



COLEGIO DE POSTGRADUADOS

INSTITUCION DE ENSEÑANZA E INVESTIGACION EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO DE HIDROCIENCIAS

EVALUACIÓN DE MODELOS EMPÍRICOS, MATEMÁTICOS Y REDES NEURONALES PARA ESTIMAR DATOS FALTANTES EN ESTACIONES METEOROLÓGICAS EN MÉXICO

GAMBOA CHEL ROBERTO OMAR

T E S I S

**PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE:**

MAESTRO EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MEXICO

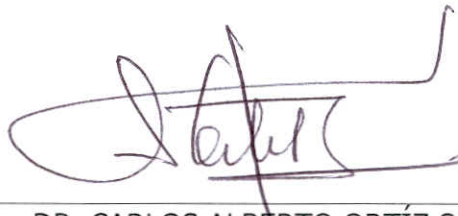
2015

La presente tesis titulada EVALUACIÓN DE MODELOS EMPÍRICOS, MATEMÁTICOS Y REDES NEURONALES PARA ESTIMAR DATOS FALTANTES EN ESTACIONES METEOROLÓGICAS EN MÉXICO realizada por el alumno Roberto Omar Gamboa Chel, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRO EN CIENCIAS
HIDROCIENCIAS

CONSEJO PARTICULAR

CONSEJERO



DR. CARLOS ALBERTO ORTÍZ SOLORIO

ASESOR



M.C. GUILLERMO CRESPO PICHARDO

ASESOR



DR. RAMÓN ARTEAGA RAMÍREZ

ASESOR



DRA. ROCÍO CERVANTES OSORNIO

Montecillo, Texcoco, Estado de México, abril de 2015



DEDICATORIA

Dedico este trabajo a las personas más importantes de mi vida:

Mi mama, María Olivia Chel Arguello, mamita te dedico esta tesis, porque siempre estuviste en cada paso durante mi formación académica y personal, además por si fuera poco gracias a ti y a tus enseñanzas me hicieron la persona que soy.

A mi Papa, Manuel de Jesús Gamboa Canul, gracias por tu apoyo para concluir esta etapa importante de mi vida.

A mis hermanos, Santos Eduardo, Clarisa Olivia, Mariana Maribel y el pequeñito de la familia Alberto Rafael, quienes están conmigo en todo momento y me apoyaron para concluir mi desarrollo profesional.

A mis abuelos, tíos, tías y primos, los cuales me echaron porras en cada instante que retornaba a mi casa.

A mi segunda familia en Chapingo “La pandillita”, Madaí, Mavet, Madián y Mijail, con los que compartí risas, enojos, tristeza y momentos inolvidables y a los cuales le agradezco mucho a Dios por haberlos puesto en mi camino.

A la familia Sánchez Galindo, por la amistad que me han mostrado en estos últimos años.

Te agradezco Dios mío, por haberme dado a esta gran familia y a estos amigos maravillosos, quienes les dedico con todo mi corazón, admiración y cariño este trabajo de tesis, Gracias familia y amigos.



AGRADECIMIENTOS

Primeramente quisiera agradecerle a Dios, por darme la vida, por estar conmigo en todo momento y que gracias a su amor y misericordia me hicieron una persona de bien, gracias Dios.

Al Colegio de Postgraduados, por darme la oportunidad de progresar y desarrollarme profesionalmente.

Los millones de mexicanos (as) que pagan impuestos, quienes, a través del Consejo Nacional de Ciencia y Tecnología (CONACYT) y el Colegio de Postgraduados, han financiado parte de mi formación.

A mi profesor consejero Dr. Carlos A. Ortiz Solorio por el apoyo mostrado, así como su disponibilidad en cada momento y por las certeras observaciones realizadas a este trabajo.

Las personas integrantes de mi Consejo Particular por el esfuerzo, la dedicación, el tiempo y el apoyo que me han brindado, pero sobre todo por la infinita paciencia para conmigo.

A Carmelita, secretaria del área de Génesis del Postgrado de Edafología por su valiosa ayuda en los trámites de la presente tesis.



EVALUACIÓN DE MODELOS EMPÍRICOS, MATEMÁTICOS Y REDES NEURONALES PARA ESTIMAR DATOS FALTANTES EN ESTACIONES METEOROLÓGICAS EN MÉXICO

Roberto Omar Gamboa Chel, M. en C.
Colegio de Postgraduados, 2015

Los datos faltantes son un problema recurrente en diferentes áreas de la investigación. La Meteorología no es la excepción y demanda que las bases se rellenen con datos generados por alguna metodología que proporcione una estimación aceptable. En el presente trabajo se propuso la evaluación de modelos empíricos (U.S. Weather Service y el de razones de distancia), la Transformada de Wavelet y las Redes Neuronales Artificiales, con el fin de determinar las metodologías más apropiadas para la estimación de datos faltantes en estaciones meteorológicas de México. Estos modelos fueron analizados en tres regiones climáticas, a saber: clima Cálido Sub-húmedo, Templado Sub-húmedo y Muy Seco Semi-cálido, que corresponden a una superficie aproximada del 50% del territorio nacional; de igual manera se estudiaron tres variables climáticas: temperatura promedio, evaporación y precipitación a nivel mensual. Se seleccionaron 42 estaciones bases (en las tres regiones climáticas) y se procedió a eliminar un 15% y 25% de los datos de cada estación, los cuales se simularon y los resultados se compararon con los valores observados. Para la evaluación estadística se utilizaron los siguientes parámetros e índices: cuadrado medio del error, raíz cuadrada del cuadrado medio del error, índice de Wilmott, diferencia relativa, error de estimación y error absoluto medio, además se realizó un análisis de regresión lineal simple con el fin de determinar el mejor modelo de estimación para datos faltantes. Los mejores modelos de estimación obtenidos para las 42 estaciones meteorológicas fueron en primer lugar la Transformada de Wavelet, como segundo el U.S. Weather Service con cinco estaciones cercanas y como tercero las Redes Neuronales Artificiales

Palabras Clave: *Datos faltantes, transformada de Wavelet, Redes neuronales artificiales, modelos empíricos, variables climáticas, regiones climáticas.*



EVALUATION OF EMPIRICAL MODELS, MATHEMATICS AND NEURAL NETWORKS TO GENERATE MISSING DATA IN WEATHER STATIONS IN MEXICO

Roberto Omar Gamboa Chel, M. en C.
Colegio de Postgraduados, 2015

Missing data are a recurrent problem in different research areas. Meteorology is no exception and databases often need to be infilled with data generated by any method that provides an acceptable estimate. This paper evaluates empirical models (U.S. Weather Service and distance ratios), the Wavelet Transform and Artificial Neural Networks to determine the most appropriate methodologies for estimating missing data in weather stations in Mexico. These models were analyzed in three climate regions, namely Warm Sub-humid, Temperate Sub-humid and Semi-warm Very Dry climates, which together account for 50% of Mexico's territory; similarly, three climate variables were studied: average monthly temperature, evaporation and precipitation. Forty-two base stations (in the three climate regions) were selected and then 15% and 25% of each station's data were removed. After that, the missing data were simulated and the results compared with the observed values. For the statistical evaluation, parameters and indices were used are: mean squared error, root mean square error, Willmott's index, difference relative, standard error of the estimate and mean absolute error. Also, a simple linear regression analysis was performed to determine the best model for estimating missing data. The best estimation models obtained for the 42 weather stations were, in descending order, the Wavelet Transform, the U.S. Weather Service with five nearby stations and lastly the Artificial Neural Networks.

Keywords: *Missing data, Wavelet Transform, Artificial Neural Networks, empirical models, climate variables and climate regions.*



CONTENIDO

Página

1. INTRODUCCIÓN	1
2. OBJETIVOS	2
2.1. OBJETIVO GENERAL	2
2.2. OBJETIVOS PARTICULARES	2
3. REVISIÓN DE LITERATURA	3
3.1. MODELO MATEMÁTICO TRANSFORMADA DE WAVELET	3
3.1.1. <i>Conceptos básicos</i>	3
3.1.2. <i>Aplicaciones de la transformada de Wavelet</i>	4
3.1.3. <i>Wavelet Madre</i>	5
3.1.4. <i>Transformada de Wavelet Continua (TCW)</i>	5
3.1.5. <i>Transformada de Wavelet Discreta (TWD)</i>	6
3.1.6. <i>Coficiente de escala y Coficiente Wavelet</i>	7
3.1.7. <i>Análisis Multiresolución</i>	7
3.1.7.1. <i>Transformada rápida de wavelet y banco de filtros</i>	8
3.2. REDES NEURONALES ARTIFICIALES (RNA)	12
3.2.1. <i>Conceptos básicos y características de las RNA</i>	13
3.2.2. <i>Modelo neuronal artificial</i>	15
3.2.3. <i>Función de transferencia o activación</i>	16
3.2.4. <i>Regla de aprendizaje</i>	18
3.2.4.1. <i>Conjunto de datos</i>	20
3.2.5. <i>Tipos de redes</i>	21
3.2.6. <i>Algoritmo Retropropagación</i>	24
3.2.7. <i>Redes neuronales en Matlab</i>	27
3.2.7.1. <i>Crear la red (newff)</i>	27
3.2.7.2. <i>Iniciando Pesos (init)</i>	27
3.2.7.3. <i>Entrenamiento</i>	28
3.2.7.4. <i>Simulación (Sim)</i>	29
3.2.7.5. <i>Preprocesamiento y posprocesamiento</i>	30
3.3. MODELOS EMPÍRICOS	31
3.3.1. <i>Razón de distancia</i>	32
3.3.2. <i>U.S. Weather Service</i>	32
3.3.3. <i>Estimación de datos faltantes mediante regresión</i>	33
3.3.3.1. <i>Coficiente de correlación</i>	34
3.3.3.2. <i>Regresión lineal simple</i>	36
3.3.3.3. <i>Regresión lineal múltiple</i>	37



3.4.	PARÁMETROS ESTADÍSTICOS	37
3.4.1.	<i>Error cuadrático medio (MSE)</i>	37
3.4.2.	<i>Raíz cuadrada del cuadrado medio del error (RMSE)</i>	38
3.4.3.	<i>Error absoluto medio (MAE)</i>	39
3.4.4.	<i>Error estándar de estimación (SEE)</i>	39
3.4.5.	<i>Índice de Wilmott de correspondencia (d)</i>	39
3.4.6.	<i>Diferencia relativa (DR)</i>	40
4.	MATERIALES Y MÉTODOS	41
4.1.	MATERIALES	41
4.1.1.	<i>Datos meteorológicos utilizados</i>	41
4.1.2.	<i>Selección de las estaciones meteorológicas de estudio</i>	42
4.1.2.1.	Diagnóstico de las estaciones meteorológicas en México	42
4.1.2.2.	Procedimiento de selección	43
A.	Selección de las estaciones base	43
B.	Selección de las estaciones circundantes	48
4.1.3.	<i>Simulación de datos faltantes</i>	49
4.1.3.1.	Muestreo Sistemático	49
4.1.4.	<i>Software utilizados</i>	50
4.2.	MÉTODOS	51
4.2.1.	<i>Modelos empíricos</i>	52
4.2.1.1.	US Weather Service (UWS)	52
4.2.1.2.	Razón de la distancia (RD)	52
4.2.1.3.	Regresión lineal simple (RL)	53
4.2.1.4.	Regresión lineal múltiple (RM)	55
4.2.2.	<i>Modelo matemático Transformada de Wavelet (TW)</i>	58
4.2.3.	<i>Modelo de Redes Neuronales Artificiales</i>	61
4.3.	DIAGRAMAS DE FLUJO DE PROCEDIMIENTOS PARA CADA MODELO	65
4.4.	ANÁLISIS ESTADÍSTICO	70
5.	RESULTADOS Y DISCUSIÓN	72
5.1.	DISTRIBUCIÓN DE LAS ESTACIONES BASES SELECCIONADAS	72
5.2.	ANÁLISIS DE DISTANCIAS DE LAS ESTACIONES CIRCUNDANTES	73
5.3.	SIMULACIÓN DE DATOS FALTANTES CON UN 15%	74
5.4.	SIMULACIÓN DE DATOS FALTANTES CON UN 25%	74
5.5.	APLICACIÓN DE LOS MODELOS DE ESTIMACIÓN DE DATOS FALTANTES	75
5.5.1.	<i>US Weather Service</i>	75
5.5.2.	<i>Razón de la distancia</i>	77
5.5.3.	<i>Transformada de Wavelet</i>	77



5.5.4.	<i>Redes neuronales artificiales</i>	85
5.6.	COMPARACIÓN GRÁFICA DE LOS MODELOS.....	87
5.7.	ANÁLISIS ESTADÍSTICO	96
5.7.1.	<i>15% de datos faltantes</i>	97
5.7.2.	<i>25% de datos faltantes</i>	98
5.8.	ESTUDIO DE INTERRELACIÓN DE VARIABLES CLIMÁTICAS	98
5.8.1.	<i>Coefficiente de correlación simple</i>	99
5.8.2.	<i>Coefficiente de correlación múltiple</i>	100
5.8.3.	<i>Comparación estadística de la aplicación de la regresión lineal simple y múltiple</i>	101
6.	CONCLUSIONES	104
7.	RECOMENDACIONES	106
8.	LITERATURA CITADA	107
9.	ANEXOS	112
9.1.	ESTACIONES METEOROLÓGICAS BASES Y VECINAS SELECCIONADAS PARA CADA REGIÓN CLIMÁTICA	112
9.2.	COMPARACIÓN DE PARÁMETROS E ÍNDICES ESTADÍSTICOS	118
9.2.1.	<i>15% de datos faltantes</i>	118
9.2.1.1.	Por variable climática.....	118
9.2.1.2.	Por región climática	120
9.2.2.	<i>25% de datos faltantes</i>	127
9.2.2.1.	Por variable climática.....	127
9.2.2.2.	Por región climática	129
9.3.	COEFICIENTES DE CORRELACIÓN.....	137
9.4.	CÓDIGO MATLAB PARA EL MODELO DE REDES NEURONALES	141



ÍNDICE DE FIGURAS

FIGURA 1.	REPRESENTACIÓN MULTINIVEL DE UNA FUNCIÓN. FUENTE: REPRESA, 2002.....	8
FIGURA 2.	DESCOMPOSICIÓN DE LA SEÑAL	9
FIGURA 3.	DESCOMPOSICIÓN MULTINIVEL DE LA SEÑAL.....	10
FIGURA 4.	RECONSTRUCCIÓN DE LA SEÑAL.....	10
FIGURA 5.	PARTES PRINCIPALES DE LA NEURONA BIOLÓGICA. FUENTE GÓMEZ, 2010.....	13
FIGURA 6.	MODELO NEURONAL ARTIFICIAL. FUENTE: GÓMEZ, 2010	16
FIGURA 7.	FUNCIÓN SIGMOIDAL.....	17
FIGURA 8.	FUNCIÓN TANGENTE SIGMOIDAL. FUENTE: MATICH, 2001	18
FIGURA 9.	RED NEURONAL MONOCAPA. FUENTE: MARTÍN ET AL, 2009	23
FIGURA 10.	ESQUEMA DE UNA RED MULTICAPA. FUENTE: MARTÍN ET AL, 2009	23
FIGURA 11.	PRIMERA FASE DEL CICLO. FUENTE: GÓMEZ, 2010	26
FIGURA 12.	SEGUNDA FASE DEL CICLO. FUENTE: GÓMEZ, 2010	26
FIGURA 13.	ESQUEMA PARA COMPLETAR DATOS FALTANTES POR DISTANCIA. FUENTE: LOBO, 2011	32
FIGURA 14.	REGIONES CLIMÁTICAS A UTILIZAR.....	41
FIGURA 15.	HISTOGRAMA DE FRECUENCIAS DE INFORMACIÓN DISPONIBLE EN LAS ESTACIONES METEOROLÓGICAS	43
FIGURA 16.	UBICACIÓN ESPACIAL DE LAS ESTACIONES METEOROLÓGICAS SELECCIONADAS EN CADA REGIÓN CLIMÁTICA.....	44
FIGURA 17.	DISTRIBUCIÓN ESPACIAL DE LAS ESTACIONES METEOROLÓGICAS SELECCIONADAS	46
FIGURA 18.	DISTRIBUCIÓN ESPACIAL DE LAS ESTACIONES SELECCIONADAS PARA EVAPORACIÓN	47
FIGURA 19.	LOCALIZACIÓN DE LAS ESTACIONES CERCANAS A LAS BASES PARA EVAPORACIÓN	48
FIGURA 20.	LOCALIZACIÓN DE LAS ESTACIONES CERCANAS A LAS BASES PARA TEMPERATURA Y PRECIPITACIÓN.....	49
FIGURA 21.	EJEMPLO DE TABLA PARA EL MODELO DE REGRESIÓN LINEAL SIMPLE	55
FIGURA 22.	EJEMPLO DE TABLA PARA EL CÁLCULO DE REGRESIÓN LINEAL MÚLTIPLE	57
FIGURA 23.	DIAGRAMA DE DESCOMPOSICIÓN DE SEÑALES. FUENTE: KOURO Y MUSALEM, 2002.....	58
FIGURA 24.	REPRESENTACIÓN DE LA ELIMINACIÓN DE INFORMACIÓN O SUBMUESTREO. FUENTE: KOURO Y MUSALEM, 2002	59
FIGURA 25.	ÁRBOL DE DESCOMPOSICIÓN WAVELET. FUENTE: KOURO Y MUSALEM, 2002	60
FIGURA 26.	EJEMPLO DE ÁRBOL DE DESCOMPOSICIÓN WAVELET	60
FIGURA 27.	ESQUEMA DE RECONSTRUCCIÓN WAVELET. FUENTE: KOURO Y MUSALEM, 2002.....	61
FIGURA 28.	ESQUEMA DE LA RED NEURONAL A UTILIZAR.....	63
FIGURA 29.	ESQUEMA DE APROXIMACIÓN DE DATOS DE ENTRADA	65
FIGURA 30.	DIAGRAMA DE FLUJO PARA LA SIMULACIÓN DE DATOS FALTANTES	65
FIGURA 31.	DIAGRAMA DE FLUJO US WEATHER SERVICE	66
FIGURA 32.	DIAGRAMA DE FLUJO DE RAZÓN DE DISTANCIA	67
FIGURA 33.	DIAGRAMA DE FLUJO DEL MODELO MATEMÁTICO TRANSFORMADA DE WAVELET	68
FIGURA 34.	CONTINUACIÓN FIGURA 33.....	69
FIGURA 35.	MODELO DE REDES NEURONALES ARTIFICIALES	70



FIGURA 36.	DISTRIBUCIÓN DE LAS ESTACIONES BASES SELECCIONADAS POR REGIÓN CLIMÁTICA	72
FIGURA 37.	DISTRIBUCIÓN DE LAS ESTACIONES CIRCUNDANTES POR REGIÓN CLIMÁTICA	73
FIGURA 38.	SOFTWARE MATLAB 2013®	85
FIGURA 39.	PROGRAMA GENERADO EN MATLAB 2013® PARA RED RETROPROPAGACIÓN DE REDES NEURONALES ARTIFICIALES....	85
FIGURA 40.	ENTRENAMIENTO DE LA RED NEURONAL ARTIFICIAL	86
FIGURA 41.	COMPARACIÓN GRÁFICA DE LOS RESULTADOS DE LA ESTACIÓN 21081 PARA LA VARIABLE CLIMÁTICA TEMPERATURA Y 15% DE DATOS FALTANTES	88
FIGURA 42.	COMPARACIÓN GRÁFICA DE LOS RESULTADOS DE LA ESTACIÓN 23154 PARA LA VARIABLE CLIMÁTICA EVAPORACIÓN Y 15% DE DATOS FALTANTES	88
FIGURA 43.	COMPARACIÓN GRÁFICA DE LOS RESULTADOS DE LA ESTACIÓN 3166 PARA LA VARIABLE CLIMÁTICA PRECIPITACIÓN Y 15% DATOS FALTANTES	89
FIGURA 44.	COMPARACIÓN GRÁFICA DE LOS RESULTADOS DE LA ESTACIÓN 30056 PARA LA VARIABLE CLIMÁTICA TEMPERATURA Y 25% DATOS FALTANTES	89
FIGURA 45.	COMPARACIÓN GRÁFICA DE LOS RESULTADOS DE LA ESTACIÓN 5036 PARA LA VARIABLE CLIMÁTICA EVAPORACIÓN Y 25% DATOS FALTANTES	90
FIGURA 46.	COMPARACIÓN GRÁFICA DE LOS RESULTADOS DE LA ESTACIÓN 13151 PARA LA VARIABLE CLIMÁTICA PRECIPITACIÓN Y 25% DATOS FALTANTES	90
FIGURA 47.	DIAGRAMAS DE DISPERSIÓN PARA LOS VALORES OBSERVADOS Y ESTIMADOS POR MODELO DE LA VARIABLE TEMPERATURA PARA LA ESTACIÓN 50036 DE CLIMA CÁLIDO SUBHÚMEDO.....	91
FIGURA 48.	CONTINUACIÓN DE LA FIGURA 47	92
FIGURA 49.	DIAGRAMAS DE DISPERSIÓN DE LOS VALORES OBSERVADOS Y ESTIMADOS POR MODELO DE LA VARIABLE EVAPORACIÓN PARA LA ESTACIÓN 5036 DE CLIMA MUY SECO SEMICÁLIDO	93
FIGURA 50.	CONTINUACIÓN DE LA FIGURA 49.....	94
FIGURA 51.	DIAGRAMAS DE DISPERSIÓN DE LOS VALORES OBSERVADOS Y ESTIMADOS POR MODELO DE LA VARIABLE PRECIPITACIÓN PARA LA ESTACIÓN 13151 DE CLIMA TEMPLADO SUBHÚMEDO	95
FIGURA 52.	CONTINUACIÓN DE LA FIGURA 51	96



ÍNDICE DE CUADROS

CUADRO 1.	FRECUENCIA DE PORCENTAJE DE DATOS PARA LAS ESTACIONES METEOROLÓGICAS.....	42
CUADRO 2.	NÚMERO DE ESTACIONES DISPONIBLES EN CADA PERIODO.....	45
CUADRO 3.	DISTRIBUCIÓN DE LAS ESTACIONES BASES SELECCIONADAS	72
CUADRO 4.	ESTACIONES CIRCUNDANTES A UNA DISTANCIA MENOR O IGUAL A 50 KM.....	73
CUADRO 5.	CANTIDAD DE ESTACIONES CIRCUNDANTES UTILIZADAS EN LA SIMULACIÓN PARA UN 25%.....	75
CUADRO 6.	ESTACIONES CIRCUNDANTES UTILIZADAS PARA EL MODELO US WEATHER SERVICE.....	75
CUADRO 7.	ESTACIONES UTILIZADAS PARA EL MODELO DE LA RAZÓN DE LA DISTANCIA.....	77
CUADRO 8.	INFORMACIÓN DE LAS ESTACIONES UTILIZADAS PARA ESTE MODELO	78
CUADRO 9.	PARTE DE LOS DATOS CLIMÁTICOS UTILIZADOS DE LA ESTACIÓN DE APOYO	78
CUADRO 10.	APLICACIÓN DEL FILTRO PASA BAJO.....	78
CUADRO 11.	APLICACIÓN DEL FILTRO PASA ALTO.....	79
CUADRO 12.	PROCEDIMIENTO DE SUBMUESTREO O ELIMINACIÓN DE DATOS	79
CUADRO 13.	RESULTADOS DE LA DESCOMPOSICIÓN EN EL NIVEL 2.....	80
CUADRO 14.	RESULTADOS DE LA DESCOMPOSICIÓN EN CADA UNO DE LOS NIVELES	80
CUADRO 15.	COEFICIENTES DE APROXIMACIÓN EN EL NIVEL 6 DE LA ESTACIÓN PROBLEMA	81
CUADRO 16.	SOBREMUESTREO EN EL NIVEL DE RECONSTRUCCIÓN UNO	82
CUADRO 17.	APLICACIÓN DEL FILTRO PASA BAJO Y PASA ALTO PARA LA RECONSTRUCCIÓN EN EL NIVEL 1.....	82
CUADRO 18.	COEFICIENTES DE APROXIMACIÓN EN EL NIVEL DE RECONSTRUCCIÓN 1.....	83
CUADRO 19.	SOBREMUESTREO EN EL NIVEL DE RECONSTRUCCIÓN DOS	83
CUADRO 20.	APLICACIÓN DEL FILTRO PASO BAJO Y PASA ALTO EN EL NIVEL DE RECONSTRUCCIÓN DOS	84
CUADRO 21.	COEFICIENTES DE APROXIMACIÓN EN EL NIVEL DE RECONSTRUCCIÓN DOS.....	84
CUADRO 22.	PARTE DE LOS RESULTADOS DE LA RECONSTRUCCIÓN DE LA ESTACIÓN PROBLEMA	84
CUADRO 23.	DATOS DE ENTRADA PARA EL ENTRENAMIENTO DE LA RED NEURONAL ARTIFICIAL	86
CUADRO 24.	DATOS DE SALIDA PARA EL ENTRENAMIENTO DE LA RED NEURONAL	86
CUADRO 25.	ARCHIVO UTILIZADO *PRN CON LA SERIE DE TIEMPO COMPLETA PARA GENERAR DATOS FALTANTES	87
CUADRO 26.	RESULTADOS DE LA APLICACIÓN DEL MODELO DE RED NEURONAL ARTIFICIAL	87
CUADRO 27.	RESUMEN DE RESULTADOS PARA 15% DE DATOS FALTANTES	97
CUADRO 28.	RESUMEN DE RESULTADOS PARA 25% DE DATOS FALTANTES	98
CUADRO 29.	COEFICIENTES DE CORRELACIÓN SIMPLE PARA CADA VARIABLE CLIMÁTICA 15 % DE DATOS FALTANTES.....	99
CUADRO 30.	COEFICIENTES DE CORRELACIÓN SIMPLE PARA CADA VARIABLE CLIMÁTICA 15 % DE DATOS FALTANTES.....	100
CUADRO 31.	COEFICIENTES DE CORRELACIÓN MÚLTIPLE PARA CADA VARIABLE CLIMÁTICA PARA 15% DE DATOS FALTANTES	100
CUADRO 32.	COEFICIENTES DE CORRELACIÓN MÚLTIPLE PARA CADA VARIABLE CLIMÁTICA PARA 25% DE DATOS FALTANTES	101
CUADRO 33.	RESULTADOS DE LA COMPARACIÓN DE LA REGRESIÓN LINEAL SIMPLE Y MÚLTIPLE CON LOS DEMÁS MODELOS PROPUESTOS PARA 15% DE DATOS FALTANTES.....	102



CUADRO 34.	RESULTADOS DE LA COMPARACIÓN DE LA REGRESIÓN LINEAL SIMPLE Y MÚLTIPLE CON LOS DEMÁS MODELOS PROPUESTOS PARA 25% DE DATOS FALTANTES	102
CUADRO 35.	ESTACIONES METEOROLÓGICAS UTILIZADAS EN LA INVESTIGACIÓN.....	112
CUADRO 36.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA TEMPERATURA Y 15% DE DATOS FALTANTES	118
CUADRO 37.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA TEMPERATURA Y 15% DE DATOS FALTANTES	118
CUADRO 38.	RESULTADOS DE LOS PARÁMETROS E ÍNDICES ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA EVAPORACIÓN Y 15% DE DATOS FALTANTES.....	119
CUADRO 39.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA EVAPORACIÓN Y 15% DE DATOS FALTANTES.....	119
CUADRO 40.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA PRECIPITACIÓN Y 15% DE DATOS FALTANTES	119
CUADRO 41.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA PRECIPITACIÓN Y 15% DE DATOS FALTANTES.....	120
CUADRO 42.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA TEMPERATURA Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA CÁLIDO SUBHÚMEDO	120
CUADRO 43.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA TEMPERATURA Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA CÁLIDO SUBHÚMEDO	121
CUADRO 44.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA PRECIPITACIÓN Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA CÁLIDO SUBHÚMEDO	121
CUADRO 45.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA PRECIPITACIÓN Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA CÁLIDO SUBHÚMEDO	121
CUADRO 46.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA EVAPORACIÓN Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA CÁLIDO SUBHÚMEDO	122
CUADRO 47.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA EVAPORACIÓN Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA CÁLIDO SUBHÚMEDO	122
CUADRO 48.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA TEMPERATURA Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA MUY SECO SEMICÁLIDO	122
CUADRO 49.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA TEMPERATURA Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA MUY SECO SEMICÁLIDO.....	123
CUADRO 50.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA PRECIPITACIÓN Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA CÁLIDO MUY SECO SEMICÁLIDO.....	123
CUADRO 51.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA PRECIPITACIÓN Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA MUY SECO SEMICÁLIDO.....	124
CUADRO 52.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA EVAPORACIÓN Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA MUY SECO SEMICÁLIDO	124
CUADRO 53.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA EVAPORACIÓN 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA MUY SECO SEMICÁLIDO.....	124



CUADRO 54.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA TEMPERATURA Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA TEMPLADO SUBHÚMEDO	125
CUADRO 55.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA TEMPERATURA Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA TEMPLADO SUBHÚMEDO	125
CUADRO 56.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA PRECIPITACIÓN Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA TEMPLADO SUBHÚMEDO	126
CUADRO 57.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA PRECIPITACIÓN Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA TEMPLADO SUBHÚMEDO	126
CUADRO 58.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA EVAPORACIÓN Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA TEMPLADO SUBHÚMEDO	126
CUADRO 59.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA EVAPORACIÓN Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA TEMPLADO SUBHÚMEDO	127
CUADRO 60.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA TEMPERATURA Y 25% DE DATOS FALTANTES	127
CUADRO 61.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA TEMPERATURA Y 25% DE DATOS FALTANTES	128
CUADRO 62.	RESULTADOS DE LOS PARÁMETROS E ÍNDICES ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA EVAPORACIÓN Y 25% DE DATOS FALTANTES	128
CUADRO 63.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA EVAPORACIÓN Y 25% DE DATOS FALTANTES	128
CUADRO 64.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA PRECIPITACIÓN Y 25% DE DATOS FALTANTES	129
CUADRO 65.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA PRECIPITACIÓN Y 25% DE DATOS FALTANTES	129
CUADRO 66.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA TEMPERATURA Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA CÁLIDO SUBHÚMEDO	129
CUADRO 67.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA TEMPERATURA Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA CÁLIDO SUBHÚMEDO	130
CUADRO 68.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA PRECIPITACIÓN Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA CÁLIDO SUBHÚMEDO	130
CUADRO 69.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA PRECIPITACIÓN Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA CÁLIDO SUBHÚMEDO	130
CUADRO 70.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA EVAPORACIÓN Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA CÁLIDO SUBHÚMEDO	131
CUADRO 71.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA EVAPORACIÓN Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA CÁLIDO SUBHÚMEDO	131
CUADRO 72.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA TEMPERATURA Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA MUY SECO SEMICÁLIDO	132
CUADRO 73.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA TEMPERATURA Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA MUY SECO SEMICÁLIDO.....	132



CUADRO 74.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA PRECIPITACIÓN Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA CÁLIDO MUY SECO SEMICÁLIDO.....	132
CUADRO 75.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA PRECIPITACIÓN Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA MUY SECO SEMICÁLIDO.....	133
CUADRO 76.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA EVAPORACIÓN Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA MUY SECO SEMICÁLIDO	133
CUADRO 77.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA EVAPORACIÓN Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA MUY SECO SEMICÁLIDO.....	134
CUADRO 78.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA TEMPERATURA Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA TEMPLADO SUBHÚMEDO	134
CUADRO 79.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA TEMPERATURA Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA TEMPLADO SUBHÚMEDO	134
CUADRO 80.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA PRECIPITACIÓN Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA TEMPLADO SUBHÚMEDO	135
CUADRO 81.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA PRECIPITACIÓN Y 15% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA TEMPLADO SUBHÚMEDO	135
CUADRO 82.	RESULTADOS DE INDICADORES Y PARÁMETROS ESTADÍSTICOS PARA LA VARIABLE CLIMÁTICA EVAPORACIÓN Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA TEMPLADO SUBHÚMEDO	135
CUADRO 83.	ANÁLISIS DE REGRESIÓN LINEAL SIMPLE PARA EVAPORACIÓN Y 25% DE DATOS FALTANTES PARA LA REGIÓN CLIMÁTICA TEMPLADO SUBHÚMEDO	136
CUADRO 84.	COEFICIENTES DE CORRELACIÓN SIMPLE PARA EVAPORACIÓN Y 15 % DE DATOS FALTANTES.....	137
CUADRO 85.	COEFICIENTES DE CORRELACIÓN SIMPLE PARA PRECIPITACIÓN Y 15% DE DATOS FALTANTES	137
CUADRO 86.	COEFICIENTES DE CORRELACIÓN SIMPLE PARA TEMPERATURA Y 15% DE DATOS FALTANTES	137
CUADRO 87.	COEFICIENTES DE CORRELACIÓN SIMPLE PARA EVAPORACIÓN Y 25 % DE DATOS FALTANTES.....	138
CUADRO 88.	COEFICIENTES DE CORRELACIÓN SIMPLE PARA PRECIPITACIÓN Y 25% DE DATOS FALTANTES	138
CUADRO 89.	COEFICIENTES DE CORRELACIÓN SIMPLE PARA TEMPERATURA Y 25% DE DATOS FALTANTES	139
CUADRO 90.	COEFICIENTES DE CORRELACIÓN MÚLTIPLE PARA CADA VARIABLE CLIMÁTICA PARA 15% DE DATOS FALTANTES	139
CUADRO 91.	COEFICIENTES DE CORRELACIÓN MÚLTIPLE PARA CADA VARIABLE CLIMÁTICA PARA 25% DE DATOS FALTANTES	140



1. INTRODUCCIÓN

Cuando se trabaja con variables meteorológicas existe el problema que en muchos de los casos las series están incompletas o existe ausencia de datos. La falta de información se puede atribuir a varias razones; en las estaciones meteorológicas manuales se debe principalmente a la ausencia del observador, a alguna falla del aparato medidor o no registrar el dato en el momento adecuado; para las estaciones meteorológicas automáticas se atribuye principalmente a la falla del equipo y/o a la falta de energía eléctrica, ya que estas estaciones funcionan con sensores eléctricos que miden las variables climáticas de manera continua.

Los datos faltantes son un problema recurrente en todas las áreas de investigación y en la medida que las estimaciones de datos obtenidos sean los más cercanos a la realidad, se podrán aplicar a diferentes investigaciones, teniendo la certeza que se obtendrán resultados confiables del fenómeno que se estudia.

La calidad de la información meteorológica es clave al momento de analizar de la mejor manera la variabilidad climática de una zona en específico. Ante la falta de información climática se busca que las bases de datos, se completen con datos generados por alguna metodología que proporcione una buena estimación de los datos. La disponibilidad, confiabilidad y homogeneidad de una base de datos puede resultar de mucha ayuda para la toma de decisiones.

Con el fin de buscar metodologías que proporcionen buenos resultados para la estimación de datos faltantes de estaciones meteorológicas, en el presente trabajo se propuso la evaluación de métodos recientes como la Transformada de Wavelet, Redes Neuronales Artificiales y métodos empíricos: el U.S. Weather Service y el de razones de distancia. Los cuales se aplicaron para estimar datos faltantes de tres variables climáticas y para tres regiones. Las variables utilizadas fueron: temperatura, precipitación y evaporación a nivel mensual, que se obtuvieron del paquete de cómputo ERIC III

Los métodos a evaluar se seleccionaron por las siguientes razones: La transformada de Wavelet fue utilizada para estimar datos faltantes de precipitación en Perú, obteniendo resultados muy cercanos a los reales. Las redes neuronales se emplearon para estimar



datos faltantes en estaciones meteorológicas de Venezuela y en sus resultados se indicó que era un modelo útil para la reconstrucción de datos faltantes de las series; Los métodos empíricos se seleccionaron por ser métodos sencillos.

Se evaluó la capacidad predictiva de cada método, mediante un análisis estadístico de los errores obtenidos entre los valores estimados y los observados.

2. OBJETIVOS

2.1. Objetivo General

- Evaluar la capacidad predictiva de métodos empíricos (Método del U.S. Weather Service y completar por razones de distancia), redes neuronales y la transformada de Wavelet para la estimación de datos faltantes en estaciones meteorológicas en México.

2.2. Objetivos Particulares

- Estimar datos faltantes utilizando la metodología del U.S. Weather Service, completar por razones de distancia, regresión lineal simple y múltiple, redes neuronales y la transformada de Wavelet para estaciones meteorológicas localizadas en tres regiones climáticas diferentes.
- Determinar el mejor método de predicción de datos faltantes, a partir de la comparación estadística de información en estaciones bases y las generadas con cada uno de los modelos.



3. REVISIÓN DE LITERATURA

3.1. Modelo Matemático Transformada de Wavelet

3.1.1. Conceptos básicos

Una “wavelet” es el nombre dado a una “pequeña onda u ondita” que tiene su energía concentrada en un periodo de tiempo determinado, lo cual proporciona una herramienta para el análisis de fenómenos transitorios no estacionarios (Moros, 2010).

Las Wavelets son familias de funciones que se encuentran definidas en el espacio y se emplean como funciones de análisis, examinan la señal de interés en el plano tiempo-frecuencia para obtener sus características periódicas y no periódicas (Zamorano, 2010).

Una de las razones de la popularidad de las técnicas basadas en wavelets es que tiene como base funciones localizadas en tiempo y frecuencia, dando como resultado una representación más compacta y más fácil de implementar (Represa, 2002).

La gran revolución de las “wavelets” nació en la década de los 80`s a partir de la búsqueda de petróleo, el llamado “oro negro”, por el ingeniero Jean Morlet de la compañía Francesa Elf-Aquitanie quien buscaba ofrecer a los geólogos la interpretación de señales sísmicas (Moros, 2010).

Morlet, desarrolló su propia forma de analizar las señales sísmicas creando componentes que estuvieran localizados en el espacio. Estos componentes los denomino “wavelets de forma constante”, que después de cierto tiempo se conocieron como “wavelet de Morlet”. La forma de analizar una onda sísmica consistía en separarla en las “wavelet” que las componían y también volviéndolas a unir para reconstruir la onda original (proceso inverso). Tal como se hacía en el caso de la Transformada de Fourier en donde se descompone una señal en armónicos u ondas sinusoidales (Moros, 2010).

Morlet trabajó con Grossmann (físico del Centre de Physique Théorique de Marsella, Francia) de forma continua hasta lograr demostrar matemáticamente que las ondas se podían reconstruir a partir de sus componentes “wavelets”; y al lograr esto, pudieron afirmar que transformaciones de “wavelets” funcionarían mucho mejor que las transformaciones de Fourier. Los resultados de su investigación fueron publicados en un



artículo en 1984, y es el primer documento donde se identifica la palabra "wavelet" (Moros, 2010).

La transformada wavelet constituye una técnica relativamente nueva que ha sido propuesta por los investigadores como una poderosa herramienta en el análisis sobre el comportamiento local de una señal (Rupérez, 2012).

Las Transformadas Wavelet comprenden la Transformada Continua de Wavelet y la Transformada Discreta de Wavelet. Estas son las herramientas matemáticas que permiten el análisis de señales de manera muy similar como lo hace la Transformada de Fourier de Tiempo Corto dando información en el dominio del tiempo y en el dominio de la frecuencia (Sánchez, 2004).

3.1.2. Aplicaciones de la transformada de Wavelet

Las wavelets son una herramienta matemática para descomponer funciones jerárquicamente. Es por esto que las áreas de aplicación de la transformada wavelet son tan diversas y, si bien la mayoría de la teoría ya ha sido desarrollada, aún resta mucho por hacer desde el punto de vista de sus aplicaciones (Castro y Castro, 1995).

La teoría de las "Wavelet" tiene muchas aplicaciones reales y aportes en diferentes campos, siendo esta herramienta muy joven en comparación con las Transformadas de Fourier. La transformada Wavelet se utiliza con el fin de detectar discontinuidades, puntos de rupturas, identificar frecuencias puras, reducir del ruido, comprimir información, aproximar funciones, solución de ecuaciones diferenciales en señales unidimensionales, dinámica molecular, la astrofísica, la geofísica de los sismos, la óptica, el estudio de las turbulencias y la mecánica cuántica, así como en otros campos muy variados como el análisis de electrocardiogramas, el estudio del ADN, al análisis de sangre, el reconocimiento de voz, meteorología, en el campo de la biometría y análisis multifractal (Moros, 2010).



3.1.3. Wavelet Madre

Una Wavelet Madre es una función localizada, perteneciente al espacio, que contiene todas las funciones con energía finita y funciones de cuadrado integrable definidas (Nieto y Marcela, 2008).

La función matemática prototipo denominada wavelet madre debe satisfacer dos condiciones:

A. Ser finita en tiempo

Esta condición se representa por una integral en el dominio temporal (ecuación 1) que evalúa la cantidad de energía estableciendo un valor positivo definido (Zapotitla, 2011):

$$E = \int_{-\infty}^{\infty} |\psi(t)|^2 dt < +\infty \dots \dots \dots \text{(Ecuación 1)}$$

Farge, citado por Zapotitla (2011) menciona que la función que genera la onda debe de cumplir con la condición de admisibilidad en el dominio de las frecuencias, donde C es llamada constante de admisibilidad. Para ser admisible como una wavelet, esta función debe tener un promedio cero y debe estar localizada tanto en el espacio de frecuencia como en el tiempo (Ver Ecuación 2):

$$C = \int \frac{|\psi(w)|^2}{|w|} dw < + \infty \dots \dots \dots \text{(Ecuación 2)}$$

B. tener valor medio nulo

Zapotitla (2011), expresa que la función wavelet madre debe de ser oscilante (Ver Ecuación 3):

$$\int \psi(t) dt = 0 \dots \dots \dots \text{(Ecuación 3)}$$

3.1.4. Transformada de Wavelet Continua (TCW)

La Transformada Continua Wavelet intenta expresar una señal $x(t)$ continua en el tiempo, mediante una expansión de términos o coeficientes proporcionales al producto interno entre la señal y diferentes versiones escaladas y trasladadas de una función prototipo más conocida como wavelet madre (Rupérez, 2012).



