

**REDES NEURONALES ARTIFICIALES Y FUNCIONES DE TRANSFERENCIA
APLICADOS A LA ESTIMACIÓN DE HUMEDAD DE SUELO Y TEMPERATURA
DEL AIRE**

Por

CHRISTIAN HUGO CALDERON ARTEAGA

Tesis sometida como requisito parcial para obtener el título de:

MAESTRO EN CIENCIAS
En
INGENIERÍA MECÁNICA

Universidad de Puerto Rico
Recinto Universitario de Mayagüez
2007

Aprobado por:

David Dooner, Ph. D.
Miembro, Comité Graduado

Fecha

Gustavo Gutierrez, Ph. D.
Miembro, Comité Graduado

Fecha

Nazario Ramírez, Ph. D.
Presidente, Comité Graduado

Fecha

Christa von Hillebrandt Andrade, M. Sc.
Representante de Estudios Graduados

Fecha

Paul Sundaram, Ph. D.
Director del Departamento de Ingeniería Mecánica

Fecha

ABSTRACT

Statistical, mathematical and artificial neural networks techniques were used to estimate the behavior of the following variables: the air temperature (AT) and the soil moisture (SM). The AT plays an important role in the heat interchange between the land surface and the atmosphere, whereas the soil moisture helps to maintain the heat balance between the land surface and the atmosphere. These two parameters are especially important in the maintenance of the biosphere. Models for hourly estimation of the AT and hourly estimation of the SM were developed for 15 field stations in Puerto Rico. These models were generalized to regions that exhibit similar atmospheric characteristics to each station by using remote sensing information and competitive neural networks, obtaining hourly maps of AT and SM over Puerto Rico. The experimental and validation process show that the proposed methodologies are a powerful tool for the estimation of the mentioned variables and a valuable contribution in the study of the behavior of these variables.

RESUMEN

Técnicas estadísticas, matemáticas y de redes neuronales artificiales fueron utilizadas para estimar el comportamiento de dos variables a saber: la temperatura del aire y la humedad de suelo. La temperatura del aire juega un papel importante en el intercambio de calor entre el suelo y la atmósfera, mientras que a su vez la humedad de suelo ayuda a mantener el equilibrio térmico entre el suelo y la atmósfera. Estos dos parámetros son especialmente importantes en el mantenimiento de la biosfera. Modelos de estimación horaria de la temperatura del aire y de estimación horaria de la humedad de suelo fueron desarrollados para 15 estaciones de humedad de suelo en Puerto Rico. Estos modelos fueron generalizados a regiones con características atmosféricas similares a cada estación haciendo uso de información de sensores remotos y de redes neuronales competitivas, obteniendo como resultados finales mapas de temperatura del aire horaria y de humedad de suelo horaria para Puerto Rico. Los resultados experimentales y la validación de los modelos desarrollados muestran que las metodologías aquí propuestas son una herramienta potencial para la estimación de las variables mencionadas y una contribución valiosa para el estudio del comportamiento de dichas variables.

A Judy.

AGRADECIMIENTOS

En primer lugar quiero agradecer a mis padres Hermes y Alicia pues recibí de ellos su apoyo y confianza incondicional además de los mejores consejos. Y a mi hermano Hermes E., por sus palabras de aliento y su gran apoyo.

Quiero expresar un especial agradecimiento al profesor Nazario D. Ramírez por sus directrices en el desarrollo de este trabajo, por brindarme su amistad y todo su apoyo.

Mis mas sinceros agradecimientos al profesor Eric Harmsen, por su valiosa colaboración y sus comentarios acerca de este trabajo.

Debo agradecer también a mis compañeros de aventura, Oswaldo y Joan, por sus constantes aportes, su amistad sincera y todas aquellas conversaciones constructivas durante estos años.

Es necesario también agradecer a todos los estudiantes subgraduados que hicieron parte del proyecto y que en algún momento colaboraron en la obtención de los datos de campo.

Un especial agradecimiento al Recinto Universitario de Mayagüez de la Universidad de Puerto Rico, por brindarme la oportunidad de seguir un programa graduado, así como al proyecto NASA-EPSCoR grant N^o: NCC5-595 pues sin dicho apoyo no habría sido posible realizar este trabajo.

Por último, pero no menos importante, expreso mi más sincero agradecimiento a todas aquellas personas e instituciones que de una u otra manera participaron en el desarrollo de este trabajo.

TABLA DE CONTENIDO

ABSTRACT.....	II
RESUMEN	III
AGRADECIMIENTOS.....	V
LISTA DE TABLAS.....	VIII
LISTA DE FIGURAS.....	IX
1 INTRODUCCIÓN	10
1.1 JUSTIFICACIÓN	12
1.2 ESTUDIOS PREVIOS	13
1.2.1 Estudios sobre Humedad de Suelo.....	13
1.2.2 Estudios sobre Temperatura del Aire.....	17
1.3 RESUMEN DE LOS SIGUIENTES CAPÍTULOS.....	18
2 MARCO TEÓRICO	19
2.1 REGRESIÓN LINEAL MÚLTIPLE	19
2.1.1 Modelo de Regresión Lineal Múltiple (MRLM).....	19
2.1.2 Problemas con la Regresión	21
2.1.3 Herramientas para el Análisis de Regresión Lineal Múltiple.....	22
2.2 REDES NEURONALES ARTIFICIALES (RNA)	25
2.2.1 Generalidades.....	25
2.2.2 Mecanismo de Aprendizaje.....	27
2.2.3 Dimensionamiento de la Red	31
2.2.4 Algoritmos de Entrenamiento	31
2.3 FUNCIONES DE TRANSFERENCIA (FT)	32
2.3.1 Introducción.....	32
2.3.2 Modelos de función de transferencia con una sola entrada.....	32
2.3.3 Procedimiento General Para la Identificación de la FT.....	37
3 METODOLOGÍA.....	42
3.1 INTRODUCCIÓN	42
3.2 DESCRIPCIÓN GENERAL DE LA METODOLOGÍA	43
3.3 DISEÑO DE LA BASE DE DATOS	44
3.3.1 Descripción de los Datos	44
3.3.2 Base de Datos.....	51
3.4 DESCRIPCIÓN DEL MÉTODO DE RNA.....	53
3.4.1 RNA Supervisada.....	53
3.4.2 RNA No Supervisada.....	59
3.5 DESCRIPCIÓN DEL MRLM.....	61
3.5.1 Identificación del modelo de estimación de temperatura horaria.....	61
3.5.2 Estimación de parámetros.....	64
3.5.3 Evaluación del MRLM.....	66

3.5.4	<i>Verificación y Corrección del Modelo de Temperatura.</i>	66
3.6	DESCRIPCIÓN DEL MÉTODO DE FT	67
3.6.1	<i>Identificación del modelo de función de transferencia.</i>	67
3.6.2	<i>Estimación de parámetros.</i>	70
3.6.3	<i>Evaluación de la FT.</i>	72
3.6.4	<i>Verificación y Corrección de la FT.</i>	73
4	RESULTADOS Y VALIDACIÓN.	74
4.1	REGIONES CLIMATOLÓGICAS DE PUERTO RICO.	74
4.2	ESTIMACIÓN DE LA TEMPERATURA HORARIA.	78
4.3	ESTIMACIÓN DE LA HUMEDAD DE SUELO.	89
5	CONCLUSIONES Y TRABAJO FUTURO	104
6	REFERENCIAS	106
ANEXO A	ALGORITMOS DE ENTRENAMIENTO RNA	110
ANEXO A1	ALGORITMO DE <i>LEVENBERG–MARQUARDT</i> .	110
ANEXO A2	RED COMPETITIVA	114
ANEXO B	PROGRAMAS EN MATLAB	119
ANEXO B1	REGIONES CLIMATOLÓGICAS.	119
ANEXO B2	ESTIMACIÓN DE LA HUMEDAD DE SUELO HORARIA	121
ANEXO B3	ESTIMACIÓN DE LA TEMPERATURA HORARIA.	126

LISTA DE TABLAS

<i>Tabla 3.1 Ejemplo de arreglo matricial para modelaje.</i>	53
<i>Tabla 3.2 Datos provenientes del periodograma.</i>	62
<i>Tabla 4.1 Región a la que corresponde cada estación.</i>	76
<i>Tabla 4.2 medidas de tendencia central para cada región.</i>	77
<i>Tabla 4.3 Validez estadística de la red autoorganizada.</i>	78
<i>Tabla 4.4 Parámetros estimados para el MRLM (Ecuación 3.6).</i>	79
<i>Tabla 4.5 Precisión de las estimaciones de temperatura.</i>	83
<i>Tabla 4.6 Precisión de las estimaciones de temperatura con datos corregidos de satélite.</i>	86
<i>Tabla 4.7 Precisión de las estimaciones de humedad de suelo promedio mensual.</i>	91
<i>Tabla 4.8 Resultados del proceso de estimación de parámetros de la FT.</i>	98
<i>Tabla 4.9 Precisión del modelo de FT.</i>	98
<i>Tabla 4.10 Precisión del modelo de FT en regiones similares.</i>	100

LISTA DE FIGURAS

<i>Figura 2.1 Estructura general de una RNA.</i>	26
<i>Figura 2.2 Caracterización del procesamiento de una neurona.</i>	27
<i>Figura 2.3 Esquema de procesamiento neuronal.</i>	27
<i>Figura 2.4 Input y Ouput a partir de un sistema dinámico.</i>	35
<i>Figura 2.5 Estructura de la función de transferencia con el modelo del ruido.</i>	36
<i>Figura 3.1 Metodología usada en este estudio.</i>	43
<i>Figura 3.2 Localización de las estaciones de medición de humedad de suelo.</i>	45
<i>Figura 3.3 Rango mensual de temperatura usando datos de MODIS para agosto 2006.</i>	47
<i>Figura 3.4 NDVI, usando datos de MODIS para agosto 2006.</i>	48
<i>Figura 3.5 Lluvia de radar, usando datos de NEXRAD para agosto 2006.</i>	49
<i>Figura 3.6 Características texturales para Puerto Rico.</i>	50
<i>Figura 3.7 Características topográficas (alturas) para Puerto Rico.</i>	51
<i>Figura 3.8 Estructura de la base de datos.</i>	52
<i>Figura 3.9 Periodograma para la temperatura en la estación Jamaica (Agosto 2006).</i>	62
<i>Figura 3.10 Variación diaria de la Temperatura (estación Colegio, Agosto 2006)</i>	63
<i>Figura 3.11 Variación temporal de la temperatura, humedad de suelo y precipitación.</i>	69
<i>Figura 4.1 Clasificación de regiones climáticas en Puerto Rico.</i>	75
<i>Figura 4.2 Histograma para las regiones climáticas en Puerto Rico.</i>	75
<i>Figura 4.3 Tendencias de Temperatura.</i>	80
<i>Figura 4.4 Ajuste y validación del modelo de temperatura en la estación Colegio en Agosto del 2006.</i>	81
<i>Figura 4.5 Validación en la estación Yunke para agosto del 2006.</i>	82
<i>Figura 4.6 Comparación datos de MODIS (rojo). Estación Cadena agosto 2006.</i>	84
<i>Figura 4.7 Comportamiento del modelo de temperatura con datos de satélite (rojo).</i>	85
<i>Figura 4.8 Estimación final de la temperatura horaria por regiones similares.</i>	86
<i>Figura 4.9 Temperatura estimada en las primeras doce horas del 01 de Agosto del 2006. ..</i>	87
<i>Figura 4.10 Estimación de la temperatura del aire entre las 12 y las 23 horas del 01 de Agosto del 2006.</i>	88
<i>Figura 4.11 Validación del modelo de promedios de humedad de suelo mensuales</i>	90
<i>Figura 4.12 Estimación de la humedad de suelo promedio mensual sin corregir para agosto del 2006.</i>	92
<i>Figura 4.13 Promedio mensual estimado de humedad de suelo para agosto del 2006.</i>	93
<i>Figura 4.14 Tendencias de humedad de suelo.</i>	94
<i>Figura 4.15 Ajuste y validación de la función de transferencia en la estación Ramón.</i>	95
<i>Figura 4.16 Ajuste y validación de la función de transferencia para la humedad de suelo. .</i>	97
<i>Figura 4.17 Validación de la FT para puntos similares.</i>	101
<i>Figura 4.18 Estimación final de la humedad de suelo para una hora en particular.</i>	102
<i>Figura 4.19 Estimación final de la humedad de suelo de la hora 97 a la 108.</i>	103
<i>Figura 4.20 Estimación final de la humedad de suelo de la hora 109 a la 120.</i>	103
<i>Figura A.1 Capa competitiva.</i>	117

1 INTRODUCCIÓN

La temperatura del aire en la capa baja de la atmósfera controla el intercambio de calor entre el suelo y la atmósfera (Sun y Pinker, 2004a). Para estudiar los procesos que ocurren en la superficie de la tierra, es necesario contar con información de temperatura del aire y de humedad de suelo. Información precisa de ambas variables es necesaria para inicializar correctamente los modelos numéricos de predicción del clima, lo que se traduce en una mejora en las predicciones. A pesar de la gran cantidad de literatura existente, el estudio de la humedad de suelo sigue siendo un campo abierto a la investigación, debido a la gran variabilidad espacial y temporal de dicho parámetro y a la cantidad de factores que lo determinan y que se influyen mutuamente (Maneta y Schnabel, 2003). La mayor parte de los esfuerzos encaminados a la estimación de la humedad de suelo son en el campo de sensores remotos (Jiang y Cotton, 2004); estas investigaciones utilizan datos de microondas de las bandas *L* y *C*. Sin embargo, la resolución de los sensores capaces de medir estas bandas es del orden de varios kilómetros, lo que hace difícil obtener estimaciones confiables en los puntos intermedios o para áreas pequeñas. Por otro lado, los datos provenientes de estos sensores no son buenos cuando existe una cubierta de vegetación significativa. Se sabe que la humedad de suelo está ligada a el relieve (Maneta y Schnabel, 2003), a la vegetación existente en el punto y la temperatura del aire (Sun y Pinker, 2004a), la lluvia (Pan et al, 2003), además de otros parámetros como las características texturales del suelo (Cemex et al 2004). El estudio de la temperatura del aire es también un campo rico en literatura mientras que la gran mayoría de los esfuerzos en el campo de sensores remotos son realizados para

estimar la temperatura en la superficie de la tierra usando información de sensores remotos. Se conocen dos tipos de satélite, los de órbita polar y los geoestacionarios. La mayoría de los estudios sobre la temperatura superficial están enfocados en el uso de satélites de órbita polar (Sun y Pinker, 2004b). La frecuencia temporal de medición de estos satélites es de aproximadamente 2 veces diarias, lo cual puede ser inadecuado para muchas aplicaciones. Por otro lado, los satélites geoestacionarios pueden medir la variación diaria de la temperatura pero a su vez tienen el problema de que el volumen de los datos a manejar sería gigantesco, además, se han realizado pocos trabajos y a la vez complicados conducentes a la estimación de la temperatura superficial con datos de satélites geoestacionarios (Sun y Pinker, 2004b). Geerts (2002) encontró que la tendencia de la temperatura superficial está relacionada con la altura del sitio, mientras que Martin (1995), encontró que es posible obtener estimados horarios de temperatura de aire a partir de los datos diarios de temperatura máxima y mínima.

Este trabajo, está encaminado a obtener estimaciones horarias de la temperatura de aire y de la humedad de suelo. Para cumplir con el primer objetivo, se ha usado información de 15 estaciones de medición en el campo, extrayendo los datos máximos y mínimos, con los que luego se generan modelos de estimación horaria del comportamiento de la temperatura del aire. Para el cumplimiento del segundo objetivo, también se usaron datos de las 15 estaciones de medición, obteniéndose modelos de estimación de promedios mensuales de humedad de suelo y de estimación de la humedad de suelo horaria. Ambos modelos usan información de sensores remotos y otros datos de características del terreno. Para la obtención de dichos modelos se han usado herramientas estadísticas, matemáticas y de redes

neuronales artificiales, las cuales han sido consideradas de acuerdo a las necesidades de cada problema.

1.1 Justificación

Las estimaciones de humedad de suelo mediante modelos estadísticos son escasas, pues la mayoría de los trabajos se centran en los modelos hidrológicos y el uso de microondas. Los modelos hidrológicos están basados en ecuaciones de balance de material, pero aunque son buenos estimadores, requieren de una parametrización que es difícil de obtener. Los modelos de microondas cuentan con la ventaja de que pueden ser usados en grandes extensiones de terreno, pues la información de sensores remotos puede cubrir todo el globo, aunque las desventajas de este método son la interferencia de la vegetación, pues esta enmascara la humedad de suelo, los tamaños de muestreo del satélite (macro escala) y la escasa disponibilidad de información de campo con la que se pueda validar dichos modelos. Los métodos estadísticos en cambio presentan mayor flexibilidad para modelar puntos cercanos e igualmente pueden abordar grandes extensiones de terreno, incluso pueden ser usados para modelar perfiles en profundidad. A partir de datos históricos de humedad de suelo y otras variables de importancia (como la lluvia, la temperatura, etc.) correlacionadas espacial y temporalmente, es posible obtener estimaciones precisas de humedad de suelo. La necesidad de obtener datos de humedad de suelo a escalas menores radica en que la humedad de suelo es uno de los parámetros más importantes en las capas bajas de la atmósfera, razón por la cual los modelos regionales necesitan de esta información para mejorar las predicciones del clima.

Las mediciones de temperatura del aire son mas comunes que la información de humedad de suelo, aunque no necesariamente indica esto que hay mayor disponibilidad de datos, pues aunque muchos de los satélites en órbita pueden realizar dichas mediciones, no todos pueden recrear el comportamiento horario de la temperatura de aire. Aquellos que pueden realizar mediciones horarias, no han sido muy estudiados aún y es por esto que los datos de temperaturas horarias no se encuentran disponibles al publico en general. Al momento no existe bibliografía que trate el tema de la estimación indirecta de la temperatura del aire de manera horaria, aunque existe la necesidad de contar con estimados correctos de temperatura del aire, puesto que es una de las variables climatológicas que mas influye en los fenómenos que ocurren en la biosfera, y al igual que sucede con la humedad de suelo, los modelos regionales necesitan de estimados precisos de la temperatura del aire para parametrizar los procesos que ocurren en la atmósfera baja.

1.2 Estudios Previos

1.2.1 Estudios sobre Humedad de Suelo.

Existen diversos modelos de estimación de humedad de suelo, la mayoría de ellos se basan en ecuaciones generales reconocidas por la comunidad científica y a partir de las cuales se han ramificado los diversos estudios del comportamiento de la humedad de suelo. Algunos de los modelos más conocidos son:

1.2.1.1 Modelo del Ciclo Hidrológico.

Este modelo esta basado en el balance de agua y la ecuación básica de balance hidrológico es (Thorntwaite, 1948; Mather, 1978):

$$\Delta w = P - R - U - E \quad 1.1$$

Donde: P: es la precipitación; R: es el escurrimiento superficial del agua o esorrentía y el subterráneo; U: es el drenaje o precolación; E: es la evapotranspiración (planta + suelo); Δw : es la variación de la cantidad de agua en el suelo. Δw también es usado para medir la humedad de suelo y puede ser aproximado usando solo precipitación y evapotranspiración; la evapotranspiración puede ser estimada usando datos de temperatura del aire y la duración de la luz del día (Thornthwaite, 1948). Así, la humedad de suelo puede ser estimada solo con datos de temperatura y precipitación (Huang et al, 1996). En este modelo se puede apreciar la importancia de la lluvia y la temperatura en la estimación de la humedad de suelo.

1.2.1.2 *Modelo de Constante Dieléctrica.*

La constante dieléctrica global de un material está compuesta por las constantes dieléctricas de cada uno de sus componentes a través de la ecuación de Hastead (Hastead, 1973):

$$\varepsilon_m = \left[\varepsilon_h^{1/3} + p \left(\varepsilon_{inc}^{1/3} - \varepsilon_h^{1/3} \right) \right]^3 \quad 1.2$$

Donde ε_m es la constante o permisividad dieléctrica global, ε_h es la constante del sustrato y ε_{inc} la del material añadido en una proporción p . Como la constante dieléctrica del suelo es baja (entre 2 y 4) y la del agua es muy alta (>78), la constante dieléctrica de un suelo húmedo estará muy condicionada por el contenido de agua de dicho suelo. Conocidas las constantes dieléctricas del material añadido y del sustrato, se mide la constante dieléctrica global ε_m , y con esta se determina la proporción del material añadido (agua) en contenido volumétrico (Reinolds, 1970). Wang y Schmugge (1980), explican que la constante dieléctrica del suelo puede estar relacionada con el tipo de suelo o mejor con su característica textural y presentan

un modelo empírico para varios tipos de suelo con un rango de humedad entre 0 y 50%. El modelo de la constante dieléctrica para conversión a la humedad de suelo es usado por el fabricante de los sensores de humedad de suelo usados en este trabajo.

1.2.1.3 *Modelo de Microondas.*

Este modelo se basa en la estimación de la constante dieléctrica del suelo a partir de medidas de sensores remotos, en este modelo, es necesario conocer las características textuales del suelo, como el porcentaje de arcilla, arena, además de otras como la densidad aparente así que la humedad de suelo se puede estimar por medio de sensores remotos a partir de mediciones de la temperatura de brillantez (Sun y Pinker, 2004a; Wang y Schmugge, 1980; Schmugge y Wang, 1982), con la cual se puede conocer la constante dieléctrica del suelo, a partir de la inversión de la ecuación de Fresnel (Cruzado, 2004; Jakson y Le Vine, 1999). Una de las dificultades de estos modelos es la escasa disponibilidad de los datos de microonda, además de la interferencia de factores como la vegetación (Schmugge y Wang, 1982) y las ondas de radio. Las bandas usadas típicamente para medir la humedad de suelo son la *L* y la *C*, la banda *L* tiene un ancho de barrido de 23.5cm y puede ser encontrada en radares como *Synthetic Aperture Radar* (SAR), este radar no provee datos históricos por el tamaño excesivo de sus archivos; el *Advanced Land Observing Satellite* (ALOS) fue lanzado en enero del 2006, y se espera que sea posible obtener datos de la banda *L*. La banda *C*, que tiene un ancho de barrido de 5.3cm puede ser encontrada en RADARSAT, un satélite canadiense que fue lanzado en 1995, además el satélite *Aqua* cuenta con el sensor *Advanced Microwave Scanner Radiometric* (AMSR) del *Earth Observing Sistem* (EOS) y provee

información de humedad de suelo a partir de datos de la banda *C*. La banda *X* con un ancho de barrido de 3cm es poco usada por la escasa calidad de sus datos.

1.2.1.4 Otros Modelos.

Pan, Et al, (2003), proponen un modelo que utiliza datos de lluvia para estimar humedad de suelo por medio de ecuaciones diferenciales parciales estocásticas. Maneta y Schnabel, (2003), usaron redes neuronales para determinar la distribución espacial de la humedad de suelo, ellos plantearon un problema físico de movimiento de masas en relación con la altura del terreno.

Los modelos de constante dieléctrica y de microondas están estrechamente relacionados, pues la información de la microonda ofrece la temperatura de brillantes, la cual esta relacionada con el comportamiento de la constante dieléctrica del suelo, por lo que el modelo de microondas depende íntimamente del modelo de la constante dieléctrica. En los modelos del ciclo hidrológico es posible usar métodos estadísticos con el fin de predecir el comportamiento de la humedad de suelo. En este estudio se pretende relacionar el efecto de la temperatura del aire y la lluvia (componentes del ciclo hidrológico) además de otras variables como el índice normalizado de vegetación (NDVI), topografía del terreno (la cual está relacionada también con el ciclo hidrológico) y textura del suelo con la humedad de suelo; haciendo uso de modelos estadísticos que relacionen las variables involucradas y la humedad del suelo.

1.2.2 Estudios sobre Temperatura del Aire.

La obtención de modelos de estimación de temperatura superficial es muy importante en el estudio del calentamiento global. Sun y Pinker (2003) desarrollaron dos algoritmos para la estimación de la temperatura superficial usando datos provenientes del *geostationary operational environmental satellite* (GOES – 8). Ellos desarrollaron una metodología que utiliza múltiples variables climatológicas en un algoritmo de *split window* que estima la temperatura durante el día y la noche de manera separada. Geerts (2002), desarrolló modelos empíricos de estimación del rango de temperatura mensual promedio y diario. El encontró que la temperatura está fuertemente correlacionada con la ubicación geográfica y el relieve, además de otras variables climatológicas. El aspecto negativo del trabajo de Geerts se centra en la cantidad de variables utilizada y su escasa disponibilidad para la aplicación de los modelos. Liang (2001), desarrolló un algoritmo capaz de separar la emisividad de la superficie de la tierra y la temperatura superficial usando datos de imágenes multiespectrales termales de infrarrojo. Liang estableció una ecuación empírica basada en el espectro de emisividad medido, sin embargo, el no reportó los resultados de la validación puesto que requiere de un esfuerzo muy grande para la recolección de los datos de campo para validación. La estimación de la temperatura del aire a partir de datos de estaciones de campo, no ha sido abordada aun, quizás por que la comunidad científica se ha interesado más en las mediciones directas de sensores remotos. Una de las características importantes de la estimación de la temperatura del aire a partir de datos de campo es que no se requiere de grandes campañas de muestreo ni de la movilización de gran cantidad de personal, además de que los datos para validación se obtienen simultáneamente. Martín (1995), desarrolló

modelos de estimación de la temperatura del aire partiendo de sus valores máximos y mínimos. El usó información de los máximos y mínimos de temperatura y a partir del conocimiento empírico del comportamiento del ciclo diario de la temperatura desarrolló dos ecuaciones polinómicas con las que estimó la temperatura de día y la temperatura de noche. Aunque la aplicación de su modelo no requiere de datos especializados, el problema de dicho modelo es que su aplicación está visualizada para la construcción de edificios, por lo cual dichas temperaturas podrían ser a alturas considerablemente grandes, lo que complicaría la obtención de parámetros de corrección. Por otro lado, es posible utilizar modelos conjuntos de estimación de temperatura del aire, como el desarrollado en este trabajo, que utiliza datos de campo para desarrollar el modelo y datos de sensores remotos para su aplicación.

1.3 Resumen de los Sigüientes Capítulos

Primero se desarrollan las bases teóricas necesarias en el Capitulo 2. En el Capitulo 3 se presenta la metodología usada, allí se hace una breve descripción de las variables de interés, y la forma como en que se desarrollaron los modelos. En el Capitulo 4 se presentan los resultados y el análisis adecuado de los mismos. Las conclusiones y algunas recomendaciones para futuros trabajos son presentadas en el Capitulo 5.

2 MARCO TEÓRICO

Para describir la metodología propuesta, es necesario ofrecer un fundamento teórico que relate la teoría usada en cuanto a regresión lineal, redes neuronales artificiales y funciones de transferencia.

2.1 Regresión lineal Múltiple

Se estudiará la situación en la que el comportamiento de la variable y (variable dependiente) es explicado mediante una relación lineal en función de las variables x_1, x_2, \dots, x_k (variables independientes, regresoras o explicativas)

2.1.1 Modelo de Regresión Lineal Múltiple (MRLM)

Sea y una variable respuesta y x_1, x_2, \dots, x_k , variables regresoras; para describir la relación existente entre la variable dependiente y las variables independientes, si es que la hay, es necesario que se cumpla lo siguiente:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i, \quad \text{para } i=1, 2, \dots, n \quad 2.1$$

Donde: y_i : es la variable respuesta cuantitativa para la i -ésima observación; x_i : es la variable regresora i -ésima que se supone esta relacionada con y_i ; β_j : para $j = 0, 1, 2, \dots, k$, son los parámetros o coeficientes de regresión, siendo n el número de observaciones. Este modelo es llamado modelo de regresión lineal múltiple (MRLM), y ε_i es el error aleatorio y tiene las siguientes propiedades:

$$\begin{aligned}
1) \quad E(\varepsilon_i) &= 0 \\
2) \quad V(\varepsilon_i) &= \sigma^2 \\
3) \quad Cov(\varepsilon_i, \varepsilon_j) &= 0 \quad \forall i \neq j
\end{aligned}
\tag{2.2}$$

La primera propiedad indica que el promedio de los errores es igual a cero, la segunda que la varianza de cada uno de ellos es constante (homocedástico) y la tercera que los errores son independientes entre si. Es más conveniente expresar el modelo (2.1) en notación matricial así:

$$y = X\beta + \varepsilon \tag{2.3}$$

Donde:

$$\begin{aligned}
y &= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad x = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \\
\beta &= \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}
\end{aligned}$$

En general, y es un vector de observaciones $n \times 1$, X es una matriz $n \times p$ de variables regresoras, β es un vector $k \times 1$ de coeficientes de la regresión y ε es un vector $n \times 1$ de errores aleatorios (Montgomery y Peck 2001). Como el objetivo es estimar los parámetros del modelo 2.1 (equivalente al modelo 2.3), se define una nueva variable \hat{y}_i como el valor estimado de y_i , tal que:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_k x_{ik}, \text{ para } i=1, 2, \dots, n \tag{2.4}$$

Donde: $\hat{\beta}^T = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$: es el vector de parámetros estimados. Los residuos e_i son la contraparte muestral de ε_i y están dados por:

$$e_i = y_i - \hat{y}_i, \text{ para } i=1, 2, \dots, n \quad 2.5$$

El estimador de β se obtiene usando el método de mínimos cuadrados, que consiste en minimizar la suma de cuadrados del error y está dado por:

$$SCE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 = e^T e \quad 2.6$$

Substituyendo $e = y - X\hat{\beta}$ en (2.6), se obtiene la ecuación normal de mínimos cuadrados (Montgomery y Peck 1991):

$$(X^T X)\hat{\beta} = X^T y \quad 2.7$$

Suponiendo que los regresores son linealmente independientes, la solución del sistema lineal será:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad 2.8$$

2.1.2 Problemas con la Regresión

Algunos de los problemas más comunes cuando se aplica el MRLM son:

- Omisión de variables: puede causar que los coeficientes se sobreestimen o subestimen y por lo tanto puede proporcionar estimadores sesgados.
- Relación no lineal entre la variable dependiente y las variables independientes, puede ocasionar problemas de inestabilidad en la varianza ocasionando problemas de inferencia

- Cuando hay autocorrelación entre los errores, igualmente la inferencia estadística tendrá problemas serios.
- La no normalidad de los errores, puede invalidar los resultados de la inferencia estadística, ya que las pruebas t y F requieren que los errores tengan distribución de probabilidad normal.

2.1.3 Herramientas para el Análisis de Regresión Lineal Múltiple

Para validar el MRLM, existe una serie de métodos para detectar los problemas que puedan estar ocurriendo o lo que en estadística se llama diagnóstico del modelo. Una revisión muy amplia de las técnicas para el análisis de MRLM se puede encontrar en (Montgomery y Peck 2001 y Weisberg 1980).

2.1.3.1 Diagrama de Dispersión

Es un gráfico bidimensional, que sirve para determinar visualmente la relación existente entre dos variables cuantitativas, una dependiente y la otra independiente.

2.1.3.2 Coeficiente de Determinación

Para construir una medida descriptiva del ajuste global del modelo se usa el cociente entre la variabilidad explicada por la regresión y la variabilidad total, a esta medida se le llama coeficiente de determinación y esta dado por:

$$r^2 = \frac{\text{variabilidad explicada}}{\text{variabilidad total}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad 2.9$$

Este coeficiente esta acotado, es decir, $0 \leq r^2 \leq 1$ y se expresa en forma porcentual. Cuanto más cercano está a 1, indica que el ajuste global es de buena calidad. Al valor r se le

denomina coeficiente de correlación entre x e y . La desventaja de r^2 es que su valor se incrementa cuando se introducen nuevos regresores en el modelo, por tanto resulta engañoso para el análisis.

2.1.3.3 *Coeficiente de Determinación Ajustado*

Para evitar la desventaja de r^2 , este coeficiente se define como un cociente de varianzas, en lugar de un cociente de sumas de cuadrados, al que se le llama coeficiente de determinación corregido por los grados de libertad; y esta dado por:

$$r_a^2 = 1 - (1 - r^2) \left(\frac{n-1}{n-k-1} \right) \quad \mathbf{2.10}$$

Su interpretación es similar al coeficiente de determinación, descrito en la sección anterior. Sin embargo, este tiene la propiedad de que al tener regresores innecesarios disminuye considerablemente.

