato observado en  $t=n$ ,  $k=1, 2, 3 \dots$

**Arquitectura del Perceptrón Multicapa:** Diseño que implica la determinación de la función de activación a emplear, el número de neuronas y el número de capas de la red. La elección de la función de activación se hace en base al recorrido deseado.

**ANEXO N° 3**  
**CÓDIGO FUENTE Y GRÁFICOS PARA LA SIMULACIÓN**  
**(MATLAB 2015)**

```
%% 1. IDENTIFICACION
% *****
% Archivo: proyecto_tesis.m
% Autor: Raul Trinidad Jacinto Herrera
% Universidad Nacional Federico Villarreal
% Escuela Universitaria de Posgrado
% Maestria en Ingenieria de Sistemas
% Proyecto de Tesis
% Fecha: 30/09/2018

%% 2. VACIADO DE MEMORIA
% *****
clear all
close all

%% 2. CARGA DE DATOS INICIALES
% *****
% Se carga archivo original de columnas de datos
% extraidos para cada contaminante
load('PM10');
load('SO2');
load('NO2');
load('CO');
load('PM25');

% Creacion de matriz de datos (5 filas x 2888 columnas)
% Equivalente a 4 meses de datos horarios

H=[PM10 PM25 SO2 NO2 CO];

%% ELIMINACION DE LAS FILAS DE IDENTIFICACION
% *****
% Eliminacion de 8 filas de identificacion
% Esto escribe NaN en las 8 primeras filas
% que contienen datos de cabecera.
x=1;
y=1;
c=size(H);
numrows=c(1,1);
numcols=c(1,2);

for x=1:8
    for y=1:numcols
        H(x,y)=NaN;
    end
end

% Esta linea elimina los elementos con NaN
% H(~any(~isnan(H), 2), :)=[];

%% ELIMINACION DE FILAS QUE TIENEN AL MENOS 1 NaN
% *****
x=1;
y=1;
c=size(H);
numrows=c(1,1);
numcols=c(1,2);
```

```

for x=1:numrows
    for y=1:numcols
        % disp(a(x,y))
        if isnan(H(x,y))
            for y=1:numcols
                H(x,y)=NaN;
            end
        end
    end
end
end

%% CREACION DE LA MATRIZ INICIAL DE TRABAJO
% Eliminacion de Filas que tengan NaN en cada columna
% *****
H(~any(~isnan(H), 2), :)=[];

% RECONSTRUCCION DE DATOS PARA EL MODELO RNA
% Se obtiene la matriz 820x5 double
% *****
PM10_1=H(:,1);
PM25_1=H(:,2);
NO2_1=H(:,3);
CO_1=H(:,4);
SO2_1=H(:,5);
H_1=[PM10_1 PM25_1 NO2_1 CO_1 SO2_1];

%% TRANSFORMACION LINEAL PARA PM10
% *****
% Seleccion de la columna 1 (PM10)
T1=H_1(:,1);
% Transformacion Lineal - Calculamos valores maximo y minimo
% de los datos de entrada
max1 = max(T1);min1 = min(T1);
% Transformacion Lineal - Definimos valores maximo y minimo
% de los datos de salida
min2 = -1;max2 = 1;
% Transformacion Lineal - Creamos PM10_B[] array ordinario
% con datos transformados
% *****
for i= 1:length(T1)
    PM10_B(i)=((T1(i)-min1)/(max1-min1))*(max2-min2)+min2;
end
% Transposicion de PM10_B
PM10_B=PM10_B';
% Fin de transformacion lineal para PM10
%% TRANSFORMACION LINEAL PARA PM25
% *****
% Seleccion de la columna 2 (PM25)
T2=H_1(:,2);
% Transformacion Lineal - Calculamos valores maximo y minimo

% de los datos de entrada
max1 = max(T2);min1 = min(T2);
% Transformacion Lineal - Definimos valores maximo y minimo
% de los datos de salida

```

```
min2 = -1;max2 = 1;
% Transformacion Lineal - Creamos PM25_B[] array ordinario
% con datos transformados
% *****
for i= 1:length(T2)
    PM25_B(i)=((T2(i)-min1)/(max1-min1))*(max2-min2)+min2;
end
% Transposicion de PM25_B
PM25_B=PM25_B';
% Fin de transformacion lineal para PM25

%% TRANSFORMACION LINEAL PARA NO2
% *****
% Seleccion de la columna 3 (NO2)
T3=H_1(:,3);
% Transformacion Lineal - Calculamos valores maximo y minimo
% de los datos de entrada
max1 = max(T3);min1 = min(T3);
% Transformacion Lineal - Definimos valores maximo y minimo
% de los datos de salida
min2 = -1;max2 = 1;
% Transformacion Lineal - Creamos NO2_B[] array ordinario
% con datos transformados
% *****
for i= 1:length(T3)
    NO2_B(i)=((T3(i)-min1)/(max1-min1))*(max2-min2)+min2;
end
% Transposicion de NO2_B
NO2_B=NO2_B';
% Fin de transformacion lineal para NO2

%% TRANSFORMACION LINEAL PARA CO
% *****
% Seleccion de la columna 4 (CO)
T4=H_1(:,4);
% Transformacion Lineal - Calculamos valores maximo y minimo
% de los datos de entrada
max1 = max(T4);min1 = min(T4);
% Transformacion Lineal - Definimos valores maximo y minimo
% de los datos de salida
min2 = -1;max2 = 1;
% Transformacion Lineal - Creamos CO_B[] array ordinario
% con datos transformados
% *****
for i= 1:length(T4)
    CO_B(i)=((T4(i)-min1)/(max1-min1))*(max2-min2)+min2;