La Transformada Wavelet se define como la descomposición de una función $f(t)$ en un conjunto de funciones $\psi_{s,r}(t)$ que forman una base (Ver Ecuación 4) (Zapotitla, 2011).

$$W_f(s, r) = \int f(t)\psi^*(t)dt \dots\dots\dots \text{(Ecuación 4)}$$

- Donde Ψ^* representa el complejo conjugado de Ψ , esto para eliminar la parte imaginaria.

La base es generada por la dilatación s y la traslación τ de la wavelet madre definida como (Ver Ecuación 5) (Zapotitla, 2011):

$$\psi_{s,r}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-r}{s}\right) \dots\dots\dots \text{(Ecuación 5)}$$

3.1.5. Transformada de Wavelet Discreta (TWD)

La Transformada Wavelet depende de dos variables, una encargada del escalamiento de la función wavelet y otra encargada de la traslación de la función wavelet. Mediante las variables de escalamiento y traslación la transformada Wavelet es capaz de hacer un análisis tiempo frecuencia con una resolución variable, es decir, utiliza ventanas de diferente ancho durante el análisis de la señal. La función escala es la encargada de analizar el comportamiento general de la señal, mientras que la función wavelet se encarga de analizar el comportamiento del detalle de la señal (Rupérez, 2012).

Nieto y Marcela (2008), mencionan que por la complejidad en el tratamiento numérico de la TWD, debido a la variabilidad en forma continua de los parámetros de escala como de traslación, es indispensable contar con una herramienta que permita la discretización de esta. Es así que se pasará de un mapeo continuo a un espectro o conjunto finito de valores, a través del cambio de la integral por una aproximación con sumatorias. La discretización permite representar una señal en términos de funciones elementales acompañadas de coeficientes (Ver Ecuación 6).

$$f(t) = \sum C_\lambda \varphi_\lambda \dots\dots\dots \text{(Ecuación 6)}$$

En los sistemas Wavelet traen consigo las Wavelet madre $\Psi(t)$ y unas funciones de escala, las primeras son las encargadas de representar los detalles finos de la función, mientras las funciones de escala realizan una aproximación. Es posible entonces

representar una señal (Ecuación 7) como una sumatoria de funciones wavelet y funciones de escala (Nieto y Marcela, 2008):

$$f(t) = \sum_k \sum_j c_{j,k} \phi_{j,k}(t) + \sum_k \sum_j d_{j,k} \psi_{j,k}(t) \dots \dots \dots \text{(Ecuación 7)}$$

3.1.6. Coeficiente de escala y Coeficiente Wavelet

Nieto y Marcela (2008), describen que para representar una señal $f(t)$ es necesario encontrar los valores de los coeficientes $(c_{j,k})$ y $(d_{j,k})$, los cuáles permiten finalmente hacer la aproximación de la señal. Estos son producto de una multiplicación vectorial entre la función $f(t)$ y la función de escala (ϕ) y wavelet (ψ) . Para los coeficientes de escala tenemos (Ecuación 8).

$$c_{j,k} = \langle f(t), \phi_{j,k}(t) \rangle = \int_{-\infty}^{\infty} |f(t)\phi_{j,k}(t)| dt \dots \dots \dots \text{(Ecuación 8)}$$

$$\langle f(t), \phi_{j,k} \rangle = c_{j,-\infty} \langle \phi_{j,-\infty}(t), \phi_{j,k}(t) \rangle + \dots + c_{j,k} \langle \phi_{j,k}(t), \phi_{j,k}(t) \rangle + \dots + c_{j,\infty} \langle \phi_{j,\infty}(t), \phi_{j,k}(t) \rangle \dots \dots \dots \text{Ecuación 9)}$$

Ya que las funciones wavelet y de escala cumplen la propiedad de ortonormalidad, es posible asegurar que uno de los productos vectoriales sea diferente de cero $(\langle \phi_{k(t),\phi m(t)} = \delta(k - m))$ o $(\langle \psi_{k(t),\psi m(t)} = \delta(k - m))$ por lo tanto (Nieto y Marcela, 2008).

$$c_{j,k} = \langle f(t), \phi_{j,k}(t) \rangle = \int_{t_1}^{t_2} f(t)\phi_{j,k}(2^j t - k) dt \dots \dots \dots \text{(Ecuación 10)}$$

De igual manera para los coeficientes Wavelet

$$d_{j,k} = \langle f(t), \psi_{j,k}(t) \rangle = \int_{t_1}^{t_2} f(t)\psi_{j,k}(2^j t - k) dt \dots \dots \dots \text{(Ecuación 11)}$$

3.1.7. Análisis Multiresolución

El concepto de análisis en multiresolución fue desarrollado por Meyer y Mallat, y como su nombre sugiere, aplicando este tipo de análisis podemos descomponer una función complicada en diferentes funciones más simples y estudiarlas separadamente. Para entender mejor la noción de análisis en multiresolución consideremos una función como la representada en la siguiente Figura (Represa, 2002).

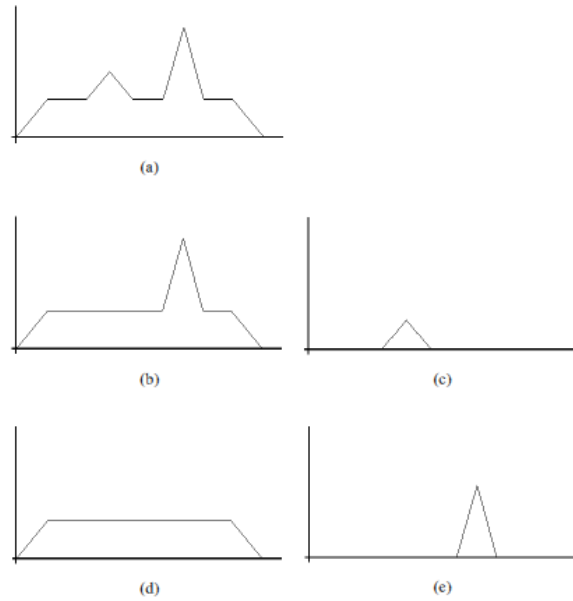


Figura 1. Representación multinivel de una función. Fuente: Represa, 2002

Esta función se puede representar de una forma más gruesa si eliminamos ciertas características. El resultado es el representado en la Figura 1b, donde la Figura 1c representa el trozo que hemos suprimido de la función original. Esta aproximación de la función original se puede a su vez subdividir en dos partes, eliminando aún más detalles. El resultado se encuentra en la Figura 1d, siendo la Figura 1e los detalles que hemos eliminado. El proceso de ir eliminando detalles de la función original es lo que se conoce como proceso de descomposición dentro del análisis de multiresolución. Observando la Figura 1 se aprecia que la función original puede ser nuevamente recuperada si vamos añadiendo a la última aproximación obtenida todos los detalles que se han eliminado. Esto sería lo que se conoce como proceso de reconstrucción (Represa, 2002).

Rupérez (2012), menciona que el análisis multiresolución consiste básicamente en aproximar una función $f(t)$ en distintos niveles de resolución, lo que nos entrega una descomposición multiescala de la forma (Ver Ecuación 12):

$$f(t) = f_0(t) + \sum_{j \geq 0} g_j(t) \dots \dots \dots \text{(Ecuación 12)}$$

3.1.7.1. Transformada rápida de wavelet y banco de filtros

Para muchas señales, el contenido de baja frecuencia es la parte más importante: es lo que aporta identidad a la señal. El contenido de alta frecuencia, por otro lado, tiene un

carácter secundario. Consideremos como ejemplo a la voz humana: si eliminamos los componentes de alta frecuencia, la voz suena diferente, pero podemos entender todavía lo que se está diciendo. No obstante, si eliminamos suficientes componentes de baja frecuencia, comenzamos a no entender el habla (Zamorano, 2010).

Es por esta razón que, en análisis Wavelet, se habla a menudo de aproximaciones (A) y detalles (D). Las aproximaciones son los componentes de gran escala y baja frecuencia de la señal. Los detalles son los componentes de pequeña escala y alta frecuencia (Zamorano, 2010).

Rupérez (2012), menciona que en la mayoría de las señales son los componentes de baja frecuencia los que le otorgan a la señal la mayor parte de su información. Mientras que los componentes de alta frecuencia se encargan de incorporar características más particulares. Es por ello que se subdividen las componentes de una señal en dos categorías:

- Aproximaciones (baja frecuencia)
- Detalles (alta frecuencia)

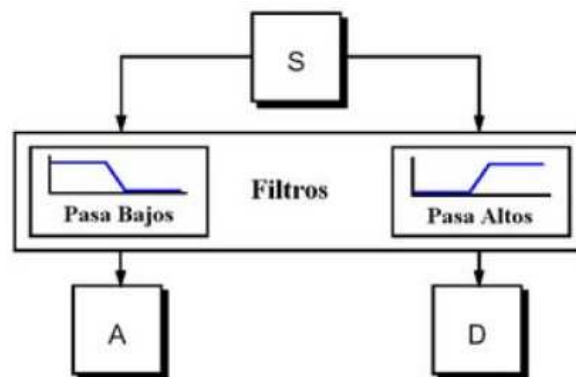


Figura 2. Descomposición de la señal

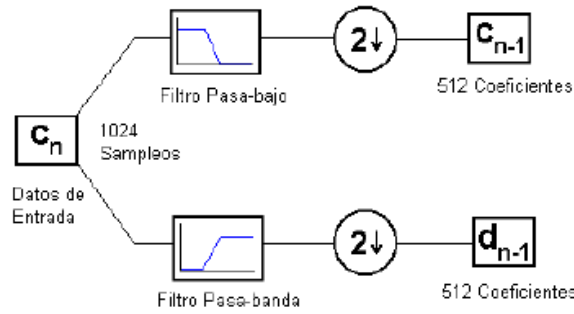


Figura 3. Descomposición multinivel de la señal

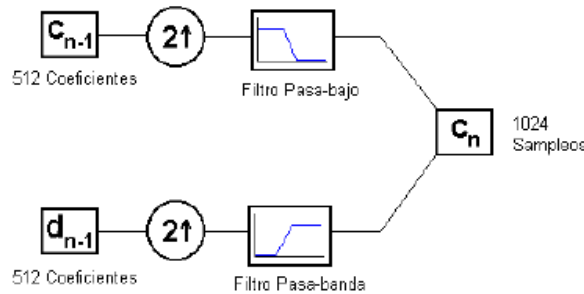


Figura 4. Reconstrucción de la señal

Coefficientes de descomposición

$$c_{j-1,k} = \frac{1}{\sqrt{2}}(c_{j,2k} + c_{j,2k+1}) \dots \dots \dots \text{(Ecuación 13)}$$

$$d_{j-1,k} = \frac{1}{\sqrt{2}}(c_{j,2k} - c_{j,2k+1}) \dots \dots \dots \text{(Ecuación 14)}$$

Coefficientes de reconstrucción

$$c_{j,2k} = \frac{1}{\sqrt{2}}(c_{j-1,k} + d_{j-1,k}) \dots \dots \dots \text{(Ecuación 15)}$$

$$c_{j,2k-1} = \frac{1}{\sqrt{2}}(c_{j-1,k} - d_{j-1,k}) \dots \dots \dots \text{(Ecuación 16)}$$

Gómez (2012), señala que el concepto de banco de filtros se crea a partir de la transformada de wavelet discreta, como se observa cuando se analizó la descripción cualitativa donde la señal es dividida en dos señales que en resumen son dos filtros uno pasa alto y otro pasa bajo. Con el de baja frecuencia se obtienen datos de aproximación y con el de alta frecuencia se obtienen datos de detalle.



Como se submuestra la señal en un factor de dos, esto implica que la señal tendrá la mitad de los puntos, esto no afecta la escala y la mitad restantes de los puntos puede ser descartada con una eliminación de redundancia en un factor de dos (Gómez, 2012).

Además debemos considerar los siguientes conceptos:

- **Downsampling (Decimado):** Se aplica para no obtener matrices del mismo tamaño a la original, consiste en extraer 1 de cada dos datos. Obtenemos que los coeficientes de detalle constan principalmente del ruido de alta frecuencia y los coeficientes de aproximación contienen menos ruido que la señal original (Gómez, 2012).
- **Upsampling (undecimado):** es el proceso alargador de una señal, insertando un cero por cada dos datos de la señal (Gómez, 2012).

A partir de los conceptos anteriores podemos explicar el proceso que realiza para transformar una señal al dominio del tiempo y frecuencia. El proceso consiste en una serie de filtrados usando el concepto de decimado, utilizando el filtro pasa bajo y el filtro pasa alto, posteriormente el resultado se vuelve a filtrar bajo las mismas características, obteniendo la señal con la mitad de resolución pero sin cambio de escala, se le conoce como banco de filtros a este proceso al cual surge el concepto importante que es el análisis Multiresolución (Gómez, 2012).

Para terminar, señalemos que este algoritmo recursivo está basado en lo que se denomina un algoritmo piramidal y que al par de filtros asociados a la descomposición se los denomina par de filtros espejo en cuadratura por la manera en que se relacionan entre sí (Castro y Castro, 1995)

3.2. Redes Neuronales Artificiales (RNA)

En los últimos años también se ha abordado el problema de los datos faltantes mediante redes neuronales artificiales (Navarro y Losilla, 2000).

Las Redes Neuronales Artificiales (Artificial Neural Networks) están inspiradas en las redes neuronales biológicas del cerebro humano. Están constituidas por elementos que se comportan de forma similar a la neurona biológica en sus funciones más comunes. Estos elementos están organizados de una forma parecida a la que presenta el cerebro humano (Basogain, 1998).

Las Redes Neuronales tratan de hacer una analogía del cerebro humano, pero esto es algo realmente difícil de lograr debido a las grandes limitantes que existían y que existen, así como el poco entendimiento que tenemos del cerebro. Las RNA se fueron desarrollando en diversos temas particulares, donde podían competir contra el ser humano procesando cantidades impresionantes de información, como es la clasificación de patrones, el control y por supuesto la predicción. La parte central de las Redes Neuronales es la neurona, la cual se considera como un elemento de procesamiento simple, y al trabajar e interactuar junto con otras, permite aproximar funciones específicas. Las Redes Neuronales son una clase de arquitectura de procesamiento en paralelo (Landassuri, 2006).

Las neuronas están constituidas principalmente de cuatro partes, (ver Figura 5), dendritas, el cuerpo de la neurona, sinapsis y el axón. (Gómez, 2010)

- **Dendrita:** son el receptor de la neurona, son ramificaciones de fibras nerviosas las cuales recibe las señales eléctricas de otras neuronas y ésta pasa a través del cuerpo de la neurona que se encarga de la suma de estas señales de entrada.
- **El axón:** es una fibra larga cuyo objetivo es llevar la señal desde el cuerpo de la neurona hacia otras neuronas mediante el contacto de la dendrita de estas (Gómez, 2010). Cambia de las entradas a las salidas (Landassuri, 2006).
- **Sinapsis:** La conexión con otras neuronas (Landassuri, 2006).

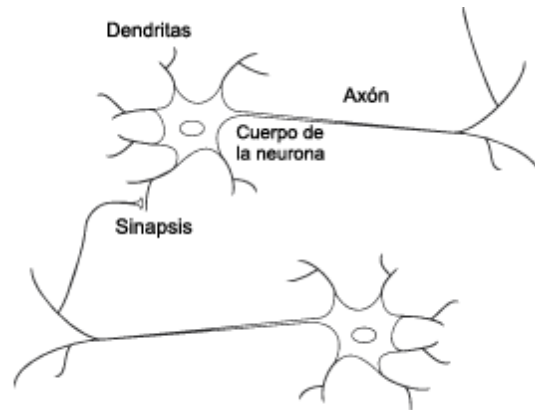


Figura 5. Partes principales de la neurona biológica. Fuente Gómez, 2010

Algunos autores usan dichos términos al referirse a las partes de una neurona computacional, en este caso no se emplean dichos nombres, únicamente se utiliza: nodos o neuronas, conexión entre nodos, función de transferencia, entradas y salidas (Landassuri, 2006).

La red neuronal se asemeja al cerebro en dos aspectos principales:

- El conocimiento es adquirido por la red a través de un proceso de aprendizaje.
- Los fuerza de conexión entre las neuronas, llamado pesos sinápticos, son usados para el almacenamiento del conocimiento adquirido (Gómez, 2010).

3.2.1. Conceptos básicos y características de las RNA

No existe una definición general de red neuronal artificial, existiendo diferentes según el texto o artículo consultado (Martín et al, 2009), desde las definiciones cortas y genéricas hasta las que intentan explicar más detalladamente qué son las redes neuronales (Matich, 2001). Por ejemplo:

- Una Red Neuronal Artificial es una gran cantidad de procesadores paralelos distribuidos, formados por una única unidad de procesamiento, la cual tiene la cualidad natural de almacenar el conocimiento experimental y tener ese conocimiento disponible para usar (Córdova, 2012).

- Un sistema de computación compuesto por un gran número de elementos simples, elementos de procesos muy interconectados, los cuales procesan información por medio de su estado dinámico como respuesta a entradas externas (Matich, 2001).
- Las RNA o sistemas conexionistas son sistemas de procesamiento de la información cuya estructura y funcionamiento están inspirados en las redes neuronales biológicas. Consisten en un conjunto de elementos simples de procesamiento llamados nodos o neuronas conectadas entre sí por conexiones que tienen un valor numérico modificable llamado peso (Montaño, 2002).

Las definiciones expuestas son una muestra de que cada autor las define de una manera. Parece ser que en todas ellas aparece el componente de simulación del comportamiento biológico (Martín et al, 2009).

Las RNA al margen de "parecerse" al cerebro presentan una serie de características propias del cerebro. Por ejemplo las RNA aprenden de la experiencia, generalizan de ejemplos previos a ejemplos nuevos y abstraen las características principales de una serie de datos (Basogain, 1998).

Es decir tienen que:

- **Aprender:** adquirir el conocimiento de una cosa por medio del estudio, ejercicio o experiencia. Las RNA pueden cambiar su comportamiento en función del entorno. Se les muestra un conjunto de entradas y ellas mismas se ajustan para producir unas salidas consistentes (Basogain, 1998).
- **Generalizar:** extender o ampliar una cosa. Las RNA generalizan automáticamente debido a su propia estructura y naturaleza. Estas redes pueden ofrecer, dentro de un margen, respuestas correctas a entradas que presentan pequeñas variaciones debido a los efectos de ruido o distorsión (Basogain, 1998).

- **Abstraer:** aislar mentalmente o considerar por separado las cualidades de un objeto. Algunas RNA son capaces de abstraer la esencia de un conjunto de entradas que aparentemente no presentan aspectos comunes o relativos (Basogain, 1998).

3.2.2. Modelo neuronal artificial

En todo modelo artificial de neurona se tienen cuatro elementos básicos (Martín et al, 2009):

- **Un conjunto de conexiones, pesos o sinapsis que determinan el comportamiento de la neurona.** Martín et al (2009) señalan que estas conexiones pueden ser excitadoras (presentan un signo positivo), o inhibitoras (conexiones negativas). Gómez (2010), menciona que cada una está caracterizada por un peso o fuerza. Una señal de entrada p_j de la sinapsis j está conectada con la neurona k la cual es multiplicada por el peso sináptico w_{kj} .
- **Un sumador o combinación lineal.** Combina las señales de entrada ponderándolas por las respectivas sinapsis de la neurona (Gómez, 2010). Se encarga de sumar todas las entradas multiplicadas por las respectivas sinapsis (Martín et al, 2009).
- **Una función de activación f .** La cual limita o normaliza la amplitud de la salida de la neurona (Gómez, 2010).
- **Un umbral exterior.** Determina el umbral por encima del cual la neurona se activa (Martín et al, 2009).

Esquemáticamente, una neurona artificial quedaría representada por la Figura 6.

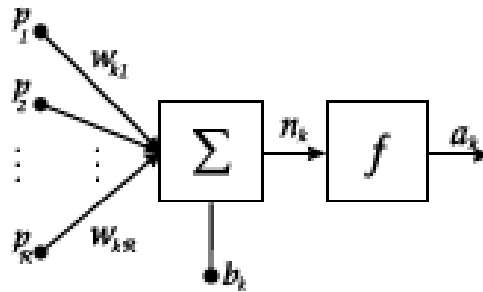


Figura 6. Modelo neuronal artificial. Fuente: Gómez, 2010

3.2.3. Función de transferencia o activación

Una neurona biológica puede estar activa (excitada) o inactiva (no excitada); es decir, que tiene un “estado de activación”. Las neuronas artificiales también tienen diferentes estados de activación; algunas de ellas solamente dos, al igual que las biológicas, pero otras pueden tomar cualquier valor dentro de un conjunto determinado (Matich, 2001).

Las funciones de transferencia también son llamadas como funciones de activación y estas determinan la relación entre los nodos de entrada y salida (Landassuri, 2006). Además define la salida de la neurona en términos de la entrada neta n_k . La elección de esta función dependerá si satisface alguna especificación del problema que se desea resolver (Gómez, 2010)

Existe una gran variedad de funciones de transferencia sin embargo, las siguientes funciones que se describen a continuación son las más usadas (Gómez, 2010).

- **Función de transferencia limite estricto (*hard limit*) o función umbral.** Establece la salida de la neurona como 0 si el argumento es menor que 0 o 1 si su argumento es mayor o igual a 0. Este tipo de función es utilizada para clasificar a la entrada de la neurona en dos categorías diferentes (Gómez, 2010).

$$f(n_k) = \begin{cases} 0, & \text{si } n_k < 0 \\ 1, & \text{si } n_k \geq 0 \end{cases} \dots\dots\dots \text{(Ecuación 17)}$$

En la práctica, es aconsejable tener una función de transferencia el cual tome valor de -1 o 1 , en tal caso, la función de transferencia *hardlim* se toma como una función

simétrica respecto al origen. De aquí que la función *hardlims* se define como la función *hardlim* simétrica respecto al origen (Ver Ecuación 18), es decir (Gómez, 2010):

$$f(n_k) = \begin{cases} -1, & n_k < 0 \\ 1, & n_k \geq 0 \end{cases} \dots\dots\dots \text{(Ecuación 18)}$$

- **Función lineal.** La salida para esta función de transferencia será igual a la entrada neta (Gómez, 2010).

$$f(n_k) = n_k \dots\dots\dots \text{(Ecuación 19)}$$

- **Función sigmoideal.** Este tipo de función es comúnmente usada en la construcción de redes neuronales artificiales. Un ejemplo de la función sigmoideal es la función log-sigmoideal el cual está definida mediante la Ecuación 20 (Gómez, 2010).

$$f(n_k) = \frac{1}{1+e^{-n_k}} \dots\dots\dots \text{(Ecuación 20)}$$

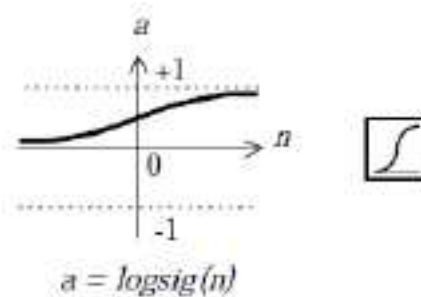


Figura 7. Función sigmoideal.

Mientras que la función *hardlim* toma valores de 0 o 1 la función sigmoideal toma un rango continuo de valores entre 0 y 1, además la función *hardlim* es una función no diferenciable mientras que la sigmoideal lo es (Gómez, 2010).

- **Función tangente sigmoideal.** Esta función está definida como (Gómez, 2010):

$$f(n_k) = \frac{e^{n_k} - e^{-n_k}}{e^{n_k} + e^{-n_k}} \dots\dots\dots \text{(Ecuación 21)}$$

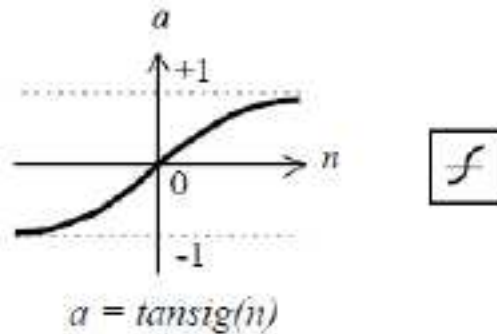


Figura 8. Función tangente sigmoideal. Fuente: Matich, 2001

3.2.4. Regla de aprendizaje

El aprendizaje de una red neuronal consiste en hallar los valores precisos de los pesos de sus conexiones para que pueda resolver un determinado problema. El proceso general consiste en ir introduciendo una serie de datos patrón y ajustar los pesos siguiendo un determinado criterio (Moreno, 2009).

La serie de tiempo es dividida en dos conjuntos: el conjunto de entrenamiento y el conjunto de prueba. El conjunto de entrenamiento sirve para obtener los pesos de la Red, mientras que el conjunto de prueba sirve para validar la eficacia de la misma. Debido a que se quiere minimizar el error de una función, el entrenamiento se puede llevar a cabo de diferentes formas (Landassuri, 2006). Los algoritmos de entrenamiento o los procedimientos de ajuste de los valores de las conexiones de las RNA se pueden clasificar en dos grupos: supervisado y no supervisado (Basogain, 1998).

- **Entrenamiento Supervisado:** se caracteriza porque el proceso de aprendizaje se realiza mediante un entrenamiento controlado por un agente externo (supervisor, maestro) que determina la respuesta que debería generar la red a partir de una entrada determinada (Matich, 2001). El entrenamiento consiste en presentar un vector de entrada a la red, calcular la salida de la red, compararla con la salida deseada, y el error o diferencia resultante se utiliza para realimentar la red y cambiar los pesos de acuerdo con un algoritmo que tiende a minimizar el error (Basogain, 1998).

En el aprendizaje supervisado, se tiene un conjunto de ejemplos llamados conjunto de entrenamiento los cuales son un grupo de entradas con salidas esperadas de la forma de la Ecuación 22 (Gómez, 2010):

$$\{p_1, t_1\}, \{p_2, t_2\}, \dots, \{p_Q, t_Q\}, \dots \quad (\text{Ecuación 22})$$

Donde p_q es una entrada o patrón de la red y t_q es la correspondiente salida esperada llamado objetivo. Cada salida que la red proporcione por cada patrón de entrada es comparada con las salidas esperadas, de aquí que la regla de aprendizaje ajusta los pesos y las ganancias de la red de tal forma que las salidas que proporcione la red sean cercanas a los objetivos esperados (Gómez, 2010).

- **Entrenamiento No Supervisado:** Los sistemas no supervisados son modelos de aprendizaje más lógicos en los sistemas biológicos (Basogain, 1998). No requieren influencia externa para ajustar los pesos de las conexiones entre sus neuronas. La red no recibe ninguna información por parte del entorno que le indique si la salida generada en respuesta a una determinada entrada es o no correcta (Matich, 2001). El conjunto de vectores de entrenamiento consiste únicamente en vectores de entrada. El algoritmo de entrenamiento modifica los pesos de la red de forma que produzca vectores de salida consistentes. El proceso de entrenamiento extrae las propiedades estadísticas del conjunto de vectores de entrenamiento y agrupa en clases los vectores similares (Basogain, 1998).

En cuanto a los algoritmos de aprendizaje no supervisado, en general se suelen considerar dos tipos, que dan lugar a los siguientes aprendizajes (Matich, 2001):

- Aprendizaje hebbiano.
- Aprendizaje competitivo y comparativo (Matich, 2001).

3.2.4.1. Conjunto de datos

Las variables de entrada pueden tener diferencias de valores de varios órdenes de magnitud de forma que el aprendizaje de la red se verá influenciado por estas diferencias ya que el incremento de pesos de una neurona es proporcional a su entrada. Una manera de evitar el problema es asignar a todas las entradas valores parecidos normalizando las entradas (Martín et al, 2009): Existen varias posibilidades:

$$x_k^* = \frac{x_k - \bar{x}_k}{\sigma_{x_k}} \dots\dots\dots \text{(Ecuación 23)}$$

Siendo:

- x_k^* =El valor normalizado de la entrada x_k
- x_k = Una variable de entrada
- \bar{x}_k =El valor medio de x_k
- σ_{x_k} =La desviación estándar de x_k

Suele utilizarse esta normalización porque presenta valor medio cero y varianza uno, con lo que evita tendencias de las entradas fuera de los rangos considerados como normales. Otra posible normalización de los datos de entrada es el siguiente (Martín et al, 2009):

$$x_k^* = \left[\frac{x_k - \text{min1}}{\text{max1} - \text{min1}} \right] [\text{max1} - \text{min1}] + \text{min2} \dots\dots\dots \text{(Ecuación 24)}$$

Donde entran en juego los valores máximos y mínimos antes y después de la transformación (que se denotan respectivamente por 1 y 2). La ventaja principal de este método es que al tratarse de una transformación lineal, se mantienen las relaciones existentes entre los patrones de entrenamiento originales. También se utiliza una normalización no lineal que hace uso de las funciones sigmoide o tangente hiperbólica, dependiendo de que se codifiquen los datos en el rango entre 0 y 1 o en el rango entre -1 y +1 respectivamente. La normalización vendría dada por la siguiente expresión (Martín et al, 2009):

$$x_k^* = \frac{1 - e^{-t}}{1 + e^{-t}} \text{ siendo } t = \frac{x_k - \bar{x}_k}{\sigma_{x_k}} \dots\dots\dots \text{(Ecuación 25)}$$

Con esta normalización, se tendrá que los valores cercanos al valor medio se transformarán en la zona lineal mientras que los alejados al valor medio lo harán en la zona plana por lo que es especialmente útil cuando alguna de las entradas presenta un rango de variación muy grande (Martín et al, 2009).

El conjunto de entrenamiento es el encargado de determinar los pesos de la Red Neuronal (aprendizaje), mediante un entrenamiento de la misma, el tamaño de este conjunto suele ser del 80% del conjunto de datos totales. No existe una regla determinada en cuanto al tamaño del conjunto de entrenamiento, pero entre mayor sea este, mayor información se tendrá del sistema; la desventaja es que entre más grande sea este conjunto, más tiempo le toma al algoritmo de entrenamiento ajustar los pesos de la Red, la ventaja es que ajusta mejor los pesos (Landassuri, 2006).

Una solución para evitar estos problemas es usar datos de validación, de esta manera la muestra es dividida en tres partes: los datos de entrenamiento que son los usados para el cálculo de gradientes y así actualizar pesos y bias (neuronas activas), los datos de validación son los encargados de monitorear el proceso durante el entrenamiento, el error de validación durante las primeras iteraciones tiende a bajar pero llega un momento cuando la red empieza a tener sobreajuste y el error de los datos de validación se incrementan, es aquí donde se detiene el entrenamiento evitando el sobreajuste y una buena generalización que será corroborada con el tercer segmento de la data, los datos de test o prueba (Sánchez, 2011).

3.2.5. Tipos de redes

Según el tipo de problema, es el tipo de representación utilizada, para lo cual existen dos grandes grupos (Landassuri, 2006):

- Redes feed-forward (no tienen ciclos)
- Redes Recurrentes (tienen ciclos)

Estos dos tipos de topologías pueden ser clasificadas en: multicapa, únicamente una capa, aleatoriamente conectados, localmente conectados, conexiones dispersas, completamente conectados, etc.

Las Redes multicapa son comúnmente usadas para realizar la predicción y debido a que no hay ningún teorema para su construcción, muchos autores usan de 1 capa hasta 3 o 4 para realizar la predicción.

Giudici, citado por Sánchez (2011), define al término arquitectura como la organización de la red neuronal: el número de capas, el número de unidades (neuronas) que siguen en cada capa, y la manera en que éstas son conectadas.

Respecto al tipo de capas que puede tener una red neuronal, existen tres tipos (Sánchez, 2011):

- **Entrada:** Son las encargadas de recibir sólo la información del ambiente externo, cada neurona en este caso corresponde a una variable explicatoria, en esta capa no se realiza ningún tipo de cálculo (Sánchez, 2011).
- **Salida:** Es la capa que produce los resultados finales, los cuales son enviados al ambiente exterior (Sánchez, 2011).
- **Ocultas:** Son las capas que se encuentran entre las capas input y output, y reciben este nombre porque no tienen contacto con el ambiente externo, son capas utilizadas exclusivamente para el análisis (Sánchez, 2011).

Según el número de capas:

- **Redes neuronales monocapas:** Se corresponde con la red neuronal más sencilla ya que se tiene una capa de neuronas que proyectan las entradas a una capa de neuronas de salida donde se realizan diferentes cálculos (Martín et al, 2009).

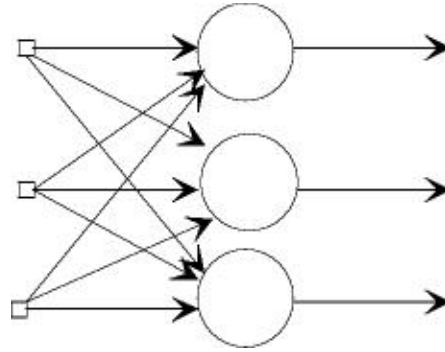


Figura 9. Red neuronal monocapa. Fuente: Martín et al, 2009

- **Redes neuronales multicapa:** Es una generalización de la anterior existiendo un conjunto de capas intermedias entre la entrada y la salida (capas ocultas). Este tipo de red puede estar total o parcialmente conectada (Martín et al, 2009).

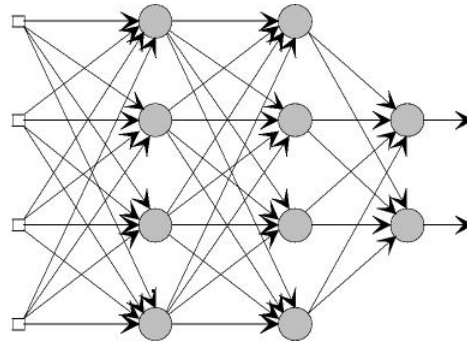


Figura 10. Esquema de una red multicapa. Fuente: Martín et al, 2009

Según el tipo de conexiones:

- **Redes neuronales no recurrentes:** En esta red la propagación de las señales se produce en un sentido solamente, no existiendo la posibilidad de realimentaciones. Lógicamente estas estructuras no tienen memoria (Martín et al, 2009).
- **Redes neuronales recurrentes:** Esta red viene caracterizada por la existencia de lazos de realimentación. Estos lazos pueden ser entre neuronas de diferentes capas, neuronas de la misma capa o, más sencillamente, entre una misma neurona. Esta estructura recurrente la hace especialmente adecuada para estudiar la dinámica de sistemas no lineales (Martín et al, 2009).

3.2.6. Algoritmo Retropropagación

Este algoritmo de **Retropropagación** es el algoritmo básico de aprendizaje que usa el perceptrón multicapa. Es el algoritmo mediante el cual se van adaptando todos los parámetros de la red. El tipo de entrenamiento que sigue este tipo de red es supervisado (Moreno, 2009).

El nombre de Retropropagación resulta de la forma en que el error es propagado hacia atrás a través de la red neuronal, en otras palabras el error se propaga hacia atrás desde la capa de salida. Esto permite que los pesos sobre las conexiones de las neuronas ubicadas en las capas ocultas cambien durante el entrenamiento (Matich, 2001).

El aprendizaje de la red se plantea como un problema de minimización de una determinada función de error. En general se usa como función del error, el error medio cuadrático (Moreno, 2009).

La importancia de la red retropropagación consiste en su capacidad de autoadaptar los pesos de las neuronas de las capas intermedias para aprender la relación que existe entre un conjunto de patrones de entrada y sus salidas correspondientes. Es importante la capacidad de generalización, facilidad de dar salidas satisfactorias a entradas que el sistema no ha visto nunca en su fase de entrenamiento. La red debe encontrar una representación interna que le permita generar las salidas deseadas cuando se le dan entradas de entrenamiento, y que pueda aplicar, además, a entradas no presentadas durante la etapa de aprendizaje para clasificarlas (Valencia et al, 2006).

El cambio de los pesos en las conexiones de las neuronas además de influir sobre la entrada global, influye en la activación y por consiguiente en la salida de una neurona. Por lo tanto, es de gran utilidad considerar las variaciones de la función activación al modificarse el valor de los pesos. Esto se llama sensibilidad de la función activación, de acuerdo al cambio en los pesos (Matich, 2001).

La invención del algoritmo Retropropagación ha desempeñado un papel vital en el resurgimiento del interés de las redes neuronales artificiales. Retropropagación es un método de entrenamiento de redes multicapa. Su potencia reside en su capacidad de



entrenar capas ocultas y de este modo supera las posibilidades restringidas de las redes de una única capa (Basogain, 1998).

El algoritmo puede dividirse en dos fases:

- **Propagación hacia delante:** se propagan las señales desde la capa de entrada hasta la de salida, determinándose la salida de la red y el error cometido al comparar ésta con el valor de la salida deseada que se le facilita a la red durante la etapa de aprendizaje (Martín et al, 2009).
- **Propagación hacia atrás:** En función de los errores cometidos en la capa de salida, el algoritmo se encarga de optimizar los valores de los pesos sinápticos que determinan las conexiones entre las neuronas mediante la Retropropagación del error desde la capa de salida a la de entrada a través de las sucesivas capas ocultas (Martín et al, 2009).

El algoritmo más simple es el que contiene una capa de entrada, una de salida y a lo más una capa intermedia de neuronas, pero no forzosamente se tiene que construir así, podemos aumentar el número de capas intermedias según lo requiera nuestro problema. Existen muchas variantes o paradigmas del algoritmo de Retropropagación para entrenar Redes feed-forward y una de las desventajas o ventajas, como se quiera ver, de esta configuración (feed-forward-retropropagación), es que se requiere de un entrenamiento supervisado con una gran cantidad de entradas-salidas (datos de entrenamiento), aparte de que no se puede entender bien el procedimiento de mapeo interno realizado por la Red (caja negra), lo que no garantiza de que el sistema converja a una solución aceptable (Landassuri, 2006).

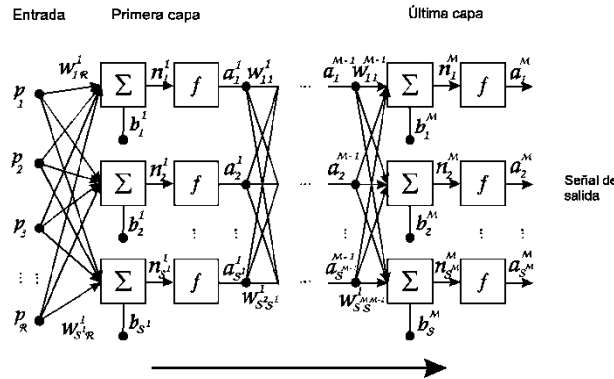


Figura 11. Primera fase del ciclo. Fuente: Gómez, 2010

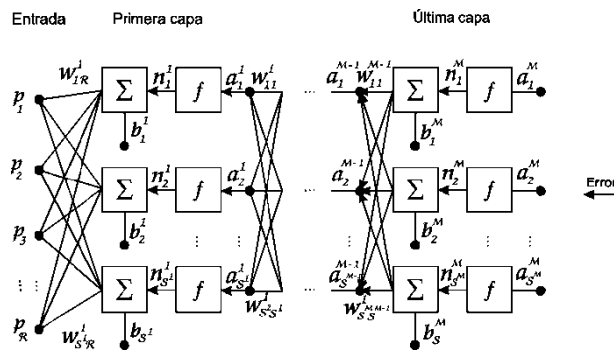


Figura 12. Segunda fase del ciclo. Fuente: Gómez, 2010

Existen variantes del algoritmo Retropropagación los cuales se dividen en dos categorías, en la primera involucra el desarrollo de técnicas heurísticas que surgen de un estudio del desempeño que tiene el algoritmo Retropropagación. Este tipo de técnicas incluyen ideas tales como variación del factor de aprendizaje y momentos (Gómez, 2010).

Por otra parte, la segunda categoría está formada por técnicas de optimización numérica, los cuales son variaciones del método de Newton tales como el algoritmo del gradiente conjugado y el algoritmo de Levenberg-Marquardt (Gómez, 2010).

El algoritmo de Levenberg-Marquardt es una técnica iterativa que localiza los mínimos de una función, la cual es expresada por una suma de cuadrados de una función no-lineal. Esta técnica fue diseñada alrededor de 1944 por K. Levenverg para tener un método que resolviera los problemas no-lineales con mínimos cuadrados. En 1963 D. W. Marquardt publica el artículo titulado “*An Algorithm for the Least-Squares Estimation of Nonlinear Parameters*” (Landassuri, 2006).

3.2.7. Redes neuronales en Matlab

El siguiente texto fue tomado del guía de usuario del Neural Network Toolbox For Use with MATLAB ® de Mathworks.

3.2.7.1. Crear la red (*newff*)

El primer paso en la formación de una red Retropropagación es crear el objeto de red. La función *newff* crea una red Retropropagación. Requiere cuatro entradas y devuelve el objeto de red. La primera entrada es una matriz R por 2 de valores mínimo y máximo para cada uno de los elementos R de vector de entrada. La segunda entrada es una matriz que contiene los tamaños de cada capa. La tercera entrada es un conjunto de células que contiene los nombres de las funciones de transferencia para ser utilizados en cada capa. La entrada final contiene el nombre de la función de entrenamiento para ser utilizado.

Por ejemplo, el siguiente comando crea una red de dos capas. Hay un vector de entrada con dos elementos. Los valores para el primer elemento de la gama de vectores de entrada entre -1 y 2, los valores del segundo elemento de la gama de vectores de entrada entre 0 y 5. Hay tres neuronas en la primera capa y una neurona en la segunda capa (de salida). La función de transferencia en la primera capa es tan-sigmoide, y la función de transferencia capa de salida es lineal. La función de formación se *traingd* (que se describe en una sección posterior).

$$net = newff([-1 2; 0 5], [3,1], {'tansig', 'purelin'}, 'traingd')$$

Este comando crea el objeto de red y también inicializa los pesos y sesgos de la red; Por lo tanto, la red está preparada para la formación. Hay momentos en que es posible que desee para reiniciar el peso, o para realizar una inicialización personalizada. La siguiente sección explica los detalles del proceso de inicialización.