2.1.3.4 *Error Estándar de Estimación*

Es la raíz cuadrada de la varianza residual, que está dado por $\sqrt{S_e^2}$, donde S_e^2 es la varianza residual y está dada por.

$$S_e^2 = \frac{SCE}{n-k-1} = \frac{\sum_{i=1}^n e_i^2}{n-k-1} \quad \mathbf{2.11}$$

Por lo tanto: S_e es el error estándar de estimación y es un estimador de la desviación estándar de los errores.

2.1.3.5 *Promedio del error absoluto (MAE)*

Es el promedio de los valores absolutos de cada uno de los errores y esta dado por:

$$MAE = \left(\frac{1}{n} \right) \sum_{i=1}^n \|e_i\| \quad 2.12$$

Donde n : es la cantidad de datos.

2.1.3.6 Suma de los errores al cuadrado (SSE)

Es la sumatoria de cada uno de los errores elevado al cuadrado y esta dado por:

$$SSE = \sum_{i=1}^n e_i^2 \quad 2.13$$

Donde n : es la cantidad de datos.

2.1.3.7 Análisis de Varianza

Se usa para probar la hipótesis:

$$\begin{aligned} H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_1 : \text{algún } \beta_j \neq 0, j = 1, 2, \dots, n \end{aligned} \quad 2.14$$

Sirve para verificar la significancia estadística de los coeficientes de regresión de manera conjunta, en otras palabras indica si alguna o algunas variables x ayudan a explicar la variable y .

2.1.3.8 Pruebas de Hipótesis sobre los Coeficientes Individuales

Se usa para probar las hipótesis:

$$\begin{aligned} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{aligned} \quad 2.15$$

Sirve para identificar las variables útiles para expresar el comportamiento de y y se usa el siguiente estadístico:

$$t = \frac{\hat{\beta}_j}{S'_{\hat{\beta}_j}} \sim t_{(n-k-1)} \quad 2.16$$

Donde $S'_{\hat{\beta}_j}$ es el estimado de la varianza de $\hat{\beta}_j$ y t es una variable aleatoria que sigue la distribución *t-student* con $n-k-1$ grados de libertad.

2.2 Redes Neuronales Artificiales (RNA)

2.2.1 Generalidades

Una red neuronal artificial no es más que una emulación del comportamiento del cerebro humano. Consiste en una asociación de sus componentes principales o unidades procesadoras llamadas neuronas. Esta asociación de neuronas depende de manera intrínseca de las interconexiones entre las neuronas. Dichas interconexiones son la base fundamental de las redes neuronales. En las redes biológicas, estas interconexiones son llamadas sinapsis y tanto en las redes biológicas como en las redes neuronales artificiales (RNA), su número es elevado, es decir que las neuronas de una RNA están altamente interconectadas (Hagan et al., 1996). Una RNA determina una relación empírica entre las variables de entrada y las de salida de un sistema dado, en el que las entradas son variables independientes (x_1, x_2, \dots) y las salidas son variables dependientes (y_1, y_2, \dots). Por lo tanto, es necesario que el diseñador de la red posea un buen entendimiento del tema para poder explicar la relación identificada (Rodríguez 2006). La Figura 2.1 muestra la estructura general de una red neuronal.

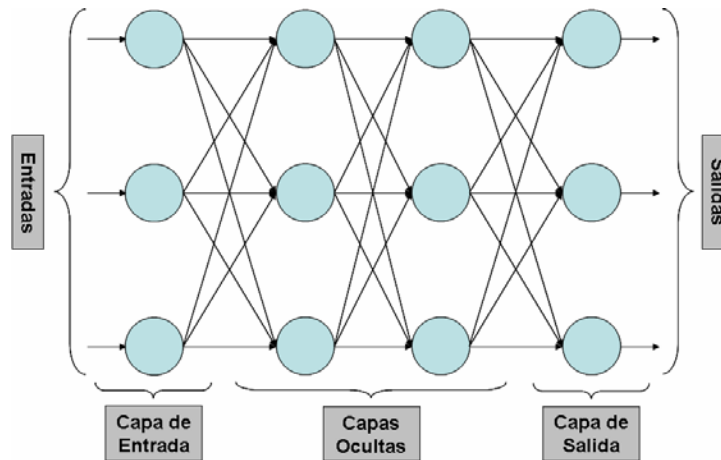


Figura 2.1 Estructura general de una RNA.

Las RNA tienen diferentes características, las cuales pueden usarse dependiendo del campo en el que vaya a ser aplicada la red. La característica principal a usarse aquí es la no linealidad de las neuronas (Hilera y Martines 1995). Las conexiones entre las neuronas de una RNA tienen asociado un peso w_{ji} , que a su vez está asociado a una regla determinada. Esta regla es llamada función de activación o transferencia. Generalmente, los pesos son creados o inicializados de manera aleatoria y luego por medio de un cierto algoritmo de entrenamiento, son actualizados para optimizar la respuesta de la red. La función de activación se encarga de asignar un valor de salida a cada neurona de la red en cada iteración (ver Figura 2.2). La entrada neta que recibe la neurona j -ésima (net_j), es la suma de cada señal individual multiplicada por el valor de la sinapsis (el peso) que la conecta con otra neurona, así:

$$net_j = \sum_i y_i w_{ji} \quad 2.17$$

Donde: y_i es el valor de salida de una neurona i en un instante dado.

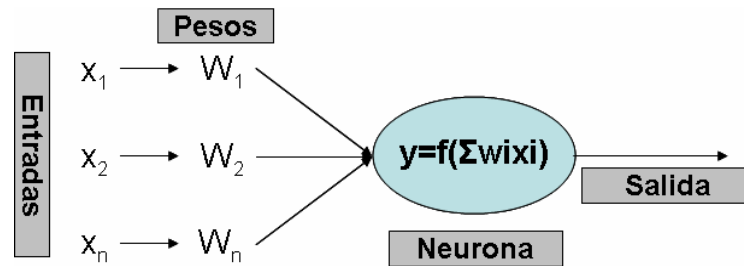


Figura 2.2 Caracterización del procesamiento de una neurona.

En la Figura 2.3 se puede observar que los vectores de entrada son afectados por sus respectivos pesos. Internamente, la neurona realiza las operaciones correspondientes, actuando sobre ella una cierta función de activación, la cual es la encargada de entregar la salida de la neurona.

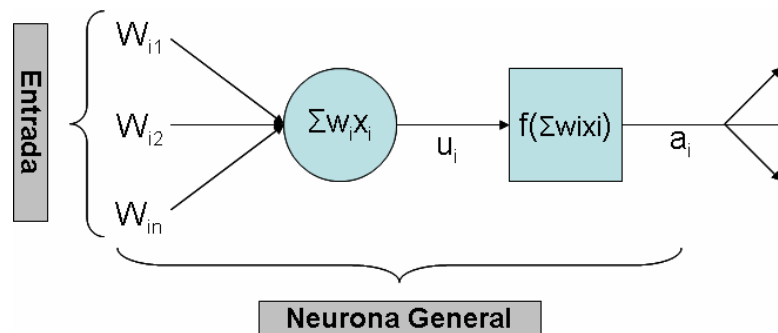


Figura 2.3 Esquema de procesamiento neuronal.

2.2.2 Mecanismo de Aprendizaje

Se considera que la red aprende modificando los valores de sus pesos. El mecanismo de aprendizaje de la red neuronal modifica sus pesos como respuesta a la información presentada. El proceso de aprendizaje termina cuando los valores de los pesos permanecen estables es decir, cuando el gradiente de la función objetivo con respecto a los pesos es cero. Esto es se consigue un punto estacionario. La regla de aprendizaje esta determinada por los criterios para cambiar el valor de las conexiones. Se conocen dos tipos de aprendizaje: aprendizaje supervisado y aprendizaje no supervisado.

2.2.2.1 Redes con Aprendizaje Supervisado

El entrenamiento es controlado por un agente externo (supervisor o maestro), el cual le presenta a la red los diferentes casos, determinando la respuesta a generar por la red a partir de una entrada determinada. Una revisión detallada puede encontrarse en (Hilera, 1995 y Hagan, 1996).

Algoritmo de aprendizaje del gradiente descendente

Este algoritmo sigue una técnica de optimización no lineal conocida como método del gradiente. En forma simplificada se puede expresar así (Hagan 1996):

$$w_{i+1} = w_i - \alpha \frac{\partial F}{\partial w_i} \quad 2.18$$

donde: w_i : Representa el los pesos de las conexiones en la iteración i ; α : es el desplazamiento o velocidad de aprendizaje; $\frac{\partial F}{\partial w_i}$: es la dirección de búsqueda para la obtención de los nuevos pesos, siendo $F = \sum (y_i - a_i)^2$, y_i es el valor objetivo u observado y a_i es la salida de la red. Valor de salida obtenido para la neurona j .

Regla delta

Un algoritmo muy conocido que permite un aprendizaje rápido fue propuesto por (Widrow 1960), llamado regla *delta* o regla del mínimo error cuadrado medio (LMSE: Least-Mean-Squared Error), que se aplicó en las redes desarrolladas por los mismos, conocidas como ADALINE y MADALINE). (Widrow 1960) definieron una función que permite cuantificar el error global cometido en cualquier momento del proceso de entrenamiento, lo cual es

especialmente importante, ya que cuanto más información se tenga sobre el error cometido, más rápido aprenderá la red. Este error medio se expresa de la siguiente forma:

$$Error_{global} = \frac{1}{2P} \sum_{k=1}^P \sum_{j=1}^N (y_j^{(k)} - d_j^{(k)})^2 \quad 2.19$$

Donde: N : es el número de neuronas de salida; P : número de datos que debe aprender la red;

$\frac{1}{2} \sum_{j=1}^N (y_j^k - d_j^k)^2$: es el error cometido durante el aprendizaje de la información k -ésima.

La idea es encontrar los pesos de las conexiones que conlleven a minimizar el error. Luego, el ajuste de los pesos se debe hacer de forma proporcional a la variación relativa del error que se obtiene al variar el peso correspondiente así:

$$\Delta w_{ji} = \alpha \frac{\partial Error_{global}}{\partial w_{ji}} \quad 2.20$$

Mediante este procedimiento, se obtiene un conjunto de pesos con los que se consigue minimizar el error medio. Otro algoritmo de aprendizaje por corrección de error es la regla delta generalizada o algoritmo de retro propagación del error (error backpropagation), el cual es una generalización de la regla delta aplicado a redes con conexiones hacia adelante (feedforward) con capas o niveles ocultos.

2.2.2.2 Redes con Aprendizaje No Supervisado

En el aprendizaje no supervisado no se requiere presentar patrones de salida deseados. Sin embargo, el algoritmo y la regla de modificación de las conexiones producen vectores de salida consistentes. Se considera que el proceso de aprendizaje es capaz de extraer ciertas propiedades estadísticas de los ejemplos de aprendizaje y agruparlos en categorías o clases de patrones similares. No se conoce a priori que salida corresponderá a cada tipo o grupo de

patrones de entrada, ni que atributos usará para clasificarlos. La red encontrará las características, regularidades, correlaciones o categorías que se puedan establecer entre los datos que se presentan en su entrada. Normalmente se tienen tres algoritmos de aprendizaje no supervisado: a) Aprendizaje Hebbiano; b) Aprendizaje por componentes principales; c) Aprendizaje competitivo y cooperativo. Los dos primeros no serán tenidos en cuenta, por lo tanto, si el lector desea encontrar una revisión detallada puede remitirse a (Hilera, 1995 y Hagan, 1996).

Aprendizaje Competitivo y Cooperativo.

En este las neuronas compiten (y cooperan) unas con otras con el fin de llevar a cabo una tarea asignada. La competencia entre neuronas se realiza en todas las capas de la red, existiendo en estas neuronas conexiones recurrentes de auto excitación y conexiones de inhibición por parte de neuronas vecinas. Si el aprendizaje es cooperativo, estas conexiones con las vecinas serán de excitación. Las informaciones similares son clasificadas formando parte de la misma categoría, y por tanto deben activar la misma neurona de salida. Las categorías deben ser creadas por la misma red, puesto que se trata de aprendizaje no supervisado, a través de las correlaciones entre los datos. El aprendizaje afecta sólo a las neuronas ganadoras, redistribuyendo el peso total entre todas las conexiones que llegan a la neurona vencedora y repartiendo esta cantidad por igual entre todas las conexiones procedentes de unidades activas. Por tanto, la variación del peso de una conexión entre una unidad i y otra j será nula si la neurona j no recibe excitación por parte de la neurona i y se modificará si es excitada por dicha neurona i .

2.2.3 Dimensionamiento de la Red

No existen reglas concretas para determinar el número de neuronas o el número de capas de una red con el fin de resolver un problema concreto. Respecto al número de capas de la red, en general tres capas son suficientes (entrada-oculta-salida). Sin embargo, algunos problemas son más fáciles de resolver con más de una capa oculta. El tamaño de las capas, tanto de entrada como de salida, lo determina la naturaleza de la aplicación. El número de neuronas ocultas interviene en la eficiencia de aprendizaje y de generalización de la red. No hay ninguna regla que indique el número óptimo, por lo cual es necesario ensayar en cada problema. Un eficiente procedimiento para identificar la estructura apropiada de la red está dado por Ramírez y Montes (2002).

2.2.4 Algoritmos de Entrenamiento

Existen varios algoritmos para entrenar una RNA, algunos de los mas usados son: *Perceptron*, *Hebbian*, *Widrow-Hoff*, *Backpropagation*, *Levenverg–Marquard* y *Competitive*; estos dos últimos son los mas convenientes para esta investigación. El algoritmo de *Levenverg–Marquard* es una modificación del método de Newton, diseñado para minimizar funciones que son la suma de los cuadrados de otras funciones no lineales (Hagan et al., 1996). Mientras que el algoritmo *Competitive* está diseñado para encontrar similitudes entre los datos de entrada a partir de comparaciones de sus distancias. Una descripción mas detallada de estos dos algoritmos se encuentra en el Anexo A.

2.3 Funciones de Transferencia (FT)

2.3.1 Introducción.

En los modelos de función de transferencia, el objetivo es relacionar dos ó más series de tiempo elaborando modelos causales de predicción o estimación. Se considera a priori que existe una causalidad unidireccional desde las variables de entrada hacia la variable respuesta. Los modelos de FT son muy utilizados en todos los campos científicos para evaluar respuestas dinámicas. Si las variables de entrada son controlables, estos modelos permiten simular y evaluar políticas alternativas. Si no lo son, ofrecen la posibilidad de estudiar ciertos "escenarios", definidos por posibles evoluciones de la variable explicativa, y su efecto en la variable respuesta. Además estos modelos son muy útiles para elaborar predicciones, o estimaciones. La construcción de los modelos de FT consta de las siguientes etapas: a) Identificación; b) Estimación; c) Verificación y d) Predicción.

2.3.2 Modelos de función de transferencia con una sola entrada.

2.3.2.1 Conceptos Generales.

El modelo de función de transferencia más sencillo es el que tiene una sola entrada, y se puede extender sin complicaciones a modelos con varias entradas. En un sistema lineal de una variable de entrada (*input*) y una variable respuesta (*output*), las series *input* y *output* se relacionan a través de un filtro lineal de la siguiente forma:

$$Y_t = v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \dots + N_t(l) = v(B) X_t + N_t \quad \mathbf{2.21}$$

Donde $v(B) = v_0 + v_1 B + v_2 B^2 + \dots$ se refiere a la FT del filtro de Box y Jenkins (Box y Jenkins 1976 y Wei 1990); N_t es el ruido del sistema que es independiente de la serie *input* X_t ;

B es un operador de retardo que indica cuantas veces se debe retrasar la variable a la que se le aplique. Los coeficientes de $v(B)$ se conocen como la función de respuesta al impulso. Para que el sistema (2.21) sea estable se debe cumplir que una variación finita en el *input* genere una variación también finita en el *output*. Esto es deberá cumplirse que:

$$\sum_{j=0}^{\infty} v_j = g \quad 2.22$$

Siendo g finito. El valor de g representa el cambio total en Y_t debido a un cambio unitario en X_t mantenido indefinidamente en el tiempo. Los propósitos del modelaje con FT son identificar y estimar la FT $v(B)$ y el modelo del ruido N_t sobre la base de la información que proporciona el *input* y la serie *output*. El modelo dado en (2.21) es inestimable, ya que posee un número infinito de parámetros. Este problema se resuelve al expresar la FT como el cociente de dos polinomios finitos:

$$v(B) = \frac{w(B)B^b}{\delta(B)} \quad 2.23$$

Donde: $w(B) = w_0 - w_1B - w_2B^2 - \dots - w_sB^s$; $\delta(B) = 1 - \delta_1B - \delta_2B^2 - \dots - \delta_rB^r$ y b es un parámetro de retardo que representa el tiempo que transcurre antes de que un impulso en el *input* produzca un efecto en el *output*. Se asume que las raíces de $\delta(B) = 0$ caen fuera del círculo unitario para que el sistema sea estable. Sustituyendo (2.23) en (2.21), se obtiene:

$$Y_t = \frac{w(B)B^b}{\delta(B)} X_t + N_t \quad 2.24$$

Por otro lado, el término de error no es necesariamente un ruido blanco. Se puede suponer, sin pérdida de generalidad, que N_t sigue un proceso ARIMA (Brockwell, 2002), aunque sigue siendo independiente de la variable *input* X_t , así:

$$\phi(B)(1-B)^d N_t = \theta(B)u_t \Rightarrow N_t = \frac{\theta(B)}{\phi(B)(1-B)^d} u_t \quad 2.25$$

Con $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ de manera que todas las raíces de ambos polinomios caen fuera del círculo unitario siendo $(1-B)^d$ el operador diferencias consecutivas utilizado para inducir que el proceso sea estacionario y u_t es un ruido blanco. Sustituyendo (2.25) en (2.24), se obtiene:

$$(1-B)^d Y_t = \frac{w(B)B^b}{\delta(B)} (1-B)^d X_t + \frac{\theta(B)}{\phi(B)} u_t \quad 2.26$$

A partir de (2.26) puede observarse que: si el proceso del término error no es estacionario (esto quiere decir que debe diferenciarse d veces para conseguir que sea estacionario), ese mismo orden de diferenciación recae tanto sobre la variable dependiente u *output* como sobre la variable explicativa o *input*. Sin embargo, en la práctica, para construir un modelo de FT, es necesario que tanto la variable dependiente como la variable explicativa sean estacionarias, por lo cual, ambas variables no necesariamente requerirán el mismo orden de diferenciación para lograr que sea estacionario. Por otra parte, una vez logrado que las variables sean estacionarias, el proceso N_t será un ARMA(p, q) estándar (Brockwell, 2002), con lo cual el modelo de función de transferencia se puede escribir como sigue:

$$y_t = \frac{w(B)B^b}{\delta(B)} x_t + \frac{\theta(B)}{\phi(B)} u_t \quad 2.27$$

Donde $y_t = (1-B)^{d'} Y_t$ y $x_t = (1-B)^d X_t$, siendo d' el orden de diferenciación de Y_t y d el orden de diferenciación de X_t . La estructura de la FT determina la naturaleza de la influencia de la variable explicativa sobre la variable dependiente. En la Figura 2.4 puede verse una representación esquemática de la FT.

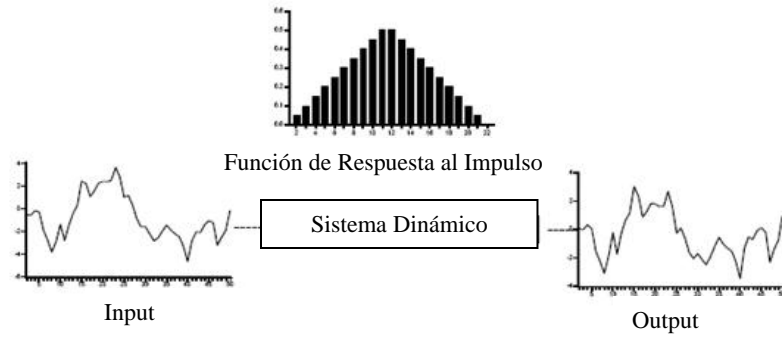


Figura 2.4 Input y Output a partir de un sistema dinámico.

La conjunción de la estructura de la función de impulso y el modelo de los errores integran el modelo de función de transferencia completo que puede verse en la Figura 2.5. A partir de (2.27), una vez que conocidos los valores de $w(B)$, $\delta(B)$ y b , se obtendrán los coeficientes de respuesta al impulso, v_j , igualando los coeficientes de B^j de ambos lados de la siguiente ecuación:

$$\delta(B)v(B) = w(B)B^b \quad 2.28$$

Que también puede escribirse como:

$$(1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r)(v_0 + v_1 B + v_2 B^2 + \dots) = (1 - w_1 B - w_2 B^2 - \dots - w_s B^s) \quad 2.29$$

Un análisis más completo de la función de respuesta al impulso puede encontrarse en (Box y Jenkins 1976 y Wei 1990).

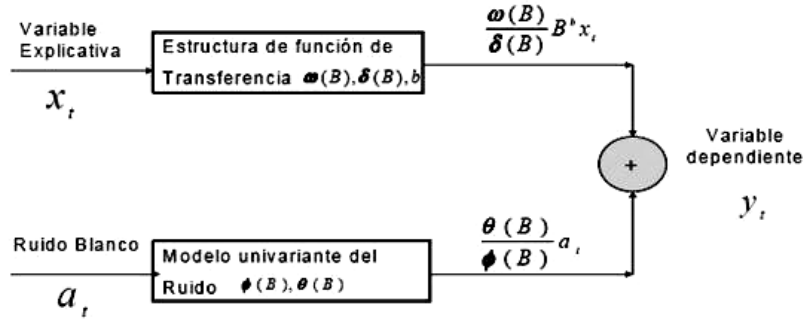


Figura 2.5 Estructura de la función de transferencia con el modelo del ruido.

2.3.2.2 Modelo de FT con Múltiples Entradas.

El modelo de FT se puede generalizar a:

- La existencia de más de un *input*
- Que el ruido viene generado por un proceso estacional multiplicativo

Para el caso de más de una entrada, si se tienen k *inputs*, el modelo de FT sería el siguiente:

$$\begin{aligned}
 y_t &= \frac{w_1(B)}{\delta_1(B)} x_{1,t-b_1} + \frac{w_2(B)}{\delta_2(B)} x_{2,t-b_2} + \dots + \frac{w_k(B)}{\delta_k(B)} x_{k,t-b_k} + N_t \\
 &= \frac{w_1(B)}{\delta_1(B)} B^{b_1} x_{1t} + \frac{w_2(B)}{\delta_2(B)} B^{b_2} x_{2t} + \dots + \frac{w_k(B)}{\delta_k(B)} B^{b_k} x_{kt} + N_t \\
 &= \sum_{i=1}^k \frac{w_i(B)}{\delta_i(B)} B^{b_i} x_{it} + N_t
 \end{aligned} \tag{2.30}$$

Donde: $y_t = (1-b)^d Y_t \quad \wedge \quad x_{it} = (1-B)^d X_{it}$

La estimación de parámetros para los modelos con más de una entrada no es una tarea fácil (Ramírez et al 2007b), puesto que comprende el uso de diferentes técnicas de optimización.

Para el desarrollo de este trabajo fue necesario utilizar las técnicas conocidas como *sequential quadratic programming* (SQP) y *Hooke and Jeeves* (H&J) (ver Anexo B).

2.3.3 Procedimiento General Para la Identificación de la FT.

2.3.3.1 El Preblanqueo.

Se le llama preblanqueo a las operaciones necesarias para conseguir que una serie de tiempo sea estacionaria (Box y Jenkins 1976, Brockwell 2002 y Wei 1990). Para ello, si se escribe el modelo de función de transferencia como:

$$y_t = v(B)x_t + N_t \quad 2.31$$

y se hace la suposición de que x_t viene generada por un proceso ARMA del tipo:

$$\phi_x(B)x_t = \theta_x(B)\alpha_t \quad 2.32$$

Donde $\phi_x(B)$ y $\theta_x(B)$ son los polinomios auto regresivo y de media móvil del proceso ARMA para la serie *input* que cumplen las condiciones de estacionalidad e invertibilidad usuales (Box y Jenkins 1976, Brockwell 2002 y Wei 1990) y α_t es un proceso de ruido blanco. Así, si se escribe α_t como:

$$\alpha_t = \frac{\phi_x(B)}{\theta_x(B)}x_t \quad 2.33$$

La ecuación 2.33 es llamada “serie *input* preblanqueada”. Aplicando la misma transformación de preblanqueo a la serie de output, se obtiene:

$$\beta_t = \frac{\phi_x(B)}{\theta_x(B)}y_t \quad 2.34$$

De esta forma, si en el modelo de función de transferencia (2.31) se premultiplica a ambos lados por $\frac{\phi_x(B)}{\theta_x(B)}$, se obtiene

$$\frac{\phi_x(B)}{\theta_x(B)} y_t = v(B) \frac{\phi_x(B)}{\theta_x(B)} x_t + \frac{\phi_x(B)}{\theta_x(B)} N_t \quad 2.35$$

Si se toma $\varepsilon = \frac{\phi_x(B)}{\theta_x(B)} N_t$ se el resultado sería:

$$\beta_t = v(B) \alpha_t + \varepsilon_t \quad 2.36$$

Ecuación a partir de la cual se obtiene fácilmente que:

$$v_j = \frac{\sigma_\beta}{\sigma_\alpha} \rho_{\beta\alpha}(j) \quad 2.37$$

Ya que α_t es un proceso de ruido blanco. Así pues, si se conocen los elementos poblacionales $\rho_{\beta\alpha}(j)$, σ_β y σ_α , se podrían obtener los valores de v_j , y de acuerdo con las características de los coeficientes de la función de respuesta al impulso, sería posible identificar el modelo de FT pertinente. En la práctica, al no conocer los valores poblacionales de $\rho_{\beta\alpha}(j)$, σ_β y σ_α es necesario estimar estos coeficientes a partir de datos muestrales.

En General, para obtener una serie estacionaria, basta con remover su tendencia y periodicidad. La tendencia o la parte determinística, es posible removerla restándole a la serie su primera o segunda diferencia o removiéndole una función lineal o cuadrática. El componente periódico también se puede remover usando diferencias; luego de conocer los periodos asociados a la serie. Los periodos se reconocen por medio del periodograma. Además pueden ser removidos usando funciones sinusoidales para restarlas a la serie original.

2.3.3.2 Identificación de la función de respuesta al impulso.

Con el objetivo de ilustrar al lector, se presentará a primero el procedimiento general para identificar la función de respuesta al impulso (Wei, 1990) y luego el procedimiento utilizado en este trabajo.

❖ Considérese el modelo de FT con un solo *input*:

$$y = v(B)x_{t-b} + N_t = \frac{w(B)}{\delta(B)}x_t + N_t \quad 2.38$$

con $N_t = \frac{\theta(B)}{\phi(B)}u_t$. La identificación de la función de respuesta al impulso $v(B)$ se debe

realizar siguiendo las siguientes etapas:

Etapla 1: Considerando que y_t y x_t son estacionarias, se preblanquea el *input* de acuerdo con su representación ARMA

$$\phi_x(B)x_t = \theta_x(B)\alpha_t \Rightarrow \alpha_t = \frac{\phi_x(B)}{\theta_x(B)}x_t \quad 2.39$$

Donde α_t es un ruido blanco con media cero y varianza σ_α^2

Etapla 2: Filtrar la serie del output utilizando el mismo filtro que preblanqueó el *input*, obteniendo:

$$\beta_t = \frac{\phi_x(B)}{\theta_x(B)}y_t \quad 2.40$$

Etapla 3: Calcular la función de correlación cruzada muestral entre β_t y α_t para estimar v_j de acuerdo con la expresión:

$$\hat{v}_j = \frac{\hat{\sigma}_\beta}{\hat{\sigma}_\alpha} \hat{\sigma}_{\beta\alpha}(j) \quad 2.41$$

Etapa 4: Identificar r (orden del polinomio $\delta(B)$) y s (orden del polinomio $w(B)$) de acuerdo a las reglas descritas por Box y Jenkins (1976). Una vez que se escogen b , r y s , se pueden encontrar estimaciones preliminares de los coeficientes de $\delta(B)$ y $w(B)$ a partir de sus relaciones con v_j . Así se obtiene la estimación preliminar de la función de transferencia:

$$\hat{v}(B) = \frac{\hat{w}_\beta}{\hat{\delta}_\alpha} B^b \quad 2.42$$

Las etapas descritas se realizan bajo las siguientes suposiciones:

- Se identifica a $v(B)$ considerando que no hay error, esto quiere decir que $N_t = 0$.
 - Se parte de que las series y_t y x_t son estacionarias.
 - ❖ El procedimiento utilizado para encontrar los parámetros de las funciones de impulso usado en este trabajo esta resumido en los siguientes pasos (Ramírez et al 2007b):
- Paso 1: Aplicar el periodograma y la función de autocorrelación con el fin de determinar si la serie es o no un proceso estacionario. De ser necesario se remueve la tendencia tomando la primera $(1-B)h_t$ o segunda $(1-B)^2 h_t$ diferencia a la serie o removiendo una función paramétrica.

Paso 2: Identificar el orden de las funciones de impulso. Una buena técnica para identificar el orden es el algoritmo de búsqueda H&J.

Paso 3: El tercer paso consiste en realizar una búsqueda aleatoria con el fin de determinar los coeficientes iniciales de las funciones de impulso. Generar 100 puntos

con una distribución uniforme sobre un rango específico es una buena opción. Luego utilizar un criterio de error para identificar el punto inicial más conveniente.

Paso 4: Usando SQP estimar los parámetros de las funciones de impulso. El algoritmo de SQP se utiliza cuando se tienen no linealidades en la función de transferencia a determinar, es decir, cuando se trata de obtener una función de transferencia no lineal (MathWorks, 2000).

3 METODOLOGÍA

3.1 Introducción

Este capítulo introduce la metodología para el desarrollo de los algoritmos conducentes a la estimación de la temperatura del aire y de la humedad de suelo. Esta metodología fue desarrollada bajo el marco de información de estaciones de medición *in-situ*, información de sensores remotos, modelos redes neuronales artificiales, modelos de regresión múltiple y modelos de función de transferencia. Esto envuelve seis principales pasos: a) La metodología es brevemente descrita considerando sus limitaciones; b) Una base de datos es diseñada para almacenar la información tanto de los sensores remotos como de las estaciones de medición; c) Una red neuronal con el algoritmo *Levenberg-Marquardt* es usada para estimar los promedios mensuales de humedad de suelo usando datos de lluvia acumulada, amplitud mensual de temperatura, índice de vegetación, características texturales y datos topográficos; d) Se desarrollan funciones de transferencia para estimación horaria de la humedad de suelo usando datos de lluvia y temperatura horaria de las diferentes estaciones de medición; e) Se desarrollan modelos de estimación de temperatura horaria a partir de los valores máximos y mínimos diarios, usando datos de temperatura de las diferentes estaciones de medición; f) Una red neuronal competitiva es entrenada para clasificar regiones climatológicas similares en Puerto Rico, con el fin de asociar las estaciones de medición a las regiones que presenten una marcada similaridad con ellas. La asociación de estaciones/regiones es necesaria para la aplicación de las funciones de transferencia en sitios diferentes a la estación.

3.2 Descripción General de la Metodología

Modelos de estimación de temperatura del aire y humedad de suelo son presentados en este trabajo. Algunos de los datos usados en este trabajo son obtenidos de sensores remotos, otros de estaciones de medición y otros se basan en información recopilada sobre Puerto Rico por agencias gubernamentales. Estos modelos son validos para Puerto Rico, aunque pueden ser aplicados en otras regiones con el conocimiento de las respectivas variables. Un esquema de la metodología para la estimación de temperaturas y humedad de suelo es mostrado en la Figura 3.1. En dicha figura se puede observar el establecimiento de una base de datos a partir de los datos de entrada a los modelos. Esta base de datos será explicada en el siguiente apartado. Es posible observar también en dicha figura los diferentes modelos generados en el transcurso de este trabajo, cuyos datos de entrada provienen de la base de datos creada y la información de salida es utilizada conjuntamente para la obtención de los resultados finales de estimación de humedad de suelo y temperatura del aire.