end
% Transposicion de CO_B
CO_B=CO_B';
% Fin de transformacion lineal para CO

%% TRANSFORMACION LINEAL PARA SO2
% *****
% Seleccion de la columna 5 (SO2)
T5=H_1(:,5);
% Transformacion Lineal - Calculamos valores maximo y minimo
```

```
% de los datos de entrada
max1 = max(T5);min1 = min(T5);
% Transformacion Lineal - Definimos valores maximo y minimo
% de los datos de salida
min2 = -1;max2 = 1;
% Transformacion Lineal - Creamos SO2_B[] array ordinario
% con datos transformados
% *****
for i= 1:length(T5)
    SO2_B(i)=((T5(i)-min1)/(max1-min1))*(max2-min2)+min2;
end
% Transposicion de SO2_B
SO2_B=SO2_B';
% Fin de transformacion lineal para CO

%% CREACION DE MATRIZ H_2 820x5 double
% *****
H_2=[PM10_B PM25_B NO2_B CO_B SO2_B];
size(H_2);
%% CONVERSION A VECTORES
% *****
H_3=con2seq(H_2');
% Como resultado se tiene el vector H_3
% 1x820 cell
% 1 fila x 820 columnas con celdas
% cada celda contiene 5 datos
%% FORMACION DEL ARCHIVO DE ENTRADA

PM25Inputs=H_3;
% PM25Inputs es un 1x820 cell (cada celda de 5 datos)
PM25Targets=con2seq(PM25_B');

%% Hasta aqui creacion y normalizacion de:
% *****
% PM25Inputs -> 1x820 cell (celda con 5 datos de entrada)
% PM25Targets -> 1x820 cell (celda con 1 dato de salida)
% *****

%% 8. DEFINICION DE DATOS PROCESADOS
% *****
X = PM25Inputs;
T = PM25Targets;

minscore=-1.2;
maxscore= 1.2;

rng('default');
% para que no varien los numeros aleatorios

%% 9. CREACION DE RED NARX EN MODO OPEN-LOOP
% *****
% Las redes dinamicas con retroalimentacion, como las NARXNET
% puede alternar entre los modos open-loop y closed-loop
% con las funciones openloop y closeloop.
% Las redes closed-loop pueden hacer multiples predicciones
% Estas redes continuan prediciendo cuando ya no existe
```

```
% retroalimentacion externa, usando la retroalimentacion interna
% Aquí una RNA es entrenada para modelar el sistema
% y simulada en el modo por defecto (open-loop).
% Funcion de transferencia
net.layers{1}.transferFcn = 'tansig';

% Ingreso de Parámetros de RNA NARX
% *****
inputDelays = input('inputDelays (default 2): ')
feedbackDelays = input('feedbackDelays (default 2): ')
hiddenLayerSize = input('hiddenLayerSize (default 16): ')

net = narxnet(1:inputDelays,1:feedbackDelays,hiddenLayerSize,'open');

inputDel = inputDelays;
feedDel = feedbackDelays;

%% 10. PREPARACION DE DATOS
% *****
[x,xi,ai,t] = preparets(net,X,{},T);

%% 11. ENTRENAMIENTO DE LA RED
% *****
% Seleccion de la Funcion de Entrenamiento (Training Function)
% Comando ntrain en Matlab

%net.trainFcn = 'trainlm';% Levenberg-Marquardt
%net.trainFcn = 'trainbr';% Bayesian Regulation
net.trainFcn = 'trainscg';% Scaled Conjugate Gradient

net = train(net,x,t,xi,ai);

%% 12. SELECCION DE LA FUNCION DE RENDIMIENTO
% *****
% (Performance Function)

net.performFcn = 'mse';% Mean squared error

% Prueba (test) de la red
% *****
y = net(x,xi,ai);
view(net)

% Ploteo de y simulado en figura 1
% *****
figure (1);

plot(cell2mat(y))
legend('NARX OPEN F1');
title('RED NO LINEAL AUTOREGRESIVA EXTERNA ABIERTA - NARX OPEN F1');
ylabel('PM25');
xlabel('Tiempo');
axis( [ 0 850 minscore maxscore ] )

%% 13. PREDICCION MULTITETAPA LAZO CERRADO
% Desde condiciones iniciales
% *****
```

```

% Una RNA puede tambien ser simulada en modo closed-loop,

netc = closeloop(net);
view(netc)

%% 14. PREPARACION DE DATOS EN LAZO CERRADO
% Los datos de entrenamiento se usan para definir las entradas x,
% los estados iniciales de entrada (xi) y
% los estados de layer delay (ai)

[x,xi,ai,t] = preparets(netc,X,{},T);
yc          = netc(x,xi,ai);

%% PLOTEO DE Y,YC hasta 820 en Figure 3
% *****
figure (3);

plot([cell2mat(y);cell2mat(yc)]')
legend('Test Open F3','Test Closed F3');
title('NARX EN MODOS OPEN-LOOP Y CLOSED-LOOP <F3>');
ylabel('PM25');
xlabel('Tiempo');
axis( [ 0 850 minscore maxscore ] )

perf3 = mse(net,y,yc)

%% Predicción Multietapa Lazo-Cerrado siguiendo Secuencia Conocida
% *****
% También puede ser útil simular una red neuronal entrenada
% hasta el presente con todos los valores conocidos de una
% serie de tiempo en modo de lazo-abierto,
% luego cambiar al modo de lazo-cerrado para continuar con
% la simulación para hacer tantas predicciones en el futuro como se deseen.