3.2.7.2. Iniciando Pesos (*init*)

Antes de la formación de una red de alimentación hacia delante, los pesos y sesgos se deben inicializar. El comando *newff* inicializará automáticamente los pesos, pero es

posible que desee reiniciar ellos. Esto se puede hacer con el comando *init*. Esta función toma un objeto de red como entrada y devuelve un objeto de red con todos los pesos y sesgos inicializado. He aquí cómo se inicializa una red (o reinicialice):

$$= \text{netos init (neto)}$$

3.2.7.3. Entrenamiento

Una vez que los pesos de la red y los sesgos que se han inicializado, la red está preparada para el entrenamiento. La red puede ser entrenada para la aproximación de función (regresión no lineal), la asociación de patrón o modelo de clasificación. El proceso de entrenamiento requiere de un conjunto de ejemplos de comportamiento de la red apropiada entradas de red (p) y salidas de destino (t). Durante el entrenamiento, los pesos y sesgos de la red son ajustados de forma iterativa para reducir al mínimo la función de rendimiento de la red *net.performFcn*. La función de rendimiento predeterminado para las redes de Retropropagación es el cuadrado medio del error *mse* el error medio cuadrado entre la red (a) y la salida generada t .

El siguiente código crea un conjunto de entrenamiento de insumos p y apunta t . Para el entrenamiento por lotes, todos los vectores de entrada se colocan en una matriz.

$$p = [-1 \ -1 \ 2 \ 2; 0 \ 5 \ 0 \ 5];$$
$$t = [-1 \ -1 \ 1 \ 1]$$

A continuación, creamos la red de alimentación hacia adelante. Aquí se utiliza la *MinMax* función para determinar el rango de las entradas para ser utilizados en la creación de la red.

$$\text{net} = \text{newff}(\text{minmax}(p), [3,1], \{\text{tansig}, \text{purelin}\}, \text{trainlm})$$

En este punto, lo que se quiere modificar algunos de los parámetros de entrenamiento predeterminado.

- `net.trainParam.show = 50;`
- `net.trainParam.lr = 0,05;`
- `net.trainParam.epochs = 300;`

- `net.trainParam.goal = 1e-5;`

Si desea utilizar los parámetros de entrenamiento por defecto, los comandos anteriores no son necesarios.

- Levenberg-Marquardt (*trainlm*). El algoritmo de Levenberg-Marquardt fue diseñado para acercarse a la velocidad de formación de segundo orden sin tener que calcular la matriz de Hesse. Cuando la función de rendimiento tiene la forma de una suma de cuadrados (como es típico en las redes de alimentación directa de formación), entonces la matriz de Hesse se puede aproximar como

$$H = J^T J$$

Y el gradiente puede calcularse como:

$$g = J^T e$$

Donde J es la matriz jacobiana que contiene la primera derivada de los errores de la red con respecto a los pesos y sesgos, y e es un vector de los errores de red. La matriz jacobiana se puede calcular a través de una técnica de Retropropagación estándar que es mucho menos complejo que el cálculo de la matriz de Hesse. Este algoritmo parece ser el método más rápido para la formación de redes neuronales Retropropagación de tamaño moderado (hasta varios cientos de pesos).

3.2.7.4. Simulación (*Sim*)

La función *sim* simula una red. *Sim* toma la entrada de la red p , y la red objeto de red, y devuelve la salida de la red a . He aquí cómo usted puede utilizar *sim* para simular la red que hemos creado anteriormente para un solo vector de entrada:

$$\begin{aligned} p &= [1; 2] \\ a &= \text{sim}(\text{net}, p) \\ a &= -0.1011 \end{aligned}$$

(Si intenta estos comandos, el resultado puede ser diferente, dependiendo del estado de su generador de números aleatorios al inicializar la red.) A continuación, se llama *sim* para calcular los resultados para un conjunto simultáneo de tres vectores de entrada.

Esta es la forma modo de lote de simulación, en el que todos los vectores de entrada son lugar en una matriz. Esto es mucho más eficiente que la presentación de los vectores de uno a la vez.

$$p = [1 \ 3 \ 2; 2 \ 1 \ 1]$$
$$a = \text{sim}(\text{net}, p)$$
$$a = -0.1011 \ -0.2308 \ 0.1955$$

3.2.7.5. Preprocesamiento y posprocesamiento

El entrenamiento de la red neuronal se puede hacer más eficiente si algunos pasos de preprocesamiento se llevan a cabo en las entradas y los objetivos de la red. En esta sección, se describen varias rutinas de preprocesamiento que se pueden utilizar.

- **Min y Max (*premnmx*, *postmnmx*, *tramnmx*).** Antes del entrenamiento, a menudo es útil para escalar las entradas y las metas de modo que siempre se encuentran dentro de un rango específico. La función *premnmx* se puede utilizar a escala entradas y objetivos para que caigan en el rango [-1,1]. El código siguiente ilustra el uso de esta función.

$$[pn, \text{min. maxp}, tn, \text{mint}, \text{maxt}] = \text{premnmx}(p, t);$$
$$\text{net} = \text{train}(\text{net}, pn, tn);$$

Las entradas y las salidas originales de la red se dan en las matrices de p y t . Las entradas y salidas normalizadas, pn y tn , se devuelven en el intervalo [-1,1]. Los vectores minp y maxp contienen los valores máximos y mínimos de las entradas originales y mínimo, y los vectores de mint y maxt contienen los valores mínimos y máximos de las salidas originales. Después que la red ha sido entrenada, estos vectores se deben usar para transformar las entradas futuras que se aplican a la red.

Si *premnmx* se utiliza para escalar tanto las entradas y las salidas, entonces la salida de la red va a estar capacitado para producir resultados en el rango [-1,1]. Si desea convertir estas salidas de nuevo en las mismas unidades que se utilizaron para los objetivos originales, entonces usted debe utilizar la rutina

postmnmx. En el siguiente código, se simula la red que fue entrenado en el código anterior, y luego convertir la salida de la red de nuevo en las unidades originales.

$$an = sim(net, pn)$$

$$a = postmnmx(an, mint, maxt)$$

La salida de la red *an* corresponderá a las salidas normalizadas *tn*. La red no normalizada de salida *a* se encuentra en las mismas unidades que las salidas originales *t*.

Si *premmx* se utiliza para procesar previamente los datos del conjunto de entrenamiento, a continuación, siempre que se utilice la red entrenada con nuevos datos que deben ser preprocesados con el mínimo y máximos que se calcula para el conjunto de entrenamiento. Esto puede ser logrado con la rutina *tramnmx*. En el siguiente código, se aplica un nuevo conjunto de entradas a la red ya hemos entrenado.

$$pnewn = tramnmx(pnew, minp, maxp);$$

$$anewn = sim(net, pnewn);$$

$$anew = postmnmx(anewn, mint, maxt);$$

3.3. Modelos Empíricos

Los métodos de relleno o para completar datos faltantes se basan en suponer que existe una estructura de dependencia espacial que permite, para cada intervalo de tiempo, completar el valor faltante o con información regional. La información regional puede abarcar otras variables relacionadas que puedan ser explicativas de la variable con datos faltantes.

Las metodologías que permiten completar las series suponen, como es lógico, una región del espacio dentro de la cual son aplicables, la extensión dependerá de la precisión que se exija para el relleno, de la variable que se trate y del nivel temporal al cual se pretende completar valores. Por lo tanto, antes de elegir la metodología de relleno para completar una serie climática o hidrológica hay que tener en cuenta:

- Nivel de agregación temporal de los vacíos (o datos erróneos a corregir)

- La unidad física o región en la que se encuentra la estación a completar.
- La exactitud o margen de error que permite el estudio (Lobo, 2011).

3.3.1. Razón de distancia

Pizarro et al (2009), menciona que este método se utiliza para la estimación de datos menores de un año, en zonas planas no montañosas; las estaciones deben tener una disposición espacial lineal como la que se muestra en la Figura 13.

Lobo (2011), señala que si en una zona plana se cuenta con las estaciones A y B completas y una estación X con observaciones incompletas, se puede rellenar está en función de la distancia.

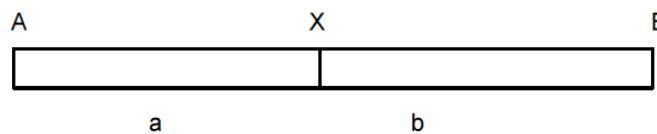


Figura 13. Esquema para completar datos faltantes por distancia. Fuente: Lobo, 2011

$$P_x = P_A + a * \frac{P_B - P_A}{(a+b)} \dots \dots \dots \text{(Ecuación 26)}$$

Donde:

- P_x : Variable climática en la estación X
- P_A : Variable climática en la estación A
- P_B : Variable climática en la estación B
- a y b: Distancias.

3.3.2. U.S. Weather Service

Este procedimiento puede ser muy confiable, y considera que el dato faltante de la estación A, puede ser estimado en base a las estaciones circundantes. Este método puede ser utilizado para estimar datos diarios, mensuales y anuales.

El método consiste en ponderar los valores observados en una cantidad W , igual al recíproco del cuadrado de la distancia D entre cada estación vecina y la estación A , El dato faltante buscado P_x será igual a (Servin, 2010):

$$P_x = \sum \frac{(P_i * W_i)}{W_i} \dots \dots \dots \text{(Ecuación 27)}$$

- P_i = Dato observado para la fecha faltante en las estaciones auxiliares circundantes
- $W_i=1/d_i^2$, siendo d_i la distancia entre cada estación circundante y la estación incompleta.

Se recomienda utilizar cuatro estaciones auxiliares (las más cercanas) y de manera que cada una quede localizada en uno de los cuadrantes que definen unos ejes coordenados que pasan por la estación incompleta, generalmente los ejes norte-sur y oriente-poniente (Servin, 2010).

3.3.3. Estimación de datos faltantes mediante regresión

Guevara (2003), menciona que la estimación de datos por regresión es ampliamente utilizada debido a la ventaja de poder emplearse en cualquier variable climática. Pueden estimarse datos faltantes o dudosos de una serie de un elemento a partir de otra serie del mismo elemento, en lugares ubicados en una misma zona climática, pero que entre ellos exista una alta relación, como puede ser la estimación de la radiación solar a partir de la insolación, estimación de la temperatura a partir de la altura o el rendimiento de un cultivo asociado a una o más variables climáticas.

Sin embargo, el ajuste o estimación por regresión requiere que se cumplan algunos requisitos para su aplicación; entre los más importantes:

- a) Que el coeficiente de correlación sea estadísticamente significativo, de lo contrario no se deberá ajustar ni estimar datos por regresión.

- b) Que los datos de cada serie sean independientes entre sí, es decir, que no exista auto correlación. De lo contrario, el valor de r y r^2 estarán influidos por los mismos datos, exagerando la correlación.

Siendo la gama de aplicación tan variada se hace por consiguiente necesario conocer el tipo de relación entre las variables:

- Relación lineal: $Y = a + bx$
- Exponencial: $Y = ab^x$; $Y = ae^{bx}$; $Y = e^{abx}$
- Logarítmica: $Y = ax^b$
- Parabólica: $Y = kx^n$; $Y = a + bx + cx^2$

3.3.3.1. Coeficiente de correlación

Sean X y Y dos variables aleatorias. Sean σ_{XY} , σ_X , y σ_Y la covarianza entre ellas y las desviaciones estándar respectivas, la correlación entre X y Y se denota por ρ_{XY} , y tiene por la Ecuación 28:

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \dots\dots\dots \text{(Ecuación 28)}$$

La correlación entre dos variables X y Y tienen las siguientes propiedades:

- $-1 \leq \rho_{XY} \leq 1$. Valores cercanos a 1 indican fuerte asociación positiva entre las variables, y valores cercanos a -1 indican fuerte asociación negativa entre ellas. Valores próximos a cero indican que no existe relación lineal entre ellas. Los valores de -1 y 1 se alcanzan cuando los valores de X y Y se encuentran sobre una línea recta, con pendiente positiva y negativa, respectivamente.
- La correlación entre X y Y no se afecta por cambios de escala en las variables. (Infante y Zarate, 2011).

Una vez calculado el valor del coeficiente de correlación interesa determinar si tal valor obtenido muestra que las variables X e Y están relacionadas en realidad o tan solo presentan dicha relación como consecuencia del azar. En otras palabras, nos preguntamos por la significación de dicho coeficiente de correlación.



Un coeficiente de correlación se dice que es significativo si se puede afirmar, con una cierta probabilidad, que es diferente de cero. Más estrictamente, en términos estadísticos, preguntarse por la significación de un cierto coeficiente de correlación no es otra cosa que preguntarse por la probabilidad de que tal coeficiente proceda de una población cuyo valor sea de cero. A este respecto, como siempre, tendremos dos hipótesis posibles:

- **Hipótesis nula:** El coeficiente de correlación obtenido procede de una población cuya correlación es cero.
- **Hipótesis alterna:** El coeficiente de correlación obtenido procede de una población cuyo coeficiente de correlación es distinto de cero.

Desde el supuesto de la Hipótesis nula se demuestra que la distribución muestral de correlaciones procedentes de una población caracterizada por una correlación igual a cero, sigue una ley de t Student con N-2 grados de libertad, de media el valor poblacional y desviación tipo:

$$S_r = \sqrt{\frac{1-(\rho_{XY})^2}{N-2}} \dots\dots\dots \text{(Ecuación 29)}$$

En consecuencia, dado un cierto coeficiente de correlación obtenido en una determinada muestra se trata de comprobar si dicho coeficiente es posible que se encuentre dentro de la distribución muestral especificada por la Hipótesis nula. A efectos prácticos, se calcula el número de desviaciones tipo que se encuentra el coeficiente obtenido del centro de la distribución, según la fórmula conocida:

$$t = \frac{\rho_{XY}-0}{\sqrt{\frac{1-(\rho_{XY})^2}{N-2}}} \dots\dots\dots \text{(Ecuación 30)}$$

Y se compara el valor obtenido con el existente en las tablas para un cierto nivel de significación α y N-2 grados de libertad ($t_{(\alpha, N-2)}$), que como se sabe, marca el límite (baja probabilidad de ocurrencia, según la Hipótesis nula de pertenencia de un cierto

coeficiente ρ_{XY} a la distribución muestra de correlaciones procedentes de una población con correlación igual a cero. De esta forma si:

- $t > t_{(\alpha, N-2)}$: Se rechaza la hipótesis nula. La correlación obtenida procede de una población cuya correlación es cero. Por tanto las variables están relacionadas.
- $t \leq t_{(\alpha, N-2)}$: Se acepta la hipótesis nula. La correlación obtenida procede de una población cuya correlación es distinta de cero. Por tanto las variables no están relacionadas.

3.3.3.2. Regresión lineal simple

Pedroza y Dikovskiy (2006), señalan que el concepto de regresión se refiere al “cuantum” o “cantidad de cambio” que experimenta una variable dependiente (Y), en relación al cambio de una unidad de una variable independiente (X). Al igual que con otros procedimientos estadísticos, destaca que en la regresión lineal se desea realizar una inferencia estadística partiendo de los valores muestrales obtenidos; por tanto, se deben cumplir ciertos requisitos, que en el caso de la regresión lineal son los siguientes:

- Normalidad y Homogeneidad de varianzas en la variable dependiente (Y) del modelo para los valores fijos de la variable independiente (X).
- Independencia de las observaciones de Y .
- Linealidad en la relación entre las variables.

El modelo de regresión simple es el siguiente:

$$Y_1 = B_0 + B_1 X_1 + e_1 \dots \dots \dots \text{(Ecuación 31)}$$

Donde:

- Y_1 = es la variable dependiente.
- B_0 = es la ordenada en el origen, o bien es el intercepto.
- B_1 = es la pendiente de la recta de regresión.
- e_1 = es el término del error, es decir la diferencia entre los valores predichos por la regresión y los valores reales.

Pizarro et al (2009), menciona que este método es uno de los más utilizados; se recomienda para la estimación de datos mensuales y anuales de la estación en estudio y las de una estación pluviométrica cercana, que cuente con una estadística consistente y observada.

3.3.3.3. Regresión lineal múltiple

El análisis de regresión múltiple puede considerarse como una extensión del análisis de regresión simple al caso en que existan dos o más variables explicativas o independientes. Se analiza lo que estas variables independientes aportan en conjunto para explicar una variable respuesta o dependiente en escala de intervalos continua y con distribución normal. Es decir, es útil para estudiar fenómenos en que intervienen múltiples factores para describir la relación entre la variable dependiente y las variables independientes o para predecir el valor de la variable dependiente para valores conocidos de las variables independientes. Para cumplir estos objetivos es necesario buscar un modelo que se ajuste a los datos disponibles (Taucher, 1997).

Álvarez (1994), menciona que conceptualmente, la única diferencia entre la regresión simple y la múltiple es que el número de variables independientes es mayor que 1. El modelo matemático, en este caso, es el siguiente:

$$Y_1 = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n + e_1 \dots \dots \dots \text{(Ecuación 32)}$$

En el modelo anterior, e_1 son los términos de residuos o errores, B_n , es el coeficiente de regresión correspondiente a la n ésima variable; indica el incremento de la variable dependiente, por aumento unitario de la n ésima variable independiente, suponiendo fijas el resto de las variables.

3.4. Parámetros estadísticos

3.4.1. Cuadrado medio del error (MSE)

Indica que para establecer la bondad de un estimador, se parte del hecho de ser deseable conocer si la estimación se encuentra lejos o cerca del valor verdadero, siempre desconocido. El error que se comete, al tomar como valor del parámetro x_i (valor estimado) el proporcionado por el estimador, es la diferencia $y_i - x_i$ (Valor observado –

Valor estimado), para eliminar el signo de las diferencias se genera su cuadrado $(y_i - x_i)^2$. Si se obtienen todas las posibles muestras y para cada una la correspondiente estimación, una media global de los errores sería la desviación cuadrática media de todos ellos o, más concretamente, su esperanza matemática, denominada error cuadrático medio, MSE, o desviación cuadrática media del estimador (Sánchez, 2006).

Fuertes (1998), describe la siguiente fórmula para calcular el error cuadrático medio:

$$\mathbf{MSE} = \frac{\sum_{i=1}^n (y_i - x_i)^2}{n} \dots\dots\dots \text{(Ecuación 33)}$$

Donde:

- y_i = Valor observado
- x_i = Valor estimado
- n = número de observaciones

Sánchez (2006), menciona que un valor pequeño del error cuadrático medio, indica que en promedio, el estimador (x_i) no se encuentra lejos del parámetro y_i , por el contrario, si toma un valor alto será señal de que dicho estimador no aproxima bien en promedio al valor desconocido.

3.4.2. Raíz cuadrada del cuadrado medio del error (RMSE)

Cai et al, citados por Aguirre (2009) presentan la siguiente fórmula para el cálculo de la RMSE.

$$\mathbf{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}} \dots\dots\dots \text{(Ecuación 34)}$$

Donde:

- RMSE = Raíz cuadrada del cuadrado medio del error

Un modelo muy adecuado es cuando el RMSE es muy cercano a cero, esto es, que el valor promedio de las desviaciones cuadradas se aproxime a cero y se eleva al cuadrado para que sea más riguroso el valor.

3.4.3. Error absoluto medio (MAE)

Willmott, citado por Fuertes (1998), utiliza la siguiente expresión para el cálculo del error absoluto medio:

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \dots\dots\dots \text{(Ecuación 35)}$$

Donde:

- MAE : Error absoluto medio

El error absoluto medio da la medida de las diferencias en promedio entre los valores pronosticados y los valores observados (Casellas, 2005).

3.4.4. Error estándar de estimación (SEE)

Irmak et al, citados por Alkaeed, et al (2006) calculan la SSE de la siguiente manera:

$$\text{SEE} = \sqrt{\left[\frac{1}{n(n-2)} \right] \left[n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2 - \frac{[n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i) * (\sum_{i=1}^n y_i)]^2}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \right]} \dots \text{(Ecuación 36)}$$

Donde:

- SEE = Error estándar de estimación

Un SEE menor implica el mejor desempeño del método aplicado.

3.4.5. Índice de Willmott de correspondencia (d)

$$d = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (y_i^* + x_i^*)^2} \dots\dots\dots \text{(Ecuación 37)}$$

Donde:

- d = Índice de Willmott de correspondencia
- $y_i^* = |y_i - \bar{y}|$
- $x_i^* = |x_i - \bar{y}|$; siendo \bar{y} la media de las observaciones



Si el valor de $d \geq 0.95$ es aceptado, pero debe ser lo más cercano a la unidad; con esto se busca que el cociente de los errores elevados al cuadrado entre la suma elevada al cuadrado de las desviaciones de los valores estimados “x” y observados “y” con respecto a la media de los valores observados (\bar{y}) sea lo más próxima a cero (Cai et al, citado por Aguirre, 2009).

3.4.6. Diferencia relativa (DR)

Cai et al, citado por Aguirre (2009) muestra la siguiente ecuación para el cálculo de DR:

$$DR = \frac{RMSE}{\bar{y}} \times 100 \dots \dots \dots \text{(Ecuación 38)}$$

Donde:

- \bar{y} = Promedio de los valores de los valores observados de la variable climática a analizar
- DR = Diferencia relativa, en %

Con esta estadística se acepta el modelo si se tiene $DR \leq 20\%$, se busca que sea lo más cercano a cero dando un rango de 20% para aceptar el modelo. Es muy adecuado para comparar la variabilidad de métodos, ya que se divide un valor promedio de los errores que existen entre los valores observados y estimados de cada método empleado entre el valor medio de los valores observados.

4. MATERIALES Y MÉTODOS

4.1. Materiales

4.1.1. Datos meteorológicos utilizados

Se utilizaron datos climatológicos de temperatura promedio, precipitación y evaporación mensual de las estaciones meteorológicas ubicadas en tres regiones climáticas en México, de acuerdo a la clasificación del Instituto Nacional de Estadística y Geografía (INEGI), los cuales son: Templado Subhúmedo, Muy Seco Semicálido y Cálido Subhúmedo. Estas regiones fueron seleccionadas porque representan el 50.10% aproximadamente del territorio nacional. En la siguiente Figura se muestran las regiones climáticas antes mencionadas.

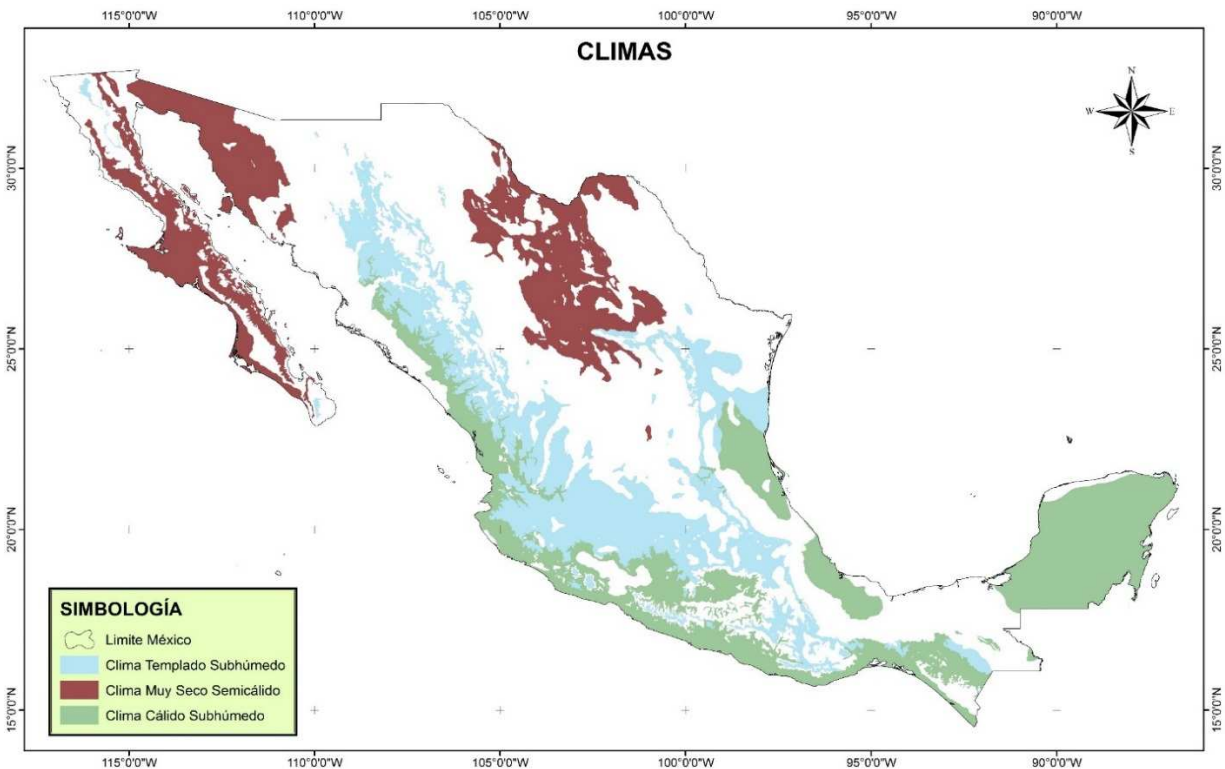


Figura 14. Regiones climáticas a utilizar

La descripción de estas regiones se presenta a continuación:

- **Templado Subhúmedo.** Observa en su mayoría temperaturas entre 10° y 18° C y de 18° a 22°C, sin embargo en algunas regiones puede disminuir a menos de 10°C; registra precipitaciones de 600 a 1,000 mm en promedio durante el año.

- **Muy Seco Semicálido.** Registra temperaturas en promedio de 18° a 22° C, con casos extremos de más de 26°C; presentando precipitaciones anuales de 100 a 300 mm en promedio.
- **Cálido Subhúmedo.** En él se registran precipitaciones entre 1,000 y 2,000 mm anuales y temperaturas que oscilan de 22° y 26°, con regiones en donde superan los 26°C.

4.1.2. Selección de las estaciones meteorológicas de estudio

4.1.2.1. Diagnóstico de las estaciones meteorológicas en México

El Extractor Rápido de Información Climática (ERIC III) v 3.0-2009, maneja un total de 6,063 estaciones y contiene información de registros diarios de variables climáticas (temperatura observada, temperatura máxima y mínima, precipitación, evaporación, tormenta, granizo, niebla y cobertura del cielo) en solo 5,466 de ellas.

De acuerdo al índice que el Servicio Meteorológico Nacional (SMN) definió para identificar que tan completos están los datos en cada una de las estaciones, podemos resumir lo siguiente:

Cuadro 1. Frecuencia de porcentaje de datos para las estaciones meteorológicas

Porcentaje de datos	Frecuencia	% acumulado
(0-10]	24	0.44%
(10-20]	37	1.12%
(20-30]	71	2.41%
(30-40]	105	4.34%
(40-50]	141	6.92%
(50-60]	204	10.65%
(60-70]	319	16.48%
(70-80]	572	26.95%
(80-90]	1043	46.03%
(90-100]	2950	100.00%
Suma	5,466	

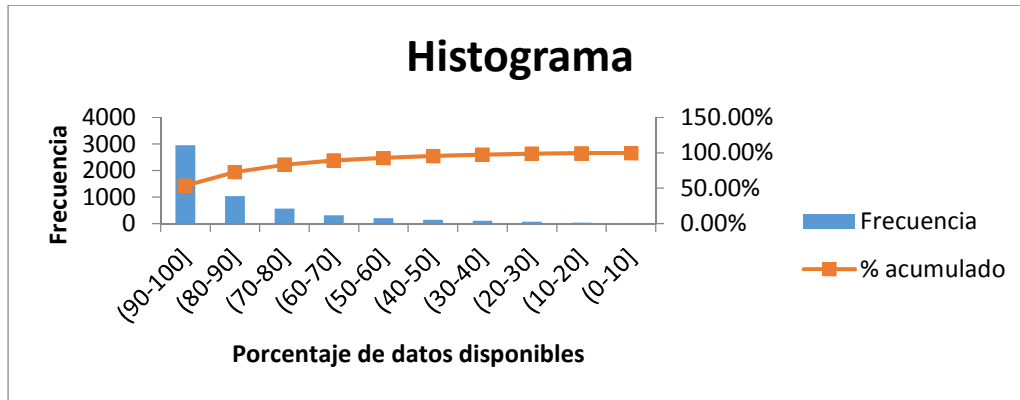


Figura 15. Histograma de frecuencias de información disponible en las estaciones meteorológicas

De acuerdo al Cuadro 1 y a la Figura 15 es importante destacar que 3,993 estaciones meteorológicas es decir el 73.05% cuenta con información del 80-100% y 2,950 estaciones que representan el 53.97% del total de las estaciones tiene información del 90-100%, según los datos reportados en el paquete de cómputo ERIC III.

Un aspecto importante a destacar es que de las 6,063 estaciones registradas, 60 estaciones no tienen coordenadas geográficas, por lo que no se ubicaron dentro del mapa y quedaron fuera del proceso de selección que se llevó a cabo.

4.1.2.2. Procedimiento de selección

A. Selección de las estaciones base

Como primer paso se seleccionaron las estaciones meteorológicas que se encuentran dentro de cada una de las regiones climáticas antes descritas y se obtuvieron los siguientes resultados que se muestra a continuación y en la Figura siguiente.

- **Cálido Subhúmedo.** Se seleccionaron un total de 1,332 estaciones meteorológicas.
- **Templado Subhúmedo.** Se seleccionaron un total de 1,493 estaciones meteorológicas.
- **Muy Seco Semicálido.** Se seleccionaron un total de 279 estaciones meteorológicas.

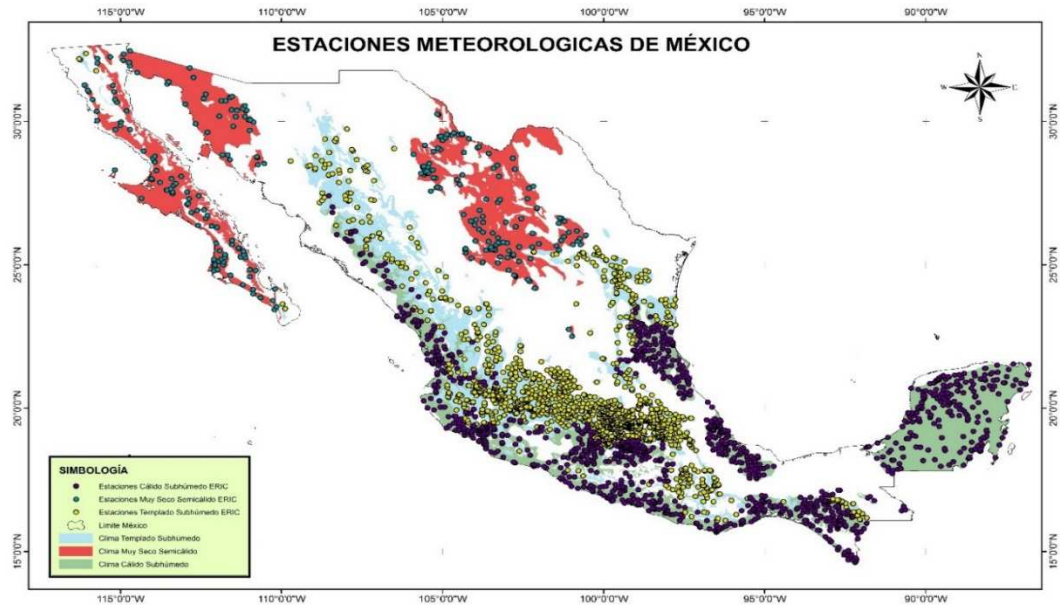


Figura 16. Ubicación espacial de las estaciones meteorológicas seleccionadas en cada región climática. Posteriormente se realizó una segunda selección tomando como criterio que la estación tenga un índice de datos disponibles entre un 99%-100% y que además tengan 15 años como mínimo de datos continuos. Cabe señalar que a pesar de que la estación tenga un índice del 100% de datos completos éstas poseen datos perdidos, por lo que fue conveniente hacer una revisión de los datos de las estaciones al momento de seleccionarlás. De esta revisión se pudo observar que los periodos disponibles de datos de temperatura y precipitación son iguales, caso contrario a la evaporación.

Los resultados de esta segunda selección se tienen lo siguiente:

- **Cálido Subhúmedo.** Se seleccionaron 17 estaciones meteorológicas.
- **Templado Subhúmedo.** Se seleccionaron 17 estaciones meteorológicas.
- **Muy seco semicálido.** Se seleccionaron ocho estaciones meteorológicas.

Se realizó una tercera selección, la cual se dividió en tres partes, primero para las variables de temperatura, la segunda la precipitación y la tercera para la evaporación. Para la primera parte se agruparon las estaciones de acuerdo a la cantidad de información disponible 15, 20 y 25 años para las variables climáticas antes mencionadas.

A continuación en el siguiente Cuadro se presentan el número de estaciones de acuerdo a los periodos de información para temperatura.

Cuadro 2. Número de estaciones disponibles en cada periodo

15 años	Cantidad	20 años	Cantidad	25 años	Cantidad
Cálido Subhúmedo (cs)					
1990-2004	7	1985-2004	5	1985-2004	3
1985-1999	10	1980-1999	7	1980-1999	6
1980-1994	8	1975-1994	7	1975-1994	4
1975-1989	8	1970-1989	5		
1970-1984	8				
Muy Seco Semicálido (mss)					
1990-2004	2	1985-2004	1	1985-2004	0
1985-1999	2	1980-1999	1	1980-1999	1
1980-1994	2	1975-1994	2	1975-1994	2
1975-1989	3	1970-1989	2		
1970-1984	4				
Templado Subhúmedo (ts)					
1990-2004	6	1985-2004	6	1985-2004	5
1985-1999	9	1980-1999	7	1980-1999	7
1980-1994	11	1975-1994	11	1975-1994	8
1975-1989	12	1970-1989	9		
1970-1984	12				

Como se observa en el Cuadro 2 no existe disponibilidad de información de manera continua y esto considerando que en el programa de cómputo ERIC III reporta 100% de disponibilidad de datos. Para la selección se utilizó el criterio de utilizar las estaciones que tengan información de 15 años de manera continua y con datos actuales (1990-2004). La ubicación espacial de las estaciones seleccionadas se presenta en el siguiente mapa, los puntos que se encuentran en un color diferente al negro representan las estaciones seleccionadas (Ver Figura 17).

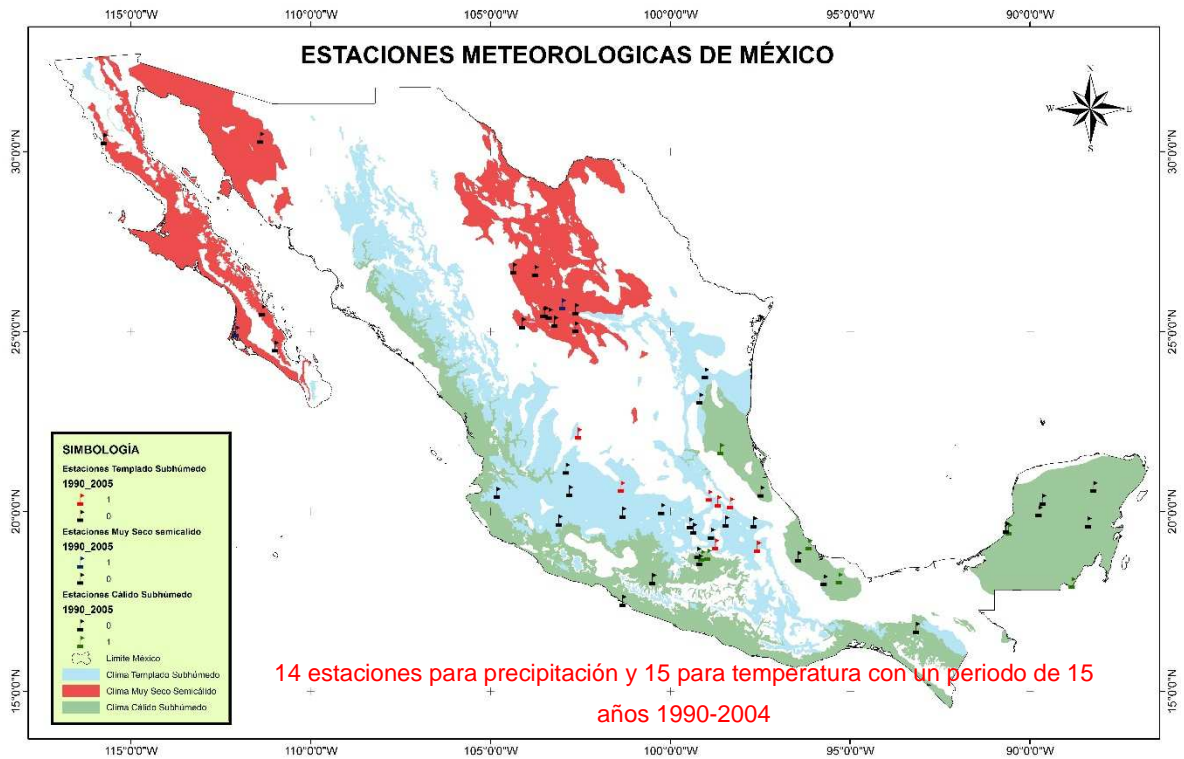


Figura 17. Distribución espacial de las estaciones meteorológicas seleccionadas

En resumen las estaciones seleccionadas para temperatura para un periodo de 15 años (1990-2004) son:

- **Templado Subhúmedo.** Seis estaciones.
- **Muy Seco Semicálido.** Dos estaciones.
- **Cálido Subhúmedo.** Siete estaciones.

En resumen las estaciones seleccionadas para precipitación para un periodo de 15 años (1990-2004) son:

- **Templado Subhúmedo.** Cinco estaciones.
- **Muy Seco Semicálido.** Dos estaciones.
- **Cálido Subhúmedo.** Siete estaciones.

Estas estaciones meteorológicas seleccionadas se utilizaron para estimar datos faltantes de precipitación y temperatura. Para la tercera parte (caso de la evaporación) se hará otra selección ya que como se comentó en los párrafos anteriores la información disponible no es la misma que para las variables precipitación y temperatura. Los resultados de la selección de las estaciones meteorológicas para estimar la variable de evaporación, se presentan a continuación.

- **Templado Subhúmedo.** Se seleccionó el periodo de 1984-1998 para un periodo de 15 años, siendo un total de siete estaciones meteorológicas.
- **Muy Seco Semicálido.** Se seleccionó el periodo de 1983-1992 para un periodo de 10 años. La razón de cambiar el periodo de análisis de 15 años a 10 años fue debido a que, de las estaciones seleccionadas no se encontró información continua de 15 años, por lo que se optó a reducir el periodo de análisis y con lo cual se encontraron que dos estaciones meteorológicas cuentan con dicho criterio.
- **Cálido Subhúmedo.** Se seleccionó el periodo de 1990-2004 para un periodo de 15 años, siendo un total de cuatro estaciones meteorológicas.

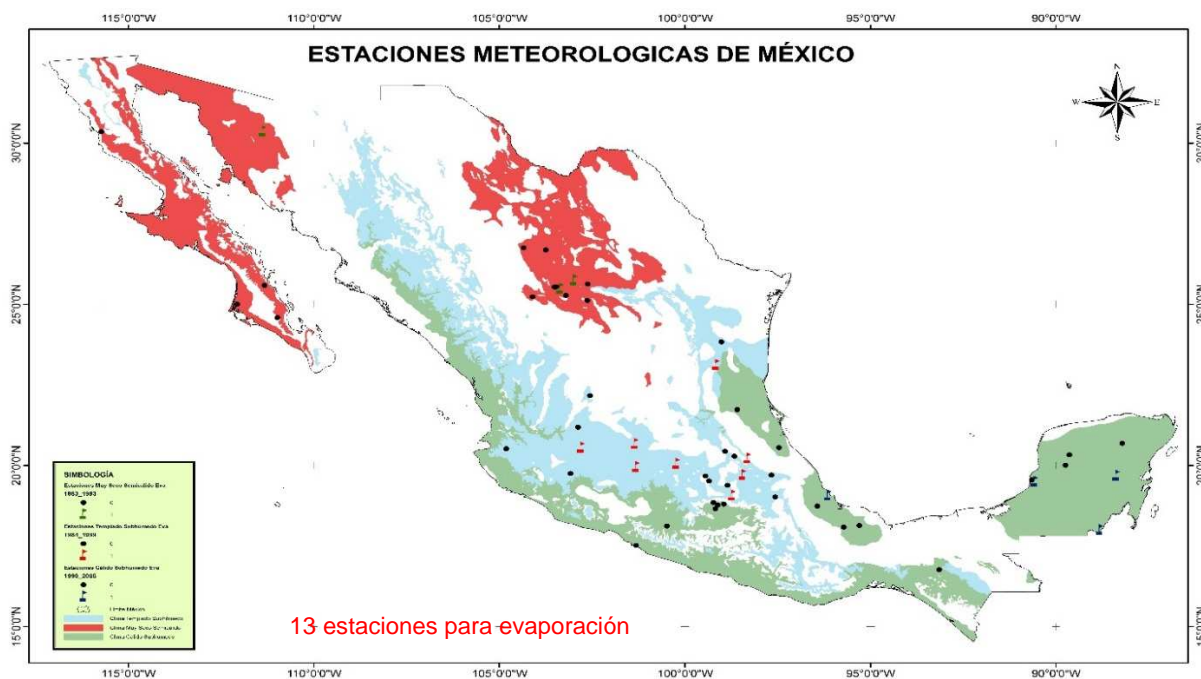


Figura 18. Distribución espacial de las estaciones seleccionadas para evaporación

Para la selección de cada una de las estaciones no se utilizaron principios estadísticos, ya que la disponibilidad de la información impidió el uso de estos.

B. Selección de las estaciones circundantes

Se realizó una selección espacial de cinco estaciones circundantes a las estaciones bases seleccionadas en el punto anterior. El principal criterio que se tomó en cuenta es la distancia a la estación base, es decir las más cercanas; otro criterio que se consideró es que tengan un 90% de datos completos y además sus periodos de información sean iguales a las estaciones base. Esta selección se realizó para las estaciones base de evaporación y las de temperatura y precipitación. Los resultados para cada uno se presentan en las siguientes Figuras.