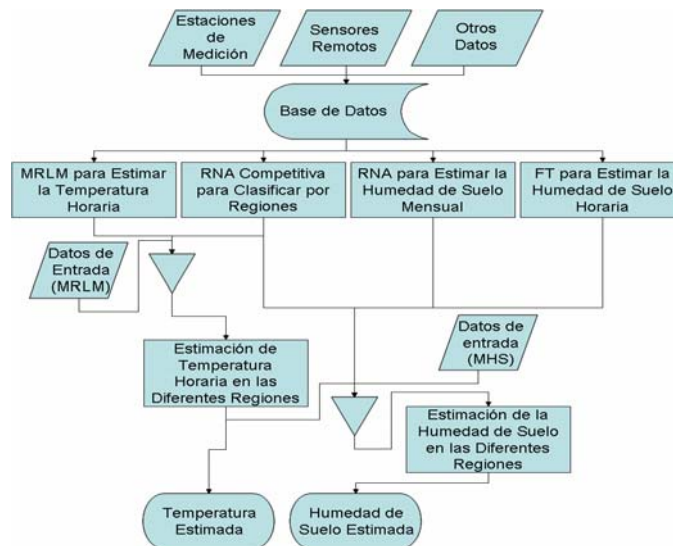


Figura 3.1 Metodología usada en este estudio.

Esta metodología esta limitada como se dijo anteriormente, por el tamaño de la base de datos, en particular por la cantidad de observaciones puntuales, que en este caso corresponde al número de estaciones usadas (15 en total). Es de anotar, que entre más observaciones se tengan, mejor será el modelo de estimación creado.

3.3 Diseño de la Base de Datos

Una base de datos fue desarrollada para almacenar la información de las estaciones de medición, de los sensores remotos y otros datos de importancia.

3.3.1 Descripción de los Datos.

3.3.1.1 Fuente de Datos Primaria.

La fuente de datos primaria para este estudio fue observaciones de humedad de suelo y temperatura proveniente de diecisiete estaciones de medición ubicadas en el oeste de Puerto Rico. La Figura 3.2 muestra la localización de dichas estaciones. Doce de estas estaciones pertenecen al proyecto llamado “Estimación de Humedad de Suelo (EHS)” y las cinco restantes pertenecen a la *Natural Resources Conservation Service* (NRCS) (para mayor información ver <http://www.wcc.nrcs.usda.gov/scan/>). Las estaciones pertenecientes al proyecto EHS cuentan con tres sensores de humedad de suelo a 20 cm. de profundidad, un sensor de temperatura ubicado a 20 cm. de altura y un sensor de lluvia. Para obtener mayor información a cerca de los dispositivos usados se puede remitir al trabajo hecho por Cruzado (2004) y a la pagina Web <http://www.decagon.com/Ech2o/>.

Las estaciones de la NRCS cuentan con sensores de humedad de suelo a diferentes profundidades, entre ellas a 20 cm., sensores de lluvia, sensores de temperatura de suelo y aire, entre otros.

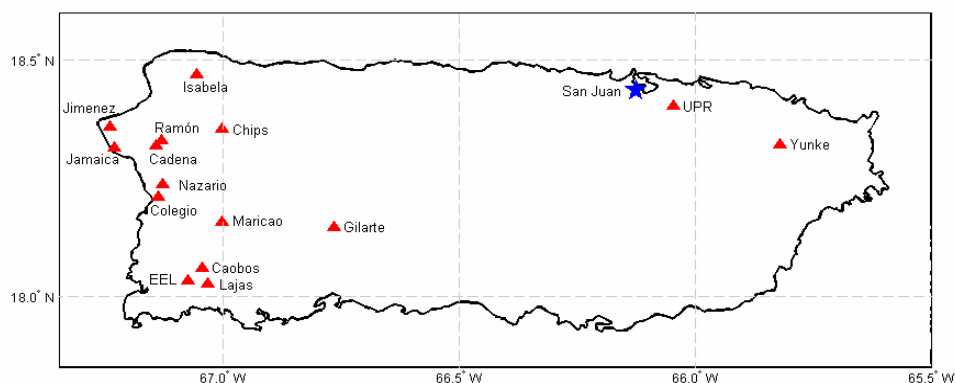


Figura 3.2 Localización de las estaciones de medición de humedad de suelo.

Los sensores de Humedad miden la constante dieléctrica del suelo y esta señal es convertida en datos de humedad de suelo (Wang y Smchmugge 1980). Las mediciones de humedad, lluvia y temperatura son en forma horaria, tanto para las estaciones del proyecto EHS como para las estaciones de la NRCS. Para este trabajo se tuvo en cuenta datos de humedad de suelo de los meses de Enero a Diciembre de los años 2005 y 2006.

3.3.1.2 Fuentes de datos de Sensores Remotos.

Datos de Temperatura

Datos de temperatura superficial son recolectados a través del *Earth Observing System* (EOS) *data gateway* (para mayor información ver <http://edcimswww.cr.usgs.gov/pub/imswelcome/> y <http://modis.gsfc.nasa.gov/about/>) que provienen del sensor *Moderate Resolution Imaging Spectroradiometer* (MODIS), el cual esta instalado en el satélite *Terra*. MODIS toma mediciones de temperatura de la superficie de la tierra de día y noche, las primeras con un intervalo de paso variable de entre las 14:30 horas y las 16:30 horas (horario de PR), mientras que las mediciones nocturnas las realiza entre la 1:30 horas y las 3:30 horas (horario de PR) aproximadamente; estas mediciones son captadas con una resolución de 1 Km². Se tienen datos diarios y mensuales de MODIS. Es decir, existen archivos diarios que

almacenan la temperatura de día y de noche, además se usan archivos con datos promedio de ocho días para obtener el promedio mensual de temperatura superficial. Los datos de temperatura superficial, tanto diarios como los de ocho días, son manipulados usando Matlab (MathWorks, 2000) y durante esta manipulación es posible que se tenga que hacer interpolaciones para encontrar datos faltantes o “*missing values*”, ver Figura 3.3. Para realizar estas interpolaciones se usa el algoritmo llamado *Kriging Near Neighbors* (KNN) (Wackernagel, 2003 y Cruzado, 2004). Se usaron las siguientes ecuaciones de corrección para estimar los valores máximos y mínimos de temperatura del aire partiendo de los datos de temperatura superficial:

$$\begin{array}{l} T_{\max} = 1.0068 \cdot T_{md} - 0.0082 \cdot H \\ T_{\min} = 1.0121 \cdot T_{mn} - 0.0801 \cdot H \end{array} \rightarrow \begin{array}{ll} MAE = 2.69 & MSE = 11.65 \\ MAE = 4.31 & MSE = 32.82 \end{array}$$

Donde T_{\max} y T_{\min} son las temperaturas máximas y mínimas respectivamente, T_{md} y T_{mn} es la temperatura medida por MODIS de día y de noche respectivamente y H es la altura en metros sobre el nivel del mar. Estas ecuaciones fueron desarrolladas usando temperaturas máximas y mínimas diarias desde el 01 de Noviembre hasta el 31 de diciembre del 2003 provenientes del sensor MODIS y de 24 de las estaciones cooperativas ubicadas sobre toda la extensión de Puerto Rico (ver <http://www.ncdc.noaa.gov/oa/hofn/coop/coop-ste-opt.html>). Se asume que los datos colectados por dichas estaciones son coherentes con los datos colectados por las estaciones del proyecto. Esto quiere decir que los datos colectados por las estaciones cooperativas equivalen a los datos de temperatura colectados por las estaciones del proyecto. Por otro lado, se obtuvieron los valores de rango mensual de temperatura del aire (R), mediante la diferencia entre el valor máximo mensual y el valor mínimo mensual.

$$R = T_{Max} - T_{Min} \quad 3.1$$

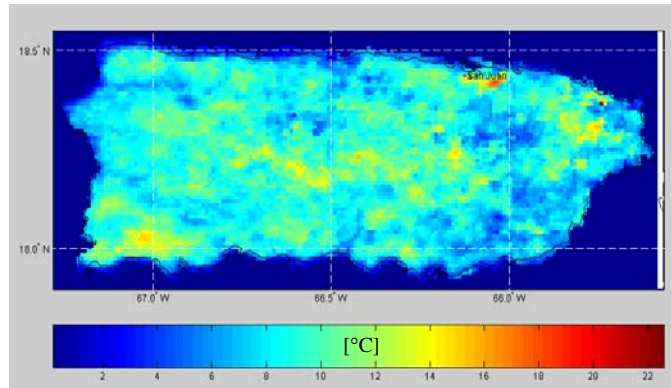


Figura 3.3 Rango mensual de temperatura usando datos de MODIS para agosto 2006.

Índice de Vegetación.

El *normalized difference vegetation index* (NDVI) permite estimar el desarrollo de una vegetación en base a la medición, de la intensidad de la radiación de ciertas bandas del espectro electromagnético que la misma emite o refleja. Además permite identificar la presencia de vegetación verde en la superficie y caracterizar su distribución espacial así como la evolución de su estado a lo largo del tiempo. Esto está determinado fundamentalmente por las condiciones climáticas. (Wang et al 2003). El NDVI se define como:

$$NDVI = \frac{(NIR - VIS)}{(NIR + VIS)} \quad 3.2$$

Donde: *NIR*: es el infrarrojo cercano; *VIS*: es el espectro visible. Los valores de NDVI oscilan entre -1 y 1, además algunos valores característicos se describen a continuación: a) El agua tiene reflectancia $VIS > NIR$, por lo tanto valores negativos de NDVI; b) Las nubes presentan valores similares de *VIS* y *NIR*, por lo que su NDVI es cercano a cero; c) El suelo

descubierto y con escasa vegetación presenta valores positivos aunque no muy elevados; d) La vegetación densa, húmeda y bien desarrollada presenta los mayores valores de NDVI. Datos del índice diferencial normalizado de vegetación, mas conocido por sus siglas en ingles son colectados a través del EOS *data gateway* (ver: <http://modis.gsfc.nasa.gov/about/> y <http://edcimswww.cr.usgs.gov/pub/imswelcome/>). Proviene del sensor MODIS con una resolución de 1 Km² y están dados mensualmente. La Figura 3.4. muestra un ejemplo del NDVI para puerto Rico.

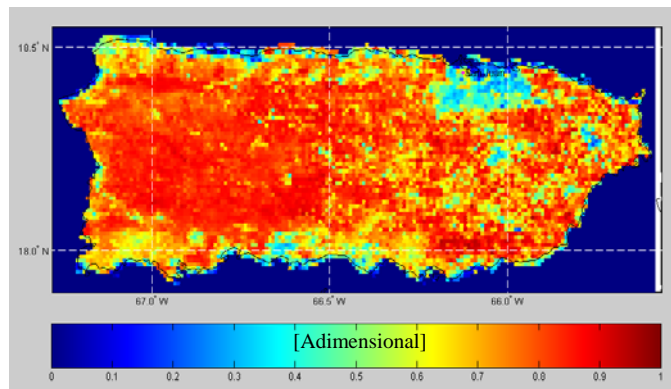


Figura 3.4 NDVI, usando datos de MODIS para agosto 2006.

Datos de Lluvia de Radar.

Datos de lluvia de radar son colectados a través de *Multisensor Precipitation Estimation* (MPE). MPE es un algoritmo que integra datos de radar y *rain gauge* con el fin de obtener datos estimados de lluvia (Marzen 2004). Estos datos están dados en forma horaria, en un formato de archivo llamado XMRG con un tamaño de grid de aproximadamente 4 x 4 Km². El radar *Next Generation Radar* (NEXRAD), localizado a $Lat=18.1156^{\circ}$ y $Lon=-66.0781^{\circ}$ (Cayey, Puerto Rico) provee esta información; la que a su vez es colectada a través del *National Weather Service* (para mayor información diríjase a la pagina Web

http://dipper.nws.noaa.gov/hdsb/data/nexrad/sjrfc_mpe.php). Fue necesario realizar una interpolación con KNN para llevar estos datos a una resolución de 1Km, ver Figura 3.5. Se conocen algunas dificultades asociadas con la estimación del radar (Hunter 1996, Fulton et al. 1998, y Westrick et al. 1999). Estudios realizados por Gerard et al (1999), para comparar ambas fuentes muestran que puede existir sobre estimación o subestimación de los datos por parte del radar. Altas y bajas reflectividades causadas por las regiones montañosas son las causas principales de estos problemas (Klazura et al. 1999). Sin embargo, para propósitos de esta investigación, se asume que los datos de NEXRAD son lo suficientemente precisos para la estimación de la humedad de suelo. Se están realizando estudios para verificar esta hipótesis (Ramírez et al, 2007b).

3.3.1.3 Otras Fuentes de Datos.

Datos de Características Texturales.

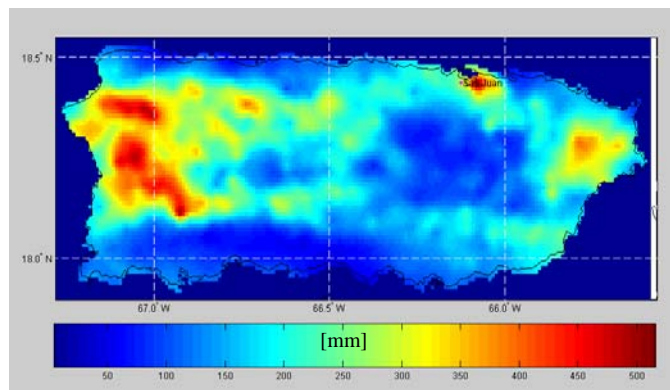


Figura 3.5 Lluvia de radar, usando datos de NEXRAD para agosto 2006.

Se usaron los resultados de 118 perforaciones distribuidas de manera irregular sobre toda la isla de Puerto Rico con el fin de crear mapas de textura de suelo con 1Km² de resolución, ver Figura 3.6. La información sobre la textura del suelo a una profundidad de 20 cm. fue

obtenida del *United States Department of Agriculture* (USDA) (para mayor información ver <http://www.usda.gov/wps/portal/usdahome>).

Datos de Características Topográficas.

Se le ha llamado características topográficas a la altura en metros sobre el nivel del mar. Estos datos fueron obtenidos del *Digital Elevation Model* (DEM) a través del *United States Geological Survey* (USGS) (para mayor información ver <http://www.usgs.gov/>).

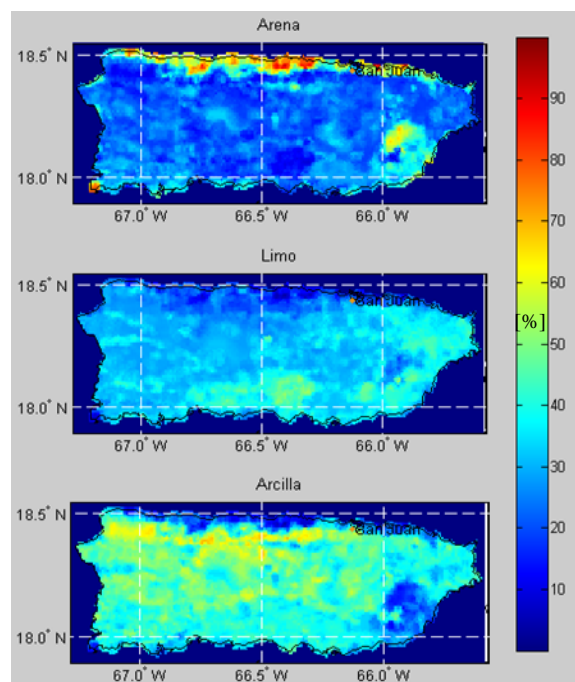


Figura 3.6 Características texturales para Puerto Rico.

Este modelo consiste en un mapa de 30 m. de resolución y fue interpolado usando la técnica de interpolación espacial llamada KNN para llevarlo a 1Km² de resolución espacial. Los resultados de dicha interpolación pueden observarse en la Figura 3.7.

3.3.2 Base de Datos.

Se diseñó una base de datos para almacenar las observaciones, e información relacionada con el presente estudio. Se cuenta con datos de humedad de suelo, lluvia y temperatura del aire desde el primero de enero del año 2005 hasta la fecha, provenientes de las estaciones de medición pertenecientes al proyecto EHS.

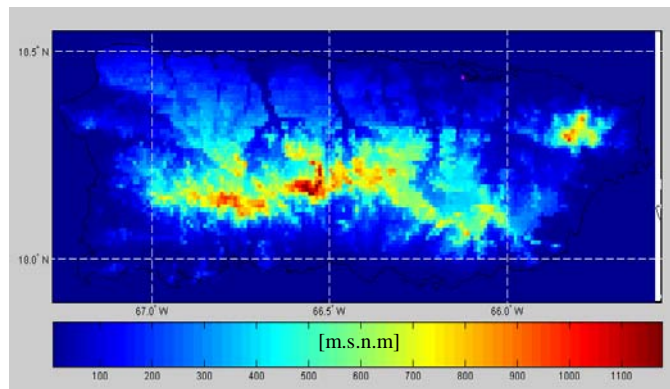


Figura 3.7 Características topográficas (alturas) para Puerto Rico.

La NRCS provee datos históricos de las mismas variables desde el 2001, aunque no todas las estaciones tienen el mismo tiempo en funcionamiento; sin embargo, se ha tomado la misma fecha de inicio para el almacenamiento de todos los datos. Los datos de sensores remotos se han almacenado de forma mensual y horaria y por ultimo, los datos de las características texturales y topográficas corresponden a archivos individuales. La principal ventaja de tener una base de datos, es que la información a cerca de las variables bajo estudio puede ser obtenida en cualquier momento y usada de acuerdo con las necesidades. La estructura de la base de datos se muestra en la Figura 3.8. En dicha figura, el campo llamado estaciones de EHS se divide en 12 archivos, uno por cada estación; al igual, el de estaciones de NRCS esta dividido en 3 archivos, cada uno de esos archivos tiene la forma interna que se describe al

lado derecho de la figura. En los campos llamados datos de sensores remotos y características texturales, los archivos tienen prácticamente la misma forma, con la diferencia de que los datos de sensores remotos están organizados de manera mensual y horaria, mientras que los de textura y topografía no varían en el tiempo.

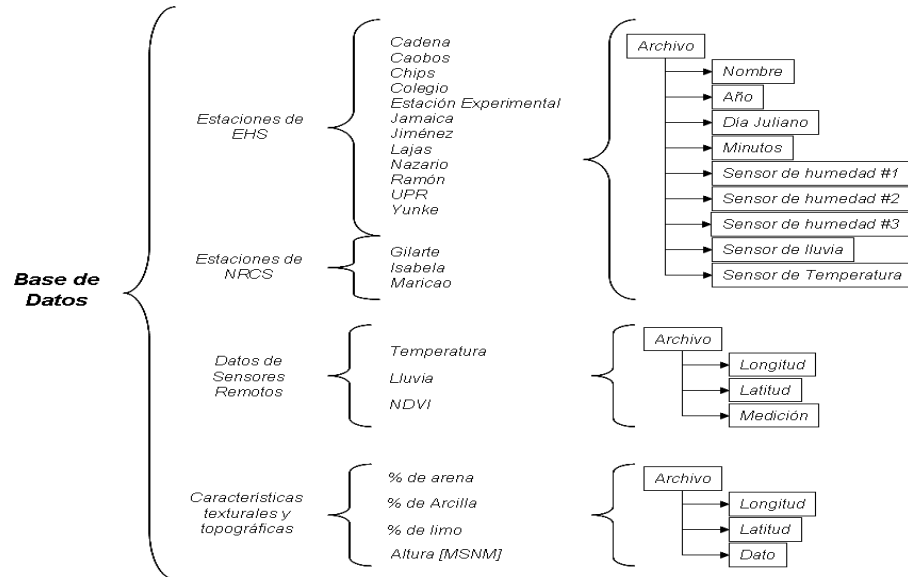


Figura 3.8 Estructura de la base de datos.

A partir de esta base de datos, se construyeron matrices de datos destinadas a la obtención de modelos de humedad promedio. Estas matrices están compuestas de datos de localización geográfica, promedios de humedad de suelo, características texturales, amplitud mensual de temperatura, NDVI, altura y milímetros de lluvia acumulados durante un mes. Un ejemplo de estos arreglos matriciales puede verse en la Tabla 3.1. Las dos primeras columnas de datos se refieren a la ubicación geográfica de las estaciones, la tercera a la humedad de suelo, la cuarta, quinta y sexta se refieren a las características texturales en valores porcentuales, la séptima se refiere a el rango de temperatura, la octava al NDVI, el cual es un parámetro

adimensional, la novena a la altura sobre el nivel del mar y la última columna a la precipitación.

Tabla 3.1 Ejemplo de arreglo matricial para modelaje.

Estación	Latitud	Longitud	HS [%]	Arcilla [%]	Limo [%]	Arena [%]	Rango de Temp. [°C]	NDVI	Altura[m]	Lluvia [mm]
Caobos	18.06	-67.05	33.54	39.75	30.56	29.69	9.23	0.83	83.36	78.00
Colegio	18.21	-67.14	21.82	30.70	8.08	61.22	19.03	0.77	9.25	238.00
Edwin	18.26	-67.08	32.29	49.01	22.28	28.72	9.75	0.86	110.48	480.00
E. Exp	18.03	-67.07	29.64	44.40	32.50	23.10	9.42	0.65	38.86	56.00
Gallo	18.03	-67.03	7.74	45.20	20.36	34.43	16.04	0.75	20.03	39.00
Nazario	18.24	-67.13	36.62	45.63	29.09	25.29	11.31	0.80	34.73	207.00
Ramon	18.33	-67.13	36.09	46.88	18.19	34.94	14.53	0.76	118.46	331.58
Isabela	18.47	-67.06	30.27	46.87	10.10	43.03	9.49	0.72	168.63	127.00
Maricao	18.16	-67.00	33.42	40.92	20.26	38.82	10.66	0.86	599.34	167.89

3.4 Descripción del Método de RNA.

Redes neuronales artificiales fueron creadas y entrenadas, con el fin de establecer modelos empíricos tanto para la estimación de la humedad de suelo promedio mensual, como para encontrar las semejanzas climáticas entre los diferentes píxeles (1 Km²) de Puerto Rico.

3.4.1 RNA Supervisada.

3.4.1.1 RNA para la estimación de la humedad de suelo promedio mensual.

Una breve descripción de lo que es una red neuronal con entrenamiento supervisado, fue realizada en la Sección 2.2.2.1. Promedios mensuales de humedad de suelo pueden ser estimados usando un modelo empírico que exprese la relación no lineal entre la humedad de suelo y otras variables que se conoce influyen en su comportamiento tanto temporal como espacial. El modelo empírico, estima en forma mensual la humedad de suelo 20 cm. profundidad. Las observaciones en las estaciones de humedad revelan que datos de características texturales del suelo, topográficos, de lluvia, temperatura del aire y NDVI pueden ser usados para estimar promedios mensuales de humedad de suelo. Un ejemplo de

los datos usados para la estimación de los promedios mensuales de humedad de suelo puede ser visto en la Tabla 3.1.

Una RNA *supervisada* fue creada y entrenada para estimar los promedios mensuales de humedad de suelo. Se selecciono una RNA porque la dinámica del contenido de agua en el suelo es un proceso no lineal y se ha demostrado que las RNA son un mecanismo eficiente para realizar aproximaciones o estimaciones, cuando las variables del sistema son altamente no lineales. En este trabajo se siguió el procedimiento general propuesto por Ramírez y Montes (2002) como se muestra a continuación: a) Establecer las variables a considerar para el modelaje. Se ha seleccionado un pequeño número de variables que están relacionadas con el comportamiento de la humedad de suelo, las cuales ya han sido descritas en la Sección 3.3.1; b) Seleccionar el algoritmo de entrenamiento. Se ha seleccionado una variación del algoritmo de BP llamada *Levenberg-Marquardt* (LM). Este es un algoritmo de optimización diseñado para minimizar funciones que son la suma de los errores cuadrados de otras funciones no lineales (Hagan et al 1996). Una de las limitaciones de este método es la gran cantidad de memoria computacional requerida para realizar los cálculos. Sin embargo, la aplicación del algoritmo de LM es la técnica mas conveniente cuando se usan pequeñas cantidades de datos para el entrenamiento de una red neuronal (Hagan 1996); c) Identificar la estructura de la red. En este paso se identifica el número de neuronas en la o las capa oculta y la función de transferencia a utilizar. Este es un proceso experimental, en el que se escoge la mejor estructura dependiendo de sus resultados. Se encontró por medio de la experimentación que la estructura que arroja los mejores resultados es la que tiene una única capa oculta y cuatro neuronas en dicha capa. La función de transferencia Sigmoidal fue

escogido a priori, basado en el hecho de que dicha función es continua, diferenciable y su rango va desde -1 hasta 1; d) Búsqueda exploratoria del punto inicial optimo. En este paso se mantiene fija la estructura de la red y se hace una búsqueda exploratoria del punto inicial. Para esto se usó un algoritmo de selección, basado en el criterio del mínimo error cuadrado. Esta técnica consiste en presentar veinte series aleatorias de valores iniciales de los pesos, los cuales optimiza el algoritmo de búsqueda de acuerdo a la información de los errores; e) Entrenar la RNA. La optimización en el paso anterior conlleva a obtener igual número de modelos probables, de los cuales se selecciona el más acertado de acuerdo a cierto criterio de minimización del error que en este caso es el MSE. Un modelo de RNA supervisada esta caracterizado por la recepción de datos de entrada para generar la identificación de un modelo. Los patrones de entrenamiento de la RNA pueden ser presentados como sigue:

$$P_{i,j} = [c_j, s_j, rt_{i,j}, r_{i,j}, v_{i,j}] \quad \wedge \quad T_{i,j} = [hm_{i,j}] \quad 3.3$$

Esta forma de presentar los patrones de entrenamiento fue introducida por Ramírez y Montes (2002). Donde, $P_{i,j}$: es el vector de patrones presentados a la red procedentes del mes i y la estación j ; c_j : es el porcentaje de arcilla en la estación j ; s_j : es el porcentaje de arena en la estación j ; $rt_{i,j}$: es el rango mensual de temperatura (Ecuación 3.1) para el mes i y la estación j ; $r_{i,j}$: es la lluvia acumulada para el mes i y la estación j ; $v_{i,j}$: es el índice de vegetación NDVI para el mes i y la estación j ; $T_{i,j}$: es el vector de patrones de salida o referencia; $hm_{i,j}$: es la humedad de suelo para el mes i y la estación j ; f) Aplicación, validación y corrección de los resultados. La aplicación consiste en generalizar los resultados obtenidos de la red neuronal. Es decir, se aplica la red obtenida a puntos diferentes de los puntos de

entrenamiento. Esto es generalizando los resultados de la red en cualquier mes, para todos los puntos de Puerto Rico usando como datos de entrada, la información de sensores remotos, y características texturales y topográficas. En la fase de verificación, se comprueba que los valores generados por la red sean acordes con la realidad, puesto que la red puede cometer errores en la generalización. En general, la humedad de suelo no sobrepasa el 50% en volumen. Sin embargo, existen ciertos valores típicos que dependen de las características texturales del suelo, a saber: *wilting point* (WP), *field capacity* (FC) y *saturation point* (SP). Estos valores corresponden al mínimo contenido de agua en el suelo, la capacidad de campo y el punto de saturación respectivamente. Cemek et al (2004), presenta en su trabajo la estimación de dichos parámetros usando funciones de pedotransferencia. Los resultados de la generalización o aplicación de la red neuronal sobre Puerto Rico se han corregido o acotado en su valor mínimo con el WP y en su valor máximo con el promedio entre el FC y el SP. Para encontrar los valores de WP, FC y SP se han usado las ecuaciones proveídas por Cemek et al (2004).

En este trabajo se tomaron cinco RNA's para cada mes. La idea es usar una medida de tendencia central, que en este caso es la mediana para obtener el estimado final. Esto quiere decir que para la estimación final del promedio mensual de un mes en particular es necesario obtener cinco RNA's. El resultado final o estimado del promedio mensual de la humedad de suelo, será la mediana de la información de salida de las cinco RNA's obtenidas. Se tomó en cuenta la mediana, puesto que de las medidas de tendencia central es la que mayor estabilidad ofrece frente a variabilidad extrema.

3.4.1.2 RNA para la estimación de la tendencia de la humedad de suelo.

La tendencia de la humedad de suelo puede ser estimada usando un modelo empírico que exprese la relación lineal entre la componente de tendencia y otras variables que se conoce influyen en su comportamiento tanto temporal como espacial. El modelos empíricos desarrollado, estima la tendencia de la humedad de suelo para cualquier punto conocidas las siguientes variables: el porcentaje de arcilla, el porcentaje de arena y la altura del punto. Es necesario también utilizar el tiempo en horas como una variable. Luego el modelo encontrará la tendencia de la humedad en cualquier momento del tiempo, en mediciones horarias, durante un mes en particular. La salida de la red, estima la función lineal que ajusta los datos de humedad de suelo, es decir ajusta una recta de tendencia a los datos de humedad suministrados, lo cual se ha tomado como estimación de la tendencia de la humedad de suelo. La red escogida para la estimación de dichos modelos tiene las mismas propiedades que la usada para la estimación de los promedios mensuales de humedad. Una RNA *supervisada* fue creada y entrenada para modelar la tendencia de la humedad de suelo. La principal diferencia con la red de la Sección anterior es que la función de activación en este caso es una función de activación lineal. Experimentalmente se encontró que el número óptimo de neuronas es para estimar tendencias es 4 y que una sola capa oculta basta para obtener el modelo. El algoritmo de entrenamiento es el mismo que el usado en la estimación de los promedios de humedad. Los vectores de entrada y salida de la red se presentan a continuación:

$$P_i = \left[\frac{H_i}{1200}, \frac{t}{1000}, Clay_i, Sand_i, cR_{j-1} \right] \quad \wedge \quad T_i = [(T_H)_i] \quad 3.4$$

Donde, P_i : es el vector de patrones presentados a la red procedentes del sitio i ; H_i : es la altura en el sitio i ; cR_{j-1} : es la lluvia acumulada durante el mes anterior ($j-1$); t : representa el tiempo en horas; $Clay_i$ y $Sand_i$: representan los porcentajes de arcilla y arena respectivamente, para el sitio i ; T_i : es el vector de patrones de salida o referencia; $(T_H)_i$: es la humedad de suelo sin su promedio para el sitio i . Se generalizó los resultados de la red para cualquier punto de Puerto Rico usando como datos de entrada, información topográfica, de características texturales y de lluvia acumulada durante el mes anterior. En la fase de validación, se comprobó que los valores generados por la red fueran acordes con la realidad, puesto que la red puede cometer errores en la generalización. Se puede asumir sin pérdida de generalidad que los valores de la humedad de suelo no sobrepasan el 50%. No se realizó corrección alguna de los datos, puesto que los valores arrojados por la red neuronal concuerdan con los valores típicos.

3.4.1.3 RNA para la estimación de la tendencia de la temperatura.

La tendencia de la temperatura puede ser estimada usando un modelo empírico que exprese la relación lineal entre la componente de tendencia y otras variables que se conoce influyen en su comportamiento tanto temporal como espacial. El modelo empírico desarrollado, estima la tendencia de la temperatura para cualquier punto conocidas las siguientes variables: la altura del punto, puesto que esta variable tiene una alta correlación con el rango de la temperatura y por ende con los valores máximos y mínimos de temperatura (Geerts B., 2002); también es necesario utilizar el tiempo como una variable. El modelo puede encontrar la tendencia de la temperatura en cualquier momento del tiempo, en mediciones horarias, durante un mes en particular. La red escogida para este propósito tiene las mismas

propiedades que la usada para la estimación de la tendencia de la humedad de suelo. Los patrones de entrenamiento de la RNA pueden ser presentados como sigue:

$$P_i = \left[\frac{H_i}{1200}, \frac{t}{1000} \right] \quad \wedge \quad T_i = \left[\frac{(T_T)_i}{100} \right] \quad 3.5$$

Donde, P_i : es el vector de patrones presentados a la red procedentes del sitio i ; H_i : es la altura en el sitio i ; t : representa el tiempo en horas; T_i : es el vector de patrones de salida o referencia; $(T_T)_i$: es el promedio entre el valor máximo y mínimo de la temperatura en el punto i . Se generalizó los resultados de la red para todo Puerto Rico usando información topográfica en el caso de la temperatura. En la fase de validación, se comprobó que los valores generados por la red fueran acordes con la realidad, puesto que la red podría cometer errores en la generalización. La temperatura del aire no posee valores típicos, aunque se puede decir sin pérdida de generalidad que raramente sobrepasa los 40 °C. No se realizó corrección alguna de los datos, puesto que los valores arrojados por la red neuronal concuerdan con los valores usuales.