%% 15. PREDICCIÓN MULTISTEP AHEAD PARA 72 ETAPAS
% *****

% En este caso se tiene un registro del comportamiento
% del sistema para 72 etapas de tiempo y se desea predecir
% hacia adelante) para 72 etapas más de tiempo.

C:\Users\Darato\Dropbox\maestria\Ingenieria_Sistemas\matlab_box\proyecto_tesis.m
Page 7
15 de junio de 2019 03:28:47 PM

% PRIMERO: definir las primeras 72 etapas de input y targets
% que representan las 72 etapas donde la salida conocida esta definida
% por el targer t. Con las siguientes 72 etapas que son definidas
% use la RNA para predecir las 72 salidas usando cada una de sus predicciones
% como retroalimentación para ayudar a la RNA a que prediga la siguiente
% etapa.

%x1 = x(1:72);
%t1= t(1:72);
%x2= x(73:144)
Numpredicciones=72;

```

```

z=length(PM25Targets);
z=z-inputDel; %z=820-4

fin2=z;
fin1=fin2-2*Numpredicciones+1;
x1 = x(fin1:fin2); %length(x1)=144
t1 = t(fin1:fin2); %length(t1)=144
x2 = x(fin1+Numpredicciones:fin2); %length(x2)=72

%% SIMULACION SOBRE ESTOS DATOS DE LA RNA LAZO-ABIERTO
% LA RNA de lazo-abierto (open-loop) se simula sobre estos datos