En la Figura 19 se muestra la localización espacial de las estaciones circundantes seleccionadas para las estaciones base de evaporación.

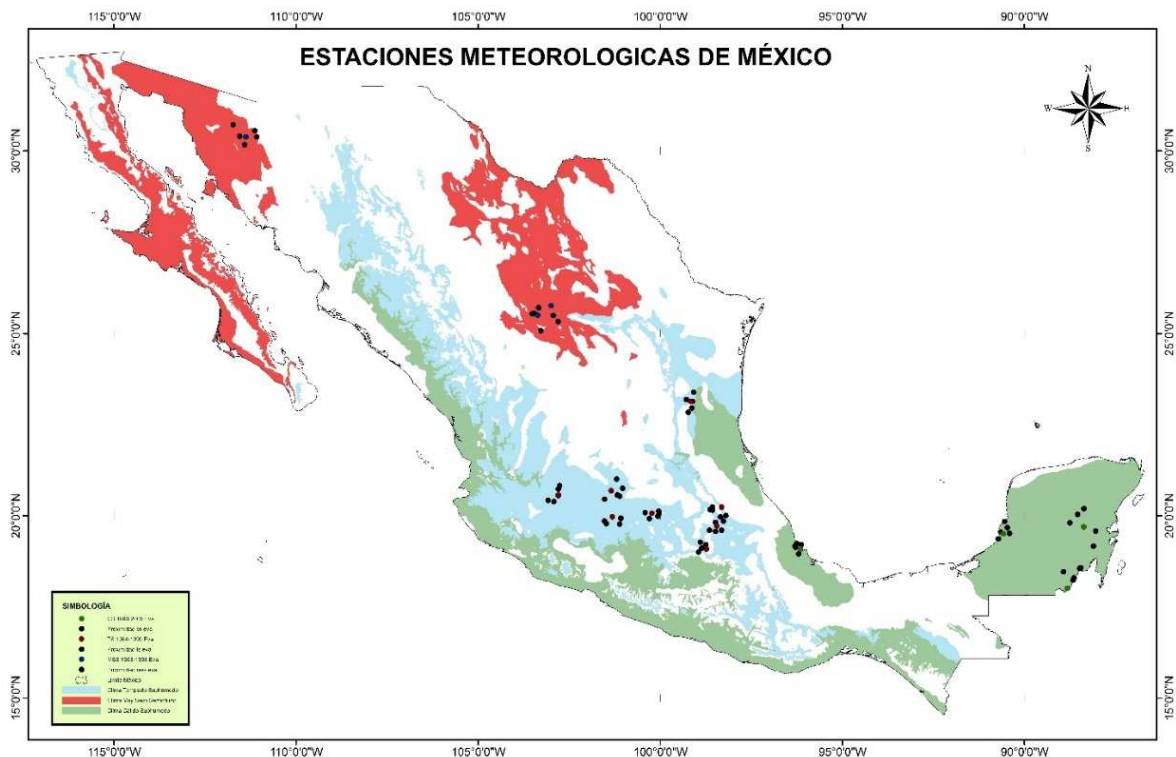


Figura 19. Localización de las estaciones cercanas a las bases para evaporación

En la Figura 20 se muestra la localización espacial de las estaciones circundantes seleccionadas para las estaciones base de temperatura y precipitación.

que componen la muestra. El primer elemento seleccionado condiciona los siguientes, que son elegidos a partir del arranque aleatorio y según un salto de amplitud constante. La selección sistemática implica un número de arranque y un intervalo de muestreo. Definidos ambos la muestra queda automáticamente establecida (Vivanco, 2005).

Desde el punto de vista operativo se procede según las siguientes etapas.

- Generar un número de arranque aleatorio. El número de arranque será un valor i que estará contenido en el intervalo con cota inferior 1 y superior igual al valor del coeficiente de elevación N (cantidad total de datos). Este elemento será el primero que compondrá la muestra.
- Seleccionar de modo sistemático los siguientes elementos a k distancia del punto de arranque aleatorio. El valor k corresponde N/n , siendo n =cantidad de datos a simular.
- En consecuencia, se establece la siguiente sucesión a partir del primer elemento i , $i+k$, $i+2k$, $i+3k$, $i+(n-1)k$. El procedimiento es sistemático porque se van seleccionando elementos que están a k lugares del i -ésimo de la lista (Vivanco, 2005).

4.1.4. Software utilizados

Para obtener los datos de temperatura promedio, evaporación y precipitación mensual se utilizó el ERIC III ®, programa que contiene información del banco de datos histórico nacional del Servicio Meteorológico Nacional (SMN) de la Comisión Nacional del Agua (CONAGUA).

Para el desarrollo del modelo de redes neuronales artificiales se utilizaron las herramientas del programa de computo MATLAB R2013 ®.

Para los modelos empíricos y el modelo transformada de Wavelet se utilizó el software Microsoft Visual Studio 2010 ® para programar el código requerido para cada modelo.



Se utilizó el software Microsoft Office Excel 2013 ® para el cálculo de los parámetros estadísticos y simulación de datos faltantes, así como regresión lineal simple y múltiple.

La elaboración de los diagramas de flujo se realizaron con el paquete de computo Microsoft Office Visio 2013 ®.

Para la elaboración de mapas y selección de las estaciones meteorológicas se utilizó el programa de cómputo Arc Map 10 ®.

4.2. Métodos

Se determinaron los datos faltantes utilizando tres modelos diferentes los cuales son: redes neuronales artificiales, modelo matemático transformada de wavelet y modelos empíricos (US Weather Service, razón de distancia, regresión lineal simple y regresión lineal múltiple). Estos modelos fueron elegidos por lo siguiente:

- La transformada de Wavelet fue utilizada para estimar datos faltantes de precipitación en Perú, obteniendo resultados muy cercanos a los reales (Carbajal et al, 2010).
- Las redes neuronales se utilizaron para estimar datos faltantes en estaciones meteorológicas de Venezuela y los resultados obtenidos ha revelado ser un modelo útil para la reconstrucción de los datos faltantes de las series (Saba et al, 2008).
- Los métodos empíricos se seleccionaron debido a que son sencillos de utilizar.

A continuación se presentan las metodologías para la determinación de datos faltantes en cada una de las regiones climáticas.

4.2.1. Modelos empíricos

4.2.1.1. US Weather Service (UWS)

Este modelo considera que el dato faltante de una estación X , puede ser estimado en base a las estaciones circundantes. En este apartado se describió el procedimiento seguido para la obtención de datos faltantes por este método.

- Se calculó la distancia existente entre la estación base y cada una de las estaciones circundantes. Para la obtención de la distancia en kilómetros se utilizó la latitud y longitud de cada una de las estaciones meteorológicas y se aplicó la siguiente ecuación:

$$D = 6378.137 * ACos(Cos(Lat1) * Cos(Lat2) * Cos(Lon2 - Lon1) + Sin(Lat1) * Sin(Lat2))..... (Ecuación 39)$$

Donde:

- Lat1 y lon1: Es la latitud y longitud en radianes de la estación base.
- Lat2 y lon2: Es la latitud y longitud en radianes de la estación circundante.
- Se ordenaron y enumeraron las estaciones circundantes de la más cercana a la más lejana.
- Se generaron los datos faltantes considerando cinco, cuatro, tres, dos y una estación cercana y aplicando la Ecuación 27.

4.2.1.2. Razón de la distancia (RD)

Si en una zona plana se cuenta con las estaciones A y B completas y una estación X con observaciones incompletas, se puede rellenar está en función de la distancia. El procedimiento seguido para la obtención de datos faltantes por este método se presenta a continuación:

- Se calculó la distancia existente entre la estación base y cada una de las estaciones circundantes. Para la obtención de la distancia en kilómetros se utilizó

la latitud y longitud de cada una de las estaciones meteorológicas y se aplicó la Ecuación 39:

- Se seleccionaron las dos estaciones circundantes más cercanas a la estación base.
- Se generaron los datos faltantes aplicando la Ecuación 26.

4.2.1.3. Regresión lineal simple (RL)

Mediante el modelo de regresión lineal simple se pretendió describir el grado de relación entre cada una de las variables climáticas, es decir estimar el dato faltante de una variable climática mediante una segunda variable climática. El procedimiento seguido para estimar los datos faltantes con regresión lineal simple mediante mínimos cuadrados se presenta a continuación:

- Se determinó la variable climática a considerar como variable independiente (X) y de igual manera se determinó la variable climática a considerar como variable dependiente (Y).
- Calcular el valor promedio de los valores de la variable independiente (\bar{X}) y la variable dependiente (\bar{Y}).
- Para cada par de valores (X, Y) se realizó lo siguiente:

- Restar el valor promedio de la variable independiente al dato X_i . [3]

$$x_i = X_i - \bar{X} \dots \dots \dots \text{(Ecuación 40)}$$

- Restar el valor promedio de la variable dependiente al dato Y_i . [4]

$$y_i = Y_i - \bar{Y} \dots \dots \dots \text{(Ecuación 41)}$$

- Elevar al cuadrado el valor de x_i (x_i^2) [5]
- Elevar al cuadrado el valor de y_i (y_i^2) [6]
- Multiplicar el valor x_i por y_i ($x_i y_i$) [7]
- Se procedió a calcular el valor de la pendiente de regresión.



$$b = \frac{\sum xy}{\sum x^2} \dots \dots \dots \text{(Ecuación 42)}$$

- Se realizó la diferencia entre el valor x_i y la pendiente de regresión (b) [8]

$$y_c = x_i - b \dots \dots \dots \text{(Ecuación 43)}$$

- Se elevó al cuadrado el valor de y_c [9]
- Se obtuvo el error entre el valor de y_i y y_c [10]

$$e = y_i - y_c \dots \dots \dots \text{(Ecuación 44)}$$

- Se elevó al cuadrado el valor del error (e) [11]
- Se calculó el valor del termino independiente, mediante la siguiente formula.

$$a = \bar{Y} - b\bar{X} \dots \dots \dots \text{(Ecuación 45)}$$

- Se calculó la desviación estándar de la regresión

$$Sr = \sqrt{\frac{\sum e^2}{n-2}} \dots \dots \dots \text{(Ecuación 46)}$$

Donde n= número total de datos a analizar

- Se calculó la desviación estándar de la pendiente de regresión (b)

$$Sb = \frac{Sr}{\sqrt{\sum x^2}} \dots \dots \dots \text{(Ecuación 47)}$$

- Se calculó la desviación estándar con respecto a x_i

$$Sx = \frac{\sum x^2}{n-1} \dots \dots \dots \text{(Ecuación 48)}$$

- Se calculó la desviación estándar con respecto a y_i

$$Sy = \frac{\sum y^2}{n-1} \dots \dots \dots \text{(Ecuación 49)}$$

- Se calculó la covarianza

$$Cxy = \frac{\sum xy}{n-1} \dots \dots \dots \text{(Ecuación 50)}$$

- Se calculó b'

$$b' = \frac{\sum xy}{\sum y^2} \dots \dots \dots \text{(Ecuación 51)}$$

- Se obtuvo el coeficiente de determinación

$$R^2 = b * b' \dots \dots \dots \text{(Ecuación 52)}$$

- Se calculó el coeficiente de correlación

$$R = \sqrt{R^2} \dots \dots \dots \text{(Ecuación 53)}$$

- La ecuación de la regresión lineal simple es:

$$Y = bX + a \dots \dots \dots \text{(Ecuación 54)}$$

- Para estimar el dato faltante (Y) se sustituyó el valor referente a la variable climática de apoyo (X) en el tiempo correspondiente.

	1	2	3	4	5	6	7	8	9	10	11	12
Número de punto	X	Y	x	y	x ²	y ²	xy	yc	yc ²	e	e ²	Y ²
1	107.50	23.40	-3.66	-2.10	13.41	4.41	7.69	-0.02	0.00	-2.08	4.33	547.56
2	97.50	27.50	-13.66	2.00	186.63	4.00	-27.31	-0.07	0.01	2.07	4.30	756.25

Figura 21. Ejemplo de tabla para el modelo de regresión lineal simple

4.2.1.4. Regresión lineal múltiple (RM)

El modelo de regresión lineal múltiple es idéntico al modelo de regresión lineal simple, con la única diferencia de que aparecen más variables explicativas.

Mediante el modelo de regresión lineal múltiple se pretendió estimar el dato faltante de una variable climática mediante dos variables climáticas distintas. El procedimiento seguido para estimar los datos faltantes con regresión lineal múltiple mediante mínimos cuadrados se presenta a continuación:

- Se determinaron las variables climáticas a considerar como variables independientes (X_1 y X_2) y de igual manera se determinó la variable climática a considerar como variable dependiente (Y).
- Se calculó el valor promedio de los valores de la variable independiente (\bar{X}_1, \bar{X}_2) y la variable dependiente (\bar{Y}).

- Para los valores de las variables independientes y dependientes (X_1, X_2, Y) se realizó lo siguiente:

- Restar el valor promedio de la primera variable independiente al dato X_{1i} [4]

$$x_{1i} = X_{1i} - \bar{X}_1 \dots \dots \dots \text{(Ecuación 55)}$$

- Restar el valor promedio de la segunda variable independiente al dato X_{2i} [5]

$$x_{2i} = X_{2i} - \bar{X}_2 \dots \dots \dots \text{(Ecuación 56)}$$

- Restar el valor promedio de la variable dependiente al dato Y_i . [6]

$$y_i = Y_i - \bar{Y} \dots \dots \dots \text{(Ecuación 57)}$$

- Se multiplicó los valores de las variables $x_{1i} * x_{2i}$ [7]
- Se elevaron al cuadrado cada uno de los valores de x_{1i} , x_{2i} y y_i [8], [9] y [10]
- Se multiplicaron los valores de $x_{1i} * y_i$ [11]
- Se multiplicaron los valores de $x_{2i} * y_i$ [12]
- Se obtuvieron las pendientes de regresión, utilizando las siguientes expresiones.

$$b_1 = \frac{\sum x_{1i} y_i \sum x_{2i}^2 - \sum x_{1i} x_{2i} \sum x_{2i} y_i}{\sum x_{1i}^2 \sum x_{2i}^2 - \sum x_{1i} x_{2i} \sum x_{1i} x_{2i}} \dots \dots \dots \text{(Ecuación 58)}$$

$$b_2 = \frac{\sum x_{1i}^2 \sum x_{2i} y_i - \sum x_{1i} x_{2i} \sum x_{1i} y_i}{\sum x_{1i}^2 \sum x_{2i}^2 - \sum x_{1i} x_{2i} \sum x_{1i} x_{2i}} \dots \dots \dots \text{(Ecuación 59)}$$

- Se obtuvo el valor de y_c [13]

$$y_c = b_1 * x_{1i} + b_2 * x_{2i} \dots \dots \dots \text{(Ecuación 60)}$$

- Se elevó al cuadrado el valor de y_{ci} [14]
- Se obtuvo el error entre el valor de y_i y y_c [15]

$$e = y_i - y_c \dots \dots \dots \text{(Ecuación 61)}$$

- Se elevó al cuadrado el valor del error (e) [16]

- Se calculó el valor del término independiente de la regresión lineal múltiple.

$$b_0 = \bar{Y} - b_1\bar{x}_1 - b_2\bar{x}_2 \dots \dots \dots \text{(Ecuación 62)}$$

- Se calculó el coeficiente de determinación:

$$R^2 = 1 - \frac{\sum e^2}{\sum y^2} \dots \dots \dots \text{(Ecuación 63)}$$

- Se calculó el error típico (Sr)

$$Sr = \sqrt{\frac{\sum e^2}{n-3}} \dots \dots \dots \text{(Ecuación 64)}$$

- Se calculó el coeficiente de correlación parcial

$$r_{1-2} = \frac{\sum x_1 x_2}{\sqrt{\sum x_1^2 \sum x_2^2}} \dots \dots \dots \text{(Ecuación 65)}$$

- Se determinó el error típico de los parámetros b1 y b2.

$$S_{b1} = \frac{Sr}{\sqrt{\sum x_1^2 (1-r_{1-2}^2)}} \dots \dots \dots \text{(Ecuación 66)}$$

$$S_{b2} = \frac{Sr}{\sqrt{\sum x_2^2 (1-r_{1-2}^2)}} \dots \dots \dots \text{(Ecuación 67)}$$

- La ecuación de la regresión lineal múltiple es:

$$Y = b_1 X_1 + b_2 X_2 + b_0 \dots \dots \dots \text{(Ecuación 68)}$$

- Para estimar el dato faltante (Y) se sustituyó el valor referente a la variable climática de apoyo (X1 y X2) en el tiempo correspondiente.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Dato	X1	X2	Y	x1	x2	y	x1*x2	(x1)*2	(x2)*2	(y)*2	x1*y	x2*y	yc	(yc)*2	e(y-yc)	e*2	Y2
1	107.50	87.20	23.400	-3.66	-37.79	-2.10	138.38	13.41	1428.44	4.41	7.69	79.39	-1.43	2.06	-0.66644	0.444141	547.56000
	97.50	147.20	27.500	-13.66	22.21	2.00	-303.36	186.63	493.07	4.00	-27.31	44.40	0.72	0.52	1.28012	1.638702	756.25000

Figura 22. Ejemplo de tabla para el cálculo de regresión lineal múltiple

4.2.2. Modelo matemático Transformada de Wavelet (TW)

A continuación se describe el algoritmo de descomposición y reconstrucción de las series climáticas de precipitación, temperatura promedio y evaporación:

- Seleccionar la estación vecina más cercana con datos completos (g), a la estación con datos faltantes (f).
- Se procede a la descomposición de la estación vecina (g):
 - Se aplicó un filtro pasa alto y un filtro pasa bajo a la serie de tiempo de la variable climática a analizar. Kouro y Musalem (2002), muestran en la Figura 23 el procedimiento anterior, Donde S es la señal que se desea analizar, A la salida de pasa bajos y D la salida del filtro pasa altos. Naturalmente, los filtros son diseñados de tal manera que sean complementarios, es decir, la suma de A y D debe ser S . Si se diseñaran los filtros en forma muy separada se perdería información, o en caso contrario se estaría amplificando la banda de entrecruzamiento. Los filtros a utilizar son los siguientes:

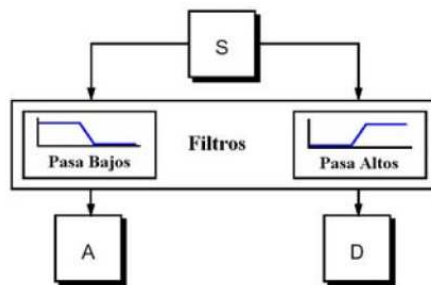


Figura 23. Diagrama de descomposición de señales. Fuente: Kouro y Musalem, 2002

$$PB = \frac{d_{i-1} + d_i}{\sqrt{2}} \dots \dots \dots \text{(Ecuación 69)}$$

$$PA = \frac{d_{i-1} - d_i}{\sqrt{2}} \dots \dots \dots \text{(Ecuación 70)}$$

Dónde: PB= Filtro pasa bajo; d_{i-1} = dato en la posición $i-1$, d_i =dato en la posición i , con $i \neq 1$; PA= Filtro pasa alto.

- Al aplicar el procedimiento anterior se duplican nuestros datos originales, pues por cada muestra de S se generan un par de muestras (A , D), por lo que se procede a realizar un submuestreo o eliminación de información. Kouro y Musalem (2002) muestran que en la Figura siguiente se representa el submuestreo en donde los círculos con flechas representan la eliminación de datos, luego, cD (coeficientes de detalle) y cA (coeficientes de aproximación) son los nuevos coeficientes obtenidos de la etapa de filtración. Intuitivamente se puede concluir que al tener cD y cA , en conjunto, se tiene la misma cantidad de datos que la señal original S .

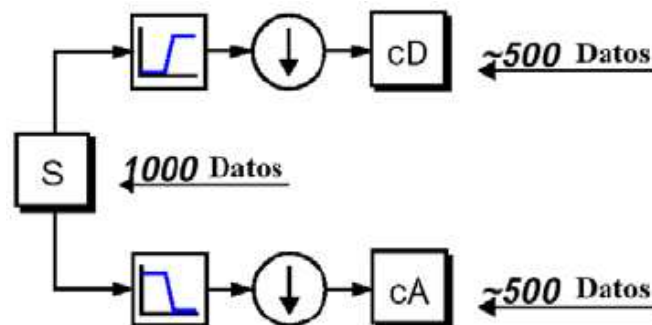


Figura 24. Representación de la eliminación de información o submuestreo. Fuente: Kouro y Musalem, 2002

- Se procede a descomponer nuevamente el coeficiente de aproximación en dos nuevas señales $cD1$ y $cA1$. Kouro y Musalem (2002), reconocen a este proceso como filtros multiniveles, es decir, aplicar el mismo procedimiento a las señales de salida de la primera etapa, y así sucesivamente hasta el nivel de precisión que se requiera. Lo anterior da origen a una descomposición multinivel conocida como ramificación o árbol de descomposición Wavelet, cuya idea es la expuesta en la Figura 25.

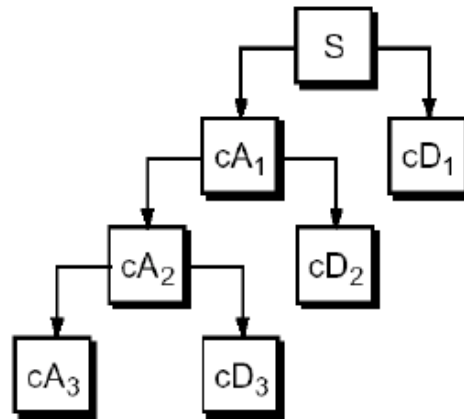


Figura 25. Árbol de descomposición Wavelet. Fuente: Kouro y Musalem, 2002

El nivel de precisión o grado de descomposición (N) que se utilizó para esta investigación fue hasta que se tuviera dos o tres coeficientes de aproximación.

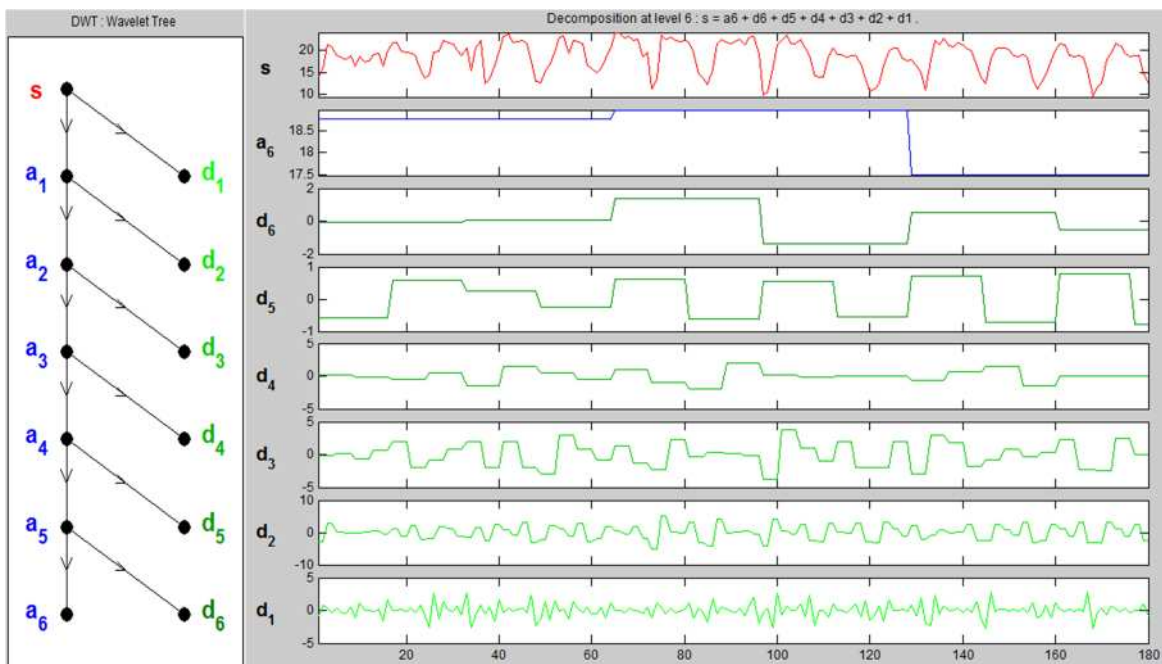


Figura 26. Ejemplo de Árbol de descomposición Wavelet

- Se promediaron los valores de f en el nivel N , cada 2^N datos. Esto debe corresponder a una aproximación de cA_N , por lo que se toma este resultado como la tendencia en la reconstrucción.
- Se tomó la tendencia de f (obtenida en el paso anterior) en el nivel N y todos los coeficientes de detalle de g de todos los niveles $N, (N-1), (N-2)$.

- Se procedió a la reconstrucción de toda la serie de la estación con datos faltantes. La metodología sigue el razonamiento en dirección contraria, es decir, a partir de los coeficientes cA_i y cD_i (i depende del número de niveles) debe obtenerse S . Lo anterior queda ilustrado en la siguiente Figura.

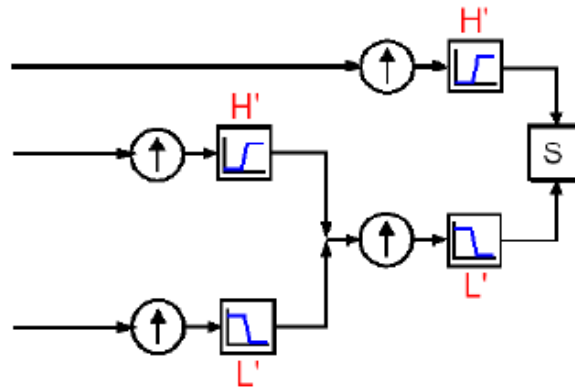


Figura 27. Esquema de reconstrucción Wavelet. Fuente: Kouro y Musalem, 2002

Los pasos seguidos para la reconstrucción son las siguientes:

- Se procedió a realizar un sobremuestreo a los coeficientes de aproximación (f) y detalle (g) del último nivel. Kouro y Musalem (2002), mencionan que se debe realizar el sobremuestreo para compensar el submuestreo realizado en el proceso de descomposición. Este proceso equivale a poner un cero entre cada dos muestras de los coeficientes de aproximación y detalle (Ferrer, 2011).
- Se aplicó nuevamente los filtros pasa bajo y filtros pasa alto, ecuación 69 y 70 respectivamente.
- Se sumaron la aproximación reconstruida (filtro pasa bajo) y los detalles reconstruidos (filtro pasa alto) para obtener el coeficiente de aproximación en el nivel $N-1$.
- El proceso de reconstrucción es iterativo hasta reconstruir la estación f .

4.2.3. Modelo de Redes Neuronales Artificiales

Para el desarrollo de este modelo se utilizaron las herramientas del programa de computo MATLAB R2013 ®. El flujo de trabajo para el proceso de diseño de redes neuronales en general tiene seis pasos principales, los cuales se presentan a continuación:

- a) Recopilar datos
- b) Crear la red
- c) Configuración de la red e Inicializar los pesos y sesgos
- d) Entrenar a la red
- e) Validar la red (análisis post-entrenamiento)
- f) Utilizar la red

Los pasos anteriores se describen a continuación:

- a) Recopilar datos.** Se creó un archivo de entrada en formato **.prn (p)* que contiene únicamente el número de mes continuo del periodo analizado. Posteriormente se creó un archivo de salida en formato **.prn (t)* que contiene únicamente el valor de la variable climática analizar correspondiente al número de mes continuo del periodo analizado.

Se procedió a normalizar los datos de entrada y salida para que se encuentren en un mismo rango de información. Utilizando el siguiente código:

$$[pn, minp, maxp, tn, mint, maxt] = premnmx(p, t);$$

Donde:

- P: Datos de entrada
- T: Datos de salida

El siguiente paso fue dividir los datos de entrada y salida para utilizarlos en el entrenamiento, validación y prueba. Se tomó un 80% de los datos para el conjunto de entrenamiento, un 10% para utilizarlos para la prueba de la red neuronal y un 10% para la validación de la red. El código que se utilizó para dicho procedimientos es el siguiente:

$$iitst=2:10:Q;$$

```
iival=10:10:Q;  
iitr=[1:10:Q 3:10:Q 4:10:Q 5:10:Q 6:10:Q 7:10:Q 8:10:Q 9:10:Q];  
val.P=pn(:,iival);  
val.T=tn(:,iival);  
test.P=pn(:,iitst);  
test.T=tn(:,iitst);  
ptr=pn(:,iitr);  
ttr=tn(:,iitr);
```

b) Crear la red. Se definió la arquitectura de la red neuronal del tipo Retropropagación, el cual fue de tres capas ocultas, con la primera capa variable en cuanto al número de neuronas (N_e) y la segunda y la tercera capa de una neurona. El entrenamiento que se utilizó es de tipo Levenberg-Marquardt (*trainlm*) ya que realiza mejor el ajuste de una función (regresión no lineal). La función de rendimiento que se utilizó es el error cuadrático medio *MSE* correspondiente al error cuadrado entre la red y la salida. La red neuronal a implementar se muestra en la Figura 28:

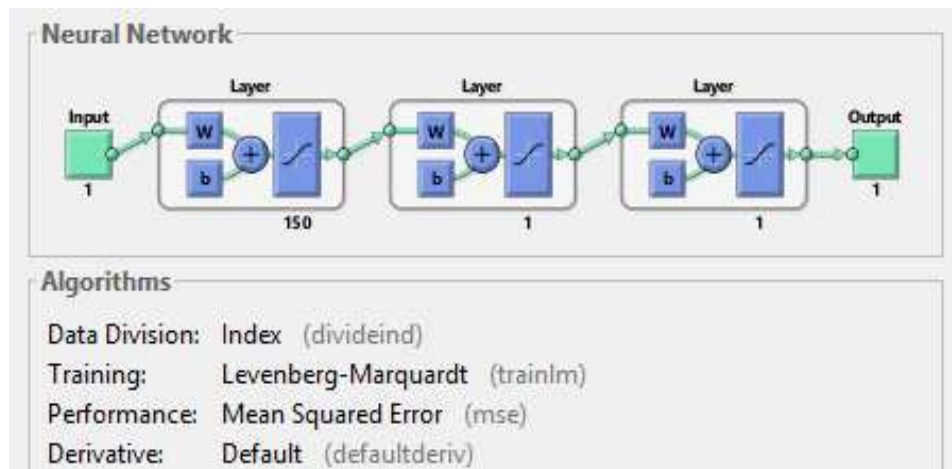


Figura 28. Esquema de la red neuronal a utilizar

Se procedió a definir el número de neuronas a ocupar en la primera capa oculta de la red neuronal.

c) Configuración de la red e Inicializar los pesos y sesgos. Posteriormente se creó la red neuronal con el siguiente código:

```
net=newff(minmax(ptr),[Ne, 1, 1],{'tansig','purelin','purelin'},'trainlm');
```

Este comando crea el objeto de red y también inicializa los pesos y sesgos de la red.

d) Entrenar la red. Se realizó el entrenamiento de la red neuronal cuya función de rendimiento es el error cuadrático medio. El entrenamiento terminó hasta que el coeficiente de correlación general, considerando datos de entrenamiento, prueba y validación, sea mayor o igual a 0.80. El código utilizado para el entrenamiento es el siguiente:

```
[net,tr]=train(net,ptr,ttr,[],[],val,test);
```

```
an=sim(net,pn);
```

e) Validar la red (análisis post-entrenamiento). Se transformaron las salidas de la red a las mismas unidades originales. El código utilizado para dicha transformación es el siguiente:

```
sal=postmnmx(an,mint,maxt);
```

f) Utilizar la red neuronal. Después del entrenamiento de la red neuronal se cargó un archivo *.prn con la serie de tiempo completa. Con la red entrenada se obtuvieron los datos faltantes de la estación meteorológica a analizar. El código utilizado para utilizar la red entrenada se presenta a continuación.

```
pnewn=tramnmx(C,minp,maxp);
```

```
ANEW=sim(net,pnewn);
```

```
SALIDA=postmnmx(ANEW,mint,maxt);
```

Donde:

- C: Son los datos de entrada completos.

Por último se exportó los resultados a una hoja de Microsoft Office Excel 2013®, utilizando el siguiente código.

```
xlswrite('SALIDA.xls', SALIDA);
```

El código completo en Matlab 2013® utilizado para el modelo de redes neuronales se presenta en el anexo 9.4.

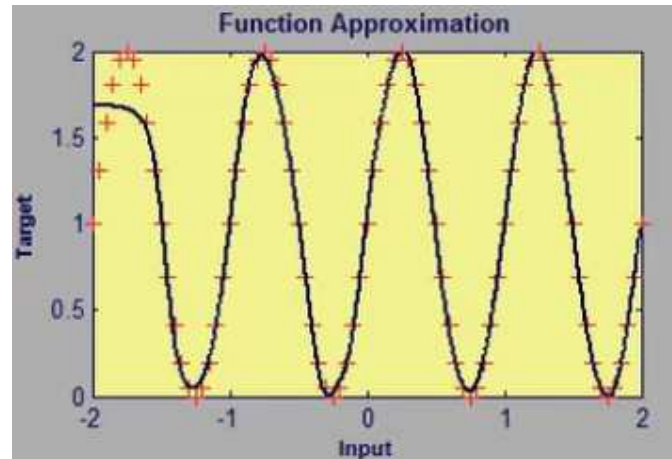


Figura 29. Esquema de aproximación de datos de entrada

4.3. Diagramas de Flujo de procedimientos para cada modelo

Para la simulación de datos faltantes y los procedimientos seguidos para cada uno de los modelos se presentan a continuación mediante diagramas de flujo.