3.4.2 RNA No Supervisada.

Una breve descripción de lo que es una red neuronal con entrenamiento no supervisado, fue realizada en la Sección 2.2.2.1. Una red neuronal competitiva es usada para identificar regiones análogas en la isla de Puerto Rico. Donde análogo esta definido como las regiones cuyo comportamiento climatológico presenta la mayor similaridad con el punto de muestra, que en este caso es cada una de las estaciones (para una descripción de la red neuronal competitiva ver anexo A2). Hagan et al, (1996) presenta una descripción muy precisa sobre una red neuronal competitiva y Veneros (2004) realiza una aplicación de esta red. La

necesidad de entrenar una red competitiva en este trabajo se debe a que se requiere obtener asociaciones entre los datos de cada uno de los píxeles de Puerto Rico. En otras palabras obtener una clasificación de las zonas con similitudes climatológicas. Esto es necesario puesto que se aplicarán los resultados obtenidos para cada una de las estaciones en las regiones a las cuales pertenezca una estación determinada. La metodología seguida para crear y entrenar la red neuronal competitiva es como sigue: a) Se escogieron cinco variables representativas de la climatología regional, a saber: índices de vegetación, lluvia acumulada mensual, temperatura promedio mensual de día, temperatura promedio mensual de noche y altura sobre el nivel del mar; b) Estas cinco variables componen la información que se provee a la red neuronal competitiva para que realice la clasificación. Cada una de estas variables contiene 8701 observaciones, una por cada píxel de Puerto Rico. Cabe notar que Puerto Rico tiene una extensión de 8701 Km², por lo cual, cada píxel representa un Km² de área dentro de la isla; c) Se seleccionó un número probable de clases igual a veinte (20). Este número fue seleccionado de manera arbitraria. Sin embargo, la red tiene la capacidad de reconocer la cantidad de clases necesaria, si esta fuera inferior a la cantidad inicial de clases. Si el número fuese mayor, la red tendría problemas para obtener la clasificación (Hilera y Martines 1995); d) El algoritmo de entrenamiento usado es conocido como la regla de Kohonen, el cual se basa en el ajuste de los pesos presentes, a partir del estado de activación anterior. Por lo tanto, las neuronas aprenden actualizando sus pesos a partir de las comparaciones de los vectores prototipo con los vectores de entrada mediante la minimización de la distancia entre dos vectores. Mientras más veces sea presentado el patrón de entrada, mejor será la clasificación (Ochoa 2004 y López 2005), aunque el tiempo computacional es una limitante

fuerte. En este trabajo se ha usado un número de épocas o iteraciones durante la fase de entrenamiento de 10.000 (López 2005).

3.5 Descripción del MRLM.

3.5.1 Identificación del modelo de estimación de temperatura horaria.

La temperatura en las capas cercanas a la superficie de la tierra es un factor importante en el intercambio de la energía entre la tierra y la atmósfera (Sun y Pinker, 2003). Datos experimentales demuestran que la temperatura muestra diferentes periodos acoplados. Estos periodos pueden ser anuales, mensuales, diarios e incluso horarios (Martín 1995). El modelo propuesto aquí aprovecha la relación lineal entre la oscilación diaria de la temperatura y las funciones trigonométricas del seno y coseno. La propiedad fundamental explotada aquí es la periodicidad de dichas funciones y la similaridad de los datos de temperatura con estas funciones periódicas. La formulación del modelo esta basada principalmente en el análisis de los datos de temperatura de las diferentes estaciones. La evaluación del modelo requiere del uso del modelo hallado anteriormente con RNA's para la tendencia de la temperatura y de datos de sensores remotos, además de datos topográficos o de alturas sobre el nivel del mar, puesto que la idea principal es aplicar los modelos desarrollados para estimar la temperatura horaria en toda la isla de Puerto Rico. Resultados experimentales muestran que el comportamiento de la temperatura durante el día y la noche, esta influenciado fuertemente por sus valores máximos y mínimos, además se ha encontrado que la temperatura horaria presenta diferentes periodos, a saber cada 6, 8, 12 y 24 horas, como se puede observar el siguiente periodograma:

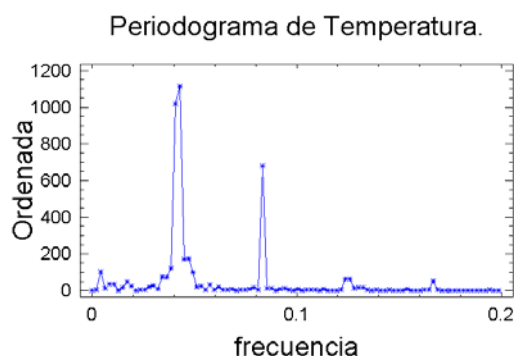


Figura 3.9 Periodograma para la temperatura en la estación Jamaica (Agosto 2006).

Para mayor claridad en la siguiente tabla se muestran los valores de la ordenada correspondientes a cada frecuencia. Cabe anotar que la frecuencia es el inverso del periodo.

La Tabla 3.2, muestra que los periodos 23.4 y 24.6316 son los más predominantes. Es claro que una interpolación entre estos dos nos entregaría el periodo de 24 horas, por lo que se obtiene así uno de los periodos en consideración, mientras que de los otros tres, los de 12 y 6 horas son claramente extraídos y el de 8.0689 es un valor bastante cercano a 8 horas.

Tabla 3.2 Datos provenientes del periodograma.

Frecuencia	Periodo	Ordenada
0.0405	24.6316	1017.63
0.0427	23.4	1111.32
0.8333	12	680
0.1239	8.0689	62.3954
0.1666	6	53.7974

Modelos de regresión lineal múltiple fueron generados para estimar el comportamiento horario de la temperatura del aire en el área de Puerto Rico, partiendo de la información conocida de sus máximos y mínimos diarios. Para estimar la temperatura del aire, en cualquier hora del día para toda la isla de Puerto Rico; se propone un modelo empírico – matemático basado en las siguientes condiciones.

- El patrón diario de temperaturas en cualquier punto se divide en un periodo de calentamiento y un periodo de enfriamiento, con temperaturas mínimas T_{Min} que generalmente se encuentran entre las 6:00 AM y las 8:00AM y temperaturas máximas T_{Max} que pueden darse entre las 11:00 AM y las 3:00 PM. Como puede observarse en la siguiente figura:

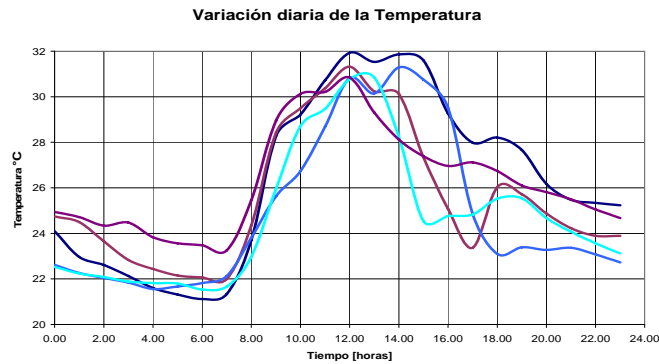


Figura 3.10 Variación diaria de la Temperatura (estación Colegio, Agosto 2006)

- La temperatura presenta tres componentes principales, a saber: a) componente periódico, el cual es debido a los periodos de calentamiento solares; b) componente estocástico debido a factores inherentes a cada punto de observación; c) componente de tendencia y persistencia, debido a cambios climáticos de largo plazo y a factores inherentes de cada punto de medición, como la altura o la posición geográfica.
- La periodicidad diaria de la temperatura y su amplitud, se ajustan a un modelo de regresión múltiple, con funciones periódicas sinusoidales como se ilustra a continuación:

$$\hat{T}_p = \sum_{i=1}^4 \left[a_i A \cdot \text{Sen} \left(\frac{2t\pi}{P_i} + \phi \right) + b_i A \cdot \text{Cos} \left(\frac{2t\pi}{P_i} + \phi \right) \right] + c, \quad P = [6, 8, 12, 24] \quad \mathbf{3.6}$$

Donde: \hat{T}_p : es la parte periódica estimada de la temperatura; a_i , b_i y c : son los coeficientes de la regresión; A : es la amplitud de temperatura diaria se define como la mitad del rango definido en la Ecuación 3.1; ϕ : es el ángulo de fase; P : es un vector de cuatro entradas, cada una de las cuales representa un periodo diferente.

- La componente de tendencia se modela por medio de una red neuronal con la altura y el tiempo como entradas y como salida, la función de tendencia \hat{T}_T , como se indicó en la Sección 3.4.

3.5.2 Estimación de parámetros.

Un modelo de regresión lineal múltiple fue identificado para cada estación, obteniendo los parámetros como se indicó en la Sección 2.1. En general, la metodología seguida para la obtención del modelo se realizó en los siguientes pasos:

- Se dividieron los datos provenientes de cada estación en dos partes, una para la construcción del modelo y otra para su validación; se extrajeron los valores máximos y mínimos de cada 24 horas. Estos valores componen un vector de temperaturas máximas y mínimas.
- Con el vector obtenido, se calculó el valor de la amplitud diaria de temperatura con la Ecuación 3.1. Con estos valores se crearon 8 vectores de datos, cuatro de ellos con la función seno y los restantes con la función coseno. Cada par de vectores seno y coseno tienen asociado un periodo correspondiente a los ya discutidos anteriormente. Además, estos vectores se obtuvieron en forma horaria usando las siguientes ecuaciones:

$$\frac{A}{2} \text{Sen}\left(\frac{2t\pi}{P_i} + \phi\right) \quad 3.7$$

y

$$\frac{A}{2} \text{Cos}\left(\frac{2t\pi}{P_i} + \phi\right) \quad 3.8$$

- El valor del ángulo de fase ϕ fue encontrado de acuerdo a un algoritmo de minimización del MSE.
- Luego de la obtención del ángulo de fase, se procede con la obtención de los parámetros del modelo dado en la Ecuación 3.6. Para esto, se le retiró la tendencia a los valores de temperatura de la estación y se entra el resultado (componente periódico y estocástico) como variable dependiente del modelo. Los vectores descritos anteriormente como de seno y coseno, son entrados como variables independientes y se procedió a la estimación de los parámetros de la regresión por medio del método de los mínimos cuadrados (Montgomery y Peck, 2001) y un algoritmo de minimización del MSE, mediante el análisis de regresión.
- En este punto, es necesario realizar una validación, para verificar el comportamiento del modelo, la cual fue hecha usando datos que el modelo no conocía y evaluando las ecuaciones obtenidas. Aquí, se usaron los datos de temperatura incluyendo su tendencia junto con el modelo empírico hallado en la Sección 3.4.

En resumen, el modelo propuesto para la estimación de la temperatura horaria a partir de los valores máximos y mínimos diarios, es el resultado de sumar el componente periódico \hat{T}_p y el componente de tendencia \hat{T}_r como se muestra en la siguiente ecuación:

$$\hat{T} = \hat{T}_p + \hat{T}_T \quad 3.9$$

3.5.3 Evaluación del MRLM.

Una vez se han estimado los parámetros de los modelos, se procedió a su aplicación. Como se dijo en la Sección 3.5.1, la aplicación de los modelos de estimación de temperatura horaria requiere del uso de datos colectados por sensores remotos. En este caso se usaron datos de temperatura superficial diaria del sensor MODIS y datos topográficos de alturas para Puerto Rico provenientes de USGS, puesto que estos datos cubren toda la extensión de Puerto Rico. Para la generalización de los modelos propuestos, es necesario usar una red neuronal artificial autoorganizada o competitiva (ver Sección 3.4.2), para identificar los puntos que exhiben características climatológicas similares a las de las estaciones de medición, puesto que a cada una de ellas le corresponde un modelo. Después de ubicar los puntos con similitudes a cada estación, se aplican los modelos correspondientes para cada región, haciendo uso de la suposición expresada en la Sección 3.6.2. El componente determinístico del modelo expresado por la tendencia, fue determinado usando RNA, donde las entradas de la red corresponden a la altura sobre el nivel del mar, y el tiempo, mientras que la salida es el componente de tendencia de la temperatura del aire para cada una de las 15 estaciones. La creación de dicha RNA fue expuesta en la Sección 3.4.1.2.

3.5.4 Verificación y Corrección del Modelo de Temperatura.

Encontrados los valores de temperatura del aire para cada grid, se procedió a verificar que estos valores se ajusten a la realidad, es decir, que no excedan ciertos valores límites históricos. Para el caso de la temperatura se cuenta promedios históricos de temperaturas

máximas y mínimas provenientes de las estaciones COOP, ver la pagina Web <http://www.nws.noaa.gov/om/coop/>. Se utilizó la técnica llamada validación cruzada que consiste en utilizar una parte de los datos disponibles para el ajuste del modelo y los datos restantes para realizar la validación. Las estimaciones finales los modelos de temperatura del aire son temperatura horaria a 1Km de resolución espacial.

3.6 Descripción del Método de FT

3.6.1 Identificación del modelo de función de transferencia.

Se ha encontrado que la precipitación y la temperatura del aire están altamente correlacionadas con la humedad de suelo (Jiang y Cotton 2004). El modelo propuesto aquí esta basado principalmente en dicha observación, la cual ha sido confirmada con medidas de campo. El modelo ha sido construido usando datos de temperatura y lluvia y su aplicación requerirá de información de sensores remotos. La formulación del modelo esta basada principalmente en el análisis de las observaciones de humedad de suelo. Resultados experimentales muestran que la humedad decrece exponencialmente luego de un evento de lluvia, aunque para eventos consecutivos de lluvia, se comporta linealmente. Durante la época seca, la humedad se ve afectada por los cambios periódicos de la temperatura y además sus pendientes son inversamente proporcionales. Partiendo del modelo del ciclo hidrológico, Huang et al (1978), dicen que la humedad de suelo puede ser estimada usando solo datos de temperatura y precipitación. Esta afirmación sugiere que la dinámica de la humedad de suelo puede ser expresada como una ecuación diferencial parcial de primer orden con coeficientes constantes como sigue:

$$\frac{\partial h(t)}{\partial t} = k_1 R(t) + k_2 T(t) \quad 3.10$$

En este punto, es necesario agregar un término a la ecuación anterior. Este término será llamado nivel de humedad h , es decir, que la magnitud de la variación de la humedad esta relacionado intrínsecamente con el nivel de humedad o estado del sistema. Luego, entonces la ecuación 3.10 al agregarle este nuevo término se convierte en:

$$\frac{\partial h(t)}{\partial t} = k_1 R(t) + k_2 T(t) + k_3 h(t) \quad 3.11$$

Donde: k_1 , k_2 y k_3 : son las constantes usadas para calibrar el modelo; $\frac{\partial h(t)}{\partial t}$: es la variación de la humedad de suelo con respecto al tiempo t ; $R(t)$, $T(t)$ y $h(t)$: son la lluvia instantánea, la temperatura instantánea en el instante y la humedad de suelo en el tiempo t , respectivamente. En las ecuaciones anteriores, se ha considerado que $h(t)$, $R(t)$ y $T(t)$ son funciones continuas, pero en la practica no lo son, puesto que solo se dispone de datos discretos h_t , R_t y T_t . Pandit y Wu (1983) demostraron que una ecuación diferencial de primer orden puede ser escrita como una ecuación en diferencias de primer orden. Y esto se puede generalizar a ecuaciones diferenciales parciales de primer orden. Por lo cual, para representar la última ecuación por medio de una ecuación en diferencias parciales de primer orden con coeficientes constantes se usa la siguiente ecuación (Pandit y Wu, 1983; Box y Jenkins, 1976; Hamilton, 1994; Brockwell y Davis, 2002; Wei, 1990):

$$h_t = \phi h_{t-1} + \theta R_t + \varphi T_t \quad 3.12$$

La forma de la Ecuación 3.12 sugiere la utilización de la función de transferencia, la cual podría proveer una mejor representación del comportamiento de la humedad de suelo. La relación entre las funciones de transferencia discreta y continua también fue discutida por

Pandit y Wu (1983) y Box y Jenkins (1976). Una descripción detallada de la función de transferencia se presentó en la Sección 2.3. De la experiencia obtenida en el manejo de los datos de humedad de suelo, lluvia y temperatura, es posible realizar la siguiente descripción (ver Figura 3.11):

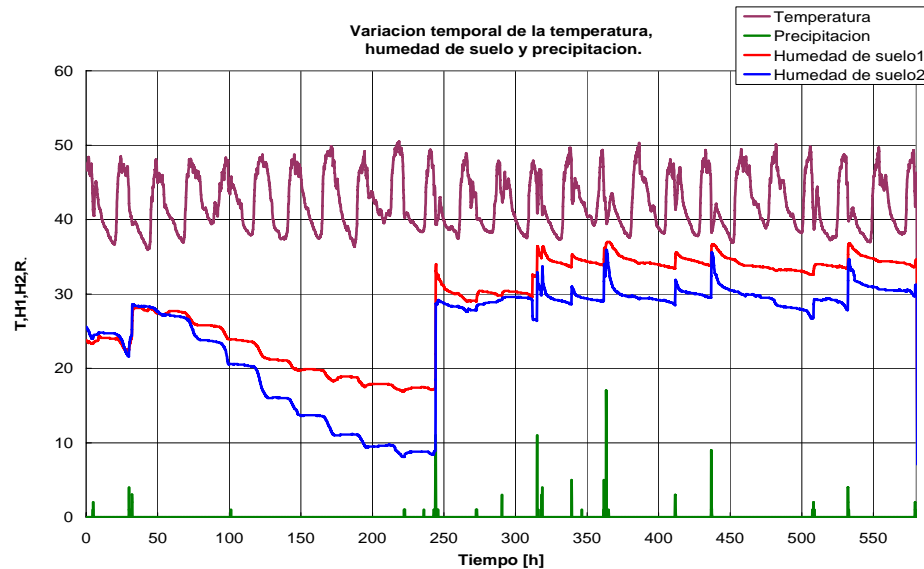


Figura 3.11 Variación temporal de la temperatura, humedad de suelo y precipitación. En el eje vertical las unidades son [%] para la humedad de suelo, [mm] par la lluvia y [°C] para la temperatura. A esta última se le sumaron 15 °C con el fin de que su grafica no interfiriera con las demás variables, esto por razones explicativas.

El nivel de la humedad de suelo, principalmente es controlado por los valores antecedentes o históricos. El comportamiento de la humedad de suelo durante una temporada sin lluvia o seca es controlado principalmente por la temperatura como puede observarse en la Figura 3.11, mientras que los cambios instantáneos de la humedad son controlados por los eventos de lluvia. En los periodos de lluvias consecutivas, la humedad puede llegar hasta un valor máximo, el cual es llamado punto de saturación, del cual desciende rápidamente para establecerse en un punto de equilibrio llamado capacidad de campo (*field capacity*). En la

Figura 3.11 puede verse claramente las observaciones hechas anteriormente. De las observaciones anteriores a cerca del comportamiento de los datos, es posible afirmar que los efectos de la lluvia y de la temperatura en la humedad de suelo pueden ser representados por funciones de respuesta al impulso. La función de respuesta al impulso para la lluvia debe ser modulada mediante un término exponencial asociado con la lluvia acumulada durante las ultimas L horas r_L . Este termino ajusta el valor de la lluvia de la siguiente manera: a) si no llueve durante los días previos su valor debe ser cero ($r_L = 0$) y la respuesta en la humedad de suelo seria completa; b) si durante los días previos ocurren eventos de lluvia, r_L debe ser grande y la respuesta en la humedad de suelo se verá atenuada por el termino exponencial. Partiendo de la descripción anterior, se propone el siguiente modelo no lineal de FT estocástica para la estimación de la humedad de suelo en forma horaria:

$$h_t = \left(\frac{\omega_{0,1} - \omega_{0,2}B}{1 - \delta_{0,1}B} \right) T_t + \left(\frac{\omega_{1,1} - \omega_{1,2}B}{1 - \delta_{1,1}B} \right) R_t \exp(-\tau_1 r_L) + \left(\frac{\omega_{2,1} - \omega_{2,2}B}{1 - \delta_{2,1}B} \right) h_{t-1} + a_t \quad \mathbf{3.13}$$

Donde: R_t : es la lluvia instantánea en el tiempo t ; T_t : es la temperatura en el tiempo t ; h_{t-1} : es la humedad de suelo en el tiempo $t-1$; r_L : es la lluvia acumulada durante las ultimas L horas; la variable entera L se obtiene durante el proceso de estimación de parámetros; ω 's y δ 's: son los parámetros de las funciones de respuesta al impulso; τ : es un parámetro de atenuación que modula la función de respuesta al impulso asociada con la lluvia.

3.6.2 Estimación de parámetros.

Un modelo de función de transferencia fue identificado para cada estación asumiendo que los coeficientes de la función de respuesta al impulso caracterizan las propiedades climatológicas

del sitio de la estación. Así, se asume que ω 's, δ 's, τ , y r_L son coeficientes que exhiben características climatológicas inherentes de un sitio determinado y consecuentemente la variabilidad espacial es expresada por los coeficientes de las funciones de respuesta al impulso. Así, evaluando la FT en sitios diferentes que exhiben características climatológicas similares a las de las estaciones de medición, se puede estimar la dinámica de la humedad de suelo. En general para estimar los parámetros de la FT se usaron los pasos descritos en la seccion2.3.3. Sin embargo, a continuación se describirá en detalle el procedimiento para estimar los parámetros de las funciones FT (Ramírez et al 2007b):

- Búsqueda aleatoria para determinar el punto inicial de la FT estocástica no lineal. Se usó la distribución de probabilidad uniforme para generar 100 puntos en un rango específico y se identificó el punto inicial mas adecuado por medio de la minimización del promedio del cuadrado de los errores mas conocido por sus siglas en ingles como MSE. Se estableció empíricamente un rango para los parámetros, el cual se muestra a continuación:

$$Min. = [-2, -2, -0.9, -2, -2, -0.9, -2, -2, -0.9, 0, 12]$$

$$Máx. = [2, 2, 0.9, 2, 2, 0.9, 2, 2, 0.9, 1, 96]$$

Donde, Min. y Máx. representan el valor mínimo y máximo respectivamente que puede asignarse a los parámetros organizados en el siguiente vector:

$$parametros = [\omega_{0,1} \quad \omega_{0,2} \quad \delta_{0,1} \quad \omega_{1,1} \quad \omega_{1,2} \quad \delta_{1,1} \quad \omega_{2,1} \quad \omega_{2,2} \quad \delta_{2,1} \quad \tau \quad L]$$

- Se utilizó el algoritmo de programación cuadrática secuencial, más conocido por sus siglas en ingles como SQP, para estimar los parámetros de las funciones de respuesta

al impulso junto con el parámetro τ , mientras que el parámetro L se mantuvo fijo (Reklaitis et al. 1983; MathWorks, 2000).

- El tercer paso consiste en mantener fijos los parámetros de la respuesta al impulso y usar el algoritmo de búsqueda de *Hooke and Jeeves* (H&J) para estimar el parámetro L (Reklaitis et al. 1983). El algoritmo H&J es una técnica de optimización que minimiza o maximiza una cierta función objetivo, mediante variación de parámetros. Esta técnica puede ser usada con números enteros o números reales en general. En este caso se utilizó programación entera, pues el número buscado debe ser un entero. Para encontrar el valor de L , se usó como función objetivo el MSE, con el fin de minimizarlo.

3.6.3 Evaluación de la FT.

Una vez se han estimado los parámetros de las funciones de transferencia, se procedió a su aplicación. Como se dijo en la Sección 3.6.1, la aplicación de los modelos de FT requiere del uso de datos colectados por sensores remotos. En este caso se usaron datos de temperatura del sensor MODIS (corregidos) y datos de lluvia del radar NEXRAD, puesto que estos datos cubren toda la extensión de Puerto Rico. En este punto, es necesario usar una red neuronal artificial autoorganizada o competitiva (ver Sección 3.4.2), para identificar los puntos que exhiben características climatológicas similares a las de las estaciones de medición. Después de ubicar los puntos con similaridades a cada estación, y teniendo en cuenta que a cada estación le corresponde una FT, se aplican los modelos de FT correspondientes para cada región, haciendo uso de la asunción expresada en la Sección 3.6.2. Cada FT fue inicializada con el valor promedio mensual de la humedad de suelo hallado en la Sección 3.4.1.1. El

componente determinístico del modelo es expresado por la tendencia, la cual fue determinada usando RNA, donde las entradas de la red corresponden a la altura sobre el nivel del mar, el porcentaje de arcilla, el porcentaje de arena y la lluvia acumulada mensual, mientras que la salida es la humedad de suelo para cada una de las 15 estaciones. La creación de dicha RNA fue expuesta en la Sección 3.4.1.2.

3.6.4 Verificación y Corrección de la FT.

Después de encontrar los valores de humedad de suelo para cada grid, se procedió a verificar que estos valores se ajusten a la realidad, es decir, que no excedan de los valores límites establecidos. Para el caso de la humedad se cuenta con ecuaciones de pedotrasferencia (Cemek et al., 2004 y Trait et al., 1979), con las que es posible estimar el valor máximo y mínimo de la humedad en un sitio conocidas las características texturales del suelo de dicho sitio. Tales ecuaciones son presentadas a continuación.

$$\omega_j = 18.74 + 0.00315(Sand_j Clay_j) - 0.146 Sand_j \quad 3.14$$

$$\phi_j = 34.00624 + 0.00386(Silt_j Clay_j) - 0.0031 Sand_j^2 \quad 3.15$$

$$\theta_j = 0.332 - 7.251 \times 10^{-4} Sand_j + 0.1276 \log_{10}(Clay_j) \quad 3.16$$

Donde: ω_j , ϕ_j y θ_j : son las estimaciones del *wilting point*, *field capacity*, y el *saturation point* para el sitio j , respectivamente; $Clay_j$, $Sand_j$ y $Silt_j$: son el porcentaje de arcilla, arena y limo para el sitio j , respectivamente. Las estimaciones finales de la FT estocástica no lineal son la humedad de suelo horaria a 1Km de resolución espacial.

4 RESULTADOS Y VALIDACIÓN.

Este capítulo presenta los resultados obtenidos por la aplicación de los métodos propuestos para estimación de temperatura horaria y humedad de suelo, tanto mensual como horaria. Estos modelos han sido construidos principalmente con datos de estaciones ubicadas en Puerto Rico, sin embargo, la metodología propuesta puede ser utilizada para la estimación de humedad de suelo y temperatura en otros lugares en donde se cuente con información semejante a la disponible en Puerto Rico. Se han utilizado 15 estaciones de medición en campo y a partir de sus datos se han establecido 15 modelos de estimación de temperatura e igual número de modelos de estimación de humedad de suelo, tanto en promedios mensuales como horaria. Estos modelos han sido determinados, mediante el análisis de información colectada y la estimación de parámetros mediante la minimización de los errores. El criterio de selección de los datos a usar para la construcción de los modelos esta basado en la experiencia obtenida durante el muestreo de campo.

4.1 Regiones climatológicas de Puerto Rico.

Para encontrar las similitudes climatológicas en Puerto Rico fue necesario entrenar una RNA no supervisada (ver Sección 3.4.2). Los resultados del entrenamiento de dicha red se muestran en la Figura 4.1. En dicha figura se puede observar la clasificación realizada por la red neuronal competitiva.

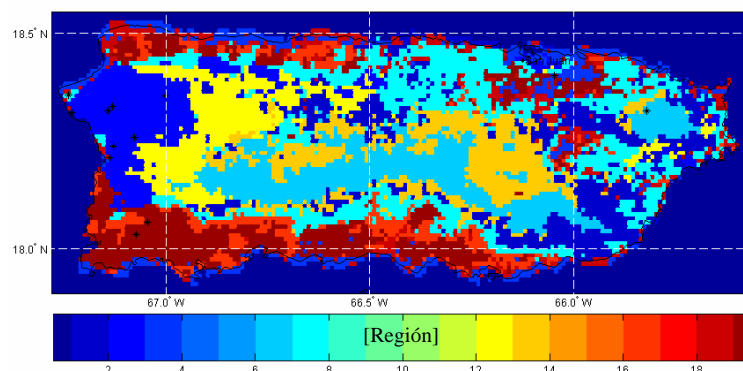


Figura 4.1 Clasificación de regiones climáticas en Puerto Rico.
Esta figura muestra el mapa de las regiones obtenidas mediante la red neuronal competitiva o RNA no supervisada. A cada región le corresponde un color según el código de colores de la barra de abajo.

En el histograma (Figura 4.2) se diferencian 11 regiones climatológicas de las 20 originalmente solicitadas a la red y en las que las similitudes climatológicas han jugado un papel muy importante para su clasificación.

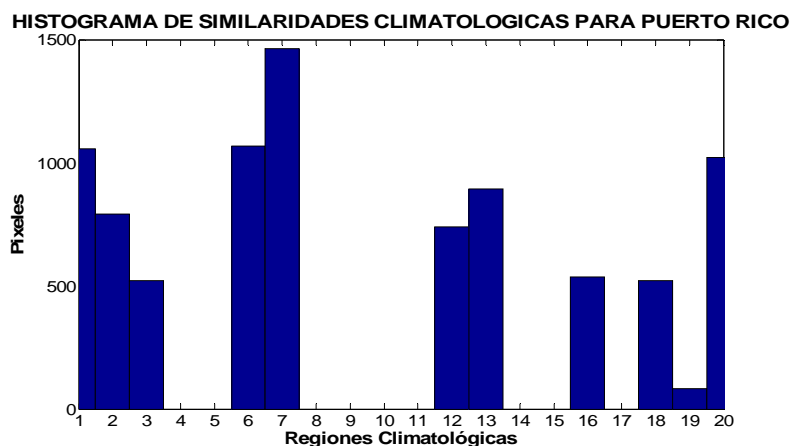


Figura 4.2 Histograma para las regiones climáticas en Puerto Rico.
Este histograma muestra que las regiones definidas por la RNA no supervisada fueron 11 de 20, que inicialmente se había solicitado a la red. Esto quiere decir que la red clasificó las similitudes climatológicas para Puerto Rico en un total de 11 regiones

Para la creación de la red neuronal se usó información de sensores remotos, e información topográfica (alturas). Estas similitudes climatológicas fueron usadas en la generalización de los resultados de los modelos de temperatura y humedad de suelo. La clasificación resultante para las estaciones se muestra en la siguiente tabla:

Tabla 4.1 Región a la que corresponde cada estación.

Estación	Cadena	Caobos	Chips	Colegio	EEL	Gilarte	Isabela	Jimenez	Lajas	Maricao	Nazario	Ramon	Jamaica	UPR	Yunke
Clase	2	20	2	2	20	6	16	2	7	20	6	2	2	3	6

Es importante destacar aquí que el resultado de esta RNA no requiere de validación puesto que es una red autoorganizada, por lo cual no es posible hablar de errores cometidos durante el proceso. Sin embargo, el hecho de que no se tenga un conjunto de datos de validación no significa que el modelo no sea valido, puesto que ya se ha probado que las RNA competitivas son una excelente herramienta para la creación de mapas que agrupen características similares de los distintos elementos en consideración.

Es posible probar que los resultados provenientes de la red neuronal autoorganizada son estadísticamente válidos usando las medidas de tendencia central como la media y la desviación estándar. Los postulados de la estadística clásica establecen que el valor promedio $\hat{\mu}$ de una muestra se encuentra dentro de un límite de confianza si es menor o igual que la media poblacional μ mas o menos dos veces la desviación estándar σ de la población ($\mu - 2\sigma \leq \hat{\mu} \leq \mu + 2\sigma$ ó mejor $-2\sigma \leq \hat{\mu} - \mu \leq 2\sigma$). Para verificar esto, se tomó como muestra el 10% de los datos de cada región obtenida por la red y se calculó el valor medio y la varianza de cada muestra. Bajo el supuesto que esos datos extraídos son una muestra representativa de cada región, se compararon las medidas de tendencia central de

cada región o población con su respectiva muestra. Resultados de este ejercicio se muestran en la siguiente tabla:

Tabla 4.2 medidas de tendencia central para cada región.