% *****

% PREPARACION DE DATOS
% *****
[x,xi,ai,t] = preparets(net,x1,{},t1);

% SIMULACION SOBRE OPEN NARX
% *****
% Para 72 datos
[y1,xf,af] = net(x,xi,ai);

%% PLOTEO DE Y1 (OPEN) SIMULADO EN FIGURE 4
% *****
figure (4);
plot(cell2mat(y1))
legend('NARX OPEN-LOOP');
title('SIMULACION RNA NARX OPEN (x1=x=y1) F4');
ylabel('PM25');
xlabel('Tiempo');
axis( [ 0 144 minscore maxscore ] )
%error('fin de figura 4');

%% 16. CONVERSION DE LA RED A LAZO-CERRADO
% Los estados finales del las capas input y layer state
% retornadas por la red se convierten a closed-loop
% Los estados finales de las capas input (ai) y layer state (af)
% se vuelven los estados iniciales xi y ai de la red closed-loop
% final input state (xf) --> initial input state (xi)
% final layer state (af) --> initial layer state (ai)

[netc,xi,ai] = closeloop(net,xf,af);

% PREPARACION DE DATOS
% *****
% Tipicamente se usa PREPARETS para definir los estados de entrada inicial
% y de layer state
% Desde que ya se han obtenido del final de la simulacion open-loop