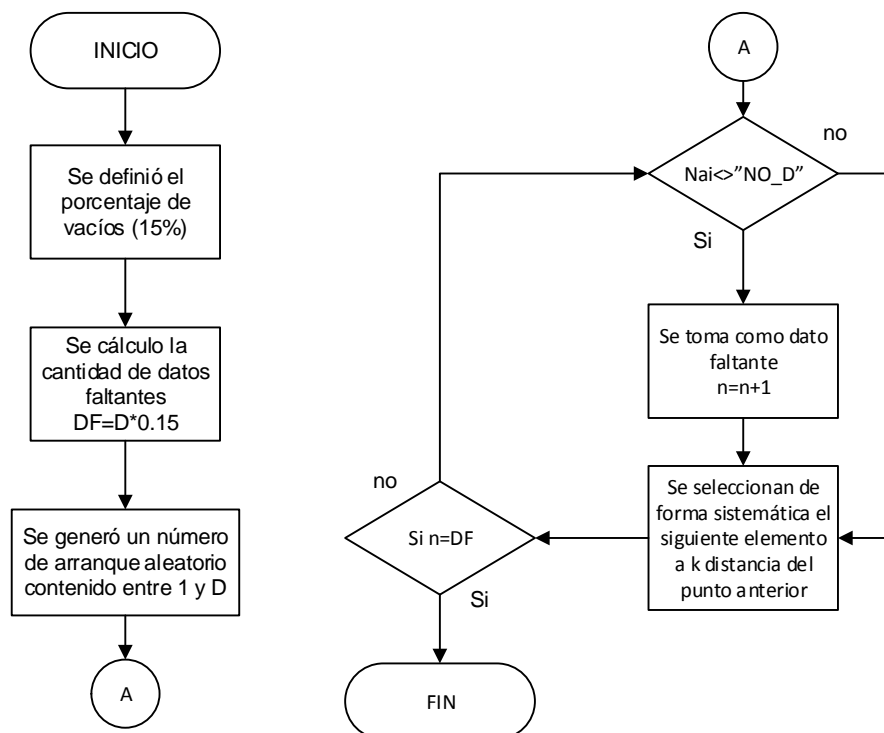


Figura 30. Diagrama de flujo para la simulación de datos faltantes

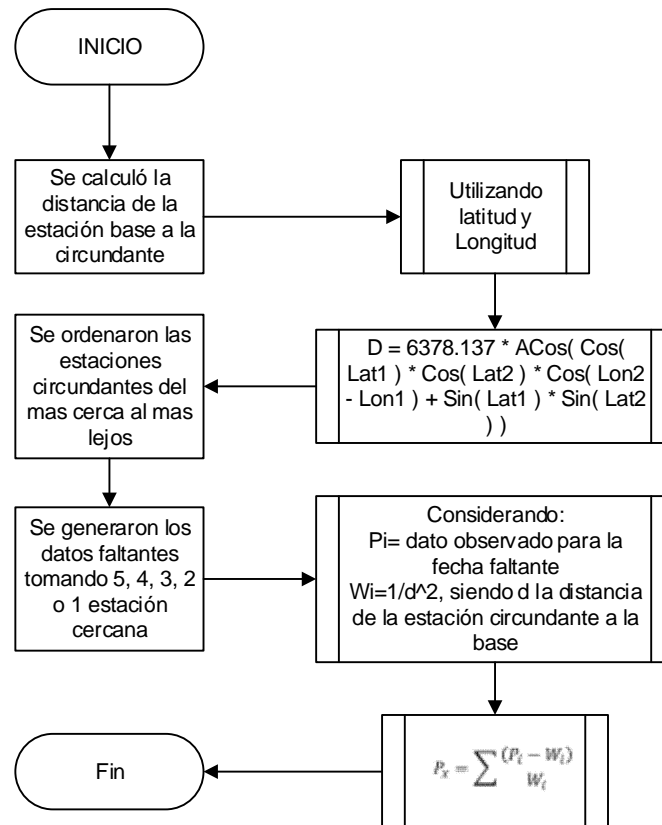


Figura 31. Diagrama de flujo US Weather Service

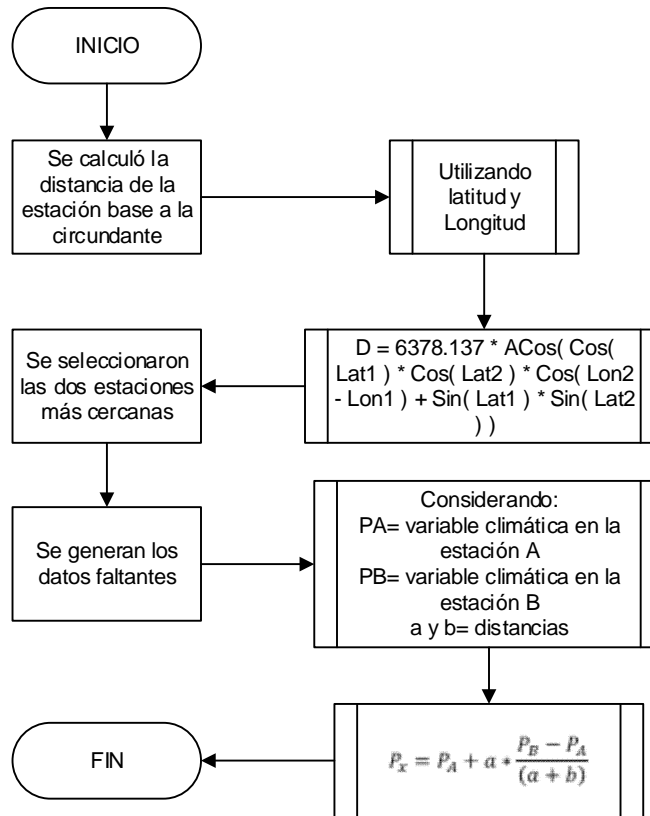


Figura 32. Diagrama de Flujo de Razón de distancia

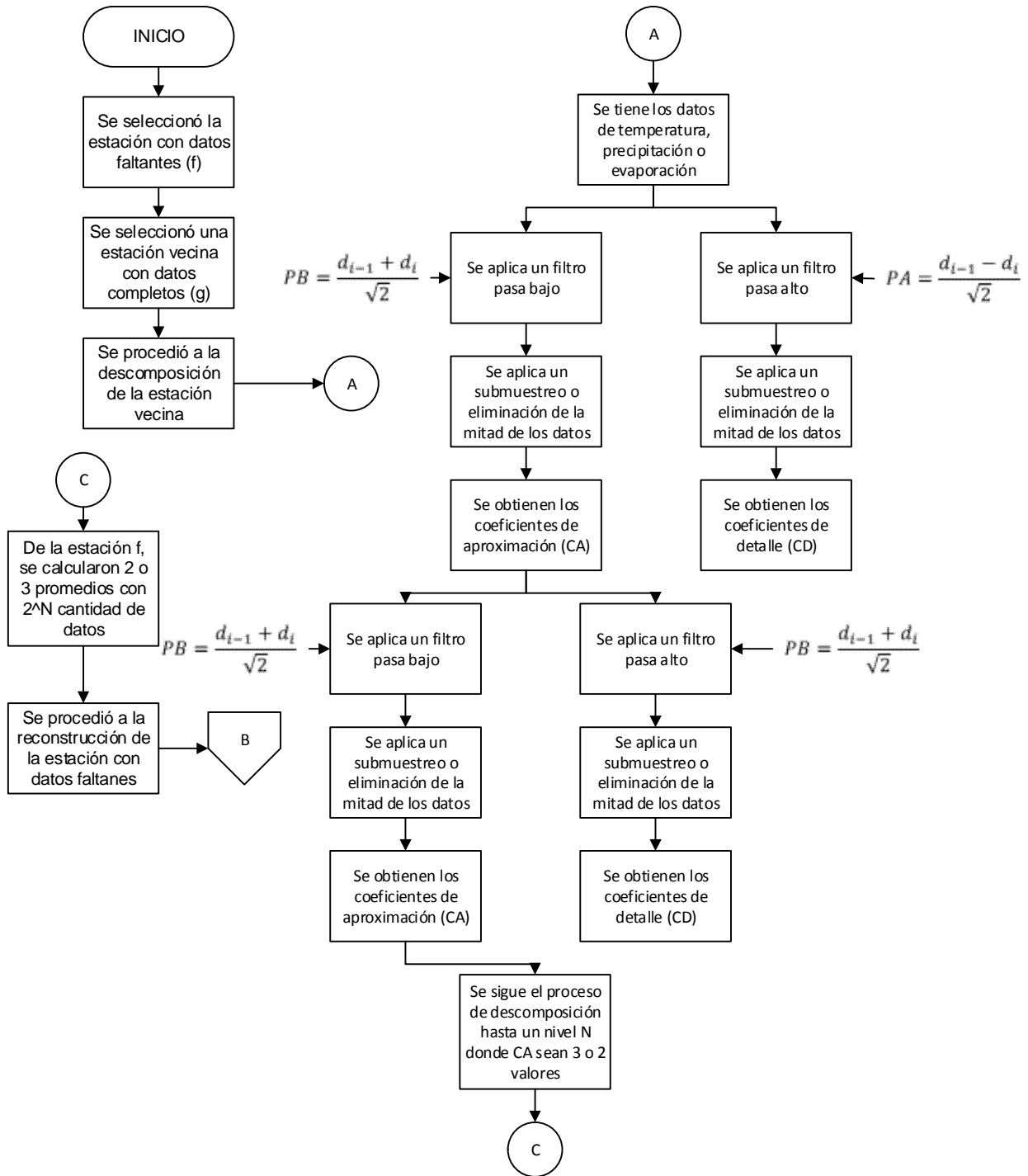


Figura 33. Diagrama de flujo del modelo matemático transformada de wavelet

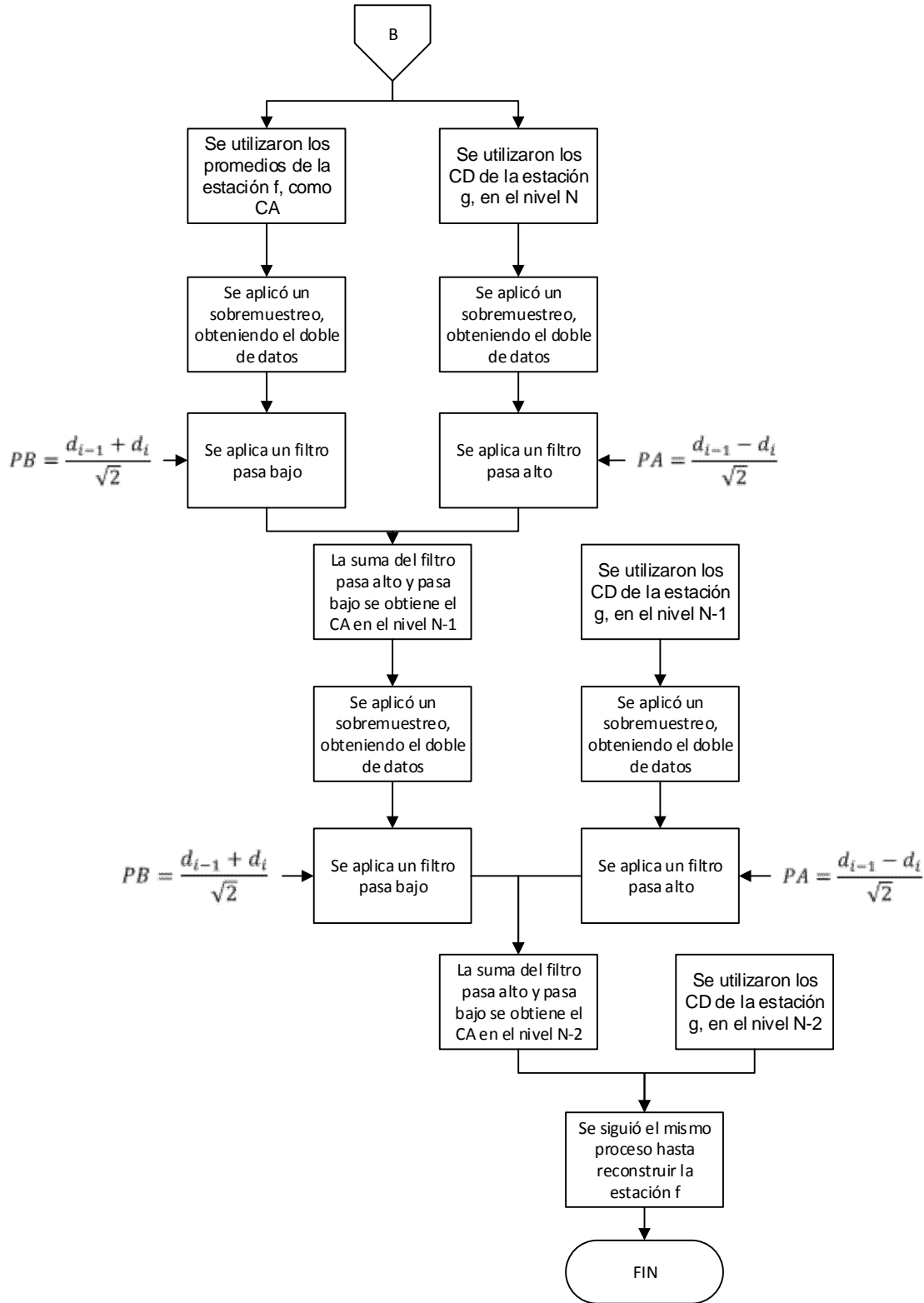


Figura 34. Continuación Figura 33

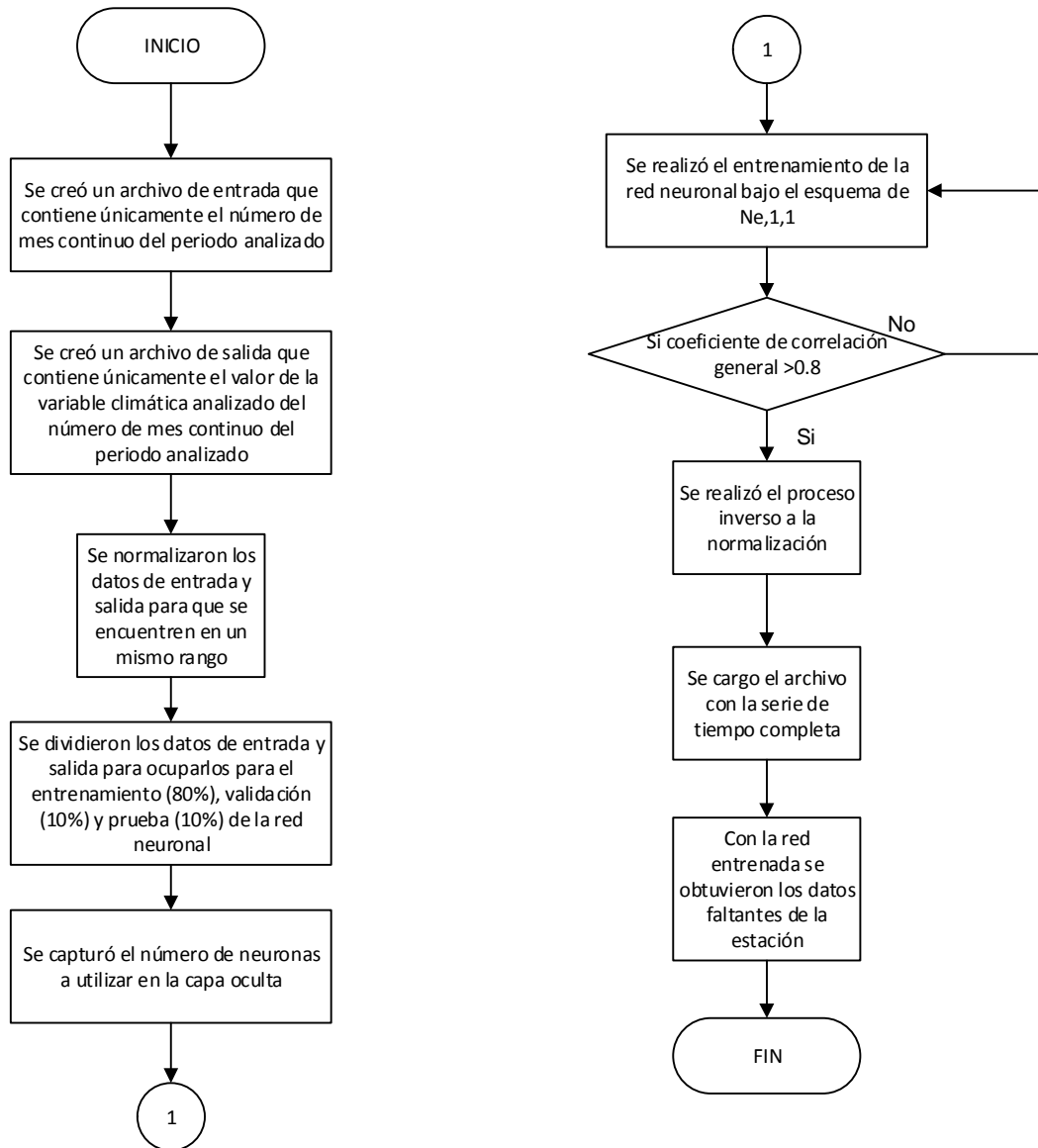


Figura 35. Modelo de redes neuronales artificiales

4.4. Análisis estadístico

Después de haber calculado los datos faltantes con cada uno de los modelos expuestos anteriormente para cada una de las estaciones meteorológicas y en cada una de las regiones climáticas se procedió a realizar la comparación estadística de las salidas de cada modelo con los valores reales de la variable climática a analizar, esto con el objeto de evaluar la capacidad predictiva de cada uno de los modelos.



Se utilizaron y aplicaron índices y errores estadísticos como son: el error cuadrático medio, raíz cuadrada del cuadrado medio del error, error absoluto medio, error estándar de estimación, índice de Willmott de correspondencia y la diferencia relativa. Estas estadísticas se aplicaron para 15% y 25% de datos faltantes para cada estación meteorológica.

De igual manera se realizó el análisis de regresión lineal simple entre las salidas de cada modelo con los respectivos valores reales de la variable climática, con la finalidad de analizar sus parámetros como son la pendiente y la ordenada al origen, de tal manera que la mejor estimación de cada modelo será cuando dichos parámetros presentan valores de 1 y 0 respectivamente.

5. RESULTADOS Y DISCUSIÓN

5.1. Distribución de las estaciones bases seleccionadas

La distribución de las estaciones bases seleccionada quedó de la siguiente manera: se seleccionaron 18 estaciones meteorológicas para la región climática cálido subhúmedo distribuyéndose en cuatro estaciones para evaporación, siete para precipitación y siete para temperatura; para la región climática muy seco semicálido se seleccionaron seis estaciones meteorológicas, dos de cada variable climática; para el clima templado subhúmedo se seleccionaron 18 estaciones meteorológicas, quedando siete para evaporación, cinco para precipitación y seis para temperatura.

En el siguiente Cuadro y Figura se presentan el número de estaciones meteorológicas por región climática.

Cuadro 3. Distribución de las estaciones bases seleccionadas

Región climática	Variable Climática	Número de estaciones	Total
Cálido Subhúmedo (cs)	Evaporación	4	18
	Precipitación	7	
	Temperatura	7	
Muy Seco Semicálido (mss)	Evaporación	2	6
	Precipitación	2	
	Temperatura	2	
Templado Subhúmedo (ts)	Evaporación	7	18
	Precipitación	5	
	Temperatura	6	
Total=		42	42

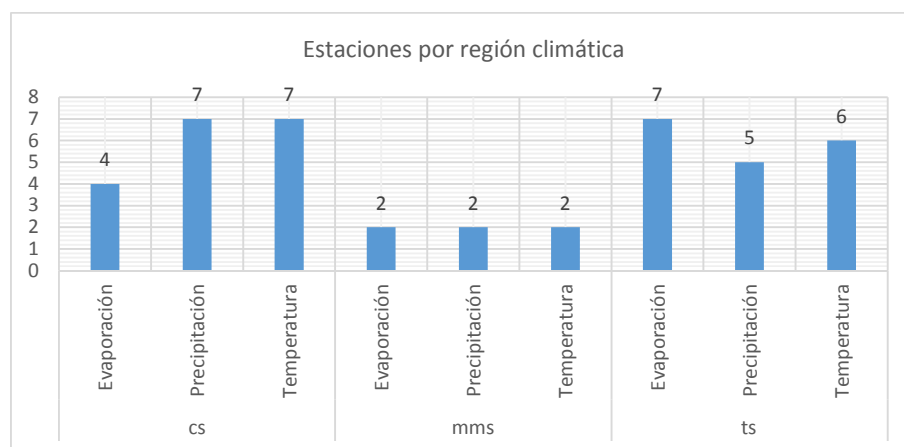


Figura 36. Distribución de las estaciones bases seleccionadas por región climática

5.2. Análisis de distancias de las estaciones circundantes

De las estaciones meteorológicas circundantes seleccionadas (5 por cada estación base) se determinaron la cantidad de estaciones que se encontraban a una distancia menor o igual a 50 km. Los resultados se presentan en el siguiente Cuadro y Figura:

Cuadro 4. Estaciones circundantes a una distancia menor o igual a 50 km

Región climática	Variable Climática	Número de estaciones	Total	Circundantes	Circundantes <=50 km
cs	Evaporación	4	18	20	13
	Precipitación	7		35	32
	Temperatura	7		35	32
mss	Evaporación	2	6	10	4
	Precipitación	2		10	6
	Temperatura	2		10	6
ts	Evaporación	7	18	35	33
	Precipitación	5		25	22
	Temperatura	6		30	26
Total=		42	42	210	174

Dónde: cs= cálido subhúmedo, ts= templado subhúmedo, mss= muy seco semicálido

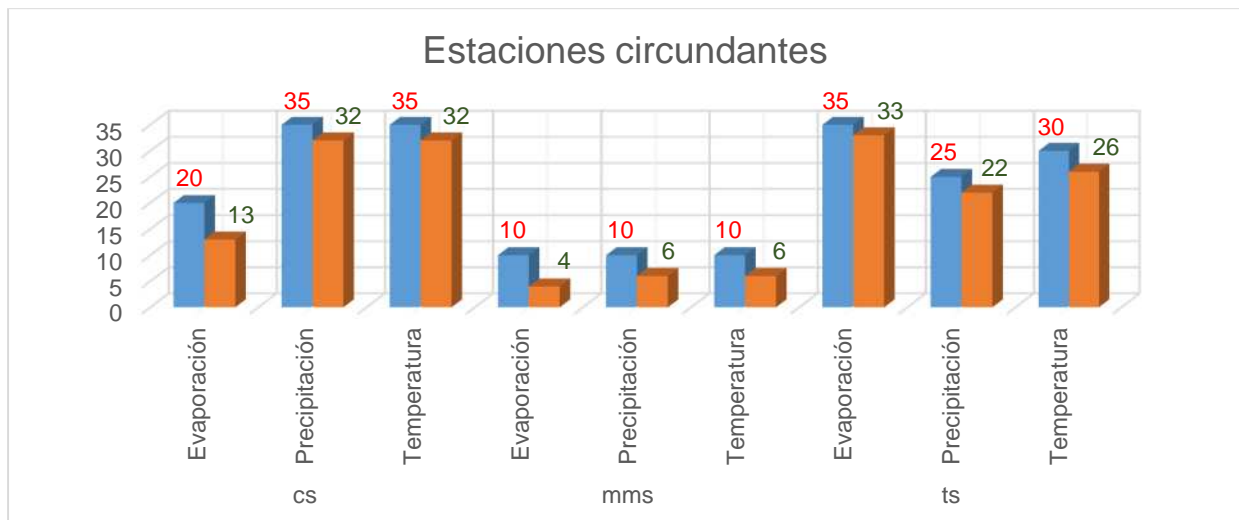


Figura 37. Distribución de las estaciones circundantes por región climática

El 82.9% de las estaciones circundantes seleccionadas se encuentran a una distancia menor o igual a 50 km, de igual manera es importante destacar que 41 estaciones bases tienen por lo menos dos estaciones cercanas que cumplen con dicho criterio y solamente una estación del clima muy seco semicálido tiene una sola estación a una distancia menor o igual a 50km.

5.3. Simulación de datos faltantes con un 15%

Para la simulación con un 15%, el dato faltante a considerar fue conocido en las cinco estaciones circundantes seleccionadas. El proceso de simulación de la estación 21081 del clima templado subhúmedo para la variable climática temperatura (0ts tp) fue de la siguiente manera:

- Para la estación se tiene 15 años de información, teniendo un total de 180 datos de temperatura.
- Se determinó la cantidad de datos faltantes con un 15% del total de datos dando como resultado 27 datos.
- Se determinó las posiciones a saltar (w) para la selección del dato faltante realizando el cociente entre la cantidad total de datos y la cantidad de datos faltantes dando un resultado de 7.
- Se determinó aleatoriamente el primer dato faltante (entre 1 y 180) el cual fue en la posición 117.
- Se verificó que el dato propuesto fuera conocido en las cinco estaciones cercanas para considerarla como dato faltante.
- Se seleccionó el siguiente dato en la posición 124 ($117+7$) y se comprobó que fuera conocido en las cinco estaciones. Se siguió el mismo procedimiento hasta obtener el 15% de datos faltantes (27).

5.4. Simulación de datos faltantes con un 25%

Para el caso de la simulación de datos faltantes con un 25% se siguió el mismo procedimiento que para el caso de 15%, con la diferencia que no se consideraron 5 estaciones circundantes en la simulación, debido a la poca disponibilidad de información. La cantidad de estaciones utilizadas para cada región y variable climática se presentan en el siguiente Cuadro.

Cuadro 5. Cantidad de estaciones circundantes utilizadas en la simulación para un 25%

Región climática	Variable Climática	Total	Estaciones					Total
			5	4	3	2	1	
cs	Evaporación	4	1	3	0	0	0	4
	Precipitación	7	6	1	0	0	0	7
	Temperatura	7	5	2	0	0	0	7
mms	Evaporación	2	2	0	0	0	0	2
	Precipitación	2	2	0	0	0	0	2
	Temperatura	2	2	0	0	0	0	2
ts	Evaporación	7	2	3	1	1	0	7
	Precipitación	5	5	0	0	0	0	5
	Temperatura	6	6	0	0	0	0	6
Total=		42	31	9	1	1	0	42

Dónde: cs= cálido subhúmedo, ts= templado subhúmedo, mss= muy seco semicálido

El 73.81% de las estaciones bases se simularon con cinco estaciones circundantes, el 21.41% se seleccionaron con cuatro, el 2.39% se seleccionaron con tres y el 2.39% seleccionaron con dos estaciones circundantes.

5.5. Aplicación de los modelos de estimación de datos faltantes

En la siguiente sección se presenta el procedimiento utilizado para estimar datos faltantes con cada uno de los modelos descritos en el presente documento. Se presenta para el caso de 15% de datos faltantes y para la estación 21081 perteneciente a la región climática templado subhúmedo y para la variable climática temperatura promedio.

5.5.1. US Weather Service

En el siguiente Cuadro se presentan las estaciones utilizadas para la estimación de datos faltantes por este método y sus correspondientes distancias (aplicando la ecuación 39) con respecto a la estación base.

Cuadro 6. Estaciones circundantes utilizadas para el modelo US Weather Service

Clave	Estado	Nombre	Latitud	Longitud	Altitud	Latitud(rad)	Longitud(rad)	Distancia (km)
21081	Pue.	SANTA CRUZ COYOTEPEC,	19.017	-97.567	2510	0.33	-1.70	
21080	Pue.	SN.SALVADOR EL SECO, DGE	19.133	-97.65	2425	0.33	-1.70	15.5884
30100	Ver.	MALTRATA, MALTRATA (DGE)	18.817	-97.267	1695	0.33	-1.70	38.6489
21007	Pue.	AHUATEPEC,TECALI DE H	18.854	-97.921	2057	0.33	-1.71	41.4564
29011	Tlax.	HUAMANTLA, HUAMANTLA	19.316	-97.911	2485	0.34	-1.71	49.1550
29160	Tlax.	ALTZAYANCA, ALTZAYANCA	19.428	-97.799	2410	0.34	-1.71	51.8455

A continuación se presenta la estimación de un dato faltante para el mes de marzo de 1990 para una, dos, tres, cuatro y cinco estaciones cercanas y aplicando la ecuación 27.

- Para una estación cercana.

$$\begin{aligned}d_1 &= 15.59 \text{ km} \\ P_1 &= 13.80^\circ\end{aligned}$$

$$P_{\text{marzo},1990} = \sum \frac{(P_i * W_i)}{W_i} = \frac{13.80 * \frac{1}{15.59^2}}{\frac{1}{15.59^2}} = 13.80^\circ$$

- Para dos estaciones cercanas.

$$\begin{aligned}d_1 &= 15.59 \text{ km} & d_2 &= 38.65 \text{ km} \\ P_1 &= 13.80^\circ & P_2 &= 13.30^\circ\end{aligned}$$

$$P_{\text{marzo},1990} = \sum \frac{(P_i * W_i)}{W_i} = \frac{13.80 * \frac{1}{15.59^2} + 13.30 * \frac{1}{38.65^2}}{\frac{1}{15.59^2} + \frac{1}{38.65^2}} = 13.73^\circ$$

- Para tres estaciones cercanas.

$$\begin{aligned}d_1 &= 15.59 \text{ km} & d_2 &= 38.65 \text{ km} & d_3 &= 41.46 \text{ km} \\ P_1 &= 13.80^\circ & P_2 &= 13.30^\circ & P_3 &= 11.80^\circ\end{aligned}$$

$$P_{\text{marzo},1990} = \sum \frac{(P_i * W_i)}{W_i} = \frac{13.80 * \frac{1}{15.59^2} + 13.30 * \frac{1}{38.65^2} + 11.80 * \frac{1}{41.46^2}}{\frac{1}{15.59^2} + \frac{1}{38.65^2} + \frac{1}{41.46^2}} = 13.52^\circ$$

- Para cuatro estaciones cercanas.

$$\begin{aligned}d_1 &= 15.59 \text{ km} & d_2 &= 38.65 \text{ km} & d_3 &= 41.46 \text{ km} & d_4 &= 49.20 \text{ km} \\ P_1 &= 13.80^\circ & P_2 &= 13.30^\circ & P_3 &= 11.80^\circ & P_4 &= 7.30^\circ\end{aligned}$$

$$\begin{aligned}P_{\text{marzo},1990} &= \sum \frac{(P_i * W_i)}{W_i} \\ &= \frac{13.80 * \frac{1}{15.59^2} + 13.30 * \frac{1}{38.65^2} + 11.80 * \frac{1}{41.46^2} + 7.30 * \frac{1}{49.20^2}}{\frac{1}{15.59^2} + \frac{1}{38.65^2} + \frac{1}{41.46^2} + \frac{1}{49.20^2}} \\ &= 13.08^\circ\end{aligned}$$

- Para cinco estaciones cercanas.

$$\begin{array}{l}
 d_1= 15.59 \text{ km} \quad d_2= 38.65 \text{ km} \quad d_3= 41.46 \text{ km} \quad d_4= 49.20 \text{ km} \quad d_5= 51.85 \text{ km} \\
 P_1= 13.80^\circ \quad P_2= 13.30^\circ \quad P_3= 11.80^\circ \quad P_4= 7.30^\circ \quad P_5= 7.40^\circ
 \end{array}$$

$$\begin{aligned}
 P_{\text{marzo},1990} &= \sum \frac{(P_i * W_i)}{W_i} \\
 &= \frac{13.80 * \frac{1}{15.59^2} + 13.30 * \frac{1}{38.65^2} + 11.80 * \frac{1}{41.46^2} + 7.30 * \frac{1}{49.20^2} + 7.40 * \frac{1}{51.85^2}}{\frac{1}{15.59^2} + \frac{1}{38.65^2} + \frac{1}{41.46^2} + \frac{1}{49.20^2} + \frac{1}{51.85^2}} \\
 &= 12.73^\circ
 \end{aligned}$$

5.5.2. Razón de la distancia

Se utilizaron las dos estaciones más cercanas para la estimación de datos faltantes por este método, en el siguiente Cuadro se presentan las dos estaciones con sus correspondientes distancias a la estación base.

Cuadro 7. Estaciones utilizadas para el modelo de la razón de la distancia

Clave	Estado	Nombre	Latitud	Longitud	Altitud	Latitud(rad)	Longitud(rad)	Distancia (km)
21081	Pue.	SANTA CRUZ COYOTEPEC,	19.017	-97.567	2510	0.33	-1.70	
21080	Pue.	SN.SALVADOR EL SECO, DGE	19.133	-97.65	2425	0.33	-1.70	15.5884
30100	Ver.	MALTRATA, MALTRATA (DGE)	18.817	-97.267	1695	0.33	-1.70	38.6489

Para la obtención del dato faltante en marzo de 1990 se utilizó la siguiente información:

$$\begin{array}{l}
 d_1= 15.59 \text{ km} \quad d_2= 38.65 \text{ km} \\
 P_1= 13.80^\circ \quad P_2= 13.30^\circ
 \end{array}$$

Aplicando la ecuación 26 se obtiene el siguiente resultado:

$$P_{\text{marzo},1990} = P_A + a * \frac{P_B - P_A}{(a + b)} = 13.80 + 15.59 \left(\frac{13.30 - 13.80}{15.59 + 38.65} \right) = 13.66^\circ$$

5.5.3. Transformada de Wavelet

Para éste modelo se requiere una estación meteorológica cercana que tuviera la mayor cantidad de datos completos para obtener sus correspondientes coeficientes de detalle y aproximación. Los coeficientes de detalle de la estación de apoyo y los coeficientes de aproximación de la estación problema fueron utilizados para reconstruir la serie de

tiempo de la estación problema. A continuación se presenta un ejemplo de la aplicación de este modelo.

Datos de entrada

Cuadro 8. Información de las estaciones utilizadas para este modelo

Clave	Estado	Nombre	Latitud	Longitud	Altitud	Latitud(rad)	Longitud(rad)	Distancia (km)
21081	Pue.	SANTA CRUZ COYOTEPEC,	19.017	-97.567	2510	0.33	-1.70	
29011	Tlax.	HUAMANTLA, HUAMANTLA	19.316	-97.911	2485	0.34	-1.71	49.1550

Cuadro 9. Parte de los datos climáticos utilizados de la estación de apoyo

Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
P	3.3	5.4	NO_D	9.1	NO_D	11.2	NO_D	9.4	10.2	NO_D	4.2	NO_D	1.8	NO_D	NO_D	8.6	NO_D
A	4.6	5.8	7.3	9.2	11.7	12.1	12.4	12.4	12.3	10.7	9.8	7.1	5.9	6.2	10.6	13.3	13.5

Proceso de descomposición de la estación de apoyo

- Se procedió a realizar la descomposición en el nivel 1, para lo cual se aplicó el filtro pasa bajo (ecuación 69) a los datos de la estación de apoyo. Los primeros dos cálculos se presentan a continuación y parte de los resultados se presentan en el siguiente Cuadro.

$$PB_1 = \frac{d_{i-1} + d_i}{\sqrt{2}} = \frac{4.6 + 5.8}{\sqrt{2}} = 7.35$$

$$PB_2 = \frac{d_{i-1} + d_i}{\sqrt{2}} = \frac{5.8 + 7.3}{\sqrt{2}} = 9.26$$

Cuadro 10. Aplicación del filtro pasa bajo

Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
P	3.3	5.4	NO_D	9.1	NO_D	11.2	NO_D	9.4	10.2	NO_D	4.2	NO_D	1.8	NO_D	NO_D	8.6	NO_D
A	4.6	5.8	7.3	9.2	11.7	12.1	12.4	12.4	12.3	10.7	9.8	7.1	5.9	6.2	10.6	13.3	13.5
PB		7.35	9.26	11.67	14.78	16.83	17.32	17.54	17.47	16.26	14.5	11.95	9.19	8.56	11.88	16.9	18.95

- Se aplicó el filtro pasa alto (ecuación 70) a los datos de la estación de apoyo. Los primeros dos cálculos se presentan a continuación y parte de los resultados se presentan en el siguiente Cuadro.

$$PA_1 = \frac{d_{i-1} - d_i}{\sqrt{2}} = \frac{4.6 - 5.8}{\sqrt{2}} = -0.85$$

$$PA_2 = \frac{d_{i-1} - d_i}{\sqrt{2}} = \frac{5.8 - 7.3}{\sqrt{2}} = -1.06$$

Cuadro 11. Aplicación del filtro pasa alto

Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
P	3.3	5.4	NO_D	9.1	NO_D	11.2	NO_D	9.4	10.2	NO_D	4.2	NO_D	1.8	NO_D	NO_D	8.6	NO_D
A	4.6	5.8	7.3	9.2	11.7	12.1	12.4	12.4	12.3	10.7	9.8	7.1	5.9	6.2	10.6	13.3	13.5
PB		7.35	9.26	11.67	14.78	16.83	17.32	17.54	17.47	16.26	14.5	11.95	9.19	8.56	11.88	16.9	18.95
PA		-0.85	-1.06	-1.34	-1.77	-0.28	-0.21	0	0.07	1.13	0.64	1.91	0.85	-0.21	-3.11	-1.91	-0.14

- Se procedió a realizar un submuestreo o eliminación de datos, eligiendo un valor de cada dos para los filtros pasa bajo y pasa alto. Los resultados se presentan en el siguiente Cuadro.

Cuadro 12. Procedimiento de submuestreo o eliminación de datos

Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	N
P	3.3	5.4	NO_D	9.1	NO_D	11.2	NO_D	9.4	10.2	NO_D	4.2	NO_D	1.8	NO_D	NO_D	8.6	NO_D
A	4.6	5.8	7.3	9.2	11.7	12.1	12.4	12.4	12.3	10.7	9.8	7.1	5.9	6.2	10.6	13.3	13.5
PB		7.35	9.26	11.67	14.78	16.83	17.32	17.54	17.47	16.26	14.5	11.95	9.19	8.56	11.88	16.9	18.95
PA		-0.85	-1.06	-1.34	-1.77	-0.28	-0.21	0	0.07	1.13	0.64	1.91	0.85	-0.21	-3.11	-1.91	-0.14
CA	7.35	11.67	16.83	17.54	16.26	11.95	8.56	16.9	19.16	16.83	16.05	9.48	8.77	14.14	17.11	16.9	14.92
CD	-0.85	-1.34	-0.28	0	1.13	1.91	-0.21	-1.91	-0.07	0.42	1.2	0.28	-0.14	-1.41	-0.57	0.49	1.06

- Se realizó la descomposición en el nivel 2, aplicando el mismo procedimiento anterior pero con los coeficientes de aproximación (CA).

$$PB_1 = \frac{d_{i-1} + d_i}{\sqrt{2}} = \frac{7.35 + 11.67}{\sqrt{2}} = 13.45$$

$$PB_2 = \frac{d_{i-1} + d_i}{\sqrt{2}} = \frac{11.67 + 16.83}{\sqrt{2}} = 20.15$$

$$PA_1 = \frac{d_{i-1} - d_i}{\sqrt{2}} = \frac{7.35 - 11.67}{\sqrt{2}} = -3.05$$

$$PA_2 = \frac{d_{i-1} - d_i}{\sqrt{2}} = \frac{11.67 - 16.83}{\sqrt{2}} = -3.65$$

Cuadro 13. Resultados de la descomposición en el nivel 2

Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	N
P	3.3	5.4	NO_D	9.1	NO_D	11.2	NO_D	9.4	10.2	NO_D	4.2	NO_D	1.8	NO_D	NO_D	8.6	NO_D
A	4.6	5.8	7.3	9.2	11.7	12.1	12.4	12.4	12.3	10.7	9.8	7.1	5.9	6.2	10.6	13.3	13.5
PB		7.35	9.26	11.67	14.78	16.83	17.32	17.54	17.47	16.26	14.5	11.95	9.19	8.56	11.88	16.9	18.95
PA		-0.85	-1.06	-1.34	-1.77	-0.28	-0.21	0	0.07	1.13	0.64	1.91	0.85	-0.21	-3.11	-1.91	-0.14
CA	7.35	11.67	16.83	17.54	16.26	11.95	8.56	16.9	19.16	16.83	16.05	9.48	8.77	14.14	17.11	16.9	14.92
CD	-0.85	-1.34	-0.28	0	1.13	1.91	-0.21	-1.91	-0.07	0.42	1.2	0.28	-0.14	-1.41	-0.57	0.49	1.06
PB1		13.45	20.15	24.3	23.9	19.95	14.5	18	25.5	25.45	23.25	18.05	12.9	16.2	22.1	24.05	22.5
PA1		-3.05	-3.65	-0.5	0.91	3.05	2.4	-5.9	-1.6	1.65	0.55	4.65	0.5	-3.8	-2.1	0.15	1.4
CA1	13.45	24.3	19.95	18	25.45	18.05	16.2	24.05	17.25	14.9	26.45	19.21	18	24.95	19.35	18.45	27.8
CD1	-3.05	-0.5	3.05	-5.9	1.65	4.65	-3.8	0.15	3.85	-4.41	0.35	4.1	-4.4	0.35	3.15	-4.25	1.2

- Se realizó el mismo proceso de descomposición hasta obtener tres coeficientes de aproximación. El nivel de descomposición para esta estación fue de seis y los resultados de cada una de las descomposiciones se presentan a continuación.

Cuadro 14. Resultados de la descomposición en cada uno de los niveles

Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	N
P	3.3	5.4	NO_D	9.1	NO_D	11.2	NO_D	9.4	10.2	NO_D	4.2	NO_D	1.8	NO_D	NO_D	8.6	NO_D
A	4.6	5.8	7.3	9.2	11.7	12.1	12.4	12.4	12.3	10.7	9.8	7.1	5.9	6.2	10.6	13.3	13.5
PB		7.35	9.26	11.67	14.78	16.83	17.32	17.54	17.47	16.26	14.5	11.95	9.19	8.56	11.88	16.9	18.95
PA		-0.85	-1.06	-1.34	-1.77	-0.28	-0.21	0	0.07	1.13	0.64	1.91	0.85	-0.21	-3.11	-1.91	-0.14
CA	7.35	11.67	16.83	17.54	16.26	11.95	8.56	16.9	19.16	16.83	16.05	9.48	8.77	14.14	17.11	16.9	14.92
CD	-0.85	-1.34	-0.28	0	1.13	1.91	-0.21	-1.91	-0.07	0.42	1.2	0.28	-0.14	-1.41	-0.57	0.49	1.06
PB1		13.45	20.15	24.3	23.9	19.95	14.5	18	25.5	25.45	23.25	18.05	12.9	16.2	22.1	24.05	22.5
PA1		-3.05	-3.65	-0.5	0.91	3.05	2.4	-5.9	-1.6	1.65	0.55	4.65	0.5	-3.8	-2.1	0.15	1.4
CA1	13.45	24.3	19.95	18	25.45	18.05	16.2	24.05	17.25	14.9	26.45	19.21	18	24.95	19.35	18.45	27.8
CD1	-3.05	-0.5	3.05	-5.9	1.65	4.65	-3.8	0.15	3.85	-4.41	0.35	4.1	-4.4	0.35	3.15	-4.25	1.2
		26.69	31.29	26.83	30.72	30.76	24.22	28.46	29.2	22.73	29.24	32.29	26.31	30.37	31.32	26.73	32.7
		-7.67	3.08	1.38	-5.27	5.23	1.31	-5.55	4.81	1.66	-8.17	5.12	0.86	-4.91	3.96	0.64	-6.61
CA2	26.69	26.83	30.76	28.46	22.73	32.29	30.37	26.73	34.05	25.63	23.26	27.93	29.03	25.81	26.55	26.94	25.1
CD2	-7.67	1.38	5.23	-5.55	1.66	5.12	-4.91	0.64	5.27	-5.13	0.57	2.19	-5.13	3.11	4.77	-4.95	-0.78
		37.84	40.72	41.87	36.2	38.91	44.31	40.38	42.98	42.2	34.57	36.2	40.28	38.78	37.02	37.82	36.8
		-0.1	-2.78	1.63	4.05	-6.76	1.36	2.57	-5.18	5.95	1.68	-3.3	-0.78	2.28	-0.52	-0.28	1.3
CA3	37.84	41.87	38.91	40.38	42.2	36.2	38.78	37.82	38.49	36.63	39.56						
CD3	-0.1	1.63	-6.76	2.57	5.95	-3.3	2.28	-0.28	-3	3.68	0						

Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	N
		56.36	57.12	56.07	58.39	55.44	53.02	54.16	53.96	53.12	53.87	55.95					
		-2.85	2.09	-1.04	-1.29	4.24	-1.82	0.68	-0.47	1.32	-2.07	0					
CA4	56.36	56.07	55.44	54.16	53.12	55.95											
CD4	-2.85	-1.04	4.24	0.68	1.32	0											
		79.5	78.85	77.5	75.86	77.12											
		0.21	0.45	0.91	0.74	-2											
CA5	79.5	77.5	77.12														
CD5	0.21	0.91	-2														

Proceso de reconstrucción de la estación problema

- Para la reconstrucción de la estación problema se utilizaron todos los coeficiente de detalle de la estación de apoyo (CD_N) y se estimaron los coeficientes de aproximación de la estación problema mediante promedios que incluyen $2^N = 2^6 = 64$ datos de la estación problema. Estos promedios se obtuvieron respetando los vacíos de la estación problema y utilizando únicamente la información disponible. Los coeficientes de aproximación de la estación problema se presentan a continuación.

Cuadro 15. Coeficientes de aproximación en el nivel 6 de la estación problema

Coeficiente	Promedio1	Promedio 2	Promedio 3
CA5 preliminar	7.55	6.98	5.04

- Se multiplicaron los promedios anteriores por $2^{\frac{N}{2}}$, donde N es el nivel de descomposición.

$$2^{\frac{N}{2}} = 2^{\frac{6}{2}} = 2^3 = 8$$

$$Promedio1 = 7.55 * 2^3 = 60.4 ; \quad Promedio 2 = 6.98 * 2^3 = 55.84;$$

$$Promedio 3 = 5.04 * 2^3 = 40.32$$

- Se procedió a iniciar la reconstrucción de la estación meteorológica problema, con un nivel de reconstrucción igual a seis. Se realizó un sobremuestreo que consiste en colocar un cero entre dos datos, quedando de la siguiente manera.

Cuadro 16. Sobremuestreo en el nivel de reconstrucción uno

Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	N
CA5	60.4	55.84	40.32														
CD5	0.21	0.91	-2														
CA5'	0.00	60.40	0.00	55.84	0.00	40.32											
CD5'	0.00	0.21	0.00	0.91	0.00	-2.00											

- Posteriormente se aplicó un filtro pasa bajo para los coeficientes de aproximación y un filtro pasa alto para los coeficientes de detalle. A continuación se presentan los primeros dos cálculos para cada filtro así como en el siguiente Cuadro se presentan los resultados.

$$PB_1 = \frac{d_{i-1} + d_i}{\sqrt{2}} = \frac{0.00 + 60.40}{\sqrt{2}} = 42.71$$

$$PB_2 = \frac{d_{i-1} + d_i}{\sqrt{2}} = \frac{60.4 + 0.00}{\sqrt{2}} = 42.71$$

$$PA_1 = \frac{d_{i-1} - d_i}{\sqrt{2}} = \frac{0.00 - 0.21}{\sqrt{2}} = -0.15$$

$$PA_2 = \frac{d_{i-1} - d_i}{\sqrt{2}} = \frac{0.21 - 0.00}{\sqrt{2}} = 0.15$$

Cuadro 17. Aplicación del filtro pasa bajo y pasa alto para la reconstrucción en el nivel 1

Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	N
CA5	60.4	55.84	40.32														
CD5	0.21	0.91	-2														
CA5'	0.00	60.40	0.00	55.84	0.00	40.32	0.00										
CD5'	0.00	0.21	0.00	0.91	0.00	-2.00	0.00										
PB		42.71	42.71	39.48	39.48	28.51	28.51											
PA		0.15	-0.15	0.64	-0.64	-1.41	1.41											

- Se procedió a sumar los valores de los filtros pasa bajo y pasa alto para obtener los coeficientes de aproximación en el nivel de reconstrucción uno.

$$CA_{4,1} = PB_2 + PA_2 = 42.71 + 0.15 = 42.86$$

$$CA_{4,2} = PB_3 + PA_3 = 42.71 - 0.15 = 42.56$$

Cuadro 18. Coeficientes de aproximación en el nivel de reconstrucción 1

Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	N
CA5	60.4	55.84	40.32														
CD5	0.21	0.91	-2														
CA5'	0.00	60.40	0.00	55.84	0.00	40.32	0.00										
CD5'	0.00	0.21	0.00	0.91	0.00	-2.00	0.00										
PB		42.71	42.71	39.48	39.48	28.51	28.51											
PA		0.15	-0.15	0.64	-0.64	-1.41	1.41											
CA4	42.86	42.56	40.13	38.84	27.10	29.92												

- Se siguió la misma metodología explicada anteriormente y se obtuvieron los coeficientes de aproximación en la reconstrucción en el nivel 2. Los resultados del sobremuestreo y los coeficientes de aproximación y detalle utilizados se presentan a continuación.