Promedios Población						Desviaciones Población					
Región	NDVI	Lluvia	Altura	Temp. Día	Temp. Noc	Región	NDVI	Lluvia	Altura	Temp. Día	Temp. Noc
1	0.78	183.12	144.77	297.68	292.38	1	0.10	53.85	104.38	1.02	1.21
2	0.78	341.39	99.02	301.24	294.09	2	0.10	60.12	74.31	1.03	1.03
3	0.28	132.65	22.63	299.18	295.62	3	0.14	72.35	34.49	1.49	1.73
4	0.75	191.67	614.03	297.06	290.26	4	0.12	63.33	177.46	1.30	1.76
5	0.77	171.96	104.32	300.10	293.03	5	0.09	46.35	79.14	0.86	1.05
6	0.83	304.26	295.75	299.52	292.84	6	0.05	53.18	110.63	0.91	0.77
7	0.78	131.57	402.28	298.77	291.69	7	0.09	41.73	92.31	1.04	1.11
8	0.76	110.18	84.70	300.99	294.71	8	0.06	30.84	59.72	0.63	0.79
9	0.57	130.09	46.80	299.08	294.59	9	0.09	44.24	60.87	1.15	1.19
10	0.41	207.73	41.61	300.77	292.70	10	0.06	44.55	37.81	0.67	1.00
11	0.65	110.07	59.95	302.99	294.66	11	0.10	42.90	55.10	1.31	0.89
Promedios Muestra						Promedios Población - Promedios Muestra					
Región	NDVI	Lluvia	Altura	Temp. Día	Temp. Noc	Región	NDVI	Lluvia	Altura	Temp. Día	Temp. Noc
1	0.83	215.34	215.26	298.22	292.73	1	0.05	32.22	70.49	0.54	0.35
2	0.76	316.75	61.66	300.55	295.67	2	0.02	24.63	37.36	0.69	1.57
3	0.27	193.78	16.26	298.46	297.53	3	0.01	61.13	6.37	0.71	1.91
4	0.84	272.30	636.71	298.01	291.55	4	0.08	80.63	22.68	0.96	1.29
5	0.81	193.96	161.73	300.39	294.28	5	0.04	22.00	57.40	0.30	1.25
6	0.85	374.33	233.75	299.27	293.53	6	0.02	70.06	62.00	0.25	0.69
7	0.83	195.67	449.21	299.14	292.18	7	0.05	64.10	46.93	0.37	0.49
8	0.79	156.85	58.52	301.41	295.14	8	0.03	46.67	26.18	0.43	0.43
9	0.59	176.82	56.02	298.95	296.13	9	0.02	46.73	9.22	0.13	1.55
10	0.44	196.29	31.82	300.66	293.06	10	0.03	11.44	9.79	0.11	0.36
11	0.68	172.34	14.22	303.04	295.13	11	0.03	62.27	45.72	0.05	0.47

En esta tabla se muestran los valores de tendencia central para la muestra y la población en cada una de las once regiones. En la esquina superior izquierda de la tabla se muestran los promedios obtenidos para cada región o población diferenciados por cada una de las variables usadas en la creación de la red; por ejemplo: el promedio para el NDVI en la región 1 es de 0.78 y el de temperatura de día en la región 11 es de 302.99 K. En la esquina superior derecha se muestra, a su vez las desviaciones estándar de dichas poblaciones. En la parte inferior izquierda de la tabla se tienen los promedios muestrales y en la esquina inferior derecha se calcularon las diferencias entre el promedio muestral y el promedio poblacional para cada región.

Tabla 4.3 Validez estadística de la red autoorganizada.

Diferencias de Promedios - 2x(Desviaciones Población)					
Region	NDVI	Lluvia	Alturas	Temp Día	Temp Noc
1	-0.07	-48.50	-115.26	-0.90	-1.49
2	-0.13	-35.35	-107.31	-0.93	0.20
3	-0.41	-61.44	-57.38	-2.19	-0.05
4	0.03	-60.26	-189.49	-1.27	0.16
5	-0.09	-73.17	-100.42	-1.77	-0.36
6	-0.02	-48.74	-61.49	-0.78	-0.62
7	-0.03	-6.61	-107.33	-1.35	-0.78
8	-0.09	-8.59	-90.60	-0.69	-1.60
9	-0.12	-54.13	-84.05	-2.29	-0.07
10	-0.08	-27.49	-48.38	-1.56	-0.76
11	-0.15	-22.26	6.23	-1.95	-1.23

La Tabla 4.3 muestra los valores que prueban la validez estadística del resultado de la red competitiva. Estos datos se obtuvieron tomando el doble de las desviaciones estándar de la población y restándoselo a las diferencias entre los promedios. Los resultados de esta operación deben ser negativos o iguales a cero, si efectivamente se cumple la condición de que $-2\sigma \leq \hat{\mu} - \mu \leq 2\sigma$ ó positivos si no se cumple. En la tabla se puede observar que todos los valores son negativos exceptuando el valor correspondiente al NDVI para la región 4, el cual tal vez podría interpretarse como que el NDVI puede no ser variable relevante. Lo anterior confirma la condición mencionada y por consiguiente la validez estadística de los resultados arrojados por la red.

4.2 Estimación de la Temperatura Horaria.

Un aspecto importante del modelo de estimación de temperatura horaria es que podría utilizar cualquier información de satélite o estaciones experimentales, en las cuales se pueda conocer las temperaturas máximas y mínimas de la superficie de la tierra. Quince modelos (uno para cada estación) para la estimación de la temperatura del aire fueron derivados. La

Tabla 4.4 muestra los resultados de la estimación de parámetros para el MRLM que estima el componente periódico de la temperatura. La primera columna muestra los parámetros del modelo, las columnas restantes muestran en su primera los resultados de la estimación de cada uno de los parámetros para cada estación.

Tabla 4.4 Parámetros estimados para el MRLM (Ecuación 3.6).

	Cadena	Caobos	Chips	Colegio	EEL
c	-1.3042	-1.2918	-1.3096	-0.823	-0.5797
a_1	0.49623	0.47097	0.67762	0.5905	0.67667
b_1	0.48958	0.65643	0.22877	0.51181	0.53539
a_2	-0.1663	-0.1509	-0.3152	-0.167	-0.2025
b_2	-0.3355	-0.2947	-0.0256	-0.2901	-0.2205
a_3	-0.0177	0.01652	0.07953	-0.0023	0.00675
b_3	0.19399	0.12303	0.0392	0.15078	0.10381
a_4	0.02122	0.0306	-0.0449	0.02781	0.0448
b_4	-0.0849	-0.037	-0.0149	-0.0377	-0.008
ϕ	3.4786	3.1426	3.3826	3.2116	3.1416
R^2	0.86292	0.89335	0.78707	0.81687	0.89059
	Gilarte	Isabela	Jamaica	Jimenez	Lajas
c	-0.7808	-0.1906	-1.7293	-0.6831	-0.2849
a_1	0.46341	0.7767	0.55002	0.67662	0.05522
b_1	0.56146	0.06828	0.43025	-0.1911	-0.8943
a_2	-0.165	-0.1546	-0.2968	-0.3328	0.27144
b_2	-0.201	-0.1559	-0.1677	0.07358	-0.082
a_3	0.01314	-0.0141	0.13886	0.12217	-0.026
b_3	0.09788	0.13416	0.0721	0.00312	0.12893
a_4	-0.0061	0.03417	-0.0988	-0.0356	-0.0817
b_4	-0.0182	-0.03	-0.0115	0.00972	0.00343
ϕ	3.1826	3.8826	3.1626	4.0796	3.1476
R^2	0.78618	0.74726	0.8503	0.76849	0.91945
	Maricao	Nazario	Ramon	UPR	Yunke
c	-1.3552	-1.0007	-1.2607	-0.7471	-1.0914
a_1	0.31352	0.56247	0.40001	0.32151	0.6614
b_1	0.526	0.44263	0.52423	0.71742	0.20367
a_2	-0.1668	-0.2384	-0.0982	0.04973	-0.3159
b_2	-0.2347	-0.2508	-0.3098	-0.2786	0.04607
a_3	0.05283	0.08889	-0.0171	-0.1105	0.04606
b_3	0.11653	0.11443	0.17322	0.01092	-0.0669
a_4	-0.0419	-0.0274	-0.0183	0.03275	0.0166
b_4	-0.0579	-0.0232	-0.0839	0.03674	0.04244
ϕ	3.1416	3.1526	3.1526	3.2096	3.8146
R^2	0.67844	0.78934	0.80566	0.79503	0.77862

Para la estimación de la componente de tendencia se entrenó una red neuronal supervisada como se explicó en la Sección 3.4.1.3. y los resultados de su estimación se muestran en la Figura 4.3, en la que al lado izquierdo (a) se muestra el conjunto de datos que se utilizó para el entrenamiento de la red y al lado derecho (b) los resultados obtenidos. En este caso no es necesario realizar una validación puesto que de antemano se sabe que el modelo obtenido no es estadísticamente significativo (esto puede observarse en la parte (b) de la Figura 4.2). La razón de utilizar este modelo es que de la experiencia con el manejo de los datos se observó que la temperatura posee cierta componente determinística y que aunque no esta fuertemente asociada, incide en su comportamiento a largo plazo.

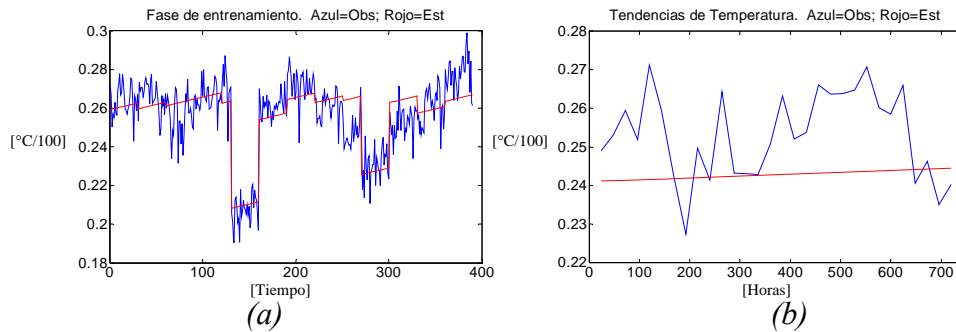


Figura 4.3 Tendencias de Temperatura.

La Figura 4.4 muestra los valores observados y estimados de la temperatura del aire por el modelo de estimación de temperatura para la estación Colegio durante agosto del 2006. Datos de temperaturas máximas y mínimas, además de alturas sobre el nivel del mar fueron usados para evaluar el modelo. Generalmente se tomaron 30 días de datos de cada estación para el ajuste del modelo y la otra parte de los datos (aproximadamente la mitad del total de los datos) fueron usados para realizar la validación del modelo.

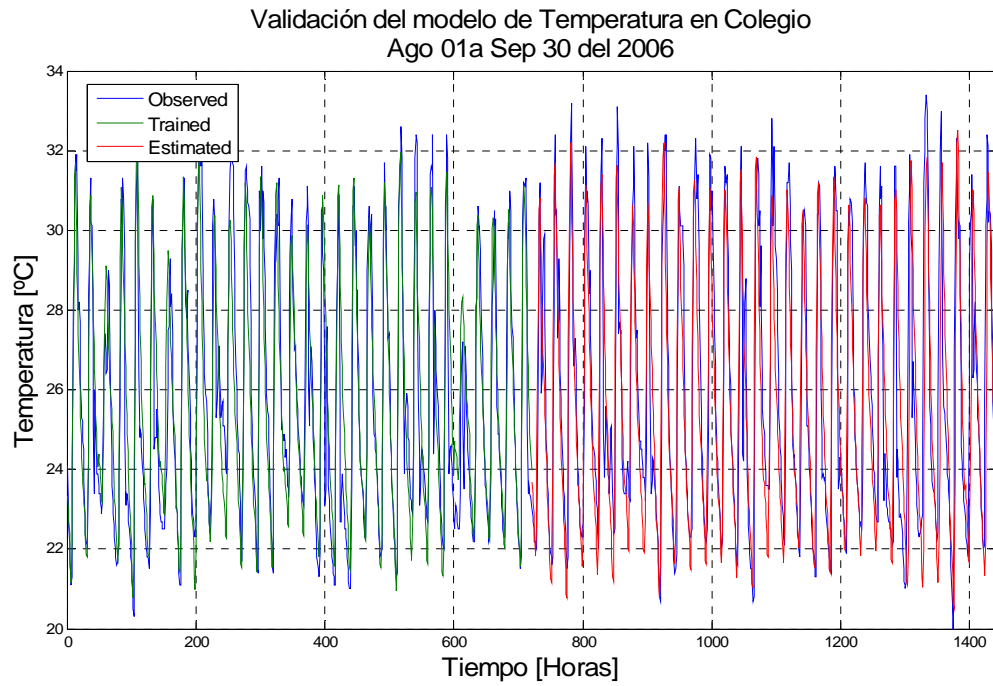


Figura 4.4 Ajuste y validación del modelo de temperatura en la estación Colegio en Agosto del 2006.

La Figura 4.5 muestra el comportamiento del modelo para la estación Yunke durante agosto del 2006. Esta figura exhibe el comportamiento del modelo de estimación de temperatura del aire durante el proceso de ajuste y la validación.

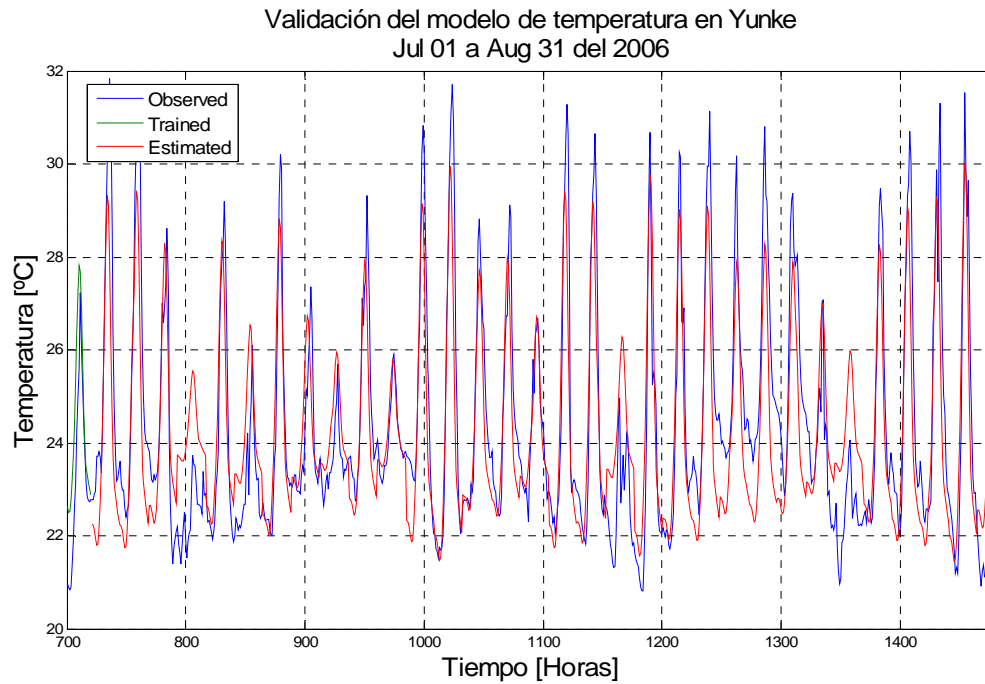


Figura 4.5 Validación en la estación Yunke para agosto del 2006.

En la figura anterior, la línea azul muestra el valor observado de la temperatura del aire en la estación, la línea verde muestra la temperatura estimada durante la construcción del modelo (parte derecha de la grafica), la línea roja es la temperatura estimada por el modelo usando los datos seleccionados para la validación, que no fueron incluidos en el ajuste del modelo. Se puede observar claramente el comportamiento del modelo estimado con respecto a los valores reales, donde las estimaciones siguen de una manera fiel a los datos reales. Los resultados de la validación del modelo son presentados en la Tabla 4.5. La primera columna muestra los nombres de las estaciones, la segunda columna representa el MAE para el ajuste del modelo, la tercera columna representa el MSE para el ajuste del modelo, la cuarta columna representa el MAE para el proceso de validación del modelo y la última columna representa el MSE para el proceso de validación del modelo. En la última fila se pueden

observar los promedios de los errores obtenidos. Del MAE del proceso de ajuste se puede observar que para los 15 modelos, en promedio se tiene un error menor de 1°C, lo cual quiere decir que los modelos son bastante aceptables. Además, si se mira el MSE del proceso de ajuste, se puede inferir que la varianza de los errores en promedio es bastante baja (Montgomery y Runger, 1996), por lo cual, se ratifica lo dicho anteriormente. Durante el proceso de validación, los errores también muestran un comportamiento bastante bueno del modelo pues el máximo valor del MAE es de 1.32°C y en promedio para las 15 estaciones el MAE fue de 1.07°C, lo cual es una buena señal de que el modelo estima correctamente la temperatura. Por otro lado, el valor promedio del MSE fue de 2.07 ratifica que el modelo se comporta de una manera consistente con respecto a los datos de temperatura reales.

Tabla 4.5 Precisión de las estimaciones de temperatura.

	MAE Ajuste	MSE Ajuste	MAE Val	MSE Val
Cadena	0.7256	0.9660	1.1679	2.5074
Caobos	0.9093	1.4042	1.2181	2.8834
Chips	0.9061	1.6072	1.2138	2.3454
Colegio	0.9771	1.8133	1.3222	3.4543
EEL	0.9468	1.4560	1.0724	2.3933
Gilarte	0.7312	0.9135	0.8732	1.2195
Isabela	0.7721	1.0710	0.8342	1.0845
Jamaica	0.8946	1.3980	1.1833	2.3114
Jimenez	0.7167	0.9746	1.0161	1.7128
Lajas	0.7290	1.0116	0.9670	1.8443
Maricao	0.8865	1.4861	0.9945	1.8014
Nazario	0.9195	1.5347	1.0623	1.9148
Ramon	0.8856	1.4349	1.1612	2.0785
UPR	0.9232	1.4499	1.1669	2.2254
Yunke	0.7477	0.8839	0.9317	1.3523
Promedios	0.8447	1.2937	1.0790	2.0752

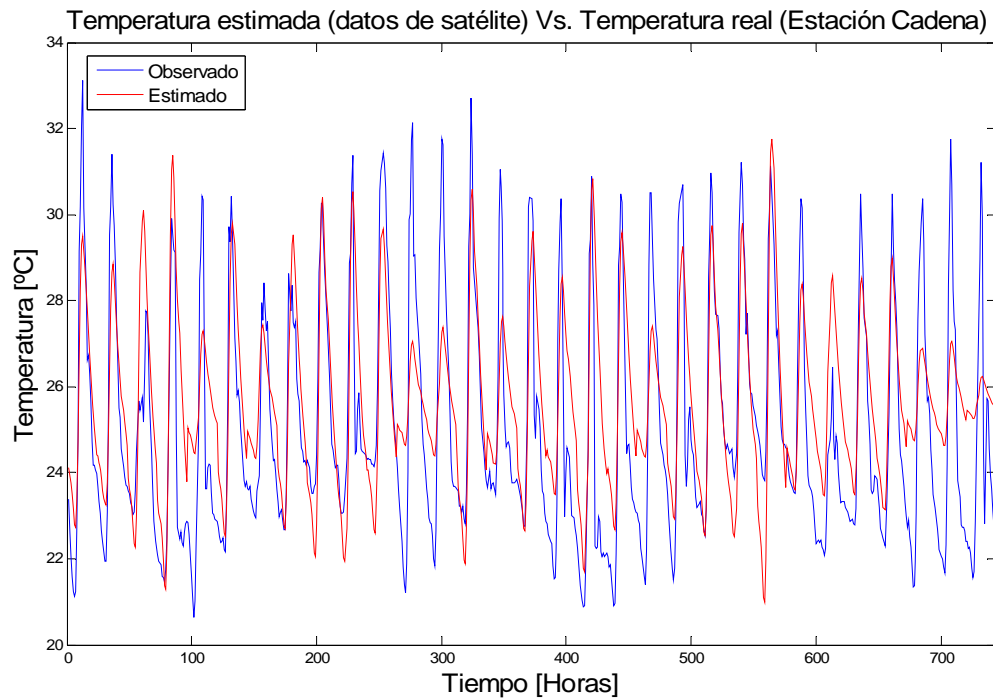


Figura 4.6 Comparación datos de MODIS (rojo). Estación Cadena agosto 2006.

La Figura 4.6 muestra la comparación entre los datos reales en la estación Cadena (azul) y los estimados con datos del satélite MODIS (rojo), para el Km^2 más cercano a dicha estación en $^{\circ}\text{C}$. Los últimos 3 días muestran que la temperatura fue subestimada, lo cual puede obedecer a problemas con la medición del satélite. Normalmente estos problemas son puntuales debido a nubosidad, puesto que la señal satelital se ve afectada por la presencia de nubes. La Figura 4.7 muestra el comportamiento del modelo de temperatura usando datos de satélite comparados con los datos reales obtenidos de la estación Isabela. Dicha comparación muestra, que la temperatura estimada sigue de cerca el patrón de temperaturas observadas, cometiendo error en los últimos días, debido a fallas de medición del satélite. La Tabla 4.7 muestra los errores cometidos por el modelo en la estimación de la temperatura del aire usando datos corregidos del satélite MODIS para 13 de las estaciones. El hecho de haber

tomado 13 estaciones de las 15 responde a que en dos de las estaciones no se usaron datos del mes de agosto, lo que imposibilitaba la comparación. Es de anotar que en promedio, el MAE tiene un valor bajo aceptable ($MAE=2.11^{\circ}C$), lo cual es un indicativo de que el modelo se comporta bien al ser evaluado con los datos provenientes de satélite.

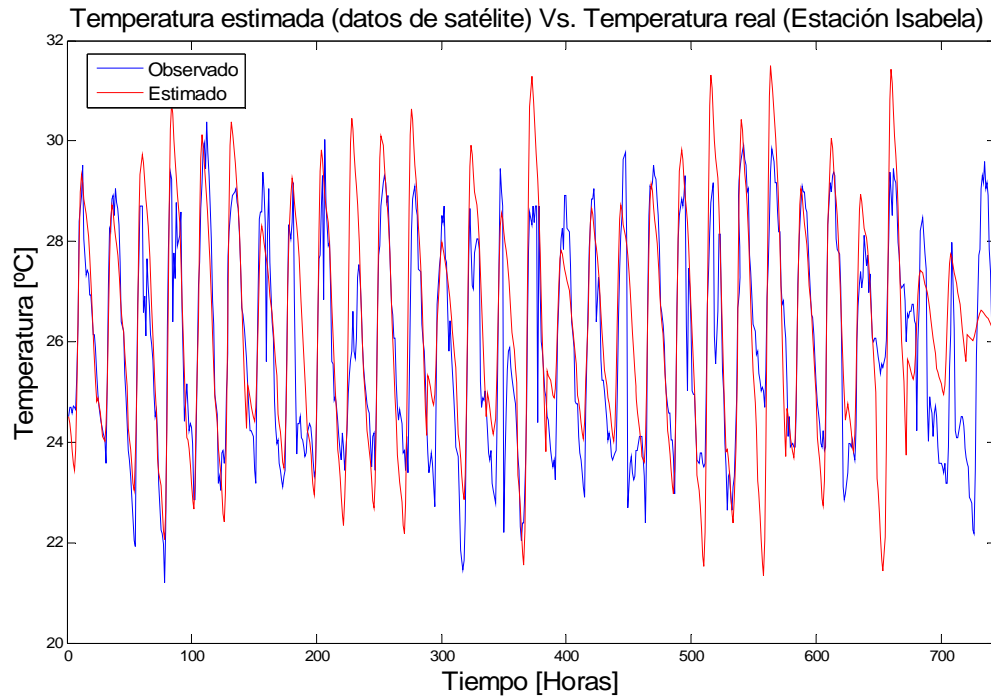


Figura 4.7 Comportamiento del modelo de temperatura con datos de satélite (rojo).

La estimación de la temperatura usando datos de satélite para la estación Gilarte fue bastante pobre, seguramente debido a problemas de nubosidad. Teniendo en cuenta que dicha estación se encuentra ubicada en la zona montañosa y además con vegetación abundante, no es raro que haya ocurrido dicho problema. En lo referente al MSE (fila 2 de la Tabla 4.6), el valor promedio ($MSE=7.53$) esta dentro de los niveles de aceptación puesto que muestra que la variabilidad del modelo es pequeña y si se tiene en cuenta que existe un valor máximo de

33.86 en la estación Gilarte y que dicho valor aumenta significativamente el promedio del MSE, entonces es notorio que dicho valor máximo enmascara el valor real del MSE.

Tabla 4.6 Precisión de las estimaciones de temperatura con datos corregidos de satélite.

	Cadena	Chips	Colegio	EEL	Gilarte	Isabela	Jamaica	Jimenez	Maricao	Nazario	Ramon	UPR	Yunke	Promedios
MAE	1.5114	2.1935	1.7215	2.0594	5.4952	1.0213	1.6562	1.2315	3.7145	1.9188	1.6248	1.7965	1.5507	2.1150
MSE	3.4472	6.8487	4.3397	6.7232	32.862	1.9556	4.6127	2.3574	16.366	5.2752	4.0451	5.1513	3.9926	7.5367

Los resultados finales del modelo de estimación de temperatura del aire consisten en la generación de mapas de temperatura horaria haciendo uso de las similitudes de las regiones climatológicas con las características climatológicas de las estaciones consideradas en este estudio. Un ejemplo de dichos mapas se puede observar en la Figura 4.8, en la que las zonas de color amarillo muestran los valores estimados de temperatura, mientras que las zonas azules dentro del mapa corresponden a los puntos en los que no se encontró ninguna relación climatológica con las estaciones de medición, por lo cual no fue posible realizar la estimación de su temperatura.

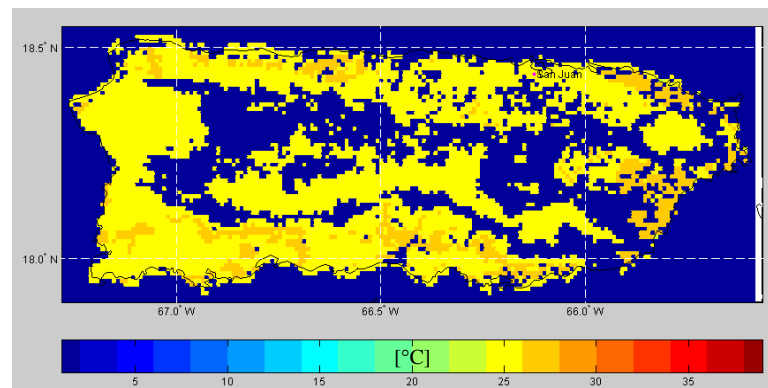


Figura 4.8 Estimación final de la temperatura horaria por regiones similares.

Este grafico muestra la variación espacial de temperatura del aire en Puerto Rico a las 8:00 PM del 20 de Agosto del 2006. La parte interna del mapa en color azul se refiere a aquellas regiones climatológicas donde no encontró similitud alguna con las estaciones de medición.

Las Figuras 4.9 y 4.10 muestran las estimaciones horarias de temperatura del aire de las primeras 24 horas del mes de agosto del 2006. Aquí es posible observar la variación horaria de la temperatura, el periodo de calentamiento y enfriamiento y el comportamiento del modelo en un plazo de 24 horas.

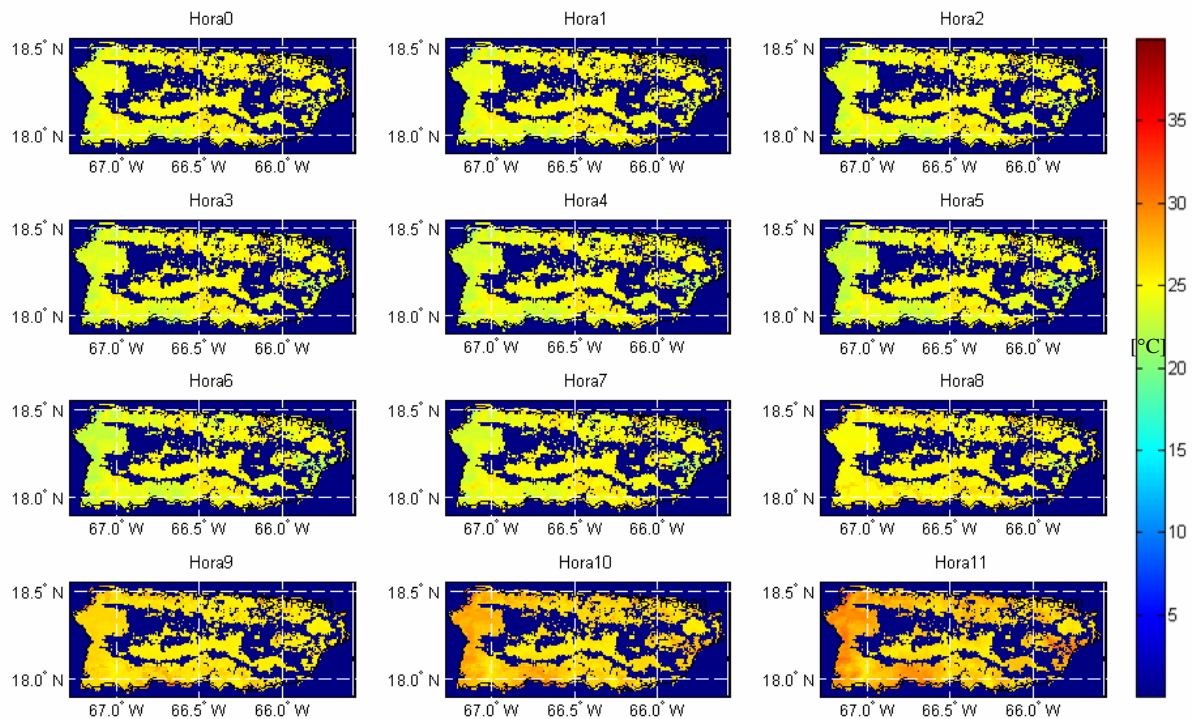


Figura 4.9 Temperatura estimada en las primeras doce horas del 01 de Agosto del 2006.

Este grafico muestra la variación espacial de temperatura del aire en Puerto Rico para el periodo comprendido entre las doce de la noche y las 11 de la mañana del 01 de Agosto del 2006. Se puede apreciar tanto la variación horaria como la variación espacial de la temperatura del aire. La barra de la derecha indica la convención de colores en °C para los mapas de temperatura, y muestra desde un valor frío (azul) hasta un valor caliente (rojo)

Es importante notar, que el modelo estima bastante bien la variación horaria de la temperatura mirándolo desde el punto de vista de los periodos de enfriamiento y calentamiento, pues al observar detenidamente la Figura 4.9 se puede notar que el final de la

fase de enfriamiento y el inicio del periodo de calentamiento se encuentra cercano a las 9:00 AM y consecutivamente, revisando la Figura 4.10 se nota que dicho periodo de calentamiento finaliza cerca de las 4:00 PM, consecuentemente con el inicio de la fase de enfriamiento. También en la zona de montaña, se observa que la temperatura es relativamente mas baja que en las zonas costeras, lo cual es un indicativo de que el modelo realiza una excelente aproximación de la variación espacial de la temperatura del aire.

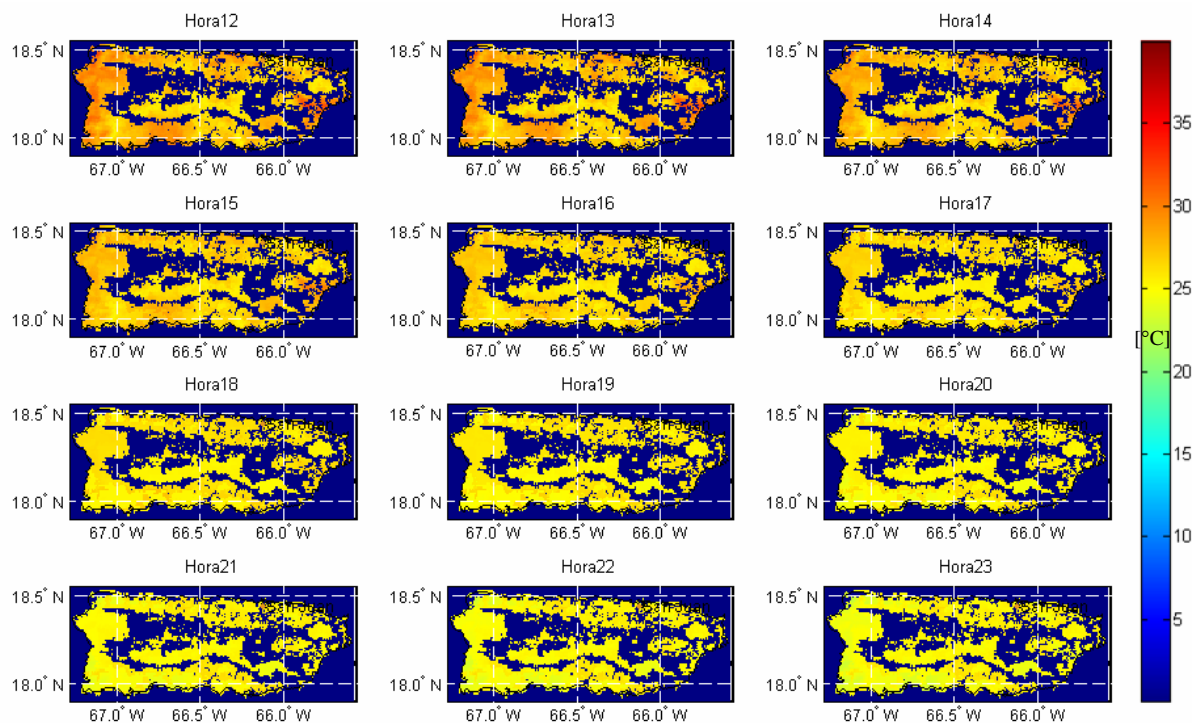


Figura 4.10 Estimación de la temperatura del aire entre las 12 y las 23 horas del 01 de Agosto del 2006.