% no se necesita PREPARETS para continuar con las proximas 72 predicciones
% de la red closed-loop

% SIMULACION SOBRE CLOSED LOOP NARX
% *****
% Para 72 datos
[y2,xf,af] = netc(x2,xi,ai);

```

```

%% PLOTEO DE Y2 (CLOSED) SIMULADO EN FIGURE 5
% *****
figure (5);
plot(cell2mat(y2))
legend('NARX CLOSED-LOOP');
title('SIMULACION RNA NARX CLOSED-LOOP x2=y2 72 <F5>');
ylabel('PM25');
xlabel('Tiempo');
axis( [ 0 1.05*Numpredicciones minscore maxscore ] )
%error('fin de figura 5');

%% PLOTEO DE Y1 (OPEN) e Y2 (CLOSED) EN FIGURE 6
% *****
figure(6);
plot([
    nan(1,inputDel),cell2mat(y1(Numpredicciones+1:end))
    ;cell2mat(y2(1:Numpredicciones))]')
%plot([cell2mat(y1(100:end)),nan(1,inputDel);cell2mat(y2(1:100))]')

legend('NARX Open-Loop <F6>','NARX Closed-Loop <F6>');
title('SIMULACION NARX OPEN-LOOP (72 DATOS) <F6>');
ylabel('PM25');
xlabel('Tiempo');
axis( [ 0 1.05*Numpredicciones minscore maxscore ] )
%error('Y1 Y2 -- fin de figura 6');
perf6 = mse(net,y,yc)
e=cell2mat(y)-cell2mat(yc);
perf7=mae(e);

% Impresion de Datos en Figura 6
% *****
text (10, 0.80, 'Input Delay: ');