Cuadro 19. Sobremuestreo en el nivel de reconstrucción dos

Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	N
CA4	42.86	42.56	40.13	38.84	27.10	29.92											
CD4	-2.85	-1.04	4.24	0.68	1.32	0											
CA4'	0.00	42.86	0.00	42.56	0.00	40.13	0.00	38.84	0.00	27.10	0.00	29.92					
CD4'	0.00	-2.85	0.00	-1.04	0.00	4.24	0.00	0.68	0.00	1.32	0.00	0.00					

- Aplicando los filtros pasa bajo y pasa alto, se obtienen los siguientes resultados.

$$PB_1 = \frac{d_{i-1} + d_i}{\sqrt{2}} = \frac{0.00 + 42.86}{\sqrt{2}} = 30.31$$

$$PB_2 = \frac{d_{i-1} + d_i}{\sqrt{2}} = \frac{42.86 + 0.00}{\sqrt{2}} = 30.31$$

$$PA_1 = \frac{d_{i-1} - d_i}{\sqrt{2}} = \frac{0.00 - (-2.85)}{\sqrt{2}} = 2.02$$

$$PA_2 = \frac{d_{i-1} - d_i}{\sqrt{2}} = \frac{-2.85 - 0.00}{\sqrt{2}} = -2.02$$

Cuadro 20. Aplicación del filtro paso bajo y pasa alto en el nivel de reconstrucción dos

Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	N
CA4	42.86	42.56	40.13	38.84	27.10	29.92											
CD4	-2.85	-1.04	4.24	0.68	1.32	0											
CA4'	0.00	42.86	0.00	42.56	0.00	40.13	0.00	38.84	0.00	27.10	0.00	29.92					
CD4'	0.00	-2.85	0.00	-1.04	0.00	4.24	0.00	0.68	0.00	1.32	0.00	0.00					
PB		30.31	30.31	30.10	30.10	28.38	28.38	27.47	27.47	19.16	19.16	21.16						
PA		-2.02	2.02	-0.74	0.74	3.00	-3.00	0.48	-0.48	0.93	-0.93	0.00						

- Sumando los filtros pasa bajo y pasa alto se obtienen los siguientes coeficientes de aproximación en el nivel de reconstrucción dos.

$$CA_{4,1} = PB_2 + PA_2 = 30.31 - 2.02 = 28.29$$

$$CA_{4,2} = PB_3 + PA_3 = 30.31 + 2.02 = 32.33$$

Cuadro 21. Coeficientes de aproximación en el nivel de reconstrucción dos

Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	N
CA4	42.86	42.56	40.13	38.84	27.10	29.92											
CD4	-2.85	-1.04	4.24	0.68	1.32	0											
CA4'	0.00	42.86	0.00	42.56	0.00	40.13	0.00	38.84	0.00	27.10	0.00	29.92					
CD4'	0.00	-2.85	0.00	-1.04	0.00	4.24	0.00	0.68	0.00	1.32	0.00	0.00					
PB		30.31	30.31	30.10	30.10	28.38	28.38	27.47	27.47	19.16	19.16	21.16						
PA		-2.02	2.02	-0.74	0.74	3.00	-3.00	0.48	-0.48	0.93	-0.93	0.00						
CA3	28.29	32.32	29.36	30.83	31.37	25.38	27.95	26.98	20.09	18.23	21.16							

- Este proceso es iterativo hasta reconstruir totalmente la serie de tiempo de la estación problema. Los resultados de la reconstrucción en el nivel seis se presentan a continuación:

Cuadro 22. Parte de los resultados de la reconstrucción de la estación problema

Est	1990												1991					N
	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abr	Mayo	
Est	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	N
P	3.3	5.4	NO_D	9.1	NO_D	11.2	NO_D	9.4	10.2	NO_D	4.2	NO_D	1.8	NO_D	NO_D	8.6	NO_D
A	4.6	5.8	7.3	9.2	11.7	12.1	12.4	12.4	12.3	10.7	9.8	7.1	5.9	6.2	10.6	13.3	13.5
Re	2.21	3.41	4.91	6.80	9.31	9.70	10.01	10.01	9.91	8.31	7.41	4.71	3.51	3.80	8.21	10.91	11.11

- El resultado del dato perdido del mes de marzo de 1990 es:

$$P_{\text{marzo},1990} = 4.91^{\circ}$$

5.5.4. Redes neuronales artificiales

Para la aplicación de este modelo se utilizó el software MATLAB 2013®, y únicamente se definieron el número de neuronas a ocupar en la primera capa oculta el cuál para la temperatura fue de 150 neuronas, 300 neuronas para evaporación y 350 neuronas para precipitación.

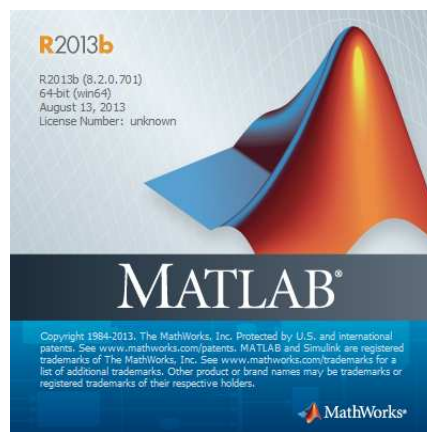


Figura 38. Software MATLAB 2013®

Se colocó el número de neuronas en el programa generado en MATLAB 2013®, como se observa en la siguiente Figura:



Figura 39. Programa generado en MATLAB 2013® para red retropropagación de redes neuronales artificiales

Se generaron los archivos de entrada y salida para el entrenamiento de la red neuronal, estos archivos se generaron de la siguiente manera:

- Se colocaron los datos de la variable climática y su número de mes continuo de todo el periodo analizado, posteriormente se fueron eliminando los datos climáticos faltantes con su respectivo número de mes continuo.
- Se generó un archivo de entrada en formato *.prn que contiene el número del mes continuo y un archivo de salida *.prn que contiene los datos de la variable climática, parte de estos archivos se presentan a continuación para la estación analizada en esta sección, cabe mencionar que esta información tiene que estar en forma vertical.

Cuadro 23. Datos de entrada para el entrenamiento de la red neuronal artificial

1	2	4	6	8	9	11	13	16	18	20	21	23	25	27	28	30	32	33	34
---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-------

Cuadro 24. Datos de salida para el entrenamiento de la red neuronal

3.3	5.4	9.1	11.2	9.4	10.2	4.2	1.8	8.6	11.5	10.3	11.9	4.0	4.9	6.6	9.3	10.4	10.5	9.4	8.2
-----	-----	-----	------	-----	------	-----	-----	-----	------	------	------	-----	-----	-----	-----	------	------	-----	-----	-------

Se cargaron estos archivos completos para el entrenamiento de la red neuronal. El entrenamiento terminará cuando el coeficiente de correlación general sea mayor o igual a 0.8 como se muestra en la siguiente Figura.

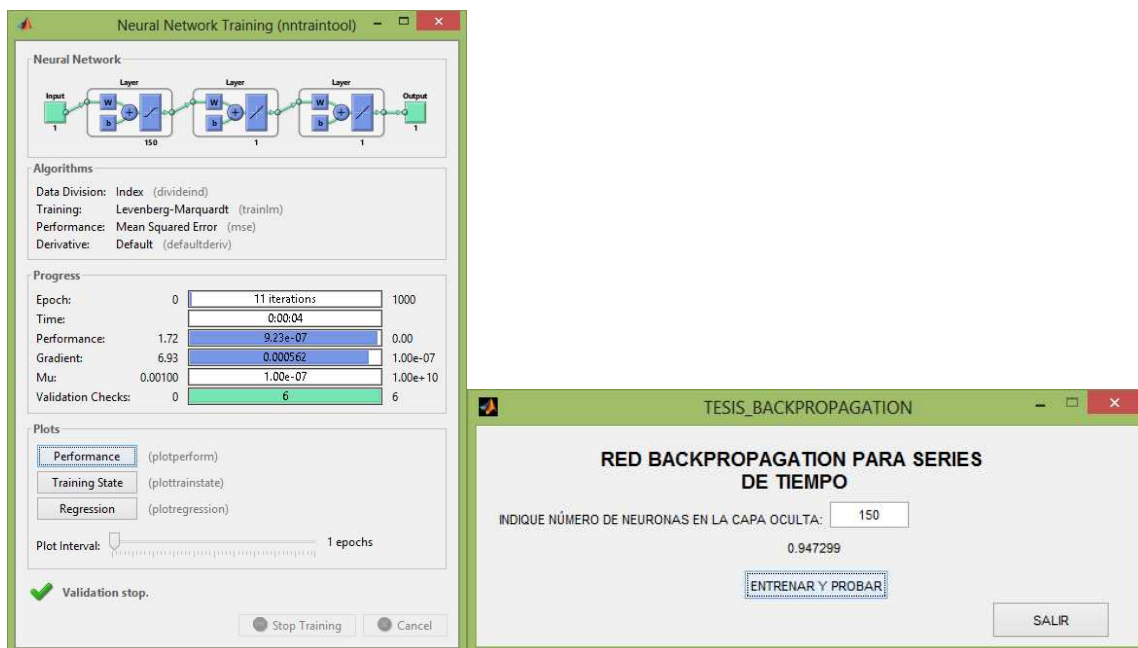


Figura 40. Entrenamiento de la red neuronal artificial

Después de entrenar a la red neuronal artificial se cargó un archivo *.prn con la serie de tiempo completa para generar los datos faltantes. Parte de este archivo se muestra a continuación y como se mencionó anteriormente estos datos deben estar en forma vertical.

Cuadro 25. Archivo utilizado *.prn con la serie de tiempo completa para generar datos faltantes

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	-------

Los resultados obtenidos con la red neuronal entrenada se presentan en el siguiente Cuadro.

Cuadro 26. Resultados de la aplicación del modelo de red neuronal artificial

1990												1991								
Ene	Feb	Mar	Abr	May	Ju	Jul	Ago	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abr	May	Ju	Jul	Ago
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3.30	5.40	NO_D	9.10	NO_D	11.20	NO_D	9.40	10.20	NO_D	4.20	NO_D	1.80	NO_D	NO_D	8.60	NO_D	11.50	NO_D	10.30
3.43	3.45	6.92	9.12	9.33	10.94	10.51	9.54	10.02	8.21	4.23	1.84	1.86	4.06	8.03	8.59	8.65	9.70	10.31	10.08

El resultado del dato perdido del mes de marzo de 1990 es:

$$P_{\text{marzo},1990} = 6.92^{\circ}$$

5.6. Comparación gráfica de los modelos

En el siguiente apartado se presentan tres estaciones con los resultados de las estimaciones de datos faltantes con cada una de las variables climáticas, los cuales son los siguientes:

- Para 15% de datos faltantes: Para la variable temperatura se presenta la estación 21081 de la región climática templado subhúmedo (Ver Figura 41). Para la variable climática evaporación se presenta la estación 23154 de la región climática cálido subhúmedo (Ver Figura 42). Para la variable climática precipitación se presenta la estación 3166 de la región climática muy seco semicálido (Ver Figura 43).
- Para 25% de datos faltantes: Para la variable temperatura se presenta la estación 30056 de la región climática cálido subhúmedo (Ver Figura 44). Para la variable climática evaporación se presenta la estación 5036 de la región muy seco

semicálido (Ver Figura 45). Para la variable climática precipitación se presenta la estación 13151 de la región climática templado subhúmedo (Ver Figura 46).

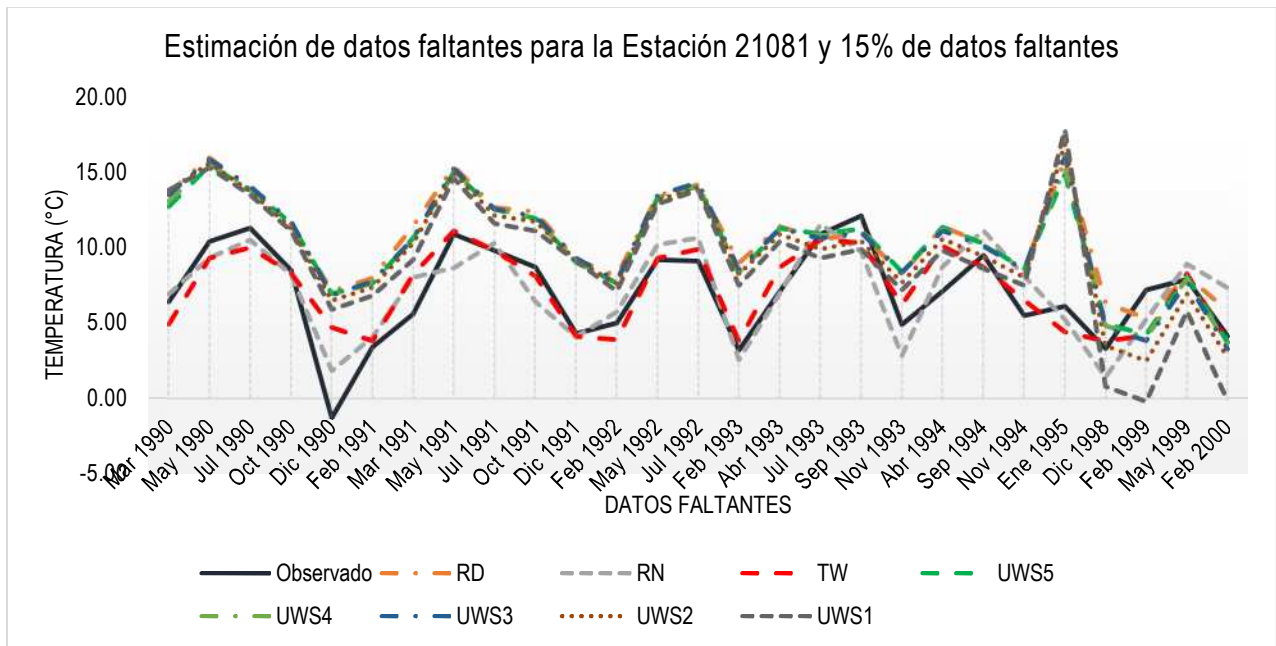


Figura 41. Comparación gráfica de los resultados de la estación 21081 para la variable climática temperatura y 15% de datos faltantes

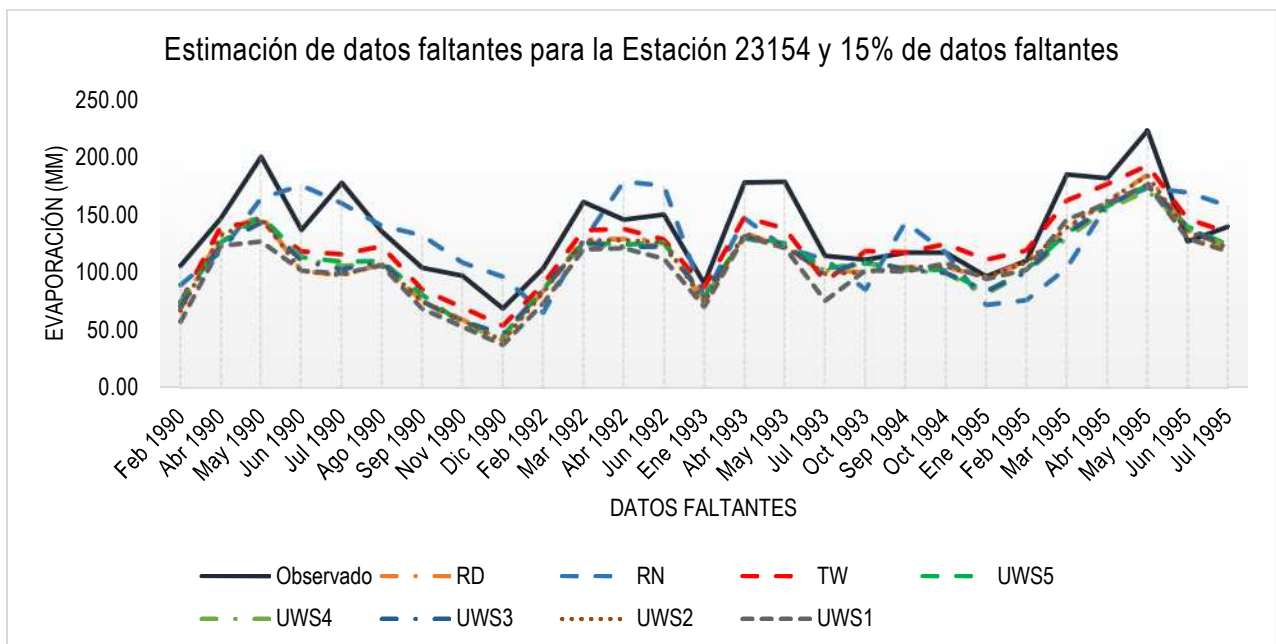


Figura 42. Comparación gráfica de los resultados de la estación 23154 para la variable climática evaporación y 15% de datos faltantes

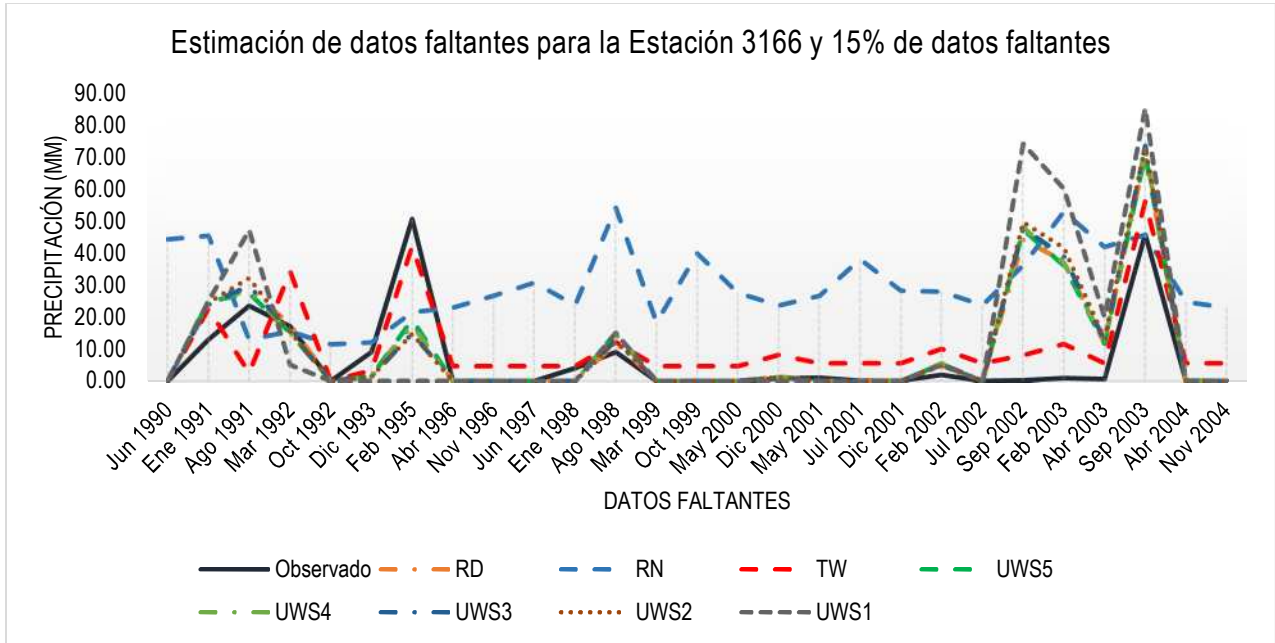


Figura 43. Comparación gráfica de los resultados de la estación 3166 para la variable climática precipitación y 15% datos faltantes

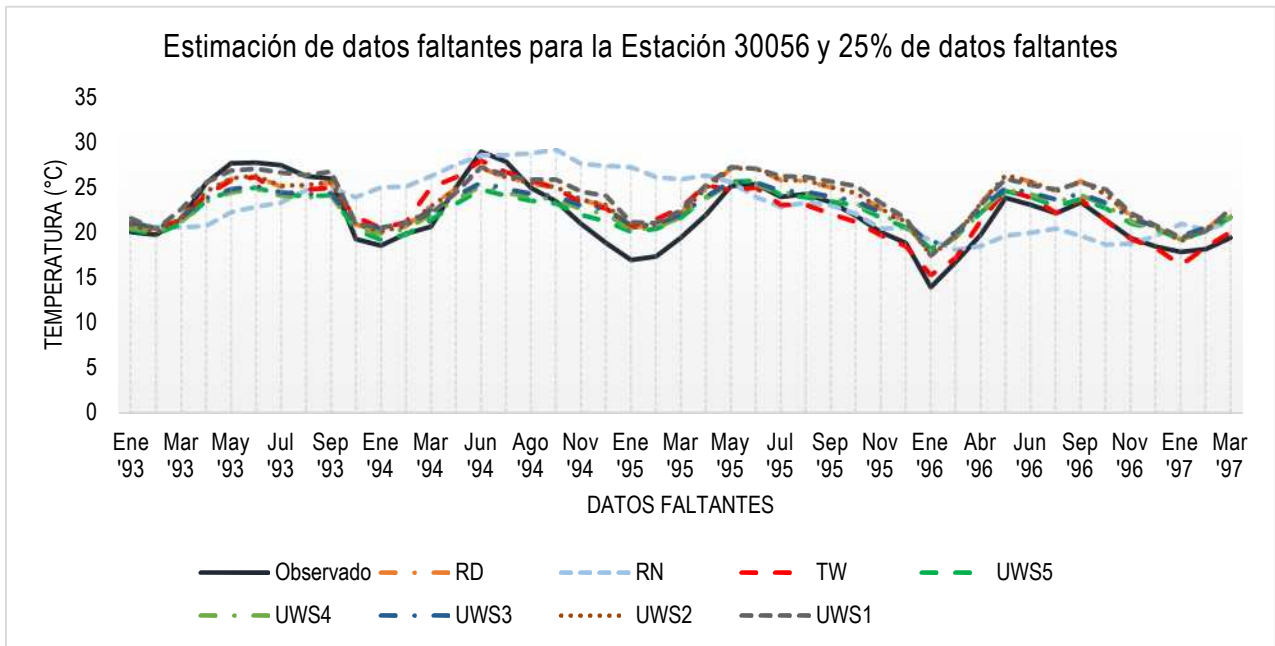


Figura 44. Comparación gráfica de los resultados de la estación 30056 para la variable climática temperatura y 25% datos faltantes

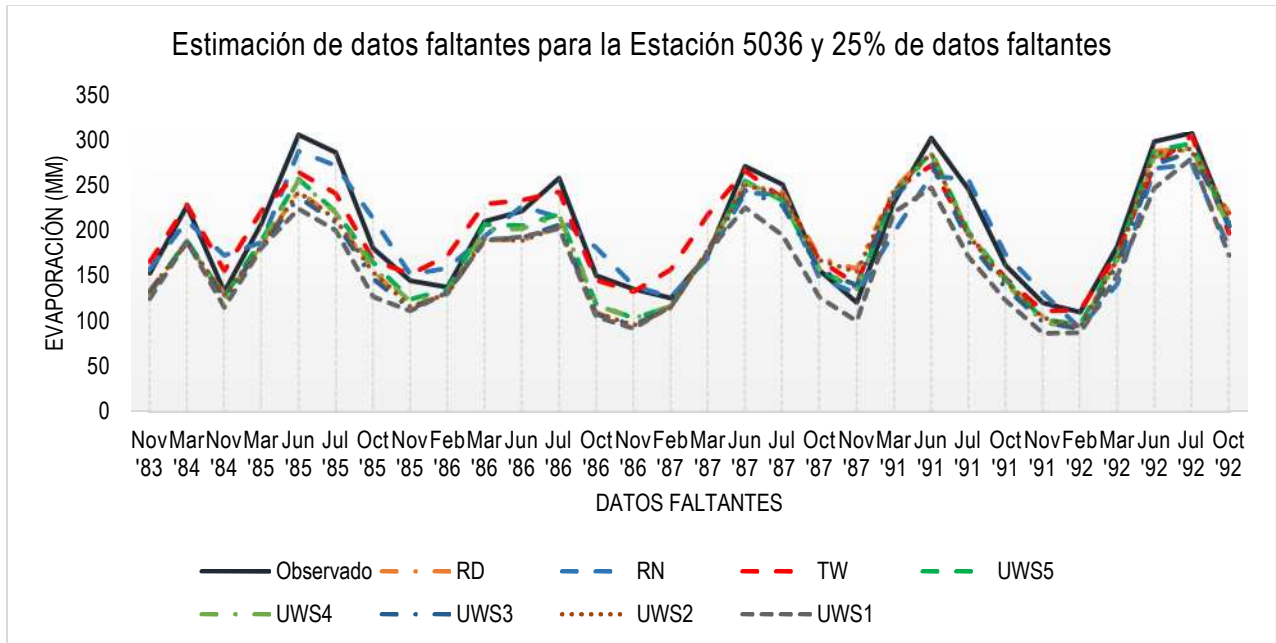


Figura 45. Comparación gráfica de los resultados de la estación 5036 para la variable climática evaporación y 25% datos faltantes

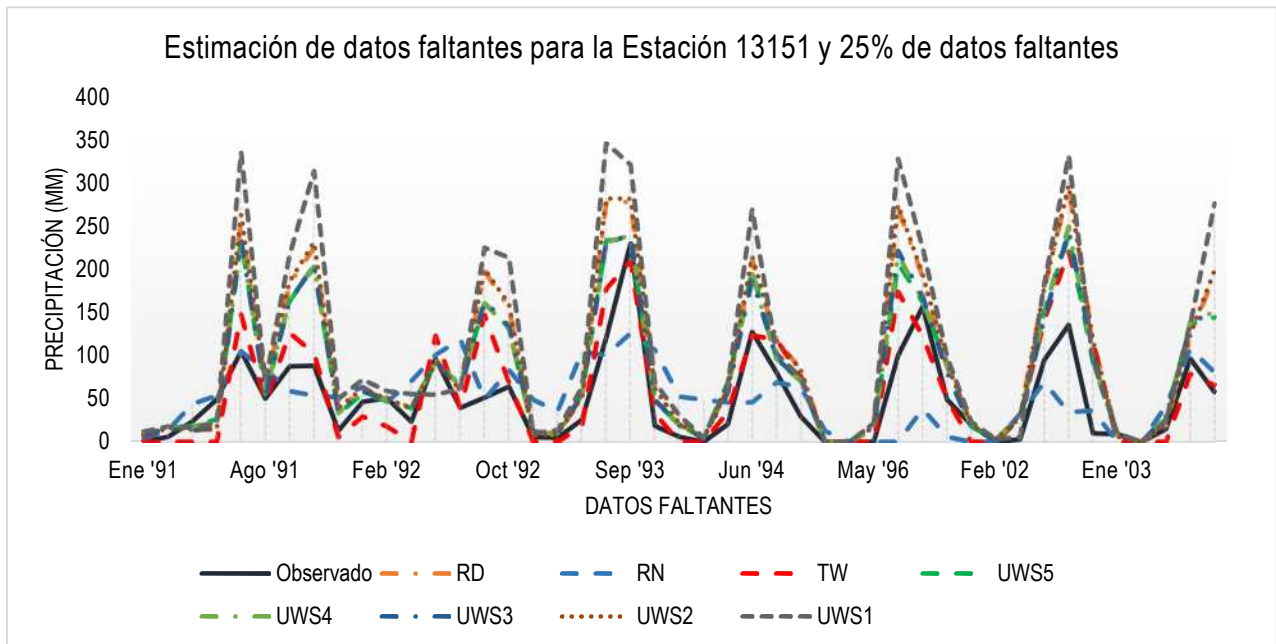


Figura 46. Comparación gráfica de los resultados de la estación 13151 para la variable climática precipitación y 25% datos faltantes

A continuación se presentan gráficas de diagramas de dispersión, por cada variable climática y para una estación meteorológica, donde se muestran los valores observados

y los valores estimados con cada uno de los modelos. Los resultados que se presentan son la estación 30056 para la variable temperatura, 5036 para la variable evaporación y 13151 para precipitación, cada uno para 25% de datos faltantes.

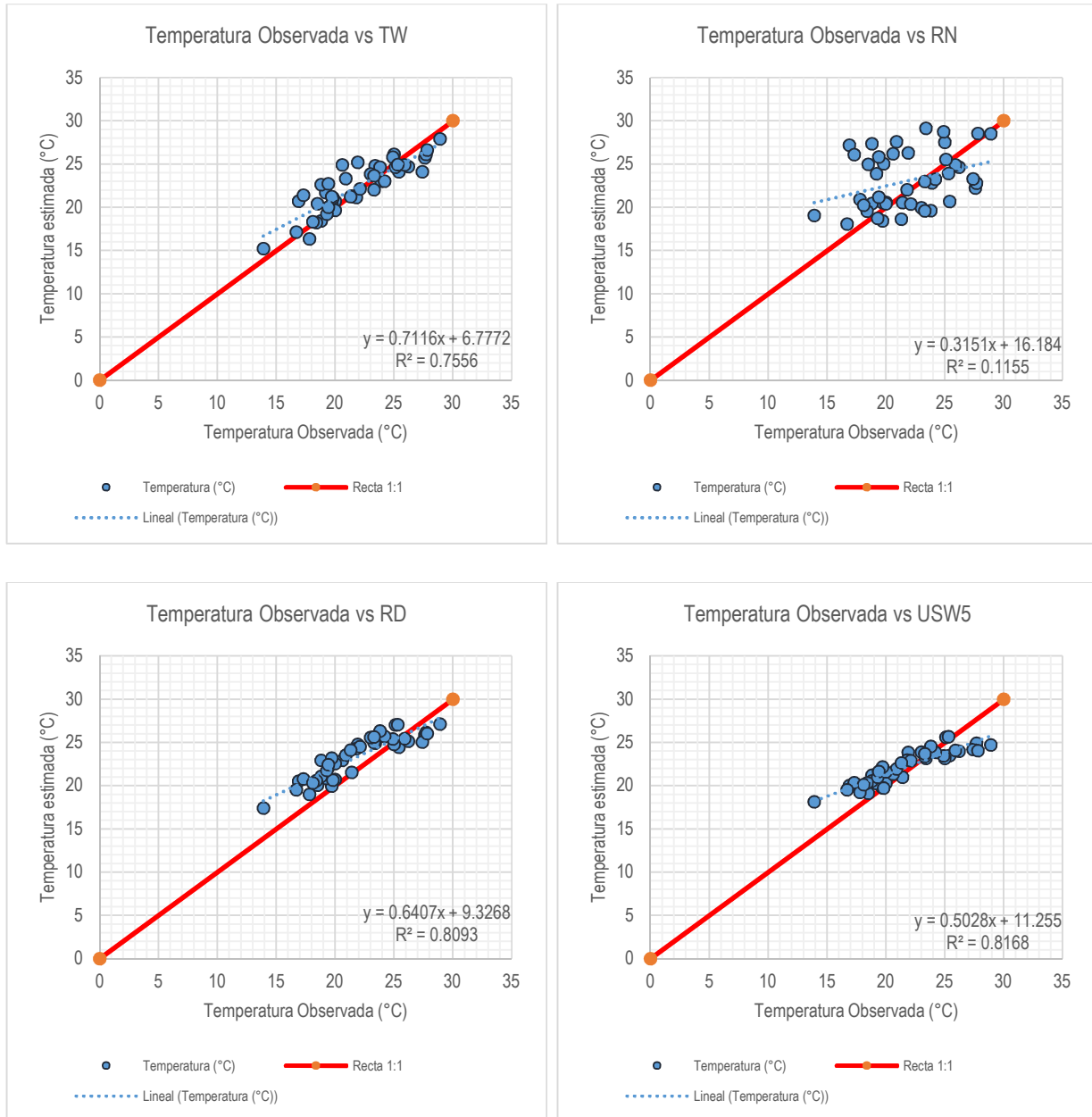


Figura 47. Diagramas de dispersión para los valores observados y estimados por modelo de la variable temperatura para la estación 50036 de clima cálido subhúmedo

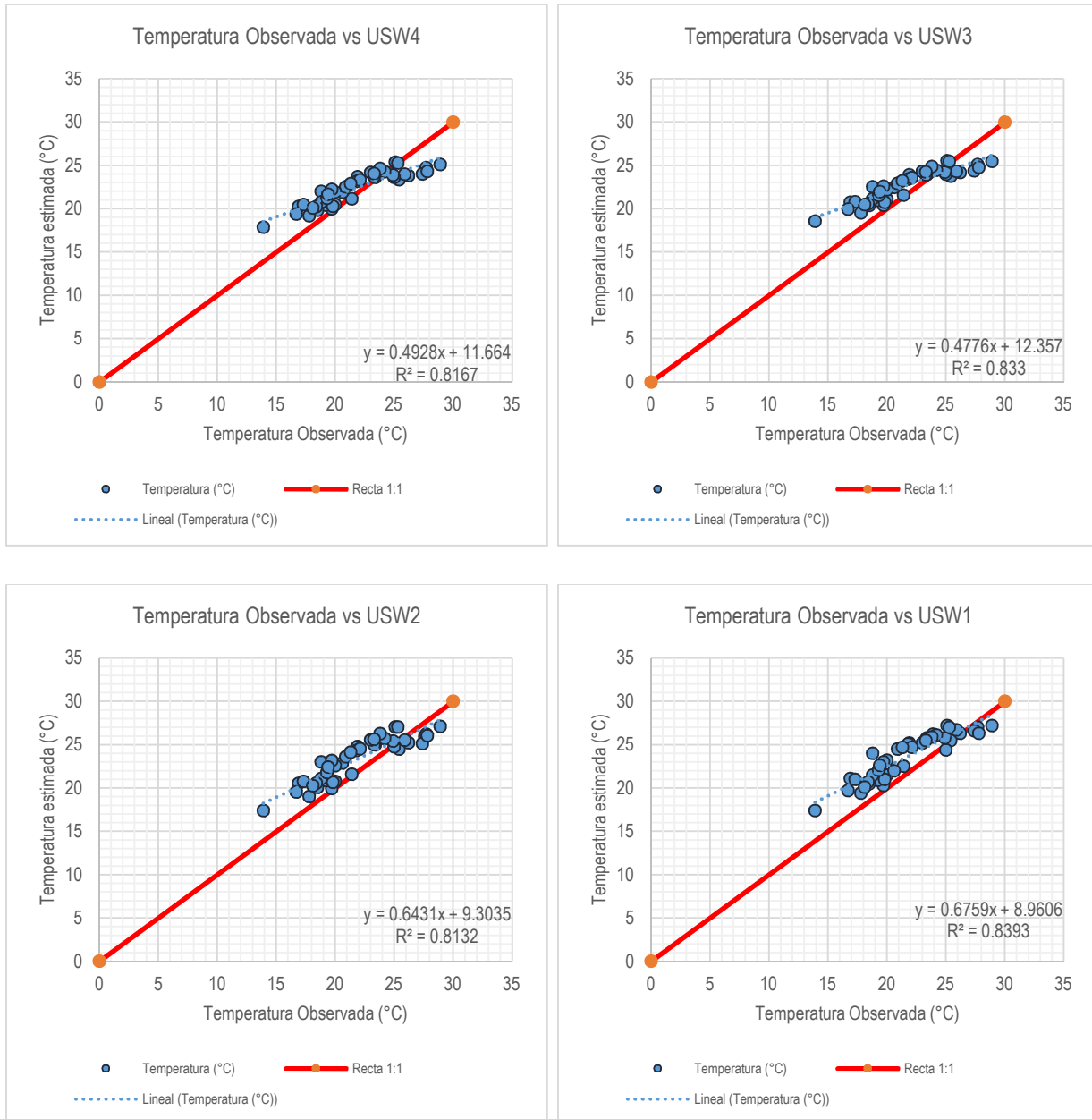


Figura 48. Continuación de la Figura 47

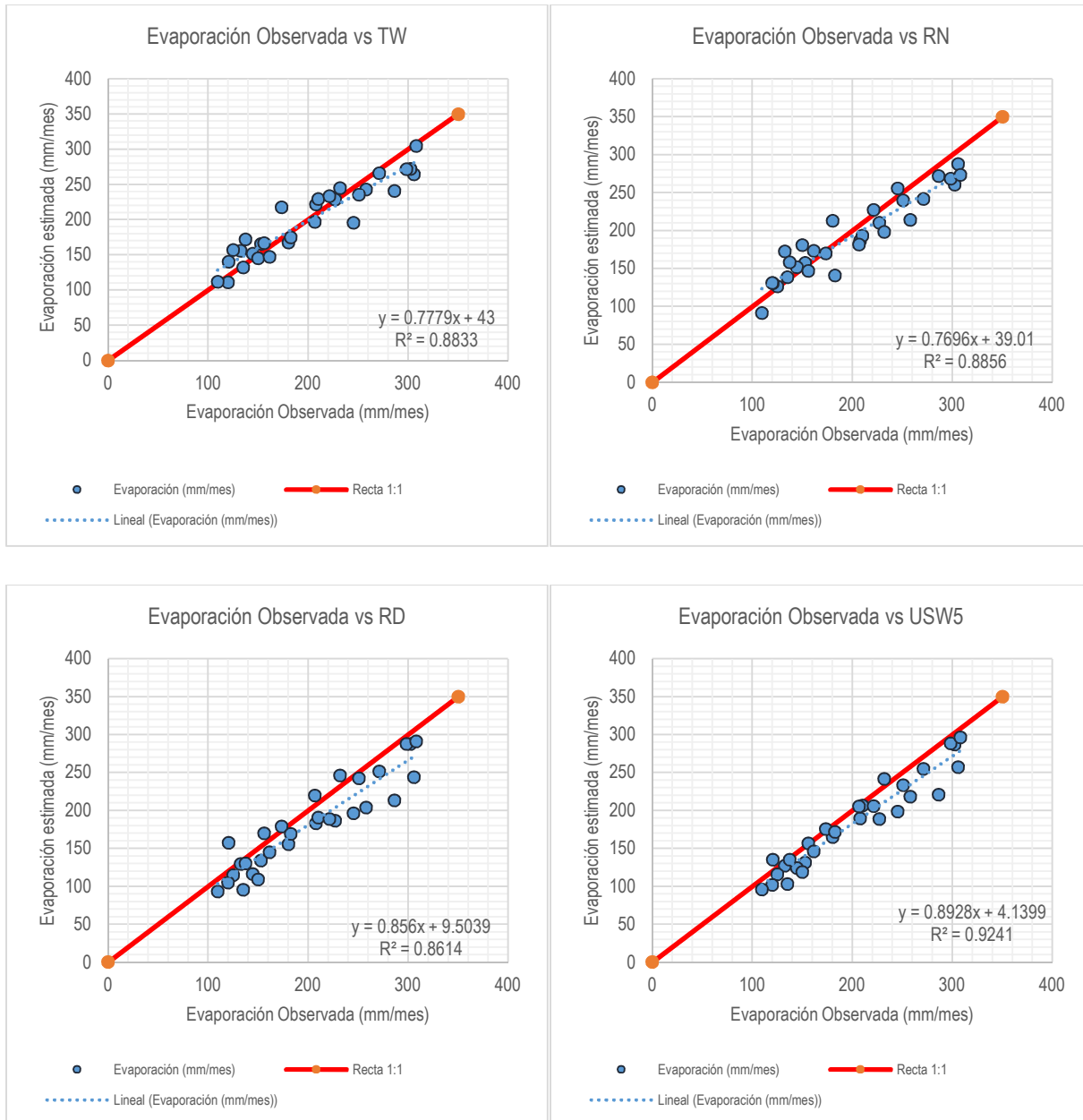


Figura 49. Diagramas de dispersión de los valores observados y estimados por modelo de la variable evaporación para la estación 5036 de clima muy seco semicálido