Este grafico muestra la variación espacial de temperatura del aire en Puerto Rico para el periodo comprendido entre las doce del día y las 11 de la noche del 01 de Agosto del 2006. Aquí se puede apreciar los periodos de calentamiento y enfriamiento diarios.

4.3 Estimación de la Humedad de Suelo

El modelo de estimación de humedad de suelo involucra variables obtenidas mediante satélite, mediciones de campo y otros parámetros como la topografía y las características texturales. Estas variables, pueden ser obtenidas fácilmente para otros sitios en los que se quiera aplicar la metodología descrita aquí. La humedad de suelo posee variación tanto espacial como temporal (Lookingbill T. y Urban D., 2004; Calderón, 2005 y Ramírez et al, 2005). A esto se le ha llamado memoria de corto y largo plazo, esto quiere decir que la humedad de suelo se puede dividir en una característica de memoria a largo plazo y una de memoria a corto plazo (Ramírez et al, 2007a y Ramírez et al, 2007b). La memoria a largo plazo esta influenciada por las características inherentes a cada punto y que poseen una escasa o nula variación temporal, como por ejemplo, las características texturales, la topografía o el NDVI, entre otras. Mientras que la memoria a corto plazo la influyen variables como la lluvia y la temperatura, las cuales, tienen una influencia inmediata en el comportamiento de la humedad de suelo. Por otro lado, existe también variación espacial, debida a los cambios en la posición geográfica. Por lo tanto, la humedad de suelo posee variación espacio – temporal y a su vez la variación temporal esta dividida en una memoria a corto plazo y una a largo plazo. Se crearon quince modelos de estimación de humedad de suelo (uno para cada estación). En cada uno de ellos se tuvo en cuenta la memoria a largo y corto plazo. La primera se definió por medio de un modelo de RNA, en el cual se estimaron los contenidos promedios mensuales de la humedad de suelo para los doce meses del año (ver Sección 3.4.1.2). Para los modelos de promedios mensuales, el modelo se desarrolla en general con 18 estaciones, aunque algunos meses pueden tener mayor o menor numero de

muestras, mientras que la validación se realizó usando cuatro estaciones para cada mes. La Figura 4.11 muestra el resultado de la validación del modelo de RNA para la estimación de los promedios mensuales de humedad de suelo para el mes de Marzo. Las barras rojas indican los valores estimados, los cuales son comparados con los valores observados (barras azules).

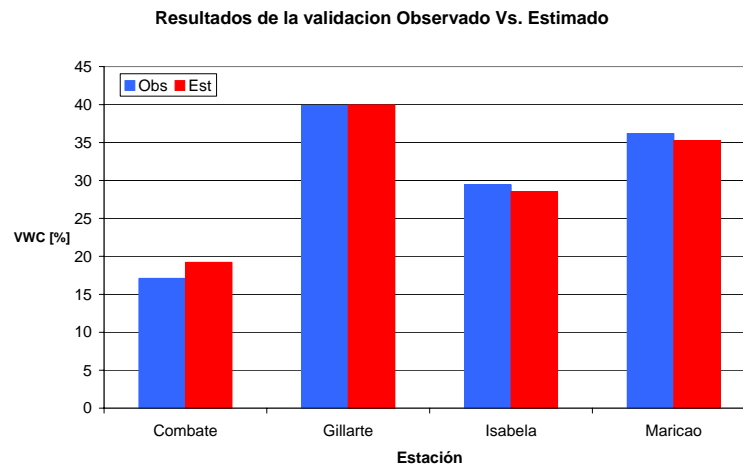


Figura 4.11 Validación del modelo de promedios de humedad de suelo mensuales

La Tabla 4.7 muestra la precisión del modelo de RNA para la estimación de la humedad de suelo promedio mensual. Las columnas 2, 3, 7 y 8 contienen los valores de los errores cometidos durante el ajuste de los modelos mensuales, las columnas 2 y 7 corresponden a los valores del MAE de ajuste y las columnas 3 y 8 equivalen a los valores del MSE de ajuste. Igualmente, las columnas 4, 5, 9 y 10 corresponden a los errores durante el proceso de validación. En este caso, las columnas 4 y 5 contienen los valores del MAE de validación, mientras que las columnas 9 y 10 contienen los valores del MSE de validación. Las filas llamadas promedio, contienen los valores promedio del error cometido por las cinco RNA.

En promedio para todo el año 2006, el valor del MAE de validación fue de 2.72%, mientras que en promedio durante el 2006 el MSE fue de 12.32. Estos valores son bastante bajos, lo cual sugiere que el comportamiento de los modelos se acerca bastante a la realidad. Estos valores fueron obtenidos promediando los resultados de las cinco RNA's para cada mes. Los resultados de los promedios mensuales indican que el mes de Mayo tiene un modelo muy preciso y que el valor máximo y bastante aceptable resultó para el mes de Julio.

Tabla 4.7 Precisión de las estimaciones de humedad de suelo promedio mensual.

	MAE Aju	MSE Aju	MAE Val	MSE Val		MAE Aju	MSE Aju	MAE Val	MSE Val
Enero	2.17E-13	1.18E-25	2.34	11.95	Julio	9.52E-13	4.98E-24	4.15	28.43
	8.07E-13	1.24E-24	2.90	10.67		1.90E-14	6.60E-28	3.88	24.18
	2.77E-11	2.30E-21	2.00	5.12		6.65E-12	9.78E-23	4.59	31.77
	1.79E-12	1.40E-23	1.67	3.55		8.12E-10	1.01E-18	4.62	32.48
	2.24E-08	1.02E-15	1.80	4.51		2.52E-10	2.24E-19	2.66	8.57
Promedios	4.49E-09	2.05E-16	2.14	7.16	Promedios	2.14E-10	2.47E-19	3.98	25.08
Febrero	7.63E-02	3.64E-02	3.53	19.96	Agosto	1.65E-11	5.74E-22	3.03	13.27
	5.50E-02	1.73E-02	3.09	26.21		1.06E-11	3.25E-22	3.45	14.91
	2.67E-02	3.66E-03	1.89	7.74		7.09E-13	1.31E-24	2.60	10.45
	4.91E-02	1.27E-02	2.34	8.55		2.64E-14	1.56E-27	3.49	12.65
	8.92E-02	5.11E-02	3.33	18.40		6.42E-09	9.21E-17	3.08	12.10
Promedios	5.93E-02	2.42E-02	2.83	16.17	Promedios	1.29E-09	1.84E-17	3.13	12.68
Marzo	4.13E-04	4.19E-07	1.38	2.39	Septiembre	1.02E-10	1.64E-19	3.65	16.03
	2.55E-04	1.19E-07	2.42	10.60		1.71E-13	6.03E-26	3.00	13.53
	2.60E-04	1.34E-07	2.77	8.93		1.74E-12	8.22E-24	2.57	7.73
	5.59E-04	9.37E-07	3.19	14.75		1.96E-13	1.31E-25	3.49	20.33
	4.19E-04	4.84E-07	2.74	10.09		1.26E-11	3.59E-22	2.57	7.56
Promedios	3.81E-04	4.18E-07	2.50	9.35	Promedios	2.34E-11	3.29E-20	3.05	13.04
Abril	1.40E-13	4.44E-26	2.67	8.26	Octubre	2.17E-13	1.18E-25	2.34	11.95
	1.63E-10	5.07E-20	1.09	1.60		7.63E-02	3.64E-02	3.53	19.96
	2.97E-13	2.65E-25	1.55	2.56		8.12E-10	1.01E-18	4.62	32.48
	7.22E-09	1.18E-16	2.18	7.33		3.20E-13	1.49E-25	1.70	4.17
	4.10E-11	2.93E-21	2.50	13.42		8.92E-02	5.11E-02	3.33	18.40
Promedios	1.48E-09	2.37E-17	2.00	6.63	Promedios	3.31E-02	1.75E-02	3.10	17.39
Mayo	4.06E-12	7.10E-23	2.41	8.44	Noviembre	1.06E-11	3.25E-22	3.45	14.91
	3.34E-12	2.70E-23	1.27	1.94		4.46E-02	5.04E-19	3.16	18.33
	3.01E-13	1.54E-25	2.40	8.33		3.37E-04	3.01E-07	2.58	10.35
	2.69E-10	1.29E-19	1.39	2.83		3.20E-04	2.76E-07	2.63	9.14
	2.34E-11	1.21E-21	2.31	9.87		2.80E-04	4.69E-07	2.93	11.51
Promedios	6.01E-11	2.60E-20	1.96	6.28	Promedios	9.11E-03	2.09E-07	2.95	12.85
Junio	2.69E-12	1.45E-23	3.12	12.11	Diciembre	4.42E-02	1.14E-02	2.10	7.70
	9.49E-10	2.05E-18	2.47	10.54		8.03E-02	4.60E-02	2.99	16.56
	3.20E-13	1.49E-25	1.70	4.17		5.34E-02	2.18E-02	2.55	14.56
	5.44E-12	5.69E-23	3.12	11.95		3.72E-04	3.77E-07	1.24	2.15
	2.72E-11	1.33E-21	3.45	17.12		2.29E-04	1.07E-07	2.18	9.54
Promedios	1.9686E-10	4.0964E-19	2.77	11.18	Promedios	0.04	0.02	2.21	10.10

Para la creación de estos modelos se usaron las siguientes variables: NDVI mensual, Rango de temperatura mensual, lluvia acumulada mensual, porcentaje de arcilla y porcentaje de arena, así como la humedad de suelo promedio mensual. La Figura 4.12 muestra las estimaciones preliminares de los promedios de humedad para el mes de agosto del año 2006 aun sin corregir. La corrección se hizo teniendo en cuenta los valores limites de la humedad dados por las Ecuaciones 3.14, 3.15 y 3.16. En este caso se usó la ecuación 3.14 que corresponde al *wilting point* como el valor del mínimo de la humedad de suelo, mientras que para el valor máximo se calculó como el *field capacity* (Ecuación 3.15). La razón para haber usado este valor recae en el hecho de que el promedio de la humedad de suelo no puede estar un periodo de tiempo tan prolongado (un mes) en su máximo valor que es el *saturation point*, por lo que el máximo debería ser un valor cercano al *field capacity*.

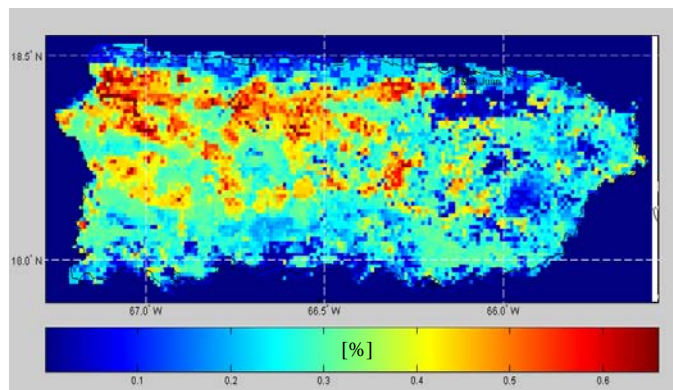


Figura 4.12 Estimación de la humedad de suelo promedio mensual sin corregir para agosto del 2006.

Aquí se puede observar los estimados de la humedad de suelo promedio mensual como los entrega la RNA. Las unidades están expresadas en porcentaje en volumen de agua en el suelo.

La figura 4.13 muestra las estimaciones finales de los promedios mensuales de la humedad de suelo ya corregidas para el mes de agosto del 2006. Las estimaciones preliminares, así

como las finales, consisten en la generalización de la RNA por medio de su aplicación a todos los puntos de la isla de Puerto Rico. En ambos casos se ha usado información de satélite, radar y características texturales para dicha generalización. En esta figura se nota la variación espacial de los promedios de la humedad de suelo.

La estimación de la componente de tendencia o determinística se realizó entrenando una red neuronal como se explicó en la Sección 3.4.1.1 y los resultados de su estimación se muestran en la Figura 4.14, en la que en la parte (a) se muestra el conjunto de datos que se utilizó para el entrenamiento de la red y en la parte (b) puede apreciarse el resultado obtenido.

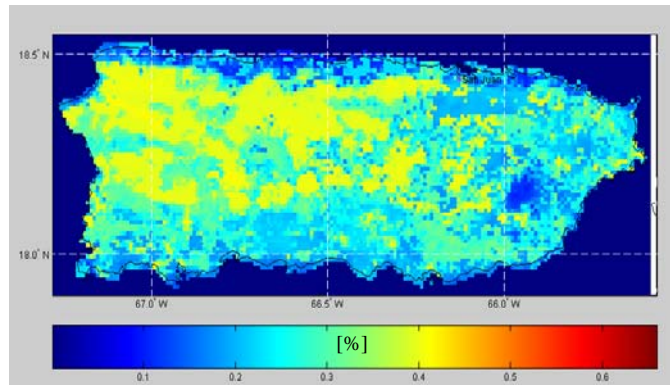


Figura 4.13 Promedio mensual estimado de humedad de suelo para agosto del 2006.

Aquí se puede observar los estimados de la humedad de suelo promedio mensual después de realizar la corrección con los valores límites establecidos.

Una línea roja indica la estimación de la tendencia de la humedad de suelo y en azul se puede ver los datos observados de humedad de suelo a los que se les ha restado su promedio. Al igual que para los modelos de temperatura, no es necesario realizar una validación puesto que estos resultados poseen validez empírica mas no estadística.

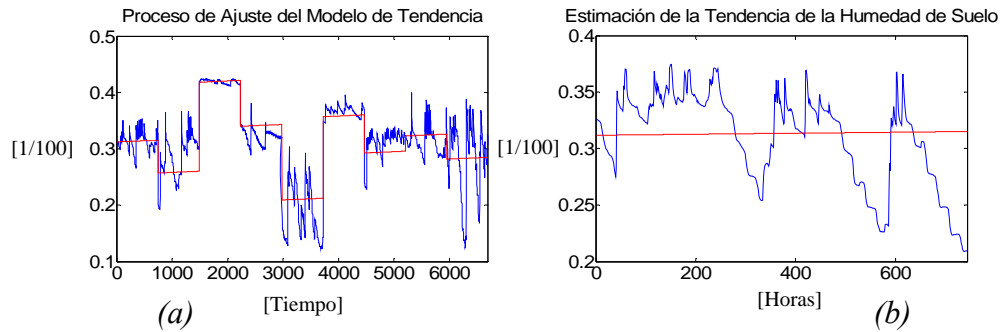


Figura 4.14 Tendencias de humedad de suelo.

En esta figura se puede observar los procesos de ajuste y validación de la tendencia de la humedad de suelo. Los valores de humedad de suelo están dados en fracciones puesto que los datos de entrenamiento de la red fueron entrados en fracciones para mejorar el rendimiento de los algoritmos usados.

La Figura 4.15 muestra el comportamiento del modelo de FT para la estimación de la humedad de suelo. En dicha figura se diferencian los valores observados y estimados de la humedad de suelo para la estación Ramón en las fechas comprendidas entre agosto 01 y septiembre 30 del 2006. Datos de temperatura horaria y de lluvia horaria fueron usados para construir el modelo. Generalmente se tomaron 30 días de datos de cada estación para el ajuste del modelo y la otra parte de los datos (aproximadamente la mitad del total de los datos) fueron usados para realizar la validación del modelo. Esta técnica de validación es llamada validación cruzada. La Figura 4.15 parte (a) muestra en conjunto, el proceso de ajuste y el de validación para el modelo de FT desarrollado en dicha estación. Allí puede observarse claramente como el modelo de FT sigue fielmente el comportamiento de los datos observados. Mientras que la parte (b) de la figura muestra un detalle ampliado del proceso de validación, en el cual se usaron datos de un mes aproximadamente. En este caso puede observarse que el modelo estima muy bien el comportamiento de la humedad de suelo.

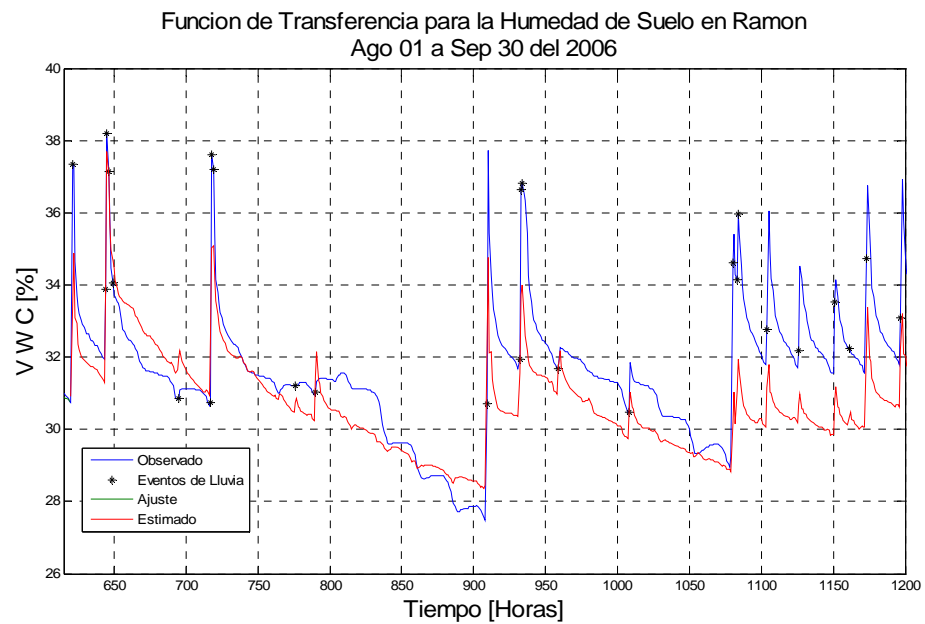
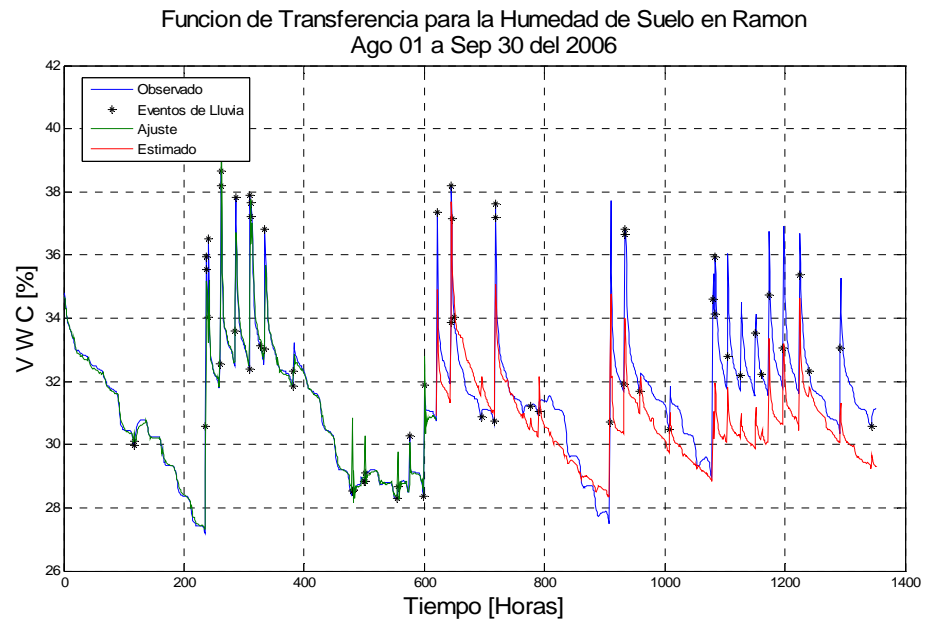


Figura 4.15 Ajuste y validación de la función de transferencia en la estación Ramón.

La Figura 4.16 muestra el comportamiento del modelo de FT en la estación Chips durante agosto y septiembre del 2006. En la parte (a) de esta figura, así como en la figura anterior se muestran conjuntamente el proceso de ajuste y validación. En la parte (b) se puede observar que el modelo de FT (Ecuación 3.13) representado por una línea roja, se comporta similarmente a los datos observados, los cuales se muestran con una línea azul. De lo anterior se puede extraer que el modelo de FT estima correctamente los valores reales de la humedad de suelo, puesto que el comportamiento de las funciones de impulso es muy similar al comportamiento de la humedad de suelo observada. Tanto en esta figura como en la anterior, se muestra el proceso de ajuste y validación. Una línea azul muestra los valores reales de la humedad de suelo, una línea verde muestra los valores estimados por el modelo durante el proceso de ajuste, una línea roja muestra los valores estimados por el modelo durante el proceso de validación y un asterisco denota el momento en que ocurrió un evento de lluvia. El comportamiento del modelo de función de transferencia durante el proceso de ajuste y el de validación puede ser visto en esta figura. Es posible verificar la respuesta al impulso durante los eventos de lluvia, en los que la función de transferencia responde estimando correctamente la humedad de suelo, aunque en algunos casos, se comete un error (visual), más adelante se mostrará que estos errores son aceptables. En la Tabla 4.8 se muestran los valores resultantes de la estimación de parámetros para cada una de las estaciones consideradas. Dicha estimación de parámetros fue realizada usando el lenguaje de programación Matlab (MathWorks, 2000) y fue descrita en la Sección 3.6.2.

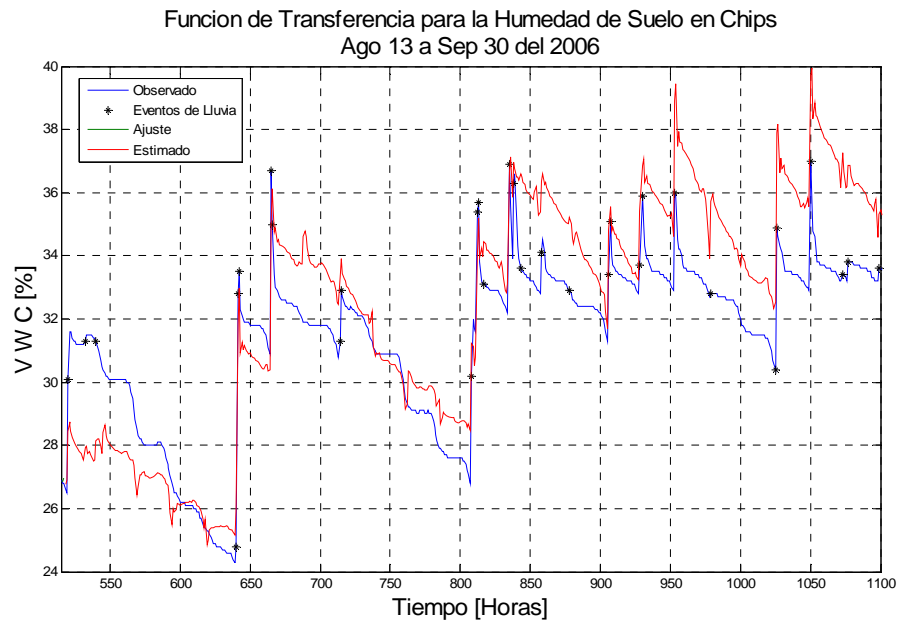
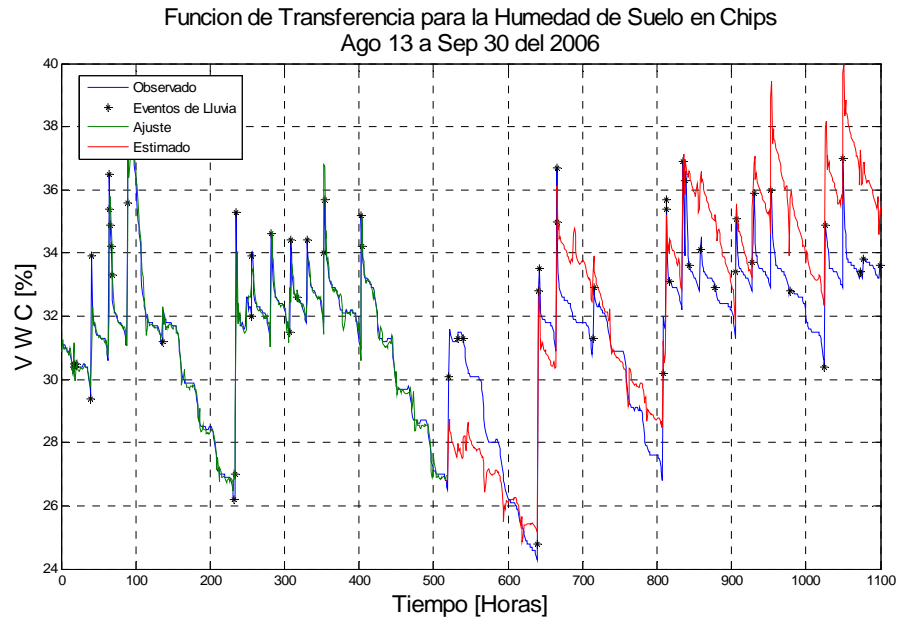


Figura 4.16 Ajuste y validación de la función de transferencia para la humedad de suelo.

Es importante notar, que el valor de L cuando τ es cero, es irrelevante pues en realidad se anularía el valor de la variable en el exponencial de la Ecuación 3.13.

Tabla 4.8 Resultados del proceso de estimación de parámetros de la FT.

	$w_{0,1}$	$w_{0,2}$	$d_{0,1}$	$w_{1,1}$	$w_{1,2}$	$d_{1,1}$	$w_{2,1}$	$w_{2,2}$	$d_{2,1}$	t	L
Cadena	-0.030401	0.042376	0.32073	0.73816	0.75267	-0.3164	1.0711	0.33119	-0.41918	0.13292	40
Caobos	-0.06532	0.042116	-0.60445	1.1781	1.217	-0.48216	1.1714	0.13634	-0.3139	0.13884	68
Chips	-0.22993	0.20148	0.20343	0.57515	0.5309	-0.50292	0.75607	-0.089511	0.32648	0.037	71
Colegio	-0.035497	0.033294	-0.19603	0.24564	0.31165	-0.23244	0.9549	0.67528	-0.64383	0.017672	12
EEL	0.039789	-0.032539	0.44959	1.3382	-0.90192	0.33163	1.431	-0.085488	-0.35381	0.0094961	12
Gilarte	-0.00228	-0.0041302	-0.21679	1.0824	0.68264	-0.62413	0.96837	0.43861	-0.41825	2.6946	15
Isabela	0.12352	-0.090313	-0.59294	0.24126	0.24637	0.11086	0.6796	-0.46518	0.78158	0.01	43
Jamaica	-0.024429	0.0065657	-0.00071226	0.17595	0.30664	-0.42279	1.1537	0.59884	-0.75885	0.10827	39
Jimenez	-0.15379	0.14763	-0.25741	0.21599	0.25172	-0.080602	0.85372	0.12207	0.01145	0	12
Lajas	0.0060503	0.00046266	0.87288	1.244	1.36	0.07406	0.56463	0.23516	0.19111	0.10381	70
Maricao	-0.00066574	-0.00072494	-0.72452	0.0068041	0.059562	0.034199	1.2271	-0.0017638	-0.23233	0	12
Nazario	-0.049073	0.047876	-0.19061	0.13764	0.11298	-0.22957	1.0564	0.21182	-0.27657	0.019038	85
Ramon	-0.060377	0.05661	0.4033	0.42853	0.52462	-0.42549	0.78096	-0.67917	0.89935	0.034875	113
UPR	-0.19293	0.1468	0	0	0.48343	-0.21722	1.0177	0.49665	-0.52161	0.03	65
Yunke	-0.025108	0.070815	-0.66986	0.38871	0.49111	-0.45939	1.203	0.20961	-0.42739	0.069271	46

La Tabla 4.9 contiene los resultados de los errores cometidos por el modelo de función de transferencia. Estos errores explican que tan buena es la precisión del modelo con respecto a la estimación de la humedad de suelo horaria.

Tabla 4.9 Precisión del modelo de FT.

	MAE Aju	MSE Aju	MAE Val	MSE Val
Cadena	0.1109	0.0775	1.2074	2.1694
Caobos	0.0937	0.0477	2.0807	7.1181
Chips	0.2130	0.1900	1.7005	3.9651
Colegio	0.1464	0.1408	2.3075	7.5967
EEL	0.1159	0.0881	3.4289	13.0460
Gilarte	0.0360	0.0032	0.2804	0.1572
Isabela	0.1626	0.1120	1.0829	2.0870
Jamaica	0.0681	0.0165	1.2912	3.1945
Jimenez	0.1756	0.1613	2.3255	7.4590
Lajas	0.1132	0.0646	1.8246	4.8246
Maricao	0.0517	0.0139	0.7786	1.2066
Nazario	0.0882	0.0244	0.7763	0.9664
Ramon	0.1530	0.1452	1.1298	1.9910
UPR	0.2270	0.3552	3.0827	13.4380
Yunke	0.1322	0.0606	1.9265	5.2865
Average	0.1258	0.1001	1.6816	4.9671

De la tabla 4.9 se puede observar que el comportamiento del modelo de función de transferencia es bastante bueno, pues los errores en general son muy bajos. Durante el

proceso de ajuste el promedio del MAE fue de 0.12%, lo cual es un buen indicador de que el modelo ajusta bien los datos, sin embargo, es necesario ver también el MSE para conocer que tan grande es la variabilidad de la respuesta del modelo. En este caso, el promedio del MSE es de 0.10, lo cual también indica que la respuesta del modelo es buena con respecto a los datos de ajuste. Durante el proceso de validación, los errores, aunque son un poco mayores que para el proceso de ajuste, siguen siendo aceptables, con un promedio del MAE de 1.68% que indica que en general, los modelos cometerían un error del 1.7% (contenido volumétrico de humedad), aproximadamente en promedio para las estimaciones del contenido de humedad de suelo. Este valor es un error aceptable si se tiene en cuenta que la variabilidad de la humedad de suelo es bastante alta. Revisando el MSE, se puede observar que en promedio es del orden de 4.96, lo cual es un resultado que aunque medianamente alto, es también aceptable. Resumiendo, la información de la Tabla 4.9 indica que el modelo de función de transferencia se comporta bien con respecto a las estimaciones de la humedad de suelo. Para la creación de mapas de humedad de suelo se usaron 6 de los 15 modelos desarrollados, pues algunas de las estaciones se encontraban ubicadas en la misma región. Las estaciones restantes se usaron para comprobar la hipótesis de que los modelos obtenidos pueden ser usados en sitios con características climatológicas similares. De la Tabla 4.1 se tiene que las estaciones Cadena, Chips, Colegio, Jiménez, Jamaica y Ramón pertenecen a la región número 2, en la que se ha usado como FT representativa la FT de la estación Chips y es claro que se habría podido escoger cualquiera de ellas, además, las estaciones Gilarte, Nazario y Yunke, pertenecen a la región 6 y en este caso se usó la FT de la estación Nazario, y para las estaciones Caobos, EEL y Maricao pertenecientes a la región 20 se utilizó la FT de

la estación Caobos. Las estaciones pertenecientes a una misma región que no fueron usadas para la estimación final de la humedad de suelo, se usaron para realizar el proceso de validación de la hipótesis de similitud. La Tabla 4.10 muestra los resultados obtenidos de dicha validación. Se hace notar que para las estaciones Isabela, Lajas y UPR no se realizó dicho proceso de validación, puesto que estas estaciones no poseen similitud alguna con otra estación. En este caso, se asume que los resultados de validación obtenidos para las otras estaciones aplican para las estaciones restantes.

Tabla 4.10 Precisión del modelo de FT en regiones similares.