text (30, 0.80, num2str(inputDelays))
text (10, 0.60, 'Feed Back Delay: ');
text (30, 0.60, num2str(feedbackDelays))
text (10, 0.40, 'Hidden Layer Size: ');
text (30, 0.40, num2str(hiddenLayerSize))

% Impresion de indicadores de rendimiento (mse, mae)
% *****
text (10, 0.20, 'Valor MSE: ');

text (30, 0.20, num2str(perf6,4))
text (10, 0.00, 'Valor MAE: ');
text (30, 0.00, num2str(perf7,4))

%% 17. INVERSION DE LA TRANSFORMACION LINEAL Y
% CALCULO DE VALORES PM25 ESTIMADOS
% (y2 es el vector de valores pronosticados)
%*****

PM25_Pronost=cell2mat(y2);
W=PM25_Pronost;
%*****
% Transformacion Lineal - Definimos valores maximo y minimo
% de los datos de salida
max01=max(T2);
min01=min(T2);

```

```
% *****  
%Transformacion Lineal - Definimos valores maximo y minimo  
% de los datos de salida  
max02=1  
min02=-1  
% *****  
% Inversion de la Transformacion Lineal  
  
PM25_Transf=((W-min02)/(max02-min02))*(max01-min01)+min01;  
W=PM25_Transf;  
  
% *****  
% creamos vector constante de valor 25 y 50  
V25=0;  
p=1;  
for p=1:72  
    V25(p) = 25;  
end  