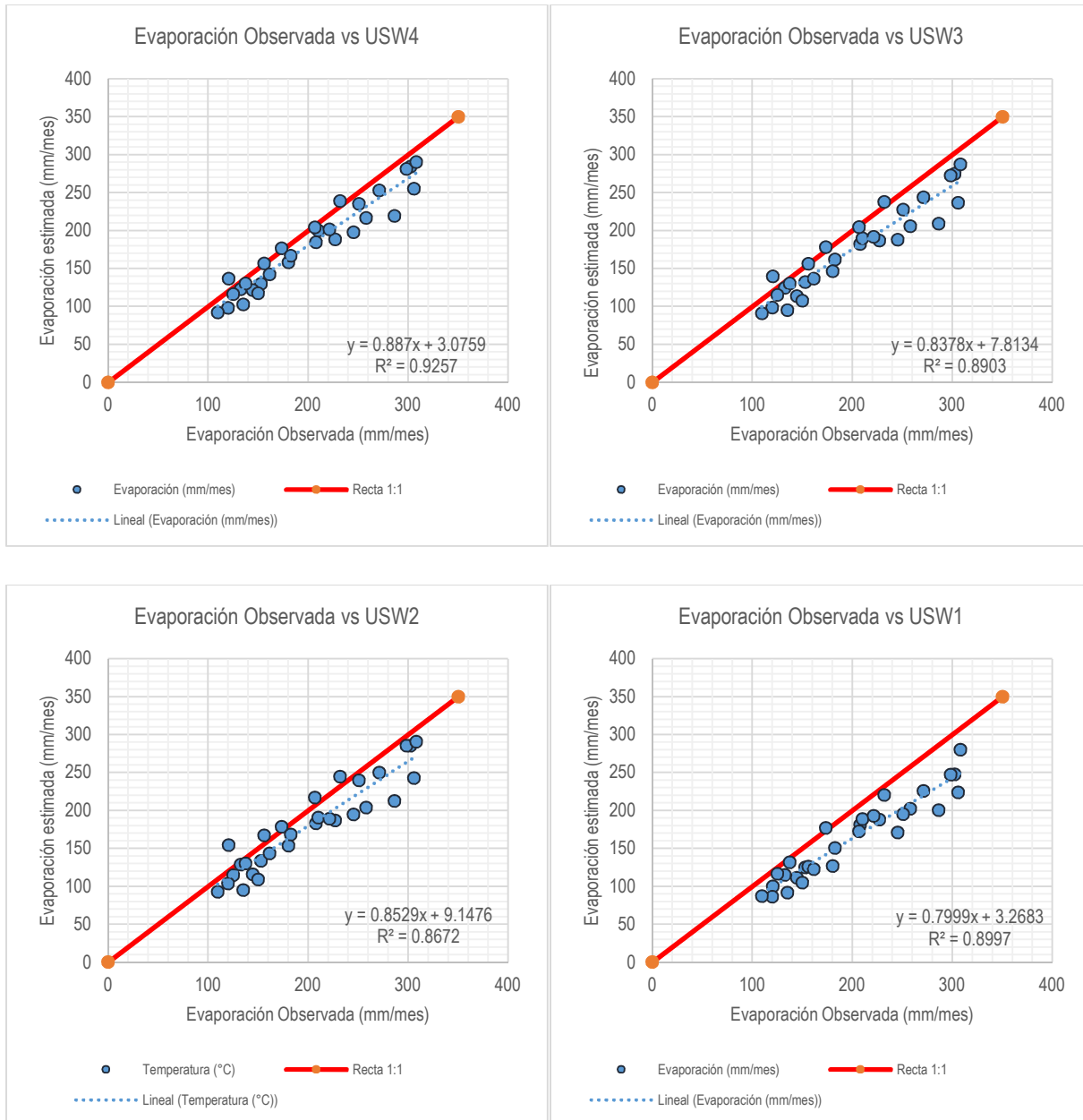


Figura 50. Continuación de la Figura 49

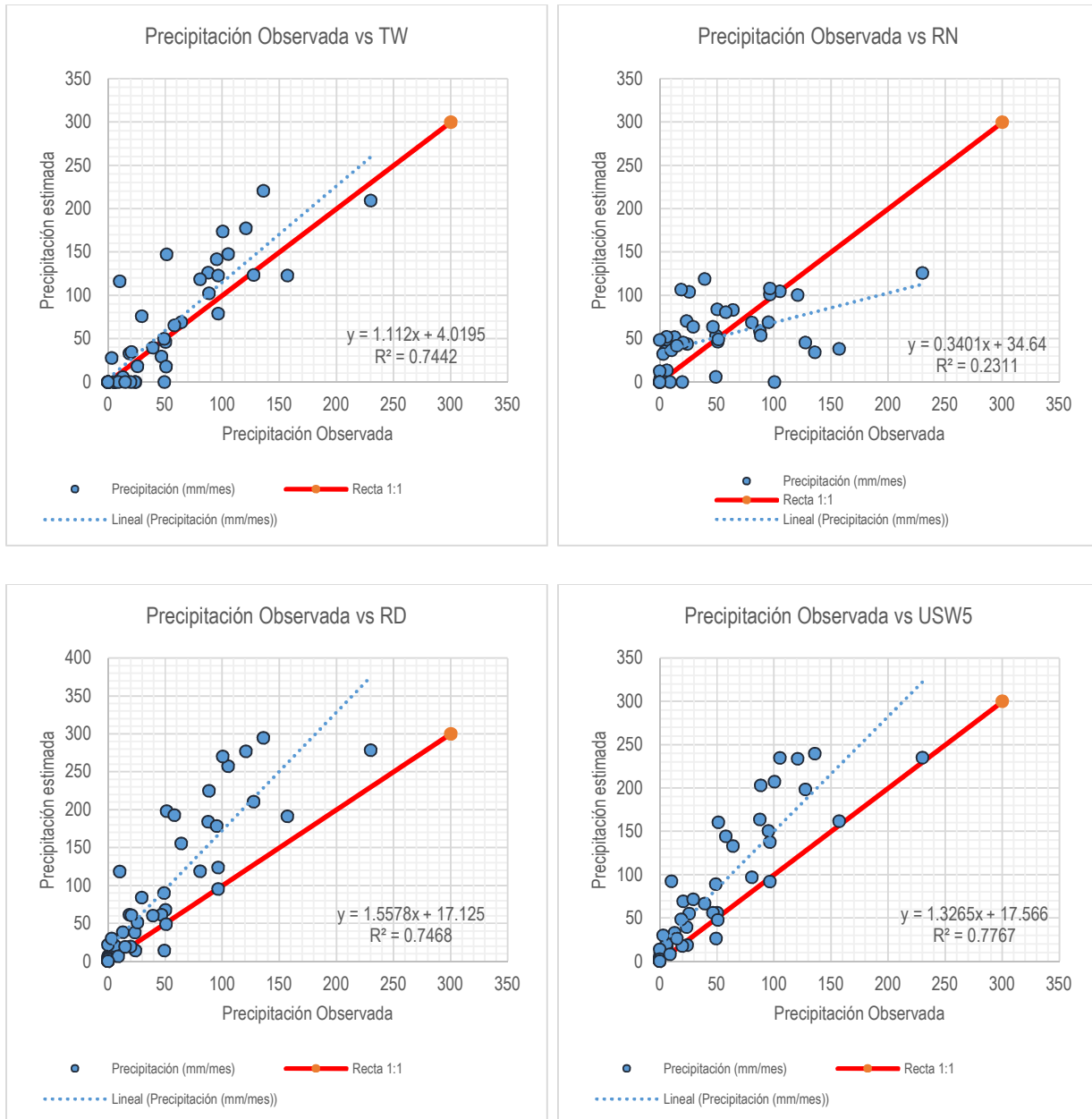


Figura 51. Diagramas de dispersión de los valores observados y estimados por modelo de la variable precipitación para la estación 13151 de clima templado subhúmedo

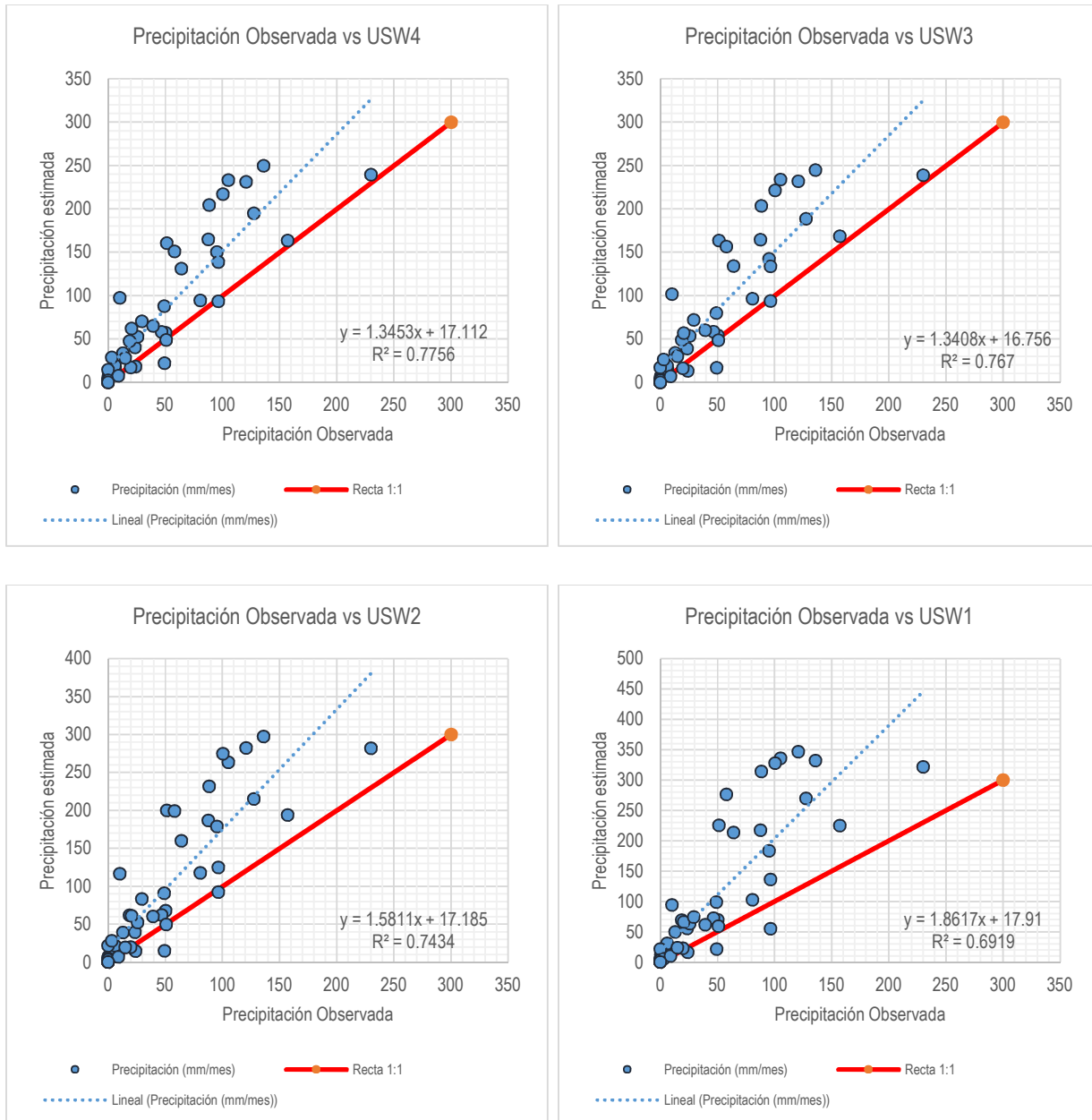


Figura 52. Continuación de la Figura 51

5.7. Análisis estadístico

En esta sección se muestra los resultados obtenidos de la aplicación de los índices y parámetros estadísticos para determinar el mejor modelo de estimación de datos faltantes. Se consideró como el mejor modelo aquel que tuviera el mayor número de índices y parámetros estadísticos que así lo demuestren. El análisis estadístico se generó de dos formas: por variable climática y por región climática, los Cuadros

comparativos para estos se encuentran en el anexo 9.2 del presente trabajo de investigación. Los resultados finales para cada porcentaje de datos faltantes se presentan a continuación:

5.7.1. 15% de datos faltantes

En el Cuadro 27 se presentan los resultados de aplicar los índices y parámetros estadísticos, en este Cuadro se observa que el mejor de estimación es la del Servicio climático de los Estados Unidos considerando cinco estaciones cercanas al tener una mejor estimación de variables en cuatro ocasiones, el segundo mejor método para este porcentaje de datos faltantes es el modelo matemático transformada de wavelet al estimar tres veces de la mejor manera y como tercer modelo de estimación se encuentran las redes neuronales artificiales al estimar de igual manera tres veces.

Cuadro 27. Resumen de resultados para 15% de datos faltantes

Modelo	RD	RN	TW	UWS5	UWS4	UWS3	UWS2	UWS1
Por variable climática								
Temperatura								
Precipitación								
Evaporación								
Por región climática								
Cálido Subhúmedo								
Temperatura								
Precipitación								
Evaporación								
Muy Seco Semicálido								
Temperatura								
Precipitación								
Evaporación								
Templado Subhúmedo								
Temperatura								
Precipitación								
Evaporación								
Total=	1	3	3	4	0	1	0	0

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas

5.7.2. 25% de datos faltantes

Como se observa en el siguiente Cuadro el mejor modelo de estimación es la transformada de Wavelet, ya que es el mejor en cinco ocasiones, como segundo mejor modelo es la del servicio climático de los Estados Unidos con cinco y tres estaciones cercanas. Los Cuadros comparativos por variable y región climática para este porcentaje de datos faltantes se presentan en la sección de anexos.

Cuadro 28. Resumen de resultados para 25% de datos faltantes

Modelo	RD	RN	TW	UWS5	UWS4	UWS3	UWS2	UWS1
Por variable climática								
Temperatura								
Precipitación								
Evaporación								
Por región climática								
Cálido Subhúmedo								
Temperatura								
Precipitación								
Evaporación								
Muy Seco Semicálido								
Temperatura								
Precipitación								
Evaporación								
Templado Subhúmedo								
Temperatura								
Precipitación								
Evaporación								
Total=	1	1	5	2	0	2	0	1

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas

5.8. Estudio de interrelación de variables climáticas

Se realizó el estudio de interrelación variables para determinar si una variable climática puede ser estimada mediante otra, siempre que exista una alta relación entre ellos. Para determinar la relación entre las variables se utilizó el coeficiente de correlación simple y múltiple, los cuales se presentan a continuación.

5.8.1. Coeficiente de correlación simple

Para estimar datos faltantes es necesario que el coeficiente de correlación describa el comportamiento del 50% de los datos, por lo que se eligió un valor de coeficiente de correlación de 0.7.

Para el caso de 15% de datos faltantes se tienen las cantidades siguientes que cumplen con la condición antes descrita:

- Para la variable evaporación
 - Cuatro estaciones se relacionan con la temperatura
 - Ninguna estación se relaciona con la precipitación.
- Para la variable precipitación
 - Ninguna estación se relaciona con la temperatura y con la evaporación
- Para la variable temperatura
 - Tres estaciones se relacionan con la evaporación
 - Ninguna estación se relaciona con la precipitación

Las estaciones que cumplen con esta condición se presentan en el Cuadro 29, para conocer el número de estación a la que se refiere cada uno de los nombres, consultar el anexo 9.1.

Cuadro 29. Coeficientes de correlación simple para cada variable climática 15 % de datos faltantes

EVA	TEMP	TEMP	EVA
0 mss e	0.8007	1 mss tp	0.8825
1 mss e	0.8716	4 cs tp	0.7215
7 ts e	0.7035	6 cs tp	0.7190
1 cs e	0.7298		

Para el caso de 25% de datos faltantes se tienen las cantidades siguientes que cumplen con la condición de mayor o igual a 0.7:

- Para la variable evaporación
 - Tres estaciones se relacionan con la temperatura
 - Ninguna estación se relaciona con la precipitación.
- Para la variable precipitación

- Ninguna estación se relaciona con la temperatura y con la evaporación
- Para la variable temperatura
 - Tres estaciones se relacionan con la evaporación
 - Ninguna estación se relaciona con la precipitación

Las estaciones que cumplen con esta condición se presentan a continuación:

Cuadro 30. Coeficientes de correlación simple para cada variable climática 15 % de datos faltantes

EVA	TEMP	TEMP	EVA
0 mss e	0.7950	1 mss tp	0.8844
1 mss e	0.8843	4 cs tp	0.7224
1 cs e	0.7260	6 cs tp	0.7320

Los coeficientes de correlación completos para las 42 estaciones se presentan en el anexo 9.3.

5.8.2. Coeficiente de correlación múltiple

En este apartado se presentan los resultados del coeficiente de correlación múltiple para las 42 estaciones bases seleccionadas para 15% y 25% de datos faltantes.

Para el caso de 15%, en el Cuadro siguiente se observa que la evaporación que se relaciona con la precipitación y temperatura de manera conjunta únicamente cumplen cuatro estaciones meteorológicas; para la precipitación que se relaciona con la evaporación y la temperatura solamente dos estaciones, destacando que las variables de manera independiente no cumplía ninguna; y para temperatura que se relaciona con precipitación y evaporación de manera simultánea cumplen seis estaciones meteorológicas.

Cuadro 31. Coeficientes de correlación múltiple para cada variable climática para 15% de datos faltantes

EVAPORACIÓN	MÚLTIPLE	PRECIPITACIÓN	MÚLTIPLE	TEMPERATURA	MÚLTIPLE
0 mss e	0.8587	5p ts tp	0.7214	3 ts tp	0.8002
1 mss e	0.8964	1p cs tp	0.7184	5 ts tp	0.8102
7 ts e	0.7469			1 mss tp	0.9108
1 cs e	0.7384			1 cs tp	0.8486
				4 cs tp	0.7738
				6 cs tp	0.8651

Para el caso del 25%, en el Cuadro siguiente se observa que la evaporación que se relaciona con la precipitación y temperatura de manera conjunta únicamente cumplen cuatro estaciones meteorológicas; para la precipitación que se relaciona con la evaporación y la temperatura solamente dos estaciones, destacando que las variables de manera independiente no cumplía ninguna para 25% de datos faltantes; y para temperatura que se relaciona con precipitación y evaporación de manera simultánea cumplen ocho estaciones meteorológicas.

Cuadro 32. Coeficientes de correlación múltiple para cada variable climática para 25% de datos faltantes

EVAPORACIÓN	MÚLTIPLE
0 mss e	0.8473
1 mss e	0.9055
7 ts e	0.7372
1 cs e	0.7344

PRECIPITACIÓN	MÚLTIPLE
5p ts tp	0.7241
1p cs tp	0.7171

TEMPERATURA	MÚLTIPLE
3 ts tp	0.8040
5 ts tp	0.8213
6 ts tp	0.7140
1 mss tp	0.9119
1 cs tp	0.8564
3 cs tp	0.7182
4 cs tp	0.7786
6 cs tp	0.8686

5.8.3. Comparación estadística de la aplicación de la regresión lineal simple y múltiple

Se realizó la comparación estadística de los modelos generados por regresión lineal simple y múltiple con los modelos propuestos en la presente investigación, destacando que esta comparación se realizó por estación meteorológica debido a que como se observó en los Cuadros anteriores no todas las estaciones bases cumplen con la condición del coeficiente de correlación simple y múltiple mayor o igual a 0.7. De igual manera de las estaciones seleccionadas algunas no se aplicaron debido a que el dato faltantes a considerar también era faltante en la otra variable climática.

Para el caso de 15% de datos faltantes, en el Cuadro siguiente se observa que el mejor modelo de estimación, analizando 10 estaciones meteorológicas es la transformada de Wavelet con cuatro estaciones meteorológicas. Es importante destacar que el modelo de regresión lineal múltiple fue el mejor modelo para la variable climática precipitación que se encuentra en la región climática templado subhúmedo.

Cuadro 33. Resultados de la comparación de la regresión lineal simple y múltiple con los demás modelos propuestos para 15% de datos faltantes

Método	RD	RN	TW	UWS5	UWS4	UWS3	UWS2	UWS1	RLV1	RLV2	RM
Evaporación											
0 mss e											
1 mss e											
7 ts e											
1 cs e											
Precipitación											
5p ts tp											
1p cs tp											
Temperatura											
3 ts tp											
5 ts tp											
1 cs tp											
4 cs tp											
Total	1	1	4	0	1	1	1	0	0	0	1

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; RLV1= Regresión lineal simple para la variable 1; RLV2= Regresión lineal simple para la variable 2; RM=Regresión lineal múltiple

En el Cuadro siguiente se observa que el mejor modelo de estimación de datos faltantes para 25% y analizando nueve estaciones meteorológicas es la transformada de Wavelet con seis estaciones meteorológicas.

Cuadro 34. Resultados de la comparación de la regresión lineal simple y múltiple con los demás modelos propuestos para 25% de datos faltantes

Método	RD	RN	TW	UWS5	UWS4	UWS3	UWS2	UWS1	RLV1	RLV2	RM
Evaporación											
0 mss e											
1 mss e											
7 ts e											
1 cs e											
Precipitación											
5p ts tp											
1p cs tp											
Temperatura											
5 ts tp											
1 cs tp											



Método	RD	RN	TW	UWS5	UWS4	UWS3	UWS2	UWS1	RLV1	RLV2	RM
4 cs tp											
Total	1	0	6	0	0	1	0	1	0	0	0

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; RLV1= Regresión lineal simple para la variable 1; RLV2= Regresión lineal simple para la variable 2; RM=Regresión lineal múltiple

6. CONCLUSIONES

De acuerdo al análisis estadístico realizado para 15% y 25% de datos faltantes se concluye que los mejores modelos de estimación de datos faltantes es la Transformada de Wavelet, seguido del modelo del servicio climático de los Estados Unidos considerando cinco estaciones cercanas, como tercer modelo se encuentran las redes neuronales artificiales. Estos modelos fueron los que en su mayoría tuvieron una buena estimación de los datos faltantes.

Estos modelos fueron determinados como los mejores considerando por lo menos tres índices o parámetros estadísticos que así lo demostraron. Es importante mencionar que estos modelos son eficientes para estimar datos faltantes considerando las siguientes ventajas y desventajas:

- US Weather Service, la desventaja principal de este modelo es que requiere de cinco estaciones cercanas, situación que algunas veces resulta complicado encontrar dichas estaciones meteorológicas. La ventaja es que únicamente requiere los datos del tiempo que se esté analizando y no de un periodo de información.
- La Transformada de Wavelet, tiene como ventaja principal que únicamente requiere una estación cercana con un comportamiento similar a la estación que se esté analizando. Este modelo hereda los valores extremos y singularidades de la estación cercana. La desventaja de este modelo con respecto al modelo anterior es que requiere de un cierto periodo de tiempo de información completa donde se encuentren los datos que se quieren estimar.
- Las Redes Neuronales Artificiales, su ventaja es que no necesita de estaciones cercanas y únicamente requiere la información de la misma estación para generar los datos faltantes. Su desventaja radica en que en la variable precipitación, este modelo no resulta ser el más apropiado y para las otras variables una buena estimación está sujeta a un buen entrenamiento de la red neuronal, situación que



hay que realizar cada vez que se quieran estimar datos faltantes para estaciones meteorológicas.

Cada uno de estos modelos requiere de cierta información que les permite estimar de la mejor manera los datos faltantes para cada variable climática y para cada región climática. En el Cuadro 27 y Cuadro 28 se resumen los mejores modelos.



7. RECOMENDACIONES

Realizar un estudio similar para obtener modelos apropiados para las regiones climáticas faltantes en la república mexicana, ya que este estudio solo se abarcó el 50% aproximadamente del territorio nacional.

Para estudios posteriores donde se requiere estimar datos faltantes se recomienda realizar estos con los modelos propuestos para cada región climática y para cada variable climática, considerando su factibilidad dependiendo de la información que se tenga.

Estos modelos considerados como mejores se realizaron únicamente para datos mensuales, por lo que se sería importante probar su capacidad predictiva para datos diarios.

8. LITERATURA CITADA

- Aguirre R., A. 2009. Comparación entre la presión de vapor actual observada y la estimada de observatorios sinópticos de México. Tesis profesional Departamento de Irrigación. Universidad Autónoma Chapingo. México. 107p.
- Alkaeed, O., Flores, C., Jinno, K., Tsutsumi, A. 2006. Comparison of several reference evapotranspiration methods for Itoshima Peninsula Area, Fukuoka, Japan. *Memoirs of the Faculty of Engineering*. Volumen 66. Número 1.15p.
- Álvarez, R. 1994. Estadística multivariante y no paramétrica con SPSS. Primera edición. Editorial Díaz de Santos. Madrid, España. 408p.
- Basogain, X. 1998. Redes neuronales artificiales y sus aplicaciones. Publicación de la escuela de ingenieros. Escuela Superior de Ingeniería de Bilbao. Bilbao, España. 79p.
- Carbajal, C., M.; Yarleque C.; Posadas A.; Silvestre E.; Mejía A.; Quiroz R. 2010. Reconstrucción de datos faltantes de precipitación pluvial diaria mediante la Transformada de Wavelet. *Revista Peruana Geo-Atmosférica*. RPGA (2), 76-88p.
- Casellas, O. 2005. Simulación de los campos de viento de la península Ibérica y el área geográfica de Catalunya con alta resolución espacial para distintas situaciones meteorológicas típicas. Tesis doctoral. Universitat Politècnica de Catalunya. Barcelona España.
- Castro, L. y Castro, S. 1995. Wavelets y sus Aplicaciones. I Congreso Argentino de Ciencias de la Computación. (En línea). < <http://hdl.handle.net/10915/24289>>. Consultado el 06 de mayo de 2014.



- Cordova J. 2012. Modelado de series en tiempo vía redes neuronales wavelet haar. Tesis Doctoral. Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional. Distrito Federal, México. 155p.
- Ferrer, L. 2011. Divertimento con wavelets. Departamento de Ingeniería Electrónica. (En línea). < http://upcommons.upc.edu/eprints/bitstream/2117/13081/1/Ferrer_Jcee2010.pdf>. Consultado el 06 de mayo de 2014.
- Fuertes L., S. 1998. Evaluación de dos métodos de determinación de la evapotranspiración de referencia en condiciones semiáridas. Tesis profesional. Universidad de Lleida. España. 202p.
- Gómez, H. 2012. Aplicación de la transformada wavelet y el método level set para el filtrado y segmentación de imágenes. Tesis profesional. Universidad Politécnica Salesiana. Cuenca, Ecuador. 109p.
- Gómez, W. 2010. Pronóstico de una serie temporal usando redes neuronales. Tesis profesional. Universidad Tecnológica de la Mixteca. Oaxaca, México. 112p.
- Guevara, J. 2003. Métodos de estimación y ajuste de datos climáticos. Publicidad Gráficas León S.R.L. Segunda edición. Caracas, Venezuela. 128p.
- Infante, S. y Zárate, G. 2011. Métodos estadísticos, un enfoque interdisciplinario. 2da Edición. Trillas. México. 643p.
- Kouro, S. y Musalem, R. 2002. Tutorial introductorio a la teoría de Wavelet. Artículo presentado como trabajo de la asignatura Técnicas Modernas en Automática. (En línea). <<http://www2.elo.utfsm.cl/~elo377/documentos/Wavelet.pdf>>. Consultado el 06 de mayo de 2014.



- Landassuri, V. 2006. Predicción de series de tiempo con descomposición empírica en nodos, algoritmo genético y redes neuronales artificiales. Tesis de maestría. Instituto Politécnico Nacional, Distrito Federal, México. 98p.
- Lobo L., D. 2011. Guía metodológica para la elaboración del mapa de zonas áridas, semiáridas y subhúmedas secas de América latina y el caribe. (En línea). <http://www.cazalac.org/documentos/Guia_Mapa_ZA_ALC.pdf>. Consultado el 26 de noviembre de 2012.
- Martín, J.; Soria, E.; Serrano, A. 2009. Redes neuronales artificiales. Material de curso. Universidad de Valencia. Valencia, España. 145p.
- Matich, D. 2001. Redes neuronales: conceptos básicos y aplicaciones. Material de curso. Universidad Tecnológica Nacional. Buenos Aires, Argentina. 55p.
- Montaña, J. 2002. Redes neuronales artificiales aplicadas al análisis de datos. Tesis Doctoral. Universidad de les Illes Balears. Mallorca, España. 275p.
- Moreno, A. 2009. Desarrollo de una interfaz gráfica de redes neuronales usando Matlab. Proyecto fin de carrera. Universidad Carlos III Madrid. Madrid, España. 126p.
- Moros, A. 2010. Caracterización de señales de precipitación mediante la transformada de Fourier y transformada de Wavelet. Tesis de maestría. Pontificia Universidad Javeriana, Bogotá, Colombia. 199p.
- Navarro, J.; Losilla, J. 2000. Análisis de datos faltantes mediante redes neuronales artificiales. Revista Psicothema. Volumen 12. Número 3. 8p.



- Nieto, N.; Marcela, D. 2008. El uso de la transformada de wavelet discreta en la reconstrucción de señales senosoidales. Revista Scientia et Technica. Año XIV. Número 38. 6p.
- Pedroza, H. y Dikovskiy, L. 2006. Sistema de análisis estadístico con SPSS. Primera edición. IICA. Managua, Nicaragua. 167p.
- Pizarro, R.; Ausensi, P.; Aravena, D.; Sangüesa, C.; León, L.; Balocchi, F. 2009. Evaluación de métodos hidrológicos para la completación de datos faltantes de precipitación en estaciones de la región del Maule, Chile. Revista Aqua Lac, Programa Hidrológico Internacional. Vol. 1. Nº 2. Pp. 172-185. Universidad de Talca, Chile.
- Represa, C. 2002. Una nueva aproximación a la multiresolución en el dominio del tiempo (MRTD) basada en la transformada discreta en Wavelets. Tesis Doctoral. Universidad de Burgos. Burgos, España. 27p.
- Rupérez, E. 2012. Wavelets. Proyecto Sistemas Informáticos. Universidad Complutense de Madrid. Madrid, España. 69p.
- Saba I.; Ortega J.; Cedeño F. 2008. Estimación de datos faltantes en estaciones meteorológicas de Venezuela vía un modelo de redes neuronales. Revista de Climatología. Vol. 8: 51-70. Universidad de Carabobo. Venezuela.
- Sánchez A., P. 2006. Métodos estadísticos aplicados. Primera edición. Universidad de Barcelona. Barcelona, España. 328 p.
- Sánchez, D. 2004. Segmentación y realce de señales de voz usando la transformada de Wavelet y DSP's. Proyecto de grado. Universidad del Quindío. Armenia, Colombia. 78p.



- Sánchez, L. 2011. Efecto de los datos atípicos discordantes y contaminantes en la implementación de las redes neuronales supervisadas backpropagation. Tesis Profesional. Universidad Nacional de Ingeniería. Lima, Perú. 155p.
- Servin A. M. 2010. Módulo de gestión de bases de datos de estaciones climatológicas. Tesis profesional. Facultad de Ingeniería Civil. Universidad Michoacana de San Nicolás de Hidalgo. Morelia, Michoacán. 58 pp.
- Taucher, E. 1997. Bioestadística. Editorial Universitaria. Segunda edición. Santiago de Chile, Chile. 310p.
- Valencia, M.; Yáñez, C.; Sánchez L. 2006. Algoritmo backpropagation para redes neuronales: conceptos y aplicaciones. Centro de investigación en computación, Instituto Politécnico Nacional. Serie: Verde. Número: 125. 14p.
- Vivanco, M. 2005. Muestreo estadístico diseño y aplicaciones. Editorial Universitaria. Primera Edición. Santiago de Chile, Chile. 209p.
- Zamorano, M. 2010. Análisis de señales mediante STFT y Wavelet, aplicación a defectología en rodamientos. Proyecto fin de carrera. Universidad Carlos III de Madrid. Madrid, España. 429p
- Zapotitla, J. 2011. Precursores de protones relativistas solares por análisis digital de señales utilizando Transformada wavelet y análisis de componentes principales. Tesis profesional. Universidad Nacional Autónoma de México. D.F. México. P 31-43.

9. ANEXOS

9.1. Estaciones meteorológicas bases y vecinas seleccionadas para cada región climática

Cuadro 35. Estaciones meteorológicas utilizadas en la investigación

Clave	Estado	Nombre	Latitud	Longitud	Altitud	Tipo	distancia(km)
10108	Dgo.	CIUDAD LERDO (DGE)	25.5	-103.367	1140	0 mss e	
5029	Coah.	PRESA LA FLOR, TORREON	25.068	-103.284	1300	0 mss e	48.8103
5036	Coah.	SAN PEDRO, SAN PEDRO	25.757	-102.996	1100	0 mss e	46.9577
10009	Dgo.	CIUDAD LERDO, LERDO (SMN)	25.533	-103.517	1135	0 mss e	15.5105
10045	Dgo.	MAPIMI (KM.29), MAPIMI	25.833	-103.98	1300	0 mss e	71.8128
10049	Dgo.	NAZAS, NAZAS	25.23	-104.107	1300	0 mss e	80.2741
5036	Coah.	SAN PEDRO, SAN PEDRO	25.757	-102.996	1100	1 mss e	
5029	Coah.	PRESA LA FLOR, TORREON	25.068	-103.284	1300	1 mss e	81.9835
5041	Coah.	VIESCA, VIESCA (DGE)	25.319	-102.808	1100	1 mss e	52.2869
10009	Dgo.	CIUDAD LERDO, LERDO (SMN)	25.533	-103.517	1135	1 mss e	57.9260
10085	Dgo.	TLAHUALILLO, TLAHUALILLO	26.101	-103.441	1100	1 mss e	58.7466
10108	Dgo.	CIUDAD LERDO (DGE)	25.5	-103.367	1140	1 mss e	46.9577
13002	Hgo.	APAN, APAN (DGE)	19.717	-98.45	2496	1 ts e	
13027	Hgo.	SAN JERONIMO, TEPEAPULCO	19.817	-98.483	1528	1 ts e	11.6564
13085	Hgo.	PRESA TEZOYO, ALMOLOYA	19.85	-98.267	2590	1 ts e	24.2210
21037	Pue.	EL PAREDON, CHIGNAHUAPAN	19.867	-98.267	0	1 ts e	25.4212
29006	Tlax.	CUAULA, CALPULALPAN	19.6	-98.65	2600	1 ts e	24.6823
29023	Tlax.	SAN CRISTOBAL	19.567	-98.483	0	1 ts e	17.0526
16027	Mich.	CUITZEO, CUITZEO	19.967	-101.317	1831	2 ts e	
16016	Mich.	CARRILLO PUERTO, A. O.	19.917	-101.083	1840	2 ts e	25.1115
16028	Mich.	CUITZILLO GRANDE,	19.767	-101.117	1840	2 ts e	30.5633
16100	Mich.	PUENTE SAN ISIDRO,	19.85	-101.533	1989	2 ts e	26.0913
16155	Mich.	COPANDARO, JIMENEZ	19.933	-101.083	1981	2 ts e	24.7764
16254	Mich.	TEREMENDO, MORELIA	19.783	-101.483	2350	2 ts e	26.8617
16124	Mich.	TEMASCALES, CONTEPEC	20.067	-100.233	2470	3 ts e	
11132	Gto.	PURUAGUA, JERECUARO	20.083	-100.417	1980	3 ts e	19.3206
11031	Gto.	JERECUARO, JERECUARO	20.15	-100.483	1787	3 ts e	27.7187
22033	Qro.	SAN PABLO, AMEALCO	20.128	-99.997	240	3 ts e	25.5891
16111	Mich.	SAN JOSE, MARAVATIO	19.917	-100.3	2100	3 ts e	18.1093
22047	Qro.	PRESA EL CAPULIN, AMEALCO	20.125	-100.044	240	3 ts e	20.7866
13099	Hgo.	METEPEC, METEPEC	20.233	-98.317	225	4 ts e	
13028	Hgo.	SAN MIGUEL ALLENDE,	20.233	-98.567	2480	4 ts e	26.1126
13100	Hgo.	OMITLAN, OMITLAN	20.167	-98.633	400	4 ts e	33.8210
13116	Hgo.	EL SEMBO, HUASCA OCAMPO	20.15	-98.567	200	4 ts e	27.7056
21066	Pue.	PUEBLO NVO., CHIGNAHUAPAN	19.95	-98.083	2700	4 ts e	39.8865
21209	Pue.	LOS HUMEROS, CHIGNAUTLA	20	-98.2	1200	4 ts e	28.6761
14090	Jal.	LA YERBABUENA, TOTOTLAN	20.567	-102.8	1570	5 ts e	
14016	Jal.	ATEQUIZA (CHAPALA),	20.417	-103.083	1521	5 ts e	33.9066
14087	Jal.	LA RED, TEPATITLAN DE M.	20.733	-102.8	1793	5 ts e	18.4790
14111	Jal.	PONCITLAN, PONCITLAN	20.385	-102.923	1530	5 ts e	23.9794
14147	Jal.	TEPATITLAN	20.817	-102.767	1746	5 ts e	28.0413



Clave	Estado	Nombre	Latitud	Longitud	Altitud	Tipo	distancia(km)
14154	Jal.	TOTOTLAN, TOTOTLAN	20.55	-102.8	1790	5 ts e	1.8924
11028	Gto.	IRAPUATO, IRAPUATO (SMN)	20.683	-101.35	1724	6 ts e	
11001	Gto.	ABASOLO, ABASOLO	20.45	-101.533	1760	6 ts e	32.1952
11035	Gto.	LA SANDIA, LEON	20.917	-101.7	1753	6 ts e	44.7787
11044	Gto.	MANDUJANO, STA. CRUZ J.R	20.75	-101.033	2066	6 ts e	33.8388
11052	Gto.	PERICOS, SALAMANCA	20.533	-101.117	1772	6 ts e	29.4657
11096	Gto.	SALAMANCA, SALAMANCA	20.567	-101.183	1722	6 ts e	21.6672
28002	Tamps.	AHUALULCO, GOMEZ FARIAS	22.95	-99.133	100	7 ts e	
28063	Tamps.	MIQUIHUANA, MIQUIHUANA	23.583	-99.783	1892	7 ts e	96.8709
28011	Tamps.	CALLEJONES, OCAMPO (DGE)	22.867	-99.5	480	7 ts e	38.7496
28218	Tamps.	LA BOCA, VICTORIA	23.65	-99.1	340	7 ts e	77.9967
28044	Tamps.	LA ENCANTADA, LLERA	23.383	-99.083	375	7 ts e	48.4722
28219	Tamps.	EL ENCINO, LLERA	23.133	-99.117	160	7 ts e	20.4373
23154	Q.R.	AGUA BLANCA, OTHON P. BLA	18.009	-88.823	94	0 cs e	
23001	Q.R.	ALVARO OBREGON, OTHON P.	18.293	-88.642	52	0 cs e	36.9603
23015	Q.R.	LOS POZOS, OTHON P. BLANC	18.55	-88.483	20	0 cs e	70.1319
23016	Q.R.	NICOLAS BRAVO, OTHON P. B	18.458	-88.932	123	0 cs e	51.2939
23018	Q.R.	PUCTE, OTHON P. BLANCO	18.232	-88.666	65	0 cs e	29.8688
23049	Q.R.	INIA-CAECHET, OTHON P. BL	18.572	-88.459	41	0 cs e	73.5395
30056	Ver.	EL TEJAR, MEDELLIN	19.083	-96.15	10	1 cs e	
30048	Ver.	EL COPITAL, MEDELLIN	18.95	-96.2	57	1 cs e	15.7128
30137	Ver.	PUENTE NACIONAL	19.333	-96.4	110	1 cs e	38.2775
30104	Ver.	MATA ANONA, SOLEDAD D.	19.133	-96.283	250	1 cs e	15.0563
30136	Ver.	PUENTE JULA, PASO OVEJAS	19.233	-96.25	13	1 cs e	19.7331
30192	Ver.	VERACRUZ, VERACRUZ	19.2	-96.133	2	1 cs e	13.1465
4072	Camp.	SIHO-CHAC, CHAMPOTON	19.506	-90.584	15	2 cs e	
4053	Camp.	SANTA CRISTINA, CAMPECHE	19.815	-90.381	10	2 cs e	40.4482
4029	Camp.	SABANCUY, CARMEN	18.973	-91.176	5	2 cs e	85.9755
4059	Camp.	TINUN, TENABO	19.961	-90.228	50	2 cs e	62.9040
4068	Camp.	CHINA I.N.I.P., CAMPECHE	19.673	-90.474	10	2 cs e	21.8790
4084	Camp.	TIXMUCUY, CAMPECHE	19.55	-90.65	0	2 cs e	8.4817
23031	Q.R.	X-PICHIL, FELIPE C. PUERT	19.694	-88.376	34	3 cs e	
23003	Q.R.	FELIPE CARRILLO PUERTO, F	19.576	-88.045	22	3 cs e	37.1070
23013	Q.R.	LA PRESUMIDA, JOSE Ma. MO	19.802	-88.754	46	3 cs e	41.3886
23024	Q.R.	TIHOSUCO, FELIPE C. PUERT	20.193	-88.366	63	3 cs e	55.5583
23150	Q.R.	SABAN, JOSE Ma. MORELOS	20.032	-88.542	63	3 cs e	41.4460
23152	Q.R.	ANDRES QUINTANA ROO, FELI	19.162	-88.104	25	3 cs e	65.7466
21081	Pue.	SANTA CRUZ COYOTEPEC,	19.017	-97.567	2510	0 ts tp	
21007	Pue.	AHUATEPEC, TECALI DE H	18.854	-97.921	2057	0 ts tp	41.4564
21080	Pue.	SN.SALVADOR EL SECO, DGE	19.133	-97.65	2425	0 ts tp	15.5884
29011	Tlax.	HUAMANTLA, HUAMANTLA	19.316	-97.911	2485	0 ts tp	49.1550
30100	Ver.	MALTRATA, MALTRATA (DGE)	18.817	-97.267	1695	0 ts tp	38.6489
29160	Tlax.	ALTZAYANCA, ALTZAYANCA	19.428	-97.799	2410	0 ts tp	51.8455
13099	Hgo.	METEPEC, METEPEC	20.233	-98.317	225	2 ts tp	
13096	Hgo.	ATOTONILCO, ATOTONILCO G	20.283	-98.667	1500	2 ts tp	36.9731
13029	Hgo.	SAN LORENZO ZAYULA,	19.967	-98.35	2350	2 ts tp	29.8113
13100	Hgo.	OMITLAN, OMITLAN	20.167	-98.633	400	2 ts tp	33.8210



Clave	Estado	Nombre	Latitud	Longitud	Altitud	Tipo	distancia(km)
13116	Hgo.	EL SEMBO, HUASCA OCAMPO	20.15	-98.567	200	2 ts tp	27.7056
21209	Pue.	LOS HUMEROS,CHIGNAUTLA	20	-98.2	1200	2 ts tp	28.6761
13116	Hgo.	EL SEMBO, HUASCA OCAMPO	20.15	-98.567	200	3 ts tp	
13099	Hgo.	METEPEC, METEPEC	20.233	-98.317	225	3 ts tp	27.7056
13100	Hgo.	OMITLAN, OMITLAN	20.167	-98.633	400	3 ts tp	7.1519
13112	Hgo.	EL CHICO, M. EL CHICO	20.25	-98.75	0	3 ts tp	22.1232
13096	Hgo.	ATOTONILCO, ATOTONILCO G	20.283	-98.667	1500	3 ts tp	18.1197
13029	Hgo.	SAN LORENZO ZAYULA,	19.967	-98.35	2350	3 ts tp	30.4939
13151	Hgo.	EL ENCINO, STGO.DE ANAYA	20.433	-98.917	2400	4 ts tp	
13116	Hgo.	EL SEMBO, HUASCA OCAMPO	20.15	-98.567	200	4 ts tp	48.2485
13096	Hgo.	ATOTONILCO, ATOTONILCO G	20.283	-98.667	1500	4 ts tp	30.9772
13100	Hgo.	OMITLAN, OMITLAN	20.167	-98.633	400	4 ts tp	41.9046
13112	Hgo.	EL CHICO, M. EL CHICO	20.25	-98.75	0	4 ts tp	26.8111
13128	Hgo.	SAN AGUSTIN TLAXIAXCA	20.1	-98.883	2355	4 ts tp	37.2390
11028	Gto.	IRAPUATO, IRAPUATO (SMN)	20.683	-101.35	1724	5 ts tp	
11001	Gto.	ABASOLO, ABASOLO	20.45	-101.533	1760	5 ts tp	32.1952
11007	Gto.	CALDERONES, GUANAJUATO	21	-101.2	1900	5 ts tp	38.5848
11044	Gto.	MANDUJANO, STA. CRUZ J.R	20.75	-101.033	2066	5 ts tp	33.8388
11052	Gto.	PERICOS, SALAMANCA	20.533	-101.117	1772	5 ts tp	29.4657
11096	Gto.	SALAMANCA, SALAMANCA	20.567	-101.183	1722	5 ts tp	21.6672
1010	Ags.	LA TINAJA , SAN JOSE DE	22.164	-102.554	2425	6 ts tp	
1091	Ags.	LOS ALISOS, CALVILLO	21.742	-102.716	2040	6 ts tp	49.8657
14026	Jal.	CASA LLANTA, COLOTLAN	22.05	-103.367	1775	6 ts tp	84.8040
14101	Jal.	MICHOACANEJO,TEOCALTICHE	21.55	-102.583	1890	6 ts tp	68.4158
14175	Jal.	COLOTLAN, COLOTLAN (DGE)	22.117	-103.267	1735	6 ts tp	73.7041
1047	Ags.	AGUA ZARCA, JESUS MARIA	21.974	-102.584	2300	6 ts tp	21.3759
3166	B.C.S.	PABELLON, COMONDU	25	-112.067	40	0 mss tp	
3027	B.C.S.	LA POZA GRANDE, COMONDU	25.75	-112.033	25	0 mss tp	83.5596
3033	B.C.S.	LAS CRUCES, COMONDU	25.35	-111.75	40	0 mss tp	50.3780
3045	B.C.S.	PUERTO SAN CARLOS (DGE)	24.783	-112.1	10	0 mss tp	24.3851
3063	B.C.S.	SANTO DOMINGO, COMONDU	25.483	-111.917	18	0 mss tp	55.8484
3124	B.C.S.	RAMADITAS, COMONDU	25.117	-111.95	30	0 mss tp	17.5738
5036	Coah.	SAN PEDRO, SAN PEDRO	25.757	-102.996	1100	1 mss tp	
5027	Coah.	PRESA CUIJE, MATAMOROS	25.697	-103.34	1120	1 mss tp	35.1385
5041	Coah.	VIESCA, VIESCA (DGE)	25.319	-102.808	1100	1 mss tp	52.2869
5139	Coah.	EMILIANO ZAPATA,E.ZAPATA	25.489	-102.942	1110	1 mss tp	30.3220
5180	Coah.	FRANCISO I. MADERO	25.769	-103.277	1100	1 mss tp	28.2031
10108	Dgo.	CIUDAD LERDO (DGE)	25.5	-103.367	1140	1 mss tp	46.9577
23154	Q.R.	AGUA BLANCA,OTHON P. BLA	18.009	-88.823	94	0 cs tp	
23001	Q.R.	ALVARO OBREGON,OTHON P.	18.293	-88.642	52	0 cs tp	36.9603
23015	Q.R.	LOS POZOS,OTHON P. BLANC	18.55	-88.483	20	0 cs tp	70.1319
23016	Q.R.	NICOLAS BRAVO,OTHON P. B	18.458	-88.932	123	0 cs tp	51.2939
23018	Q.R.	PUCTE,OTHON P. BLANCO	18.232	-88.666	65	0 cs tp	29.8688
23049	Q.R.	INIA-CAECHET,OTHON P. BL	18.572	-88.459	41	0 cs tp	73.5395
30035	Ver.	CUATOTOLAPAN	18.133	-95.3	14	1 cs tp	
30103	Ver.	MATA DE LIMONES, ISLA	18.133	-95.533	50	1 cs tp	24.6493
30143	Ver.	JUAN RODRIGUEZ CLARA	18	-95.4	148	1 cs tp	18.1990