	MAE val	MSE val	Región
Cadena	4.51	22.29	2
Colegio	1.15	1.95	2
Jimenez	2.59	12.45	2
Ramon	3.45	14.17	2
Jamaica	1.49	3.47	2
Gilarte	2.91	10.32	6
Yunke	2.36	7.44	6
EEL	5.54	35.61	20
Maricao	6.10	44.95	20
Promedios	3.34	16.96	

En esta tabla puede notarse que el MAE es de 3.34% en promedio, el cual es un valor aceptable mientras que el promedio del MSE es de 16.96, lo cual deja ver que los modelos se ajustan bastante bien a los valores observados de la humedad de suelo en las estaciones pertenecientes a una misma región. La parte (a) de la Figura 4.17 muestra los resultados de la validación con la FT de la estación Chips aplicada a los datos pertenecientes a la estación Jamaica, mientras que la parte (b) muestra la FT de la estación Nazario aplicada a los datos pertenecientes a la estación Yunke. Una línea roja muestra los datos estimados, mientras que una línea azul representa los datos observados en campo. En ambos casos se puede apreciar que cuando la función de transferencia de una estación es aplicada a un sitio con

características climatológicas similares, esta estima con una precisión aceptable la humedad de suelo observada.

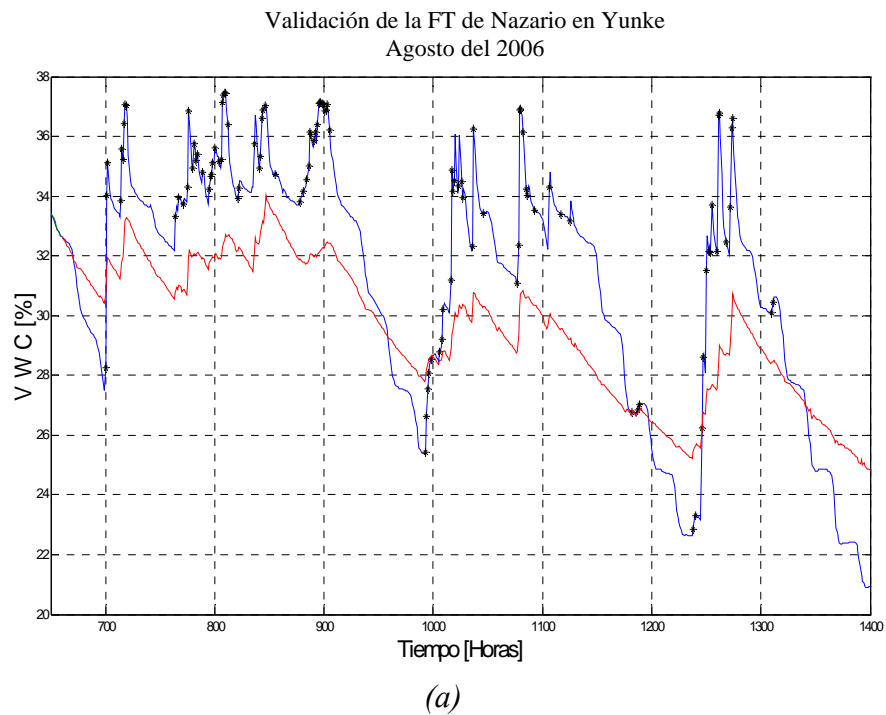
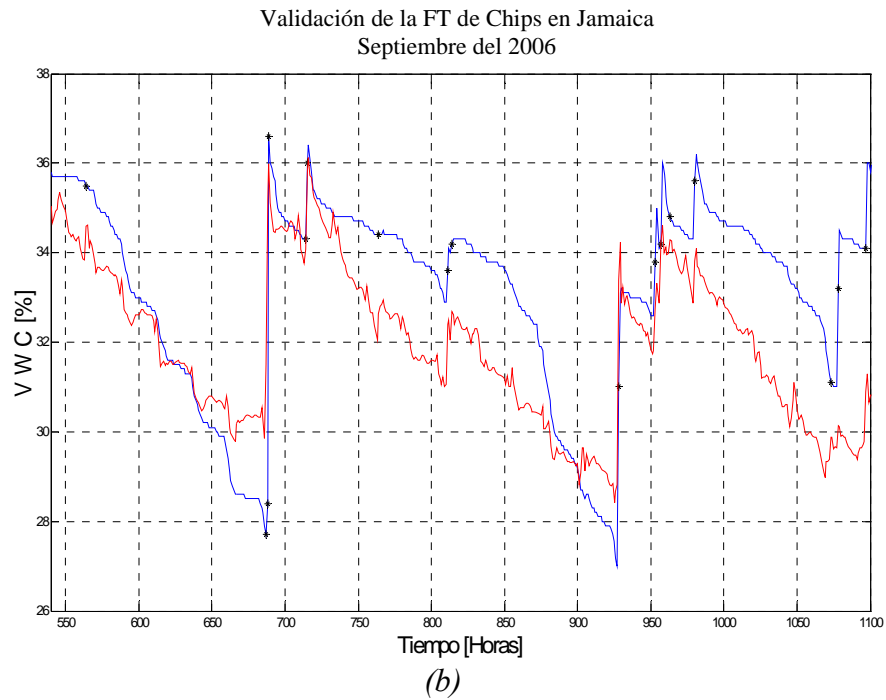


Figura 4.17 Validación de la FT para puntos similares.

La Figura 4.18 muestra los resultados finales de la estimación de la humedad de suelo para una hora en particular, en ella se puede observar la variación espacial de la humedad de suelo y el comportamiento del modelo de estimación horaria de humedad de suelo.

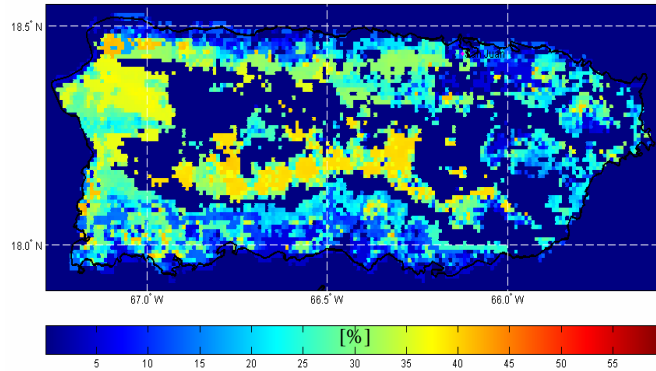


Figura 4.18 Estimación final de la humedad de suelo para una hora en particular.

Esta figura muestra el mapa de humedad de suelo para las 8:00 AM del día 5 de Agosto del 2006. Aquí se puede apreciar la variación espacial de la humedad de suelo. La parte interna del mapa en azul oscuro, se refiere a las regiones donde no se encontró similitud con las estaciones de medición.

Las figuras 4.19 y 4.20 muestran el comportamiento de la humedad de suelo durante un periodo de 24 horas en el mes de agosto del 2006 (específicamente para el día 5 de agosto). En ellas es posible observar la variación horaria de la humedad de suelo y su comportamiento a lo largo del tiempo. Pude notarse que la mayor variación ocurre en las horas de temperaturas máximas (entre las horas 9 y 14) y de temperaturas mínimas (entre las horas 17 y 22), lo cual es un indicativo de que los mapas obtenidos muestran correctamente la variación horaria de la humedad de suelo. Estas figuras corresponden a la estimación final de la humedad de suelo en forma horaria. Durante este periodo de tiempo de 24 horas la humedad sufre variaciones tanto espaciales como temporales, los cuales pueden ser vistos en dichas figuras.

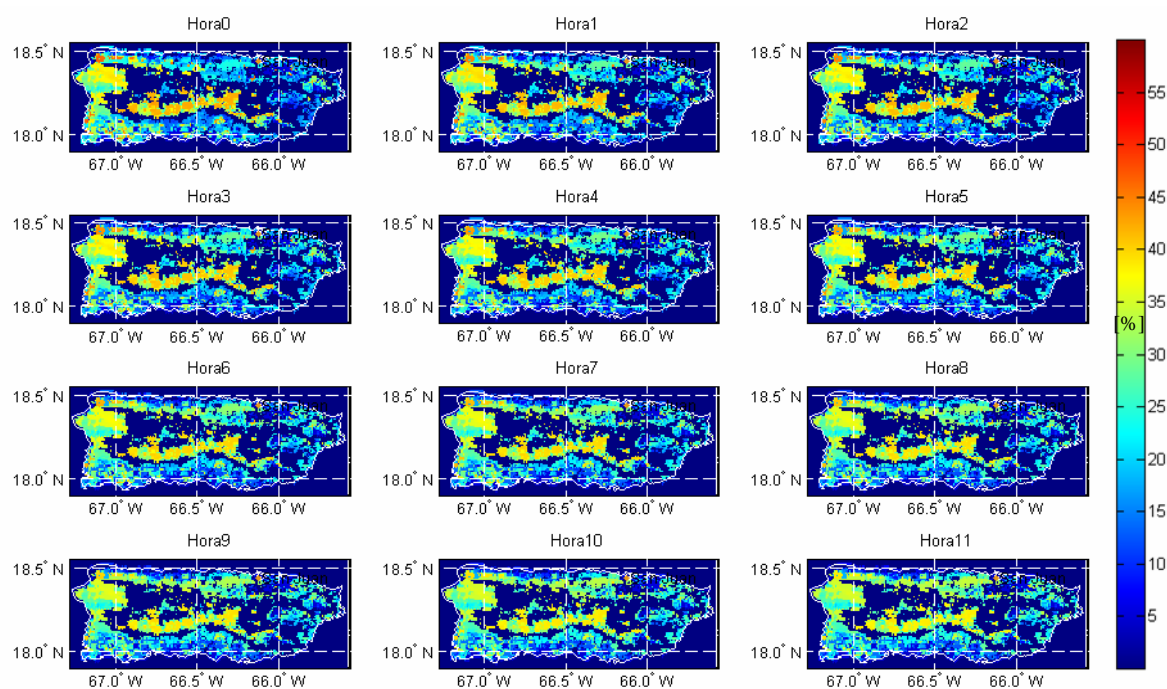


Figura 4.19 Estimación final de la humedad de suelo de la hora 97 a la 108.

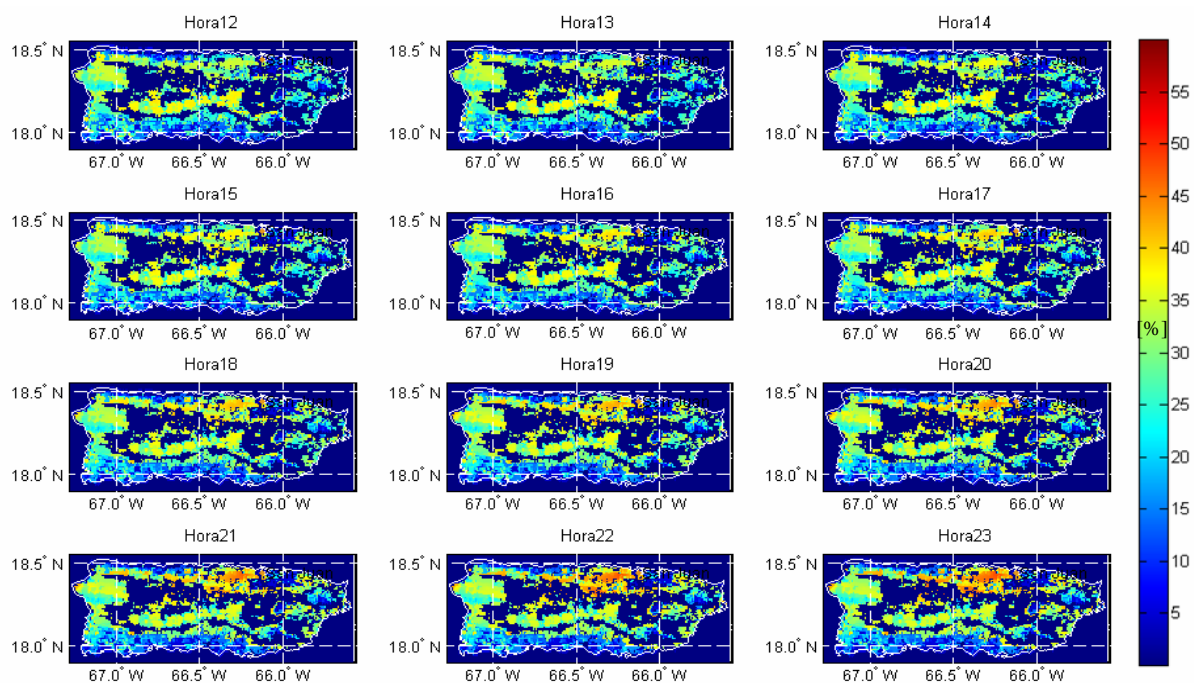


Figura 4.20 Estimación final de la humedad de suelo de la hora 109 a la 120.

5 CONCLUSIONES Y TRABAJO FUTURO

La principal contribución de este estudio es el desarrollo de modelos para la estimación de la humedad de suelo y de la temperatura del aire en forma horaria. Una combinación de técnicas estadísticas, matemáticas y de redes neuronales artificiales son usadas para la estimación de dichos parámetros. Ambos tipos de aproximaciones no han sido usados anteriormente por ningún autor.

Para el caso de la temperatura, se han usado los valores máximos y mínimos diarios de temperatura del aire con el fin de generar el patrón de comportamiento horario de la temperatura del aire. El criterio para la obtención y validación de los modelos fue la minimización del MAE y la verificación del comportamiento del MSE. La metodología aquí introducida consiste en hacer uso de la relación entre la periodicidad de la temperatura y algunas funciones trigonométricas, además de la relación entre la tendencia de la temperatura y la altura sobre el nivel del mar. El modelo desarrollado mostró un excelente comportamiento durante el proceso de validación y además se comporta de manera aceptable con los datos de satélite, lo cual era el objetivo más importante a cumplir por el modelo.

En la estimación de la humedad de suelo se ha usado una pequeña cantidad de variables que se conoce intervienen en el proceso de variación de la misma. Inicialmente se obtuvieron modelos empíricos para la estimación mensual de la humedad de suelo, los cuales fueron usados para la inicialización de los modelos de función de transferencia. Para la obtención y validación de dichos modelos se usó el criterio de minimización del MAE, teniendo en cuenta también los valores del MSE. El comportamiento del modelo final fue muy bueno, lo

cual se comprobó verificando el comportamiento de sus errores. Por otro lado, su aplicación usando datos de sensores remotos y de características texturales y topográficas, exhibe un comportamiento aceptable, lo cual es bueno, pues definitivamente, no es una tarea fácil estimar la humedad de suelo con gran precisión debido a su extrema variabilidad y sensibilidad a pequeños cambios en otras variables ambientales.

Las metodologías aquí propuestas se convierten en herramientas potenciales y poderosas para la estimación de la humedad de suelo y de la temperatura del aire. Además su aplicación no está restringida al área de Puerto Rico, pues como se expuso en las secciones anteriores, estas metodologías pueden ser aplicadas en cualquier sitio con condiciones atmosféricas similares a las de Puerto Rico y donde se cuente con la información necesaria para su aplicación.

Finalmente, para futuros investigadores, se propone el desarrollo de modelos encaminados a mejorar la estimación de la componente estocástica de la temperatura, lo cual es posible hacerse usando modelos ARMA, o alguna otra metodología con la que se pueda simular el comportamiento de un proceso estocástico. Por otro lado, durante el proceso de modelamiento de la humedad de suelo, se vislumbraron dos posibles alternativas. La primera puede ser basada en el uso de la ley de enfriamiento de Newton, relacionándola termodinámicamente hablando con las pérdidas o ganancias de energía que ocurren en el suelo durante el proceso de variación de la humedad de suelo. La segunda puede ser basada en las técnicas de ecuaciones en diferencias y su relación con las series de tiempo, suponiendo que el comportamiento de la humedad de suelo obedece a una ecuación en diferencias parciales de primer orden con coeficientes constantes.

6 REFERENCIAS

Brockwell, P. J., and R. A. Davis, (2002). Introduction to Time Series and Forecasting. Springer-Vela New York, Second Edition.

Box, G. E. P., y Jenkins G. M. (1976). Time Series Analysis Forecasting and Control. Holden-Day, 1976.

Calderón-Arteaga C., (2005). Soil Moisture Estimation Using Passive Remote Sensing Information, ADMI 2005, The symposium on computing at minority institutions,. Rincon, P. R, Oct. 13 – 15.

Cemek, B., Meral, R., Apan, M., y Merdum, H., (2004). Pedotransfer Function for the Estimation of the Field Capacity and Permanent Wilting Point. Pakistan Journal of Biological Sciences. 7(4): 535-541.

Cruzado, J. Harold, (2004). Algoritmos de Estimación e Interpolación de Parámetros Geofísicos. Tesis M. Sc. Departamento de Ingeniería Eléctrica y Computadoras, Universidad de Puerto Rico.

Fulton, R. A., J. P. Breidenbach, D. J. Seo, y D. A. Miller, (1998). The WSR-88D rainfall algorithm. Weather Forecasting, 13, 377–395.

Geerts B., (2002). Empirical estimation of the monthly mean – daily temperature range, Theoretical and Applied. Climatology, vol 000, 1–21.

Hagan, M.T., Demuth, H.B., y y Beal, M, (1996). Neural Network Design, PWS Publishing Company, Boston.

Hamilton, J. D., (1994). Time Series Analysis. Princeton University Press, Princeton, NJ.

Hastead, J.B, (1973). Aqueous Dielectrics, Chapman and Hall. London.

Hilera José y Martines Victor. (1995). Redes Neuronales Artificiales, Fundamentos, Modelos y Aplicaciones, Ed. Addisom Wesley Iberoamericana, S.A.

Huang Jin, Van Den Dool Huug y Georgakakos Constantine P. (1996). Analysis of Model – calculated soil moisture over the United States (1931 – 93) and application to long range temperature forecasts, Journal of Climate, Vol 9, No. 6, 1350 – 1362.

Hunter, S. M., (1996). WSR-88D Radar Rainfall Estimation: Capabilities, Limitations and Potential Improvements. National Weather Digest, v. 20, p. 26-38.

Jakson T. J. y Le Vine David E., (1999). Mapping soil moisture using an aircraft-based passive microwave instrument: algorithm and example, *Journal of Hydrology*, Vol 184, 85 – 99.

Jiang Hongli y Cotton William R., (2004). Soil moisture estimation using an artificial neural network: a feasibility study, *Can. J. Remote Sensing*, Vol. 30, No. 5, pp. 827–839.

Klazura Gerard E., Thomale Jessica M., D. Scott Kelly y Paul Jendrowski, (1999), A comparison of WsR-88D Rain Estimates UIT Gauge Measurements for high and Low Reflectivity Gradient Precipitation Events. National Weather Service, Honolulu, Hawaii.

Kohonen T., (1982). Self-organized Formation of Topologically correct feature Maps. *Biological Cybernetics*, 43, pp. 59-69

Lookingbill Todd y Urban Dean, (2004). An empirical approach towards improved spatial estimates of soil moisture for vegetation analysis, *Landscape Ecology* 19: 417–433.

López Hernández Juan F., (2005). Predicción de Medio Plazo de las Pendientes de Curvas de Demanda Residual de un Agente, Tesis M. Sc. Escuela Técnica Superior de Ingeniería, Universidad Pontificia Comillas

Maneta, M. y S. Schnabel, (2003). Aplicación de redes neuronales artificiales para determinar la distribución espacial de la humedad del suelo en una pequeña cuenca de drenaje. Estudios preliminares, *Estudios de la Zona No Saturada del Suelo* Vol. VI.

Martín Monroy Manuel, (1995). Comportamiento Térmico de Cerramientos Soleados, Un Modelo de Simulación por Diferencias Finitas. Tesis Ph. D. Departamento de Construcción Arquitectónica, Universidad de Las Palmas de Gran Canaria.

Marzen J. L., (2004). Development of a Florida High Resolution Multisensor Precipitation Dataset fo 1996 – 2001 –Quality Control and Verification. Tesis M. Sc. Department of Meteorology, Florida State University.

MathWorks, (2000). Optimization Toolbox for use with Matlab: User's Guide. The MathWorks, Inc.

Mather, J.R., (1978). The climatic water balance in environmental analysis. D.C. Heath and Company, Lexington, Massachussets.

Montgomery Douglas. y Peck Eelizabeth, (2001). Introduction to Linear Regression Analysis, Third edition, Wiley – Interscience.

Montgomery Douglas y Runger George, (1996), Probabilidad y Estadística aplicadas a la ingeniería, Mc. Graw Hill.

Ochoa Maria Alejandra, (2004). Herramientas Inteligentes para Explotación de Información. Trabajo Final, Especialidad en Ingeniería de Sistemas Expertos, Instituto Tecnológico de Buenos Aires.

Pan Feifei, Peters -Lidard Christa D. y Sale Michael J., (2003). Analytical method for predicting surface soil moisture from rainfall observations, water resources research, vol. 39, No. 11.

Pandit, S.M., y S.M. Wu, (1983). Time Series and Systems Analysis with Applications. John Wiley and Sons. New York.

Ramírez-Beltrán, N. D. y Montes, J. A., (2002). Neural Networks to model dynamic systems with time delays, IIE Transactions, 34, 313-327.

Ramírez-Beltrán, N. D., Vásquez R., Cruzado H. y Harmsen E., (2005). A short- and long-Term Memory model to Estimate Soil Moisture. The 31th International Symposium on Remote Sensing of Environment – 2005 “Global Monitoring for Sustainability and Security” Saint Petersburg, Rusia. Junio 20-24.

Ramírez-Beltrán, N. D., Calderón-Arteaga C. y Vásquez R. (2007a). An empirical model to estimate soil moisture over vegetated areas. 21th Conference on hydrology, 87th Annual AMS Meeting. San Antonio, TX, Enero 14-19.

Ramírez-Beltrán, N. D., Vásquez R., Harmsen E., Cruzado H. y Castro J. M., (2007b). Transfer Function Models and Neural Networks to Estimate Soil Moisture. Paper sometido en octubre del 2006 a la revista Journal of the American Water Resources Association.

Reynolds, S.G., (1970). The gravimetric method of soil moisture determination, Part I, A study of equipment and methodological problems, Journal of Hydrology 11: 258-273.

Rodríguez Harry, (2006). A Neural Networks Method to Predict Activity Coefficients for Binary Systems Based on Molecular Functional Group Contribution.. Tesis M. Sc. Departamento de Ingeniería Industrial, Universidad de Puerto Rico.

Schmugge T y .Wang J., (1982). Passive microwave sensing of soil moisture under vegetation canopies, Water Resources Research, Vol 18, No 4.

Sun Donglian y Pinker Rachel T., (2003). Estimation of land surface temperature from a geostationary operational satellite (GOES – 8), Journal of Geophysical Research, vol. 108, No. D11, 4326.

Sun Donglian y Pinker Rachel T., (2004a). Case Study of Soil Moisture Effect on Land Surface Temperature Retrieval, IEEE Geosciences and Remote Sensing Letters, Vol. 1, No. 2, April.

Sun Donglian y Pinker Rachel T., (2004b). Land Surface Temperature Estimation from the Next Generation of Geostationary Operational Environmental Satellites: GOES M–Q, Journal of Applied Meteorology, vol. 43, pp 363 – 372.

Trait, S., K.E. Saxton, y R.I. Papendick, (1979). Pressure and Hydraulic conductivity curves for various soil textures. Internal Report. USDA-ARS, Washington State University, Pullman.

Thornthwaite, C.W., (1948). An approach toward a rational classification of climate. Geographical Review

Veneros Anthony,(2004). A Neural Network Approach to Predict Hurricane Intensity in the North Atlantic Basin. Tesis M. Sc. Departamento de Ingeniería Industrial, Universidad de Puerto Rico.

Wackernagel, Hans, (2003). Multivariate Geostatistics: An introduction with applications, third edition, Springer-Verlag, Berlin Heidelberg, New York.

Wang J., Rich P.M. y Price K. P., (2003). Temporal Responses of NDVI to Precipitation and Temperature in the Central Great Plains, USA. J. Remote Sensing, Vol 24, No 11, 2345 – 2364.

Wang J. y Schmugge T., (1980). An empirical model for the complex dielectric permittivity of soils as a function of water content. IEEE Transactions on Geoscience Remote Sensing, GE-18, 288-295.

Wei, W. S., (1990). Time series analysis: Univariate and multivariate methods. Addison-Wesley Publishing Company.

Westrick, K. J., C. F. Mass, y B. A. Colle, (1999). The limitations of the WSR-88D radar network for quantitative precipitation measurement over the western United States. Bull. Amer. Meteor. Soc., 80, 2289–2298.

Widrow y M.Hoff, (1960). Adaptive switching circuits, IREWESCON Convention Record

ANEXO A ALGORITMOS DE ENTRENAMIENTO RNA

ANEXO A1 ALGORITMO DE *LEVENBERG–MARQUARDT*

Este algoritmo es una modificación del método de Newton, que fue diseñado para minimizar funciones que son la suma de los cuadrados de otras funciones no lineales; es por ello que el algoritmo de *Levenberg–Marquardt*, tiene un excelente desempeño en el entrenamiento de redes neuronales donde el rendimiento de la red esta determinado por el error cuadrático promedio.

El método de Newton para optimizar el rendimiento $e(x)$ es:

$$X_{k+1} = X_k - A_k^{-1} g_k \quad \text{A.1}$$

$$A_k \equiv \nabla^2 e(x)|_{x=x_k} \quad g_k \equiv \nabla e(x)|_{x=x_k} \quad \text{A.2}$$

Si se asume que $e(x)$ es una suma de funciones cuadráticas:

$$e(x) = \sum_{i=1}^n v_i^2(x) = v^T(x)v(x) \quad \text{A.3}$$

El gradiente puede ser escrito entonces en forma matricial:

$$\nabla e(x) = 2J^T(x)v(x) \quad \text{A.4}$$

Ajustando el método de Newton, se obtiene el algoritmo de *Levenberg–Marquardt*

$$x_{k+1} = x_k - [J^T(x_k)J(x_k) + \mu_k I]^{-1} J^T(x_k)v(x_k) \quad \text{A.5}$$

ó

$$\nabla x_k = -[J^T(x_k)J(x_k) + \mu_k I]^{-1} J^T(x_k)v(x_k) \quad \text{A.6}$$

La nueva constante μ determina la tendencia el algoritmo, cuando μ_k se incrementa, este algoritmo se aproxima al algoritmo de pasos descendientes para razones de aprendizaje muy pequeñas; cuando μ_k disminuye, este algoritmo se convierte en el método de Gauss - Newton

El algoritmo comienza con un valor pequeño para μ_k , por lo general 0.01, si en ese paso no se alcanza el valor para $e(x)$ entonces el paso es repetido con μ_k multiplicado por un factor $\vartheta > 1$. Si se ha escogido un valor pequeño de paso en la dirección de paso descendiente, $e(x)$ debería disminuir. Si un paso produce un pequeño valor para $e(x)$, entonces el algoritmo tiende al método de Gauss - Newton, el que se supone garantiza una rápida convergencia. Este algoritmo genera un compromiso entre la velocidad del método de Gauss-Newton y la garantía de convergencia del método de paso descendiente.

Los elementos de la matriz Jacobiana necesarios en el algoritmo de *Levenberg-Marquardt* son como sigue:

$$[J]_{h,l} = \frac{\partial e_{k,q}}{\partial x_l} \quad \mathbf{A.7}$$

Donde x es el vector de parámetros de la red, que tiene la siguiente forma:

$$x^T = [x_1, x_2, \dots, x_n] = [w_{1,1}^1, w_{1,2}^1, \dots, w_{S^1,R}^1, b_1^1, \dots, b_{S^1}^1] \quad \mathbf{A.8}$$

Para utilizar este algoritmo en las aplicaciones para redes multicapa, se redefinirá el término sensibilidad de forma que sea más simple hallarlo en cada iteración.

$$s_{i,h}^m \equiv \frac{\partial e_{k,q}}{\partial n_{i,q}^m} \quad \mathbf{A.9}$$

Donde $h = (q-1)S^M + k$

Con la sensibilidad definida de esta manera, los términos de la matriz Jacobiana pueden ser calculados más fácilmente:

$$[J]_{h,l} = \frac{\partial e_{k,q}}{\partial w_{i,j}^m} = \frac{\partial e_{k,q}}{\partial n_{i,q}^m} \cdot \frac{\partial n_{i,q}}{\partial w_{i,j}^m} = s_{i,h}^m \cdot \frac{\partial n_{i,q}}{\partial w_{i,j}^m} = s_{i,h}^m \cdot a_{i,q}^{m-1} \quad \mathbf{A.10}$$

y para las ganancias:

$$[J]_{h,l} = \frac{\partial e_{k,q}}{\partial b_i^m} = \frac{\partial e_{k,q}}{\partial n_{i,q}^m} \cdot \frac{\partial n_{i,q}}{\partial b_i^m} = s_{i,h}^m \cdot \frac{\partial n_{i,q}}{\partial b_i^m} = s_{i,h}^m \quad \mathbf{A.11}$$

De esta forma, cuando la entrada p_Q ha sido aplicada a la red y su correspondiente salida a_Q^M ha sido computada, el algoritmo Backpropagation de *Levenberg-Marquardt* es inicializado con:

$$S_q^M = -f^M(n_q^M) \quad \mathbf{A.12}$$

Cada columna de la matriz S_q^M debe ser propagada inversamente a través de la red para producir una fila de la matriz Jacobiana. Las columnas pueden también ser propagadas conjuntamente de la siguiente manera:

$$S_q^m = f^m(n_q^m) (W^{m+1})^T S_q^{m+1} \quad \mathbf{A.13}$$

La matrices sensibilidad total para cada capa en el algoritmo de *Levenberg-Marquardt* son formadas por la extensión de las matrices computadas para cada entrada:

$$S^m = [S_1^m \parallel S_2^m] \quad \mathbf{A.14}$$

Para cada nueva entrada que es presentada a la red, los vectores de sensibilidad son propagados hacia atrás, esto se debe a que se ha calculado cada error en forma individual, en lugar de derivar la suma al cuadrado de los errores. Para cada entrada aplicada a la red habrá S^M errores, uno por cada elemento de salida de la red y por cada error se generara una fila de la matriz Jacobiana.

Este algoritmo puede resumirse de la siguiente manera:

1. Se presentan todas las entradas a la red, se calculan las correspondientes salidas y cada uno de los errores según

$$e_q = t_q - a_q^M \quad \text{A.15}$$

se calcula después, la suma de los errores cuadrados para cada entrada $e(x)$

2. Se calculan las sensibilidades individuales y la matriz sensibilidad total y con estas, se calculan los elementos de la matriz Jacobiana.

3. Se obtiene ∇x_k

4. Se recalcula la suma de los errores cuadrados usando $x_k + \nabla x_k$. Si esta nueva suma es más pequeña que el valor calculado en el paso 1 entonces se divide μ por \mathcal{J} , se calcula $x_{k+1} = x_k + \nabla x_k$ y se regresa al paso 1. Si la suma no se reduce entonces se multiplica μ por \mathcal{J} y se regresa al paso 3.

El algoritmo debe alcanzar convergencia cuando la norma del gradiente (Ecuación A4) sea menor que algún valor predeterminado, o cuando la suma de los errores cuadrados ha sido reducida a un error que se haya trazado como meta.

Como puede verse, este algoritmo converge en menos iteraciones que cualquier otro método, aunque requiere muchos cálculos por iteración, debido a que implica el cálculo de matrices inversas. A pesar de su gran esfuerzo computacional sigue siendo el algoritmo de entrenamiento más rápido para redes neuronales cuando se trabaja con un moderado número de parámetros en la red, si el número de parámetros es muy grande utilizarlo resulta poco práctico.

ANEXO A2 RED COMPETITIVA

Teuvo Kohonen presentó en 1982 un modelo de red neuronal con capacidad para formar mapas de características como ocurre en el cerebro; el objetivo de Kohonen era demostrar que un estímulo externo por sí solo, suponiendo una estructura propia y una descripción funcional del comportamiento de la red, es suficiente para forzar la formación de los mapas. Este modelo tiene dos variantes denominadas LVQ (Learning Vector Quantization) y TPM (Topology Preserving Map) o SOM (Self Organizing Map), ambas se basan en el principio de formación de mapas topológicos para establecer características comunes entre las informaciones (vectores) de entrada a la red, siendo de una sola dimensión el de LVQ y bidimensional o tridimensional la red SOM.