V50=0;  
p=1;  
for p=1:72  
    V50(p) = 50;  
end  
  
%% CALCULA SI LOS PRONOSTICOS SUPERAN ECA  
% *****  
  
W5=W';  
  
p1=0;  
v1=0  
for i=1:24  
    if W5(i)>=25  
        p1=p1+1;  
    end  
    if W5(i)>=50  
        v1=v1+1;  
    end  
end  
  
end  
p1  
  
p2=0;  
v2=0;  
for i=25:48  
    if W5(i)>=25  
        p2=p2+1;  
    end  
    if W5(i)>=50  
        v2=v2+1;  
    end  
end  
end  
p2  
  
p3=0;  
v3=0;  
for i=49:72
```

```

    if W5(i)>=25
        p3=p3+1;
    end
    if W5(i)>=50
        v3=v3+1;
    end
end
p3

W2=[p1 p2 p3;v1 v2 v3];

% Fin calculo valores que superan ECA
% *****

%% IMPRESION DE SIMULACION 72H VS ECA EN FIG. 8
% *****
figure(7);
plot([W;V25;V50]')
legend('Simulacion 72 Horas','ECA Anual','ECA 24 Horas');
title('SIMULACION 72 HORAS VS ECA PM2.5 PARA AIRE');
ylabel('PM25');
xlabel('HORAS');
axis( [ 0 72 0 100] )

% Impresion de Datos en Figura
% *****
text (40, 53, 'ECA: PM25 = 50ppm (24 horas)');
%text (30, 80, num2str(inputDelays))
text (40, 28, 'ECA: PM25 = 25ppm (Anual)');
%text (30, 60, num2str(feedbackDelays))
text (14, 70, 'Simulacion 72 horas');
%text (30, 40, num2str(hiddenLayerSize))

%*****
% Defragmentacion en 3 dias

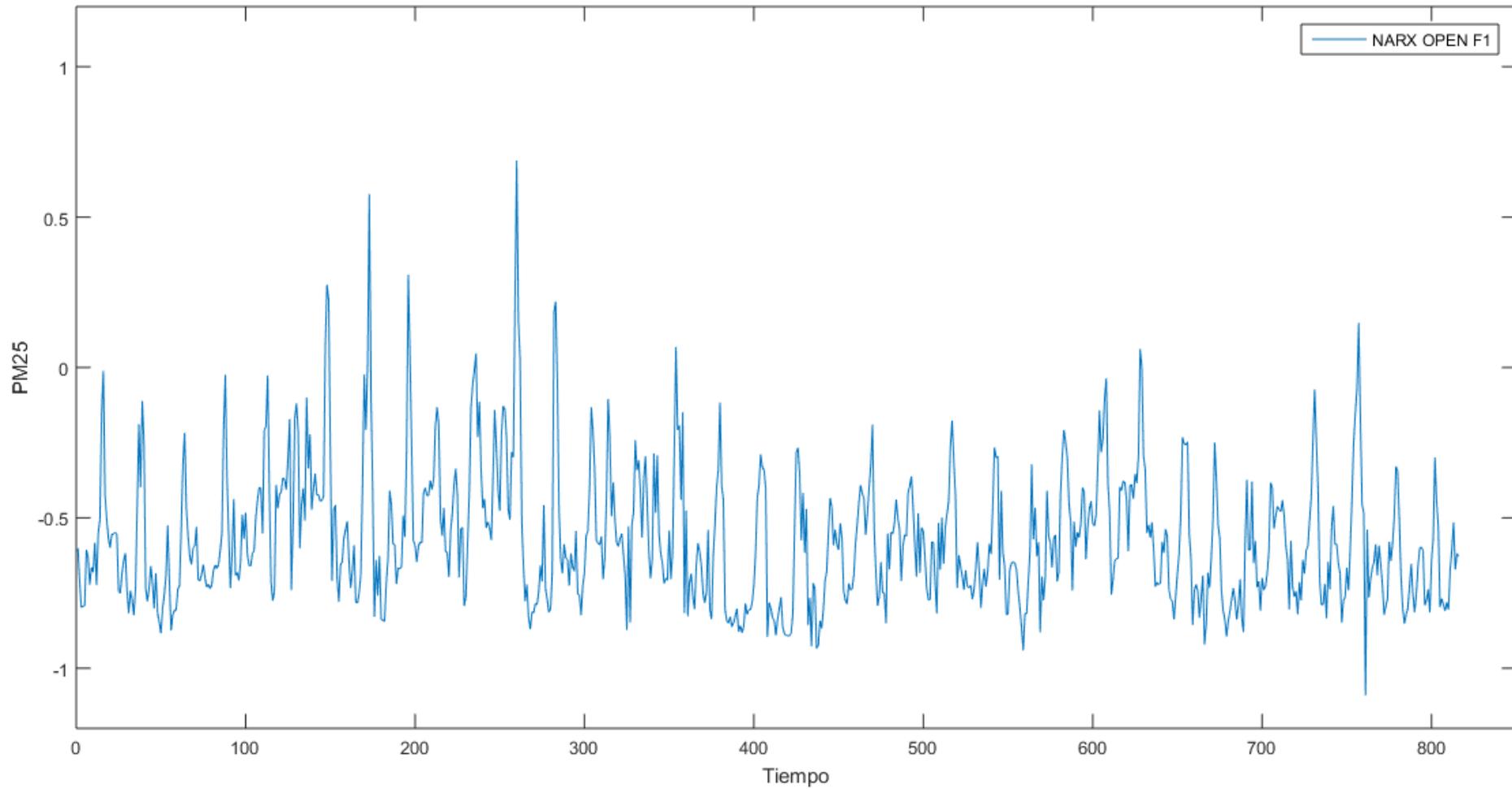
Z2=W(25:48);
Z3=W(49:72);
Z1m=mean(Z1);
Z2m=mean(Z2);
Z3m=mean(Z3);
Z4m=mean(W);
% Impresion de indicadores de rendimiento (mse, mae, sae, sse)
% *****
text (35, 90, 'Promedio Dia 1: ');
text (50, 90, num2str(Z1m,4))
text (35, 85, 'Promedio Dia 2: ');
text (50, 85, num2str(Z2m,4))
text (35, 80, 'Promedio Dia 3: ');
text (50, 80, num2str(Z3m,4))
text (35, 75, 'Prom. Simulacion 72H: ');
text (50, 75, num2str(Z4m,4));

%% FIN DEL PROGRAMA
% *****

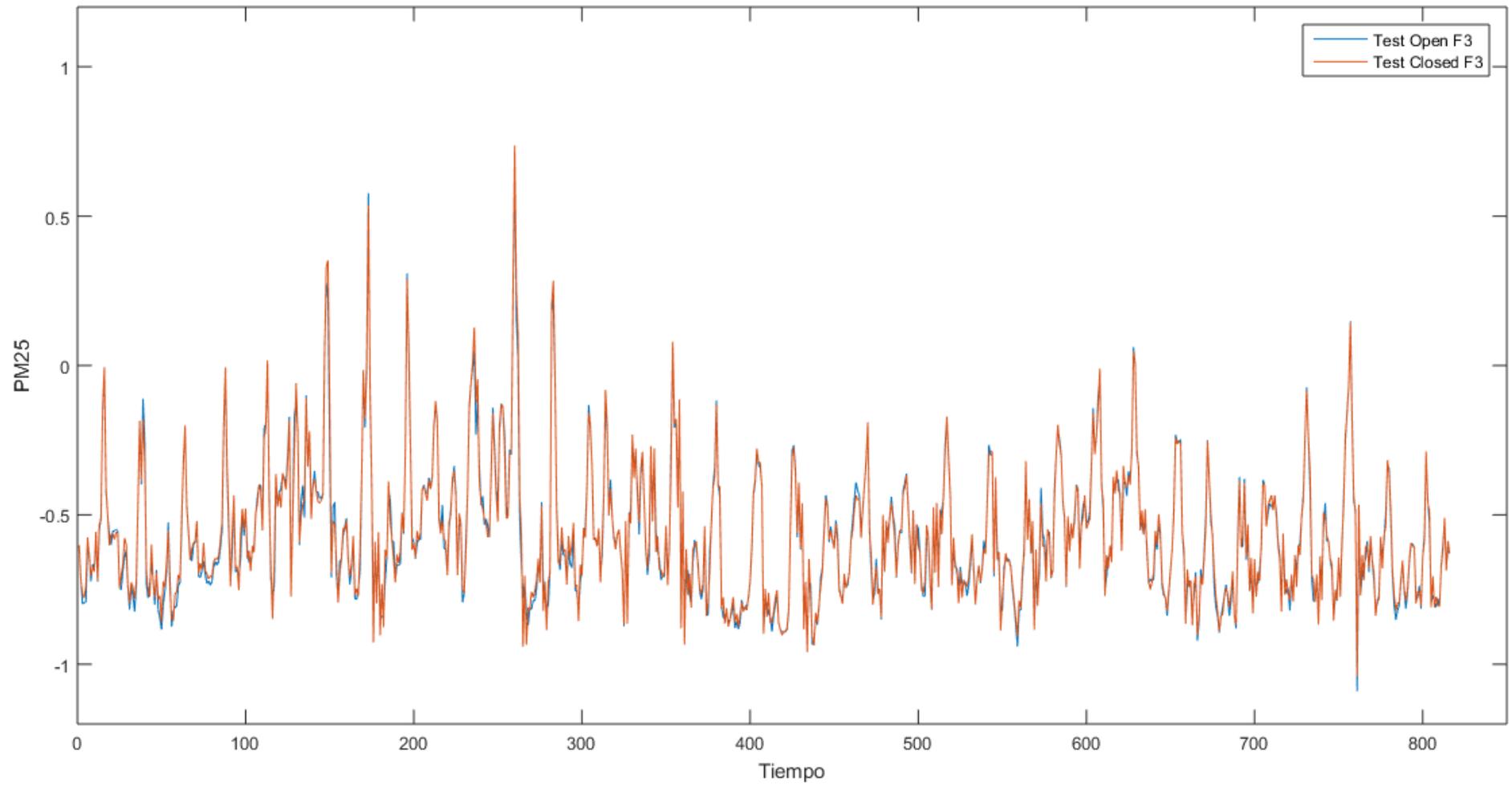
```



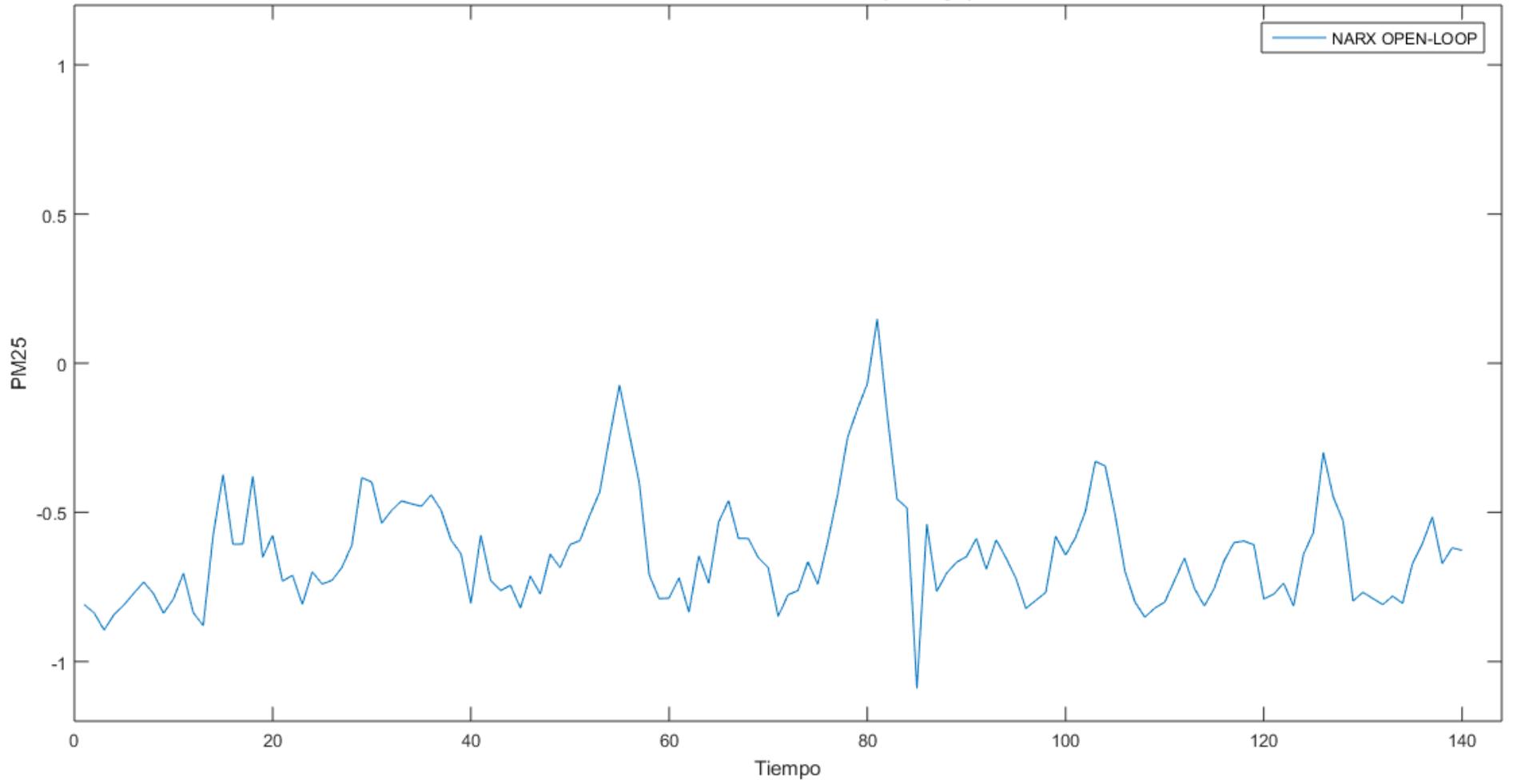
RED NO LINEAL AUTOREGRESIVA EXTERNA ABIERTA - NARX OPEN F1



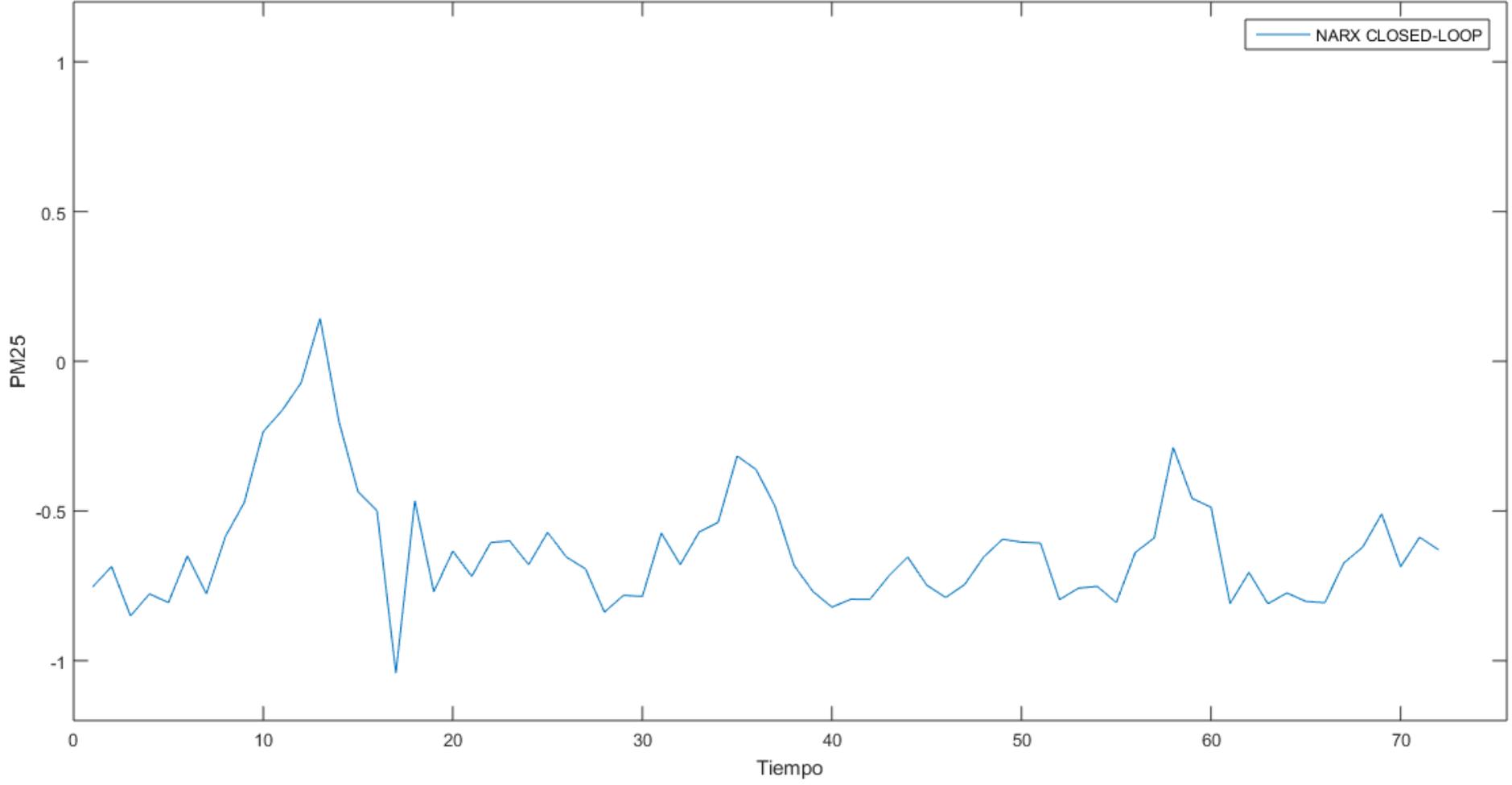
NARX EN MODOS OPEN-LOOP Y CLOSED-LOOP <F3>



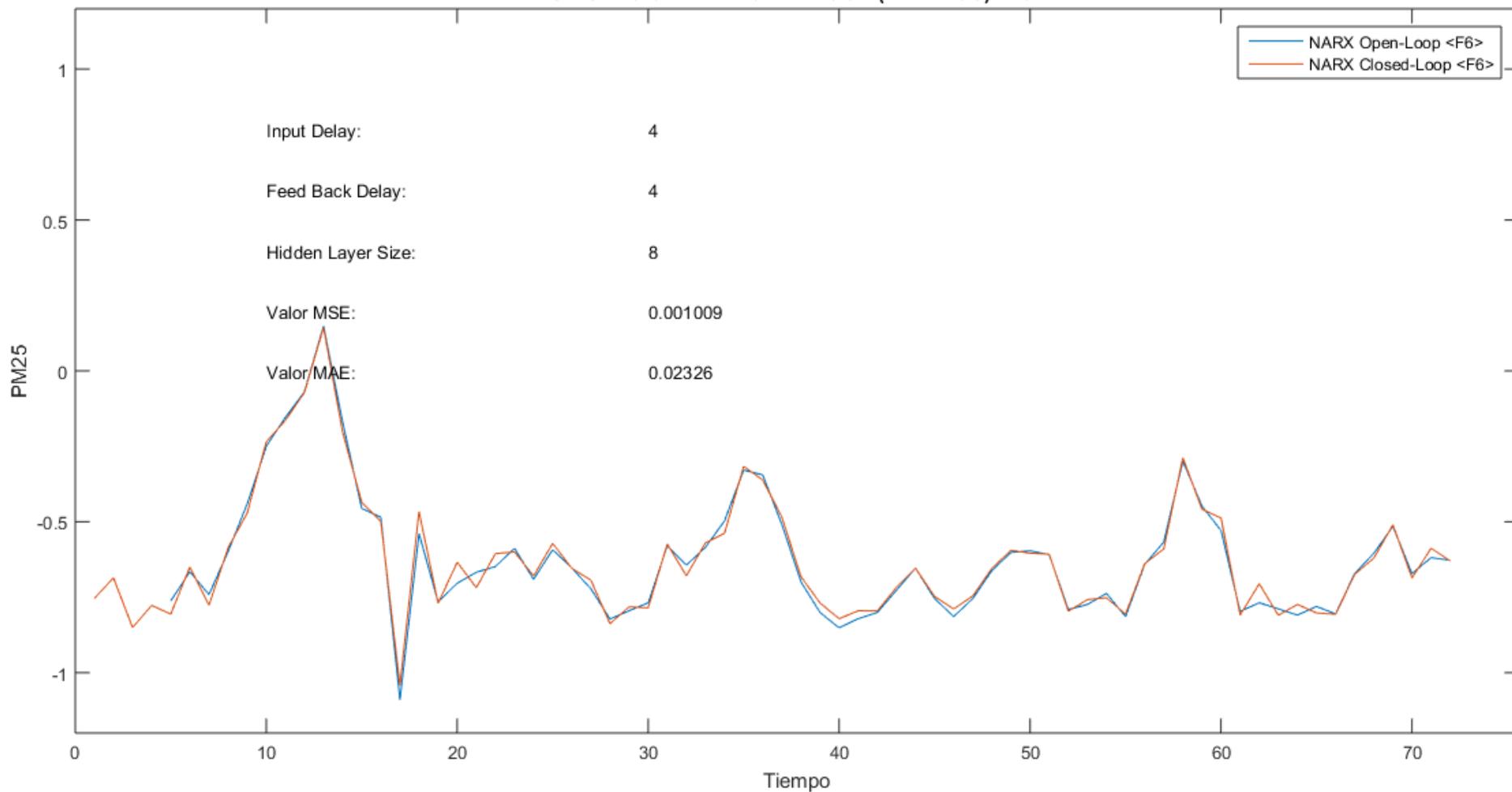
SIMULACION RNA NARX OPEN (x1=x=y1) F4



SIMULACION RNA NARX CLOSED-LOOP x2=y2 72 <F5>



### SIMULACION NARX OPEN-LOOP (72 DATOS) <F6>



### SIMULACION 72 HORAS VS ECA PM2.5 PARA AIRE