Durante la etapa de entrenamiento, se presenta a la red un conjunto de datos para establecer las categorías en función de su semejanza. Estas categorías servirán durante la fase de funcionamiento para clasificar nuevos datos que se presenten a la red. En el caso de existir más patrones de entrenamiento que neuronas de salida, más de uno deberá asociarse con la misma neurona, es decir pertenecerán a la misma clase. En este modelo el aprendizaje no concluye con presentarle una vez los patrones de entrada, sino que habrá que repetir el proceso varias veces para refinar el mapa topológico de salida, de tal forma que cuantas más veces se presenten los datos, más mas selectiva será la clasificación.

El algoritmo de aprendizaje utilizado para establecer los valores de los pesos de las conexiones entre las N neuronas de entrada y las M de salida es el siguiente:

a) se inicializan los pesos (W_{sj}) con valores aleatorios pequeños y se fija la zona inicial de vecindad entre las neuronas de salida.

b) a continuación se presenta a la red una información de entrada en forma de vector $p = (p_1, p_2, \dots, p_N)$, cuyas componentes p_i serán valores continuos.

c) puesto que se trata de un aprendizaje competitivo, se determina la neurona vencedora de la capa de salida; aquella C_s cuyo vector de pesos W_s sea el más parecido a la información de entrada p . Para ello se calculan las distancias o diferencias entre ambos vectores, considerando una por una todas las neuronas de salida, generalmente se usa la distancia euclídea o también la expresión del producto punto entre dos vectores, así:

$$net_i = \vec{p}_i \cdot \vec{W}_s + b_s \quad \text{A.16}$$

Donde: p_i : es el componente i -ésimo del vector del vector de entrada; ${}_iW_s$: es el peso de la conexión entre la neurona i de la capa de entrada y la neurona s de la capa de salida; b_s : es el *bias* asociado. Si todos los *biases* son cero, el máximo *net* que una neurona puede tener es cero.

d) una vez localizada la neurona vencedora (C^*), se actualizan los pesos de las conexiones entre las neuronas de entrada y dicha neurona, así como los de las conexiones entre las de entrada y las neuronas vecinas de la vencedora. El término $\alpha(q)$ es el coeficiente de aprendizaje o parámetro de ganancia, donde $0 \leq \alpha(q) \leq 1$ y decrece con el número de iteraciones q del proceso de entrenamiento, de tal forma que cuando se han realizado un gran numero de iteraciones durante el aprendizaje, su valor es prácticamente nulo, y la modificación de los pesos es insignificante. Para hallar α se usa una de las siguientes expresiones (Ochoa 2004 y López 2005)

$$\alpha(q) = \frac{1}{q} \quad \text{ó} \quad \alpha(q) = \alpha_1 \left(1 - \frac{q}{\alpha_2} \right) \quad \text{A.17}$$

Siendo α_1 un valor de 0.1 ó 0.2 y α_2 que decrece con el número total de iteraciones (q) del proceso de aprendizaje, que por lo general tiene un valor cercano a diez mil (10000) y no menor de quinientos (500) para obtener resultados aceptables. De tal forma, cuando se ha presentado al proceso un gran número de iteraciones el valor de α_1 es prácticamente cero y la modificación de los pesos insignificante. El proceso debe repetirse, volviendo a presentar todo el juego de patrones de aprendizaje (p_1, p_2, \dots, p_N) hasta obtener la salida deseada.

e) El aprendizaje no es proporcional a la salida de la neurona $a_s(q)$, en lugar de ello el aprendizaje ocurre cuando la neurona s sea miembro de cierta vecindad $X(q)$.

En resumen, lo que hace una red de Kohonen es realizar una tarea de clasificación, puesto que la neurona de salida activada por una entrada dada, representa la clase a la que pertenece dicha información de entrada; además ante otra entrada parecida se activará la misma neurona de salida, u otra cercana a la anterior debido a la semejanza entre las clases, así se garantiza que las neuronas topológicamente próximas sean sensibles a entradas físicamente similares; por esta causa la red es especialmente útil para establecer relaciones internas de la información que eran desconocidas previamente.

Estructura de una Red Competitiva

Definición:

Sea a una función de activación o transferencia para una red neuronal competitiva y n un vector de distancias así:

$$a = \text{Compet}(\text{net}) \quad \text{A.18}$$

Entonces:

$$a_s = \begin{cases} 1, & C_s = C^* \\ 0, & C_s \neq C^* \end{cases} \quad \text{A.19}$$

La red crea un conjunto de vectores prototipo W de tamaño igual al número de neuronas de salida.

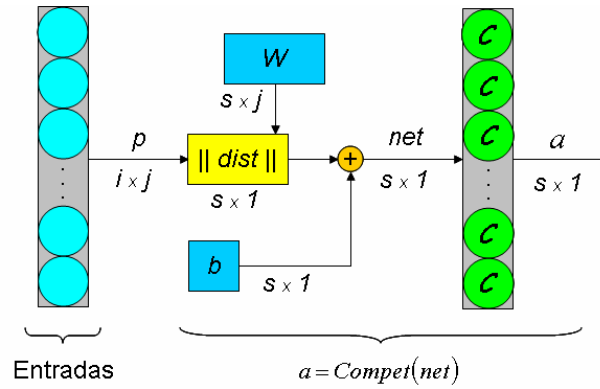


Figura A.1 Capa competitiva.

En la Figura 2.6, la caja $\|dist\|$ acepta el vector de entrada p y la matriz de pesos W y produce un vector de s elementos. La entrada neta net se calcula según la distancia entre cada vector de entrada p_i y cada uno de los vectores prototipo W_s ; siendo para cada neurona proporcional al ángulo θ_s entre el vector p_i y el vector W_s , debido a las propiedades del producto punto.

$$net_i = \left\| \vec{p}_i \right\| \cdot \left\| \vec{W}_s \right\| \cos \theta_s + b \quad \text{A.20}$$

La función de transferencia competitiva asigna una salida de 1 a la neurona cuyo vector de pesos apunte en la dirección más cercana al vector de entrada

Regla de Aprendizaje de Kohonen

Los pesos de la neurona ganadora (una fila de la matriz de pesos de entrada) son ajustados según la siguiente regla (Veneros 2004):

$${}_iW(q) = {}_iW(q-1) + \alpha(p(q) - {}_iW(q-1)) \quad \mathbf{A.21}$$

Donde los pesos de en la iteración (q) son actualizados usando los pesos de la iteración ($q-1$). Así, la neurona cuyo vector de pesos fue cercano al vector de entrada, se actualizará para ser la más cercana. El resultado es que la neurona ganadora es la más probable a ganar la competencia la próxima vez que se le presente un vector de características similares. Así, cada neurona en la capa cercana a un grupo de vectores de entrada ajustará sus pesos hacia ese grupo de vectores.

ANEXO B PROGRAMAS EN MATLAB

ANEXO B1 REGIONES CLIMATOLÓGICAS

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%ESTE PROGRAMA IDENTIFICA LAS SIMILARIDADES CLIMATOLÓGICAS
%PARA PUERTO RICO USANDO UNA RED NEURONAL COMPETITIVA
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clear all;
clc;
close all;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Longitud  Latitud  NDVI  Rain  altura  Estaciones
sensors=[
-67.14358333    18.31994444 0.81706 125 89.777; %Cadena
-67.04588889    18.06022222 0.80013 94 83.364; %Caobos
-67.00355556    18.35447222 0.71598 296 60.479; %Chips
-67.13908333    18.21166667 0.75301 227 9.2494; %Colegio
-67.07458333    18.03316667 0.63851 59 38.855; %E. Exp
-66.76727778    18.14636111 0 0 0; %gilarte
-67.0565         18.46972222 0 0 0; %isabela
-67.24083333    18.35766667 0.833 188 35.401; %Jimenez
-67.03305556    18.02666667 0 0 0; %lajas
-67.00397222    18.15733333 0 0 0; %maricao
-67.13011111    18.23708333 0.79135 348 34.729; %Nazario
-67.13222222    18.33136111 0.77308 240 118.46; %Ramon
-67.23236111    18.31638889 0.71076 160 11.29; %Jamaica
-66.047 18.403 0.36052 13.43972111 24.6888; %UPR
-65.82 18.321 0.57144 10.91080111 364.5408]; %Yunke

sensors(:,3)=sensors(:,3)*1000;
load allprpixels %Informacion de cada grid georefenciada - 1 km 8701
clear fangulo nalt
load Temp_DN; %Temperaturas de Enero a Dic del 2006 1 km - 8701
load NDVI_2006 %NDVI Enero a Dic 2006 1 km - 8701
load Rain2006 %Rain Enero a Dic 2006 1 km - 8701

for i=1:15
    peso=dist(sensors(i,1:2),geo(:,1:2)'); %Vecino mas cercano a los
    %puntos de referencia de cada estacion
    [corr,pos]=sort(peso);
    sD(i,1)=D(pos(1),8); %temperaturas de dia en agosto
    sN(i,1)=N(pos(1),8); %temperaturas de noche en agosto
    %sang(i,1)=fangulo(pos(1),1);%angulos en las estaciones
```

end

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%VARIABLES CLIMATOLÓGICAS %%%%%%%%%%

matriz=[NDVI(:,10) Rain2006(:,10)*25.4 alt D(:,8) N(:,8)];
%%%%%%%%%Estandarización de variables
V=(NDVI(:,10)-mean(NDVI(:,10)))/std(NDVI(:,10));
Rain=Rain2006(:,10)*25.4;
Rain=(Rain-mean(Rain))/std(Rain);
alt=(alt-mean(alt))/std(alt);
Dt=(D(:,8)-mean(D(:,8)))/std(D(:,8));
Nt=(N(:,8)-mean(N(:,8)))/std(N(:,8));
matriz1=[NDVI(:,10) Rain2006(:,10)*25.4 alt D(:,8) N(:,8)];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
radars=[geo V Rain alt Dt Nt];%Le estoy agregando la latitud y la longitud
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clear D N alt Rain2006 NDVI
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
stations=[radars(:,3:end)];
reg=20; %Número de clases sugeridas
ite=500*20 %Número de iteraciones que debe completar la red
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%ENTRA A RNA NO SUPERVISADA%%%%%%%%%
    stations=stations';
    u=minmax(stations);
    %net.adaptFcn to 'trains'
    net=newc(minmax(stations),reg);
    net.trainParam.epochs = ite;
    net=train(net,stations);
    grupo=sim(net,stations);
    clase =vec2ind(grupo);
    clase=clase';
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%FINALIZA CLASIFICACIÓN%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%ENCUENTRA LA REGIÓN A LA QUE PERTENECE CADA ESTACIÓN%%
    for i=1:12
        peso=dist(sensors(i,1:2),geo(:,1:2)); %Vecino mas cercano a
        %los puntos de referencia de cada estacion
        [corr,pos]=sort(peso);
        clasel=clase(1:8701,1);
        clas(i,1)=clasel(pos(1),1);
    end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%GRÁFICOS%%%%%%%%%
    c=1:1:20;
    [a b]=hist(clase,c);
    bar(a);
```



```
mapapr_ver2(clase);
save clastemp_prueba1 clase clas
```

ANEXO B2 ESTIMACIÓN DE LA HUMEDAD DE SUELO HORARIA

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%ESTE PROGRAMA CALCULA LOS PARÁMETROS DE LA FUNCIÓN
%DE TRANSFERENCIA PARA LA HUMEDAD DE SUELO.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clear all;
clc;
close all;
warning off all;

A=[ '1  = Cadena      ' ; '2  = Caobos      ' ; '3  = Chips      ' ; ...
    '4  = Colegio    ' ; '5  = EEL        ' ; '6  = Gilarte    ' ; ...
    '7  = Isabela    ' ; '8  = Jamaica    ' ; '9  = Jimenez    ' ; ...
    '10 = Lajas      ' ; '11 = Maricao     ' ; '12 = Nazario   ' ; ...
    '13 = Ramon      ' ; '14 = UPR        ' ; '15 = Yunke     ' ];

disp(A)

station_number=input('Enter the station number >')

selec=station_number;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
switch selec
    case 1
        filename='Cadena.txt';
    case 2
        filename='Caobos.txt';
    case 3
        filename='Chips.txt';
    case 4
        filename='Colegio.txt';
    case 5
        filename='EEL.txt';
    case 6
        filename='Gilarte.txt';
    case 7
        filename='Isabela.txt';
    case 8
        filename='Jamaica.txt';
    case 9
        filename='Jimenez.txt';
    case 10
        filename='Lajas.txt';
    case 11
        filename='Maricao.txt';
    case 12
```

```

        filename='Nazario.txt';
        case 13
        filename='Ramon.txt';
        case 14
        filename='UPR.txt';
        case 15
        filename='Yunke.txt';
%
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
X=load([cd '\' filename]); %Almacena las variables de humedad de suelo y
lluvia instantanea para cada estacion
sample=fix(0.5*length(X(:,1))); %muestras de estudio

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%

if selec==1;filename1=num2str('Aug 01 to Sep 30 of 2006');end
if selec==2;filename1=num2str('Oct 15 to Nov 30 of 2006');end
if selec==3;filename1=num2str('Ago 13 a Sep 30 del 2006');end
if selec==4;filename1=num2str('Aug 01 to Sep 30 of 2006');end
if selec==5;filename1=num2str('Aug 19 to Sep 05 of 2006');end
if selec==6;filename1=num2str('Jul 01 to Aug 31 of 2006');end
if selec==7;filename1=num2str('May 01 to Jun 30 of 2006');end
if selec==8;filename1=num2str('Aug 13 to Sep 30 of 2006');end
if selec==9;filename1=num2str('Aug 01 to Sep 30 of 2006');end
if selec==10;filename1=num2str('Apr 01 to Apr 30 of 2006');end
if selec==11;filename1=num2str('Jul 01 to Aug 31 of 2006');end
if selec==12;filename1=num2str('Aug 01 to Sep 30 of 2006');end
if selec==13;filename1=num2str('Ago 01 a Sep 30 del 2006');end
if selec==14;filename1=num2str('Jul 01 to Aug 31 of 2006');end
if selec==15;filename1=num2str('Jul 01 to Aug 31 of 2006');end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%

figure(1) %Mostrar los sensores de humedad de suelo
subplot(3,1,1);
plot(1:length(X(:,1)),X(:,1),1:length(X(:,1)),X(:,4));grid
title([num2str(filename(1:length(filename)-4)), ' Sensor # 1'
], 'FontSize',9)
ylabel('V W C [%]', 'FontSize',9)
xlabel('Time [Hour]', 'FontSize',9)

subplot(3,1,2);
plot(1:length(X(:,1)),X(:,2),1:length(X(:,1)),X(:,4));grid
title([num2str(filename(1:length(filename)-4)), ' Sensor # 2'
], 'FontSize',9)
ylabel('V W C [%]')
xlabel('Time [Hour]', 'FontSize',9)

```

```

subplot(3,1,3);
plot(1:length(X(:,1)),X(:,3),1:length(X(:,1)),X(:,4));grid
title([num2str(filename(1:length(filename)-4)), ' Sensor # 3'
], 'FontSize',9)
ylabel('V W C [%]', 'FontSize',9)
xlabel('Time [Hour]', 'FontSize',9)

S_N=input('See the figure and choose the sensor number (1 or 2 or 3) >')
%(ESCOJER EL SENSOR DESEADO 1 2 3)

    if S_N==1,
        filename2=num2str('S1');
    end
    if S_N==2,
        filename2=num2str('S2');
    end
    if S_N==3,
        filename2=num2str('S3');
    end

close all
% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% %Quitandole la periodicidad y tendencia a la temperatura
% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
n=1:length(X(:,5));%longitud del vector temperatura
sin24=sin(2*pi*n/24);
cos24=cos(2*pi*n/24);
sin12=sin(2*pi*n/12);
cos12=cos(2*pi*n/12);
sin8=sin(2*pi*n/8);
cos8=cos(2*pi*n/8);
sin6=sin(2*pi*n/6);
cos6=cos(2*pi*n/6);
uno=ones(length(n),1);          % observaciones en el periodo
M=[uno sin24' cos24' sin12' cos12' sin8' cos8' sin6' cos6'];
a=inv(M'*M)*M'*X(:,5);          % calcula coeficientes compo periodico
ts=M*a;                          % temperatura componente periodico
Ten=X(:,5)-ts; %Temperatura sin el componente periodico estacional
%[Ten]=Anomalias(Ten);
Ten=detrend(Ten)';%temperatura componente de tendencia
X(:,5)=Ten; %
% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
Ht=detrend(X(:,S_N)); %le quita la tendencia a la Humedad
t=1:length(X(:,S_N));
t=t';
pt=polyfit(t,X(:,S_N),1);
T=polyval(pt,t); %crea una recta de tendencia

damos=-min(Ht);
X(:,S_N)=Ht+damos; % le resta el valor minimo

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Z=[X(:,5) X(:,4) X(:,S_N)]; %Almacena las variables de humedad de suelo
%y lluvia instantanea para cada estacion

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Iniciación del modelo
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
ipesos=[2*rand(100,1) 2*rand(100,1) -0.9+1.8*rand(100,1) 2*rand(100,1) ...
        2*rand(100,1) -0.9+1.8*rand(100,1) 2*rand(100,1) 2*rand(100,1)...
        -0.9+1.8*rand(100,1) 0.01*rand(100,1)];
ihours=fix(12+84*rand(100,1));
for i=1:100
    iSSE(i)=soil_moisture1TR_ver3(ipesos(i,:),Z(1:sample,:),ihours(i));
end
[mSSE,pos]=min(iSSE);
xo=ipesos(pos,:);
inicial=ihours(pos);
%xo=pesos; %coeficientes iniciales
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Fase de Optimización con Restricciones
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
[MSE1,Yo]=soil_moisture1TR_ver3(xo,Z(1:sample,:), inicial);
options=optimset('MaxFunEvals',5000);
%options=optimset('MaxIter',1500);
%xf=fminunc(@(x) soil_moisture1TR(x,Z),xo,options)
xf=fmincon(@(x)
soil_moisture1TR_ver3(x,Z(1:sample,:),inicial),xo,[],[],[],[],[-Inf -Inf -
0.9 -Inf ...
        -Inf -0.9 -Inf -Inf -0.9 0],[Inf Inf 0.9 Inf Inf 0.9 Inf Inf 0.9
0.1]);
%[MSE2,Yf]=soil_moisture1TR_ver3(xf,Z(1:sample,:),inicial);

[SSE2,Yf,Yx,final]=soil_moisture_hooke_ver3(xf,Z(1:sample,:),inicial);
%break
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Zv=Z(1:sample,:);
Zv2=Z(1:sample,:);
Yv=Yf;
Yv2=Yo;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%PARA EVALUAR LA FT SOLO SE CAMBIA EL ARCHIVO A CARGAR.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% load TF_SM_ver3_Isabela_S1.mat;
% [MSE,Yf]=soil_moisture1TR_ver3(xf,Zv,final);
% Yv=Yf;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%

```

```

for testing=sample+1:length(Z(:,1))
    Zv=[Zv;Z(testing,1:2) 0];
    Zv2=[Zv2;Z(testing,1:2) 0];
    [SSE3,Yvf]=soil_moisture1TR_ver3(xf,Zv,final);
    Yv=[Yv;Yvf(length(Yvf),1)];
    Zv(testing,3)=Yvf(length(Yvf),1);
    [SSE3x,Yvf2]=soil_moisture1TR_ver3(xo,Zv2,final);
    Yv2=[Yv2;Yvf2(length(Yvf2),1)];
    Zv2(testing,1)=Yvf2(length(Yvf2),1);
end
Zf=Z;
e=Zf(:,3)-Zv(:,3);
SSEf=e'*e;
MAEf=sum(abs(e))/length(e)
MSEf=SSEf/length(e);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Figures
Lp=find(Zf(:,2)>0);
figure(1) %comparar funcion de transferencia con coeficientes iniciales
plot(1:length(Zf(:,3)),Zf(:,3)+T-damos,Lp,Zf(Lp,3)+T(Lp)-
damos,'k*',3:sample-inicial,Yo(3:sample-inicial)+T(3:sample-inicial)-
damos);grid
title(['Funcion de Transferencia para la Humedad de Suelo en
',num2str(filename(1:length(filename)-4))],num2str(filename1)})
ylabel('V W C [%]')
xlabel('Tiempo [Horas]')

figure(2) %comparar funcion de transferencia con coeficientes optimos
Tl=T(final+1:end);
Zf=Zf(final+1:end,:);
Lp=find(Zf(:,2)>0);
plot(1:length(Zf(:,3)),Zf(:,3)+Tl-damos,Lp,Zf(Lp,3)+Tl(Lp)-
damos,'k*',1:sample-final,Yf+Tl(1:sample-final)-damos,(sample-
final+1:length(Zf(:,3))),Yv(sample-final+1:end)+Tl(sample-final+1:end)-
damos);grid
title(['Funcion de Transferencia para la Humedad de Suelo en
',num2str(filename(1:length(filename)-
4))],num2str(filename1)},'FontSize',16)
ylabel('V W C [%]','FontSize',16)
xlabel('Tiempo [Horas]','FontSize',16)
h = legend('Observado','Eventos de Lluvia','Ajuste','Estimado',2);
% h = legend('Observed','Rainfall events','Trained','Estimated',2);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%ERRORS
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
sm_train=Zf(1:sample-final,3)+Tl(1:sample-final)-damos;
sm_fit=Yf(1:sample-final)+Tl(1:sample-final)-damos;
sm_validation=Zf(sample-final+1:end,3)+Tl(sample-final+1:end)-damos;

```

```

sm_est=Yv(sample-final+1:end)+T1(sample-final+1:end)-damos;

MAEfit=mean(abs(sm_train-sm_fit))
MSEfit=mean((sm_train-sm_fit).^2)

MAE=mean(abs(sm_validation-sm_est))
MSE=mean((sm_validation-sm_est).^2)
Errors=[MAEfit MSEfit MAE MSE]%MATRIZ QUE CONTIENE LOS ERRORES DE AJUSTE Y
VALIDACIÓN

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% save(['TF_SM_ver3_',num2str(filename(1:length(filename)-
4)),'_',num2str(filename2)],'Errors','xf','final')
%save ResChipsS1_aug2006 Errors xf; %Cambiar el nombre a cada archivo para
salvar

```

ANEXO B3 ESTIMACIÓN DE LA TEMPERATURA HORARIA

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%ESTE PROGRAMA IDENTIFICA EL MODELO PARA LA TEMPERATURA USANDO ANÁLISIS
%DE REGRESION MÚLTIPLE CON EL FIN DE ENCONTRAR MODELOS DE ESTIMACION DE
%TEMPERATURA HORARIA A PARTIR DE LOS VALORES MAXIMOS Y MINIMOS DE
%TEMPERATURA DIARIOS
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
clear all;
clc;
close all;
warning off all;

A=[ '1  = Cadena      '; '2  = Caobos      '; '3  = Chips       '; ...
    '4  = Colegio     '; '5  = EEL        '; '6  = Gilarte     '; ...
    '7  = Isabela     '; '8  = Jamaica    '; '9  = Jimenez     '; ...
    '10 = Lajas       '; '11 = Maricao     '; '12 = Nazario    '; ...
    '13 = Ramon       '; '14 = UPR        '; '15 = Yunke      '];

disp(A)

station_number=input('Enter the station number >')

selec=station_number;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
switch selec
    case 1
        filename='Cadena.txt';
    case 2
        filename='Caobos.txt';
    case 3
        filename='Chips.txt';

```

```

        case 4
filename='Colegio.txt';
        case 5
filename='EEL.txt';
        case 6
filename='Gilarte.txt';
        case 7
filename='Isabela.txt';
        case 8
filename='Jamaica.txt';
        case 9
filename='Jimenez.txt';
        case 10
filename='Lajas.txt';
        case 11
filename='Maricao.txt';
        case 12
filename='Nazario.txt';
        case 13
filename='Ramon.txt';
        case 14
filename='UPR.txt';
        case 15
filename='Yunke.txt';
%
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
x=load([cd '\' filename]); %Almacena las variables de humedad de suelo y
%lluvia instantanea para cada estacion
Samples=input('Enter the number of days for training (1 for 10 days or 2
for 20 days 3 for 30 days)>');
if Samples==1
    sample=240; %Primeras muestras para entrenamiento
end
if Samples==2
    sample=480; %Primeras muestras para entrenamiento
end
if Samples==3
    sample=720; %Primeras muestras para entrenamiento
end
training=x(1:sample,5);%Data for training
validation=x(sample:end,5);%Data for validation

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%INICIA AJUSTE %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
x=training;
[n c]=size(x);          % n=numero de observaciones, c=numero de columnas
%89.777
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
H=[ -0.5 83.364 60.479 9.2494 38.855 864.97 168.63 11.29 35.401 20.03
599.34...

```

```

34.729 118.46 24.6888 364.5408];

if selec==1;filename1=num2str('Aug 01 to Sep 30 of 2006');H=H(1);end
if selec==2;filename1=num2str('Oct 15 to Nov 30 of 2006');H=H(2);end
if selec==3;filename1=num2str('Aug 13 to Sep 30 of 2006');H=H(2);end
if selec==4;filename1=num2str('Aug 01 to Sep 30 of 2006');H=H(4);end
if selec==5;filename1=num2str('Aug 19 to Sep 05 of 2006');H=H(5);end
if selec==6;filename1=num2str('Jul 01 to Aug 31 of 2006');H=H(6);end
if selec==7;filename1=num2str('Jul 01 to Aug 31 of 2006');H=H(7);end
if selec==8;filename1=num2str('Aug 13 to Sep 30 of 2006');H=H(8);end
if selec==9;filename1=num2str('Aug 01 to Sep 30 of 2006');H=H(9);end
if selec==10;filename1=num2str('Apr 01 to Apr 30 of 2006');H=H(10);end
if selec==11;filename1=num2str('Jul 01 to Aug 31 of 2006');H=H(11);end
if selec==12;filename1=num2str('Aug 01 to Sep 30 of 2006');H=H(12);end
if selec==13;filename1=num2str('Aug 01 to Sep 30 of 2006');H=H(13);end
if selec==14;filename1=num2str('Jul 01 to Aug 31 of 2006');H=H(14);end
if selec==15;filename1=num2str('Jul 01 to Aug 31 of 2006');H=H(15);end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% MODELO DE LA TEMPERATURA
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Este modelo usa los datos de temperatura maxima y minima para encontrar
un
%modelo apropiado de temperatura en terminos de estas cantidades para
poder
%trabajar luego con los datos de modis
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%

peri=floor(n/24);          % Numero de dias en el set de datos (número
entero)
obs_peri=peri*24;          % Número de horas en el periodo

t_peri=x(1:obs_peri,:);    % toma periodos completos de dias para temp
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% en este for se consiguen las temperaturas maximas y minimas, ademas de
% las sumatorias de la lluvia
for i=1:peri
    T2(i)=max(t_peri((i-1)*24+1:i*24)); % T2=Tmax per day
    T1(i)=min(t_peri((i-1)*24+1:i*24)); % T1=Tmin per day
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%buscando la tendencia a partir de las temperaturas maxima y minima
d=T2-T1; %Amplitud diaria de la temperatura
media=(T1+T2)./2; %promedio diario

figure(1)
plot(media)

```



```

title('Daily mean temperature Training (max & Min)')
ylabel('Temperature [°C]')
xlabel('Time [Day]')

[n c]=size(t_peri);

[T_tr,TT_tr,TP_tr,ang,coeff,R2,series]=simul_templ(media,d,t_peri,obs_peri
,peri,n,H);

coeff'
ang
R2
series=[detrend(t_peri) series];
% "series" es una matriz que contiene en su primera columna a
detrend(t_peri)
%y en las siguientes columnas los valores de las series de seno y coseno
%creados a partir de maximos y minimos. Esto con el fin de verificar el
%modelo en statgraphics

%T_tr es la temperatura encontrada en el modelo
%TT_tr es la tendencia encontrada por el modelo
%TP_tr es la periodicidad encontrada por el modelo

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%En los siguientes gráficos se presenta una comparacion entre la
%temperatura observada y su componente estacional y de tendencia obtenido
%de sus maximos y %minimos diarios, la figura 5 presenta la temperatura
%sin el componente periodico ni la tendencia.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
figure(2)
plot(t_peri)
title('Temperatura Observada en la Estación')
ylabel('Temperatura [°C]')
xlabel('Tiempo [Horas]')
figure(3)
plot(T_tr)
title('Componentes Periódico y Determinístico (max & Min)')
ylabel('Temperatura [°C]')
xlabel('Tiempo [Horas]')
figure(4)
nt=length(T_tr);
time=1:nt;
plot(time,T_tr,'r',time,t_peri,'b')
title('Componentes Periódico y Determinístico (max & Min) Vs. Temperatura
Observada')
ylabel('Temperatura [°C]')
xlabel('Tiempo [Horas]')

```

```

et=t_peri-T_tr ; %verifica el error entre la temperatura observada y
estimada
MAE=mean(abs(et)) %MAE de fitting
MSE=mean(et.^2) %MSE de fitting
figure(5)
autocorr(et,100);
figure(6)
parcorr(et,100);

%save(['Model_Id_',num2str(filename(1:length(filename)-
4))], 'coeff', 'ang');

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%INICIA
VALIDATION%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%

x1=validation;
[n1 c1]=size(x1); % n=numero de observaciones, c=numero de columnas

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% MODELO DE LA TEMPERATURA
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Este modelo usa los datos de temperatura maxima y minima para encontrar
un
%modelo apropiado de temperatura en terminos de estas cantidades para
poder
%trabajar luego con los datos de modis
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%

peril=floor(n1/24); % number of days in the data set (número
entero)
obs_peril=peril*24; % number of hours in the period

t_peril=x1(1:obs_peril,:); % toma periodos completos de dias para
temp
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% en este for se consiguen las temperaturas maximas y minimas, ademas de
% las sumatorias de la lluvia
for i=1:peril
    T21(i)=max(t_peril((i-1)*24+1:i*24)); % T2=Tmax por dia
    T11(i)=min(t_peril((i-1)*24+1:i*24)); % T1=Tmin por dia
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%buscando la tendencia a partir de las temperaturas maxima y minima
d1=T21-T11; %Amplitud diaria de la temperatura
medial=(T11+T21)./2; %promedio diario

figure(7)
plot(medial)

```

```

title('Validacion Promedio de Temperatura Diaria (max & Min)')
ylabel('Temperatura [°C]')
xlabel('Tiempo [Dias]')

[n1 c1]=size(t_peril);

[T_val,TT_v,ts_v]=Val_temp(medial,d1,t_peril,obs_peril,peril,n1,ang,coeff,
H);

%T_val es la temperatura encontrada en validacion
%TT_v es la tendencia encontrada por el modelo en validacion
%TP_v es la periodicidad encontrada por el modelo en validacion

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%En los siguientes graficos se presenta una comparacion entre la
%temperatura observada y su componente estacional y de tendencia obtenido
%de sus maximos y minimos diarios en la fase de validacion.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
figure(8)
plot(t_peril)
title('Temperatura Observada en la Estación')
ylabel('Temperatura [°C]')
xlabel('Tiempo [Horas]')
figure(9)
plot(T_val)
title('Componentes Periódico y Determinístico (max & Min)')
ylabel('Temperature [°C]')
xlabel('Time [Hour]')
figure(10)
nt=length(T_val);
time=1:nt;
plot(time,T_val,'r',time,t_peril,'b')
title('Componentes Periódico y Determinístico (max & Min)) Vs Temperatura
Observada')
ylabel('Temperatura [°C]')
xlabel('Tiempo [Horas]')

et1=t_peril-T_val ; %verifica el error entre la temperatura observada y
estimada
MAE1=mean(abs(et1)) %MAE de Validacion
MSE1=mean(et1.^2) %MSE de Validacion

figure(11)
autocorr(et1,100);
figure(12)
parcorr(et1,100);

t_ori=[t_peri;t_peril]; %temperatura original
figure(13) %comparar funcion de transferencia con coeficientes optimos

```

```

plot(1:length(t_ori),t_ori,1:sample,T_tr,sample+1:length(T_val)+sample,T_val);grid
title(['Validación del Modelo de Temperatura en',num2str(filename(1:length(filename)-
4))],num2str(filename1)},'FontSize',16)
ylabel('Temperatura [°C]','FontSize',16)
xlabel('Tiempo [Horas]','FontSize',16)
h = legend('Observado','Ajustado','Estimado',2);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%ERRORES DE VALIDACION
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Errors=[MAE MSE MAE1 MSE1]
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%SALVANDO PARAMETROS DEL MODELO
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%save(['ModelId_Temp_',num2str(filename(1:length(filename)-
4))], 'Errors','coeff','ang','R2')
%save Temp_yunke_2006 th Errors coeff ang;% saving data

```