Clave	Estado	Nombre	Latitud	Longitud	Altitud	Tipo	distancia(km)
30147	Ver.	S.JUAN EVANGELISTA (DGE)	17.9	-95.133	29	1 cs tp	31.3893
30148	Ver.	SAN JUANILLO, ACAYUCAN	18.317	-95.217	75	1 cs tp	22.2837
30152	Ver.	GARRO CA. S.NICOLAS,ISLA	18.25	-95.533	20	1 cs tp	27.8714
17018	Mor.	TICUMAN, TLALTIZAPAN	18.767	-99.117	1130	2 cs tp	
17013	Mor.	TEMILPA, TLALTIZAPAN	18.7	-99.1	1000	2 cs tp	7.6707
17014	Mor.	TEMIXCO, TEMIXCO	18.85	-99.233	1060	2 cs tp	15.3226
17024	Mor.	YAUTEPEC, YAUTEPEC	18.867	-99.083	1150	2 cs tp	11.6942
17026	Mor.	ZACATEPEC A.C.A.E.	18.65	-99.183	1226	2 cs tp	14.7669
17071	Mor.	PROGRESO, JIOTEPEC	18.85	-99.15	1400	2 cs tp	9.8722
17005	Mor.	CUAUTLA, CUAUTLA (DGE)	18.8	-98.95	1309	3 cs tp	
17003	Mor.	CUAUTLA, CUAUTLA (SMN)	18.817	-98.967	1303	3 cs tp	2.6058
17007	Mor.	HUAJINTLAN, AMACUZAC	18.633	-98.967	1049	3 cs tp	18.6766
17024	Mor.	YAUTEPEC, YAUTEPEC	18.867	-99.083	1150	3 cs tp	15.8741
17054	Mor.	MOYOTEPEC,VILLA DE AYALA	18.717	-98.983	1165	3 cs tp	9.8726
17060	Mor.	ALPANOCAN, TETELA DEL V.	18.717	-98.883	2200	3 cs tp	11.6294
30056	Ver.	EL TEJAR, MEDELLIN	19.083	-96.15	10	4 cs tp	
30048	Ver.	EL COPITAL, MEDELLIN	18.95	-96.2	57	4 cs tp	15.7128
30093	Ver.	LOMA FINA,PASO DE OVEJAS	19.167	-96.3	30	4 cs tp	18.3393
30104	Ver.	MATA ANONA, SOLEDAD D.	19.133	-96.283	250	4 cs tp	15.0563
30136	Ver.	PUENTE JULA, PASO OVEJAS	19.233	-96.25	13	4 cs tp	19.7331
30192	Ver.	VERACRUZ, VERACRUZ	19.2	-96.133	2	4 cs tp	13.1465
4072	Camp.	SIHO-CHAC, CHAMPOTON	19.506	-90.584	15	5 cs tp	
4012	Camp.	HOOL, CHAMPOTON	19.513	-90.411	25	5 cs tp	18.1693
4038	Camp.	CAMPECHE, CAMPECHE	19.838	-90.544	5	5 cs tp	37.1952
4041	Camp.	CHAMPOTON,CHAMPOTON DGE	19.362	-90.72	2	5 cs tp	21.4660
4068	Camp.	CHINA I.N.I.P.,CAMPECHE	19.673	-90.474	10	5 cs tp	21.8790
4084	Camp.	TIXMUCUY, CAMPECHE	19.55	-90.65	0	5 cs tp	8.4817
24020	S.L.P.	EL FISCAL,S.V.TANCUAYALA	21.731	-98.592	30	6 cs tp	
24036	S.L.P.	LAS ADJUNTAS, TAMUIN	21.983	-98.717	25	6 cs tp	30.8825
24113	S.L.P.	SAN VICENTE	21.683	-98.617	32	6 cs tp	5.9361
24133	S.L.P.	LA GLORIA, TANLAJAS	21.817	-98.783	2100	6 cs tp	21.9435
24139	S.L.P.	TAMUIN, TAMUIN	22	-98.6	22	6 cs tp	29.9563
24149	S.L.P.	SANTA ELENA, TAMUIN	21.883	-98.85	30	6 cs tp	31.5807
21081	Pue.	SANTA CRUZ COYOTEPEC,	19.017	-97.567	2510	0p ts tp	
21007	Pue.	AHUATEPEC,TECALI DE H	18.854	-97.921	2057	0p ts tp	41.4564
21072	Pue.	SAN DIEGO TEXMELUCAN,	19.017	-97.45	2595	0p ts tp	12.3135
21080	Pue.	SN.SALVADOR EL SECO, DGE	19.133	-97.65	2425	0p ts tp	15.5884
29011	Tlax.	HUAMANTLA, HUAMANTLA	19.316	-97.911	2485	0p ts tp	49.1550
30100	Ver.	MALTRATA, MALTRATA (DGE)	18.817	-97.267	1695	0p ts tp	38.6489
13099	Hgo.	METEPEC, METEPEC	20.233	-98.317	225	2p ts tp	
13096	Hgo.	ATOTONILCO, ATOTONILCO G	20.283	-98.667	1500	2p ts tp	36.9731
13029	Hgo.	SAN LORENZO ZAYULA,	19.967	-98.35	2350	2p ts tp	29.8113
13100	Hgo.	OMITLAN, OMITLAN	20.167	-98.633	400	2p ts tp	33.8210
13116	Hgo.	EL SEMBO, HUASCA OCAMPO	20.15	-98.567	200	2p ts tp	27.7056
21209	Pue.	LOS HUMEROS,CHIGNAUTLA	20	-98.2	1200	2p ts tp	28.6761
13151	Hgo.	EL ENCINO, STGO.DE ANAYA	20.433	-98.917	2400	4p ts tp	
13116	Hgo.	EL SEMBO, HUASCA OCAMPO	20.15	-98.567	200	4p ts tp	48.2485



Clave	Estado	Nombre	Latitud	Longitud	Altitud	Tipo	distancia(km)
13096	Hgo.	ATOTONILCO, ATOTONILCO G	20.283	-98.667	1500	4p ts tp	30.9772
13100	Hgo.	OMITLAN, OMITLAN	20.167	-98.633	400	4p ts tp	41.9046
13112	Hgo.	EL CHICO, M. EL CHICO	20.25	-98.75	0	4p ts tp	26.8111
13128	Hgo.	SAN AGUSTIN TLAXIAXCA	20.1	-98.883	2355	4p ts tp	37.2390
11028	Gto.	IRAPUATO, IRAPUATO (SMN)	20.683	-101.35	1724	5p ts tp	
11001	Gto.	ABASOLO, ABASOLO	20.45	-101.533	1760	5p ts tp	32.1952
11007	Gto.	CALDERONES, GUANAJUATO	21	-101.2	1900	5p ts tp	38.5848
11044	Gto.	MANDUJANO, STA. CRUZ J.R	20.75	-101.033	2066	5p ts tp	33.8388
11052	Gto.	PERICOS, SALAMANCA	20.533	-101.117	1772	5p ts tp	29.4657
11096	Gto.	SALAMANCA, SALAMANCA	20.567	-101.183	1722	5p ts tp	21.6672
1010	Ags.	LA TINAJA , SAN JOSE DE	22.164	-102.554	2425	6p ts tp	
1091	Ags.	LOS ALISOS, CALVILLO	21.742	-102.716	2040	6p ts tp	49.8657
14026	Jal.	CASA LLANTA, COLOTLAN	22.05	-103.367	1775	6p ts tp	84.8040
14101	Jal.	MICHOACANEJO,TEOCALTICHE	21.55	-102.583	1890	6p ts tp	68.4158
14175	Jal.	COLOTLAN, COLOTLAN (DGE)	22.117	-103.267	1735	6p ts tp	73.7041
1047	Ags.	AGUA ZARCA, JESUS MARIA	21.974	-102.584	2300	6p ts tp	21.3759
3166	B.C.S.	PABELLON, COMONDU	25	-112.067	40	0p mss tp	
3027	B.C.S.	LA POZA GRANDE, COMONDU	25.75	-112.033	25	0p mss tp	83.5596
3033	B.C.S.	LAS CRUCES, COMONDU	25.35	-111.75	40	0p mss tp	50.3780
3045	B.C.S.	PUERTO SAN CARLOS (DGE)	24.783	-112.1	10	0p mss tp	24.3851
3063	B.C.S.	SANTO DOMINGO, COMONDU	25.483	-111.917	18	0p mss tp	55.8484
3124	B.C.S.	RAMADITAS, COMONDU	25.117	-111.95	30	0p mss tp	17.5738
5036	Coah.	SAN PEDRO, SAN PEDRO	25.757	-102.996	1100	1p mss tp	
5027	Coah.	PRESA CUIJE, MATAMOROS	25.697	-103.34	1120	1p mss tp	35.1385
5041	Coah.	VIESCA, VIESCA (DGE)	25.319	-102.808	1100	1p mss tp	52.2869
5139	Coah.	EMILIANO ZAPATA,E.ZAPATA	25.489	-102.942	1110	1p mss tp	30.3220
5180	Coah.	FRANCISO I. MADERO	25.769	-103.277	1100	1p mss tp	28.2031
10108	Dgo.	CIUDAD LERDO (DGE)	25.5	-103.367	1140	1p mss tp	46.9577
23154	Q.R.	AGUA BLANCA,OTHON P. BLA	18.009	-88.823	94	0p cs tp	
23001	Q.R.	ALVARO OBREGON,OTHON P.	18.293	-88.642	52	0p cs tp	36.9603
23015	Q.R.	LOS POZOS,OTHON P. BLANC	18.55	-88.483	20	0p cs tp	70.1319
23016	Q.R.	NICOLAS BRAVO,OTHON P. B	18.458	-88.932	123	0p cs tp	51.2939
23018	Q.R.	PUCTE,OTHON P. BLANCO	18.232	-88.666	65	0p cs tp	29.8688
23049	Q.R.	INIA-CAECHET,OTHON P. BL	18.572	-88.459	41	0p cs tp	73.5395
30035	Ver.	CUATOTOLAPAN	18.133	-95.3	14	1p cs tp	
30103	Ver.	MATA DE LIMONES, ISLA	18.133	-95.533	50	1p cs tp	24.6493
30143	Ver.	JUAN RODRIGUEZ CLARA	18	-95.4	148	1p cs tp	18.1990
30147	Ver.	S.JUAN EVANGELISTA (DGE)	17.9	-95.133	29	1p cs tp	31.3893
30148	Ver.	SAN JUANILLO, ACAYUCAN	18.317	-95.217	75	1p cs tp	22.2837
30152	Ver.	GARRO CA. S.NICOLAS,ISLA	18.25	-95.533	20	1p cs tp	27.8714
17018	Mor.	TICUMAN, TLALTIZAPAN	18.767	-99.117	1130	2p cs tp	
17013	Mor.	TEMILPA, TLALTIZAPAN	18.7	-99.1	1000	2p cs tp	7.6707
17014	Mor.	TEMIXCO, TEMIXCO	18.85	-99.233	1060	2p cs tp	15.3226
17024	Mor.	YAUTEPEC, YAUTEPEC	18.867	-99.083	1150	2p cs tp	11.6942
17026	Mor.	ZACATEPEC A.C.A.E.	18.65	-99.183	1226	2p cs tp	14.7669
17071	Mor.	PROGRESO, JIOTEPEC	18.85	-99.15	1400	2p cs tp	9.8722
17005	Mor.	CUAUTLA, CUAUTLA (DGE)	18.8	-98.95	1309	3p cs tp	



Clave	Estado	Nombre	Latitud	Longitud	Altitud	Tipo	distancia(km)
17003	Mor.	CUAUTLA, CUAUTLA (SMN)	18.817	-98.967	1303	3p cs tp	2.6058
17007	Mor.	HUAJINTLAN, AMACUZAC	18.633	-98.967	1049	3p cs tp	18.6766
17024	Mor.	YAUTEPEC, YAUTEPEC	18.867	-99.083	1150	3p cs tp	15.8741
17054	Mor.	MOYOTEPEC, VILLA DE AYALA	18.717	-98.983	1165	3p cs tp	9.8726
17060	Mor.	ALPANOCA, TETELA DEL V.	18.717	-98.883	2200	3p cs tp	11.6294
30056	Ver.	EL TEJAR, MEDELLIN	19.083	-96.15	10	4p cs tp	
30048	Ver.	EL COPITAL, MEDELLIN	18.95	-96.2	57	4p cs tp	15.7128
30093	Ver.	LOMA FINA, PASO DE OVEJAS	19.167	-96.3	30	4p cs tp	18.3393
30104	Ver.	MATA ANONA, SOLEDAD D.	19.133	-96.283	250	4p cs tp	15.0563
30136	Ver.	PUENTE JULIA, PASO OVEJAS	19.233	-96.25	13	4p cs tp	19.7331
30192	Ver.	VERACRUZ, VERACRUZ	19.2	-96.133	2	4p cs tp	13.1465
4072	Camp.	SIHO-CHAC, CHAMPOTON	19.506	-90.584	15	5p cs tp	
4012	Camp.	HOOL, CHAMPOTON	19.513	-90.411	25	5p cs tp	18.1693
4038	Camp.	CAMPECHE, CAMPECHE	19.838	-90.544	5	5p cs tp	37.1952
4041	Camp.	CHAMPOTON, CHAMPOTON DGE	19.362	-90.72	2	5p cs tp	21.4660
4068	Camp.	CHINA I.N.I.P., CAMPECHE	19.673	-90.474	10	5p cs tp	21.8790
4084	Camp.	TIXMUCUY, CAMPECHE	19.55	-90.65	0	5p cs tp	8.4817
24020	S.L.P.	EL FISCAL, S.V. TANCUAYALA	21.731	-98.592	30	6p cs tp	
24036	S.L.P.	LAS ADJUNTAS, TAMUIN	21.983	-98.717	25	6p cs tp	30.8825
24113	S.L.P.	SAN VICENTE	21.683	-98.617	32	6p cs tp	5.9361
24133	S.L.P.	LA GLORIA, TANLAJAS	21.817	-98.783	2100	6p cs tp	21.9435
24139	S.L.P.	TAMUIN, TAMUIN	22	-98.6	22	6p cs tp	29.9563
24149	S.L.P.	SANTA ELENA, TAMUIN	21.883	-98.85	30	6p cs tp	31.5807

9.2. Comparación de parámetros e índices estadísticos

9.2.1. 15% de datos faltantes

9.2.1.1. Por variable climática

- Temperatura

Cuadro 36. Resultados de indicadores y parámetros estadísticos para la variable climática temperatura y 15% de datos faltantes

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	5.4981	2.3448	1.8532	0.9641	14.3231	2.1856
RN	3.2209	1.7947	1.4042	0.9808	10.9628	1.7745
TW	2.7058	1.6449	1.1827	0.9836	10.0479	1.6247
UWS5	4.5641	2.1364	1.6579	0.9686	13.0499	1.8182
UWS4	4.7169	2.1718	1.6852	0.9676	13.2665	1.8528
UWS3	5.0476	2.2467	1.7389	0.9658	13.7237	1.9840
UWS2	5.3752	2.3184	1.8263	0.9652	14.1620	2.1863
UWS1	6.4109	2.5320	1.9353	0.9599	15.4664	2.4530

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 37. Análisis de regresión lineal simple para temperatura y 15% de datos faltantes

Modelo	m	b	r ²
RD	0.8702	2.3053	0.8708
RN	0.9722	0.2217	0.9273
TW	0.9575	0.7487	0.9365
UWS5	0.8311	3.0255	0.8988
UWS4	0.8301	3.0649	0.8951
UWS3	0.8412	2.8528	0.8842
UWS2	0.8798	2.0576	0.8731
UWS1	0.9012	1.5110	0.8516

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

- Evaporación

Cuadro 38. Resultados de los parámetros e índices estadísticos para la variable climática evaporación y 15% de datos faltantes

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	1256.7076	35.4501	25.9035	0.8606	23.9102	30.1295
RN	1527.9571	39.0891	27.4370	0.8292	26.3647	33.3503
TW	1076.7861	32.8144	24.2694	0.8882	22.1325	30.2198
UWS5	1056.8871	32.5098	23.9323	0.8817	21.9271	27.2143
UWS4	1092.7818	33.0573	24.3649	0.8774	22.2963	27.6880
UWS3	1122.3544	33.5016	24.7091	0.8734	22.5960	27.8711
UWS2	1262.2901	35.5287	25.8807	0.8591	23.9633	29.9607
UWS1	1262.5163	35.5319	26.5638	0.8579	23.9654	29.8724

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 39. Análisis de regresión lineal simple para evaporación y 15% de datos faltantes

Modelo	m	b	r ²
RD	0.6792	39.0210	0.5825
RN	0.6258	49.1730	0.4915
TW	0.7607	31.9610	0.6350
UWS5	0.6933	37.5000	0.6405
UWS4	0.6861	38.6990	0.6277
UWS3	0.6801	39.0220	0.6205
UWS2	0.6732	39.6070	0.5809
UWS1	0.6654	41.2540	0.5767

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

- Precipitación

Cuadro 40. Resultados de indicadores y parámetros estadísticos para la variable climática precipitación y 15% de datos faltantes

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	1914.2701	43.7524	23.8986	0.9275	66.1322	40.7588
RN	7731.2390	87.9275	52.8954	0.6601	135.9955	74.4824
TW	2648.9913	51.4684	31.2977	0.8991	77.7949	47.7928
UWS5	1898.6796	43.5738	24.2422	0.9271	65.0962	40.5968
UWS4	2067.0442	45.4648	24.1967	0.9179	67.9211	41.3165
UWS3	2078.9785	45.5958	23.8595	0.9169	68.1169	41.2179

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
UWS2	1985.2312	44.5559	23.8202	0.9240	66.5634	41.6029
UWS1	2543.0133	50.4283	26.4648	0.9089	75.3363	48.7547

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 41. Análisis de regresión lineal simple para precipitación y 15% de datos faltantes

Modelo	m	b	r ²
RD	0.8129	11.9500	0.7493
RN	0.4459	39.5690	0.2054
TW	0.7790	18.0680	0.6662
UWS5	0.8127	11.7810	0.7482
UWS4	0.7771	13.3170	0.7240
UWS3	0.7716	13.2670	0.7221
UWS2	0.8111	12.0520	0.7381
UWS1	0.8487	13.2440	0.6920

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

9.2.1.2. Por región climática

A. Cálido Subhúmedo

a. Temperatura

Cuadro 42. Resultados de indicadores y parámetros estadísticos para la variable climática temperatura y 15% de datos faltantes para la región climática cálido subhúmedo

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	4.0067	2.0017	1.6092	0.9254	9.3826	1.7463
RN	3.7021	1.9241	1.4663	0.9446	9.0189	1.9105
TW	1.6932	1.3012	0.9940	0.9735	6.0993	1.2954
UWS5	2.8494	1.6880	1.3291	0.9474	7.9123	1.4445
UWS4	3.0372	1.7428	1.3875	0.9440	8.1689	1.5047
UWS3	3.3715	1.8362	1.4739	0.9385	8.6068	1.6263
UWS2	3.9372	1.9842	1.5880	0.9276	9.3009	1.7576
UWS1	4.3275	2.0803	1.5788	0.9188	9.7510	1.8084

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 43. Análisis de regresión lineal simple para temperatura y 15% de datos faltantes para la región climática cálido subhúmedo

Modelo	m	b	r ²
RD	0.7555	5.3693	0.7535
RN	0.9651	0.4793	0.8065
TW	0.9595	0.9432	0.8996
UWS5	0.7862	4.3456	0.8287
UWS4	0.7830	4.4410	0.8156
UWS3	0.7863	4.4294	0.7924
UWS2	0.7668	5.0183	0.7566
UWS1	0.7427	5.3390	0.7337

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

b. Precipitación

Cuadro 44. Resultados de indicadores y parámetros estadísticos para la variable climática precipitación y 15% de datos faltantes para la región climática cálido subhúmedo

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	2372.8023	48.7114	27.6894	0.9370	51.8931	44.0218
RN	14296.2149	119.5668	77.9533	0.6107	125.5587	100.5797
TW	4095.6063	63.9969	43.2448	0.8880	68.1770	57.2916
UWS5	2512.3910	50.1238	28.2471	0.9331	53.3977	45.4547
UWS4	2865.7187	53.5324	30.0204	0.9210	57.0290	47.0498
UWS3	2981.7458	54.6054	30.0644	0.9176	58.1720	47.9855
UWS2	2465.7288	49.6561	28.0423	0.9346	52.8995	45.0501
UWS1	2610.4042	51.0921	28.5185	0.9347	54.4293	48.7799

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 45. Análisis de regresión lineal simple para precipitación y 15% de datos faltantes para la región climática cálido subhúmedo

Modelo	m	b	r ²
RD	0.8039	12.4500	0.7864
RN	0.3741	65.5570	0.1325
TW	0.7257	30.8370	0.6391
UWS5	0.7982	14.1370	0.7729
UWS4	0.7610	15.2610	0.7428
UWS3	0.7557	15.7400	0.7325
UWS2	0.8023	13.0330	0.7778

Modelo	m	b	r ²
UWS1	0.8477	12.6250	0.7693

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

c. Evaporación

Cuadro 46. Resultados de indicadores y parámetros estadísticos para la variable climática evaporación y 15% de datos faltantes para la región climática cálido subhúmedo

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	816.8086	28.5799	22.3506	0.8110	23.3293	25.0017
RN	637.9972	25.2586	20.7884	0.8549	20.6182	22.5716
TW	782.4113	27.9716	20.0951	0.8532	22.8328	27.5304
UWS5	842.3174	29.0227	23.3959	0.7843	23.6908	23.6121
UWS4	901.8324	30.0305	24.2536	0.7712	24.5134	24.6168
UWS3	867.2992	29.4499	23.8079	0.7834	24.0395	24.4181
UWS2	840.1605	28.9855	22.9539	0.8027	23.6604	25.0146
UWS1	1010.2981	31.7852	25.6148	0.7656	25.9457	27.3694

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 47. Análisis de regresión lineal simple para evaporación y 15% de datos faltantes para la región climática cálido subhúmedo

Modelo	m	b	r ²
RD	0.6085	47.5190	0.4445
RN	0.6790	40.5070	0.5500
TW	0.8295	20.8810	0.5508
UWS5	0.5287	56.8700	0.4038
UWS4	0.5796	57.7830	0.3756
UWS3	0.5391	55.7110	0.3970
UWS2	0.5887	49.0460	0.4279
UWS1	0.5516	52.0110	0.3542

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

B. Muy seco semicálido

a. Temperatura

Cuadro 48. Resultados de indicadores y parámetros estadísticos para la variable climática temperatura y 15% de datos faltantes para la región climática muy seco semicálido

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	3.8068	1.9511	1.5498	0.9617	11.8394	1.7921

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RN	3.0214	1.7382	1.4846	0.9719	10.5476	1.7276
TW	3.8181	1.9540	1.4365	0.9627	11.8571	1.8587
UWS5	3.9469	1.9867	1.5719	0.9578	12.0554	1.6888
UWS4	4.2706	2.0665	1.6257	0.9536	12.5400	1.7324
UWS3	4.8934	2.2121	1.7354	0.9460	13.4232	1.8174
UWS2	3.6406	1.9080	1.5048	0.9639	11.5782	1.7870
UWS1	3.0807	1.7552	1.2481	0.9736	10.6507	1.7856

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 49. Análisis de regresión lineal simple para temperatura y 15% de datos faltantes para la región climática muy seco semicálido

Modelo	m	b	r ²
RD	0.8496	2.1926	0.8672
RN	0.9316	0.9982	0.8942
TW	0.8676	2.2134	0.8636
UWS5	0.7967	3.5598	0.8661
UWS4	0.7805	3.7964	0.8550
UWS3	0.7606	4.2757	0.8358
UWS2	0.8643	2.0111	0.8718
UWS1	1.0082	-0.0430	0.9026

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

b. Precipitación

Cuadro 50. Resultados de indicadores y parámetros estadísticos para la variable climática precipitación y 15% de datos faltantes para la región climática cálido muy seco semicálido

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	153.7269	12.3987	7.2022	0.9023	133.3987	12.2132
RN	669.3012	25.8709	20.1855	0.2726	278.3475	14.1412
TW	88.2476	9.3940	6.6925	0.9270	101.0714	8.9819
UWS5	148.0216	12.1664	6.8559	0.8801	130.8998	11.5762
UWS4	162.3046	12.7399	7.1681	0.8735	137.0699	12.2755
UWS3	174.6576	13.2158	7.3681	0.8638	142.1904	12.7505
UWS2	174.4733	13.2088	7.5293	0.8893	142.1154	13.0520
UWS1	341.9487	18.4919	9.2944	0.7460	198.9560	17.7175

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 51. Análisis de regresión lineal simple para precipitación y 15% de datos faltantes para la región climática muy seco semicálido

Modelo	m	b	r ²
RD	1.0620	2.4013	0.7146
RN	-0.0210	21.5040	0.0007
TW	0.8593	3.3685	0.7519
UWS5	0.7976	4.3124	0.6111
UWS4	0.8184	4.2738	0.5954
UWS3	0.8042	4.2452	0.5684
UWS2	1.0428	2.7403	0.6788
UWS1	0.6799	5.6211	0.3278

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

c. Evaporación

Cuadro 52. Resultados de indicadores y parámetros estadísticos para la variable climática evaporación y 15% de datos faltantes para la región climática muy seco semicálido

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	569.2834	23.8597	16.3396	0.9585	14.0556	20.5909
RN	1645.0040	40.5586	29.3994	0.8921	23.8929	37.4431
TW	573.0813	23.9391	19.1148	0.9621	14.1024	23.9305
UWS5	417.8027	20.4402	14.0563	0.9696	12.0412	16.9562
UWS4	467.0965	21.6124	15.1337	0.9660	12.7318	17.6750
UWS3	671.9443	25.9219	17.1589	0.9495	15.2705	20.1504
UWS2	563.1857	23.7315	15.2791	0.9586	13.9801	18.6806
UWS1	992.3196	31.5011	20.9259	0.9256	18.5571	21.0726

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 53. Análisis de regresión lineal simple para evaporación 15% de datos faltantes para la región climática muy seco semicálido

Modelo	m	b	r ²
RD	0.8393	19.5040	0.8704
RN	0.8178	18.0550	0.6586
TW	0.9320	13.4430	0.8598
UWS5	0.8528	17.4710	0.9109

Modelo	m	b	r ²
UWS4	0.8509	16.3870	0.9036
UWS3	0.8066	21.1930	0.8663
UWS2	0.8340	17.2390	0.8896
UWS1	0.7806	17.8250	0.8473

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

C. Templado subhúmedo

a. Temperatura

Cuadro 54. Resultados de indicadores y parámetros estadísticos para la variable climática temperatura y 15% de datos faltantes para la región climática templado subhúmedo

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	7.8020	2.7932	2.2389	0.8246	26.4898	2.2175
RN	2.7262	1.6511	1.3049	0.9520	15.6586	1.6005
TW	3.5164	1.8752	1.3181	0.9336	17.7838	1.7400
UWS5	6.7703	2.6020	2.0702	0.8503	24.6764	1.9543
UWS4	6.8254	2.6125	2.0523	0.8525	24.7765	1.9911
UWS3	7.0545	2.6560	2.0493	0.8460	25.1888	2.0817
UWS2	7.6310	2.7624	2.2115	0.8368	26.1980	2.3141
UWS1	9.9516	3.1546	2.5802	0.8048	29.9173	2.8268

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 55. Análisis de regresión lineal simple para temperatura y 15% de datos faltantes para la región climática templado subhúmedo

Modelo	m	b	r ²
RD	0.5649	4.9534	0.4939
RN	0.9021	0.8001	0.8269
TW	0.8119	2.0111	0.7661
UWS5	0.6058	4.9872	0.5910
UWS4	0.6195	4.8838	0.5928
UWS3	0.6046	4.8418	0.5592
UWS2	0.6083	4.3815	0.5096
UWS1	0.6289	3.7916	0.4268

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

b. Precipitación

Cuadro 56. Resultados de indicadores y parámetros estadísticos para la variable climática precipitación y 15% de datos faltantes para la región climática templado subhúmedo

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	1976.5422	44.4583	25.2701	0.8286	88.7195	43.2577
RN	2678.0429	51.7498	35.9099	0.6312	103.2701	36.1255
TW	1648.0278	40.5959	24.4138	0.8529	81.0118	39.6132
UWS5	1703.2598	41.2706	23.7101	0.8477	82.3581	39.9059
UWS4	1753.1302	41.8704	24.0264	0.8427	83.5551	40.4837
UWS3	1621.5755	40.2688	22.9641	0.8499	80.3590	38.6563
UWS2	2084.7578	45.6591	25.6478	0.8188	91.1158	44.3888
UWS1	3434.7023	58.6063	32.4378	0.7413	116.9528	57.1236

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 57. Análisis de regresión lineal simple para precipitación y 15% de datos faltantes para la región climática templado subhúmedo

Modelo	m	b	r2
RD	0.8158	15.1790	0.5095
RN	0.3056	32.5830	0.1729
TW	0.8144	11.0080	0.5524
UWS5	0.8132	15.1880	0.5481
UWS4	0.8049	15.1580	0.5359
UWS3	0.7864	14.9590	0.5472
UWS2	0.8036	15.5510	0.4891
UWS1	0.8181	20.2830	0.3747

Dónde: m= pendiente de regresión; b= ordenada al origen; r2= coeficiente de determinación

c. Evaporación

Cuadro 58. Resultados de indicadores y parámetros estadísticos para la variable climática evaporación y 15% de datos faltantes para la región climática templado subhúmedo

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	1704.4854	41.2854	30.6663	0.8050	26.3230	34.0362
RN	2003.0637	44.7556	30.6755	0.7581	28.5355	36.2554
TW	1388.9159	37.2682	28.1274	0.8342	23.7616	31.5788
UWS5	1362.0938	36.9066	27.0606	0.8430	23.5311	30.2703
UWS4	1380.6403	37.1570	27.0647	0.8393	23.6907	30.5513

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
UWS3	1396.7889	37.3736	27.3812	0.8400	23.8289	30.9801
UWS2	1703.2512	41.2705	30.5821	0.8042	26.3134	34.2339
UWS1	1483.8399	38.5206	28.7169	0.8261	24.5602	32.7169

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 59. Análisis de regresión lineal simple para evaporación y 15% de datos faltantes para la región climática templado subhúmedo

Modelo	m	b	r ²
RD	0.6195	46.2710	0.4669
RN	0.5105	68.0550	0.3439
TW	0.6331	50.4670	0.5151
UWS5	0.6603	41.8140	0.5592
UWS4	0.6572	41.5890	0.5548
UWS3	0.6576	41.8430	0.5436
UWS2	0.6155	47.7740	0.4608
UWS1	0.6310	49.5960	0.4958

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

9.2.2. 25% de datos faltantes

9.2.2.1. Por variable climática

- Temperatura

Cuadro 60. Resultados de indicadores y parámetros estadísticos para la variable climática temperatura y 25% de datos faltantes

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	5.8756	2.4240	1.9177	0.9606	14.8878	2.1819
RN	6.7881	2.6054	1.8754	0.9571	16.0022	2.4950
TW	2.9596	1.7204	1.2249	0.9816	10.5664	1.6371
UWS4	5.2094	2.2824	1.7799	0.9631	14.0184	1.8441
UWS3	5.5666	2.3594	1.8334	0.9610	14.4911	1.9704
UWS2	5.7802	2.4042	1.8868	0.9615	14.7665	2.1801
UWS1	6.5889	2.5669	1.9216	0.9581	15.7657	2.4376

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS4=US Weather Service 4 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 61. Análisis de regresión lineal simple para temperatura y 25% de datos faltantes

Modelo	m	b	r ²
RD	0.8373	2..6545	0.8629
RN	0.8838	1.9697	0.8429
TW	0.9251	1.4395	0.9318
UWS4	0.7958	3.5266	0.8884
UWS3	0.8020	3.3801	0.8763
UWS2	0.8437	2.5436	0.8649
UWS1	0.8774	1.8410	0.8471

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

- Evaporación

Cuadro 62. Resultados de los parámetros e índices estadísticos para la variable climática evaporación y 25% de datos faltantes

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	1128.1711	33.5883	24.2321	0.8620	23.9156	30.9416
RN	1116.7264	33.4175	24.6740	0.8497	23.7940	28.7557
TW	796.1452	28.2160	20.9249	0.9130	20.0905	27.6805
UWS2	1091.2727	33.0344	23.6965	0.8658	23.5213	30.3705
UWS1	1046.1190	32.3438	23.9133	0.8734	23.0295	30.0471

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 63. Análisis de regresión lineal simple para evaporación y 25% de datos faltantes

Modelo	m	b	r ²
RD	0.7249	37.4040	0.5570
RN	0.6418	50.4180	0.5330
TW	0.8951	17.4880	0.7055
UWS2	0.7269	36.9690	0.5675
UWS1	0.7507	32.8550	0.5885

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

- Precipitación

Cuadro 64. Resultados de indicadores y parámetros estadísticos para la variable climática precipitación y 25% de datos faltantes

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	2109.6953	45.9314	25.0870	0.9379	66.5085	44.8152
RN	8728.8641	93.4284	57.3290	0.5879	135.2840	62.6418
TW	2782.6112	52.7505	29.6946	0.9128	76.3825	49.8039
UWS4	1983.3384	44.5347	24.4704	0.9409	64.4861	43.2511
UWS3	1971.4368	44.4009	24.2777	0.9416	64.2923	43.2512
UWS2	2131.6202	46.1695	25.2186	0.9378	66.8532	45.1933
UWS1	2632.0150	51.3032	27.8457	0.9312	74.2868	50.6909

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 65. Análisis de regresión lineal simple para precipitación y 25% de datos faltantes

Modelo	m	b	r ²
RD	0.8933	9.9993	0.7792
RN	0.2628	48.5940	0.1352
TW	0.8136	14.4140	0.7033
UWS4	0.8869	10.1320	0.7888
UWS3	0.8929	9.6474	0.7910
UWS2	0.9011	9.8317	0.7793
UWS1	0.9969	8.6146	0.7745

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

9.2.2.2. Por región climática

A. Cálido Subhúmedo

a. Temperatura

Cuadro 66. Resultados de indicadores y parámetros estadísticos para la variable climática temperatura y 25% de datos faltantes para la región climática cálido subhúmedo

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	4.9747	2.2304	1.7876	0.9084	10.5051	1.9643
RN	4.8239	2.1963	1.5280	0.9216	10.3446	2.0982
TW	1.7242	1.3131	1.0005	0.9726	6.1845	1.2861
UWS4	4.0121	2.0030	1.5647	0.9207	9.4341	1.5677
UWS3	4.4437	2.1080	1.6819	0.9125	9.9286	1.6916

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
UWS2	4.9812	2.2319	1.7860	0.9018	10.5119	1.8455
UWS1	4.6950	2.1668	1.6521	0.9075	10.2054	1.7751

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 67. Análisis de regresión lineal simple para temperatura y 25% de datos faltantes para la región climática cálido subhúmedo

Modelo	m	b	r ²
RD	0.7419	5.2320	0.6992
RN	0.8448	3.5381	0.7254
TW	0.9395	1.4307	0.8969
UWS4	0.7071	5.7944	0.7682
UWS3	0.7011	5.9533	0.7368
UWS2	0.6894	6.4087	0.6946
UWS1	0.6966	6.1719	0.7150

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

b. Precipitación

Cuadro 68. Resultados de indicadores y parámetros estadísticos para la variable climática precipitación y 25% de datos faltantes para la región climática cálido subhúmedo

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	3077.7892	55.4778	32.2561	0.9369	56.9658	53.6961
RN	14400.1929	120.0008	77.2756	0.5306	123.2194	76.9075
TW	4694.1156	68.5136	41.2522	0.8966	70.3512	63.3769
UWS4	3091.5103	55.6014	32.3210	0.9362	57.0926	53.6144
UWS3	3110.8058	55.7746	32.4157	0.9365	57.2705	54.0814
UWS2	3080.5263	55.5025	32.4527	0.9375	56.9911	53.9816
UWS1	3202.9508	56.5946	33.4775	0.9418	58.1126	56.2285

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 69. Análisis de regresión lineal simple para precipitación y 25% de datos faltantes para la región climática cálido subhúmedo

Modelo	m	b	r ²
RD	0.8723	13.3640	0.7755
RN	0.1931	71.2440	0.0762
TW	0.7681	24.4340	0.6578

Modelo	m	b	r ²
UWS4	0.8659	14.2720	0.7735
UWS3	0.8754	13.7340	0.7742
UWS2	0.8816	13.2070	0.7774
UWS1	0.9846	9.1332	0.8006

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

c. Evaporación

Cuadro 70. Resultados de indicadores y parámetros estadísticos para la variable climática evaporación y 25% de datos faltantes para la región climática cálido subhúmedo

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	1304.0542	36.1117	26.1164	0.6831	31.3164	30.5890
RN	605.0372	24.5975	19.2060	0.8231	21.3312	21.0793
TW	739.9226	27.2015	19.8376	0.8407	23.5894	26.9441
UWS4	1134.9439	33.6889	24.0807	0.7059	29.2154	28.2579
UWS3	1171.9467	34.2337	24.7511	0.6994	29.6878	28.8311
UWS2	1212.8883	34.8265	25.3544	0.6957	30.2020	29.4163
UWS1	1121.5992	33.4903	23.8828	0.7200	29.0431	29.5951

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 71. Análisis de regresión lineal simple para evaporación y 25% de datos faltantes para la región climática cálido subhúmedo

Modelo	m	b	r ²
RD	0.5170	67.0830	0.2357
RN	0.6098	46.7600	0.4747
TW	0.8586	17.1160	0.5229
UWS4	0.5145	65.6610	0.2636
UWS3	0.5086	66.2660	0.2514
UWS2	0.5163	66.1420	0.2495
UWS1	0.5599	57.7280	0.2787

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

B. Muy seco semicálido

a. Temperatura

Cuadro 72. Resultados de indicadores y parámetros estadísticos para la variable climática temperatura y 25% de datos faltantes para la región climática muy seco semicálido

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	4.0923	2.0230	1.6398	0.9605	12.2340	1.7329
RN	12.1747	3.4892	2.6304	0.8640	21.1013	2.5538
TW	3.9449	1.9862	1.5253	0.9636	12.0116	1.8161
UWS5	3.6405	1.9080	1.5150	0.9741	11.5389	1.3505
UWS4	4.0230	2.0057	1.5902	0.9710	12.1299	1.4201
UWS3	4.6544	2.1574	1.6833	0.9660	13.0470	1.5456
UWS2	3.1679	1.7799	1.3674	0.9783	10.7638	1.2002
UWS1	2.6093	1.6153	1.1929	0.9841	9.7688	1.1403

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 73. Análisis de regresión lineal simple para temperatura y 15% de datos faltantes para la región climática muy seco semicálido

Modelo	m	b	r ²
RD	0.8157	2.6962	0.8732
RN	0.5955	5.7792	0.6282
TW	0.8477	2.3933	0.8713
UWS5	0.7790	3.8206	0.8635
UWS4	0.7647	4.0223	0.8522
UWS3	0.7407	4.5761	0.8347
UWS2	0.8314	2.4978	0.8776
UWS1	0.9936	0.0820	0.9092

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

b. Precipitación

Cuadro 74. Resultados de indicadores y parámetros estadísticos para la variable climática precipitación y 25% de datos faltantes para la región climática cálido muy seco semicálido

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	113.3755	10.6478	5.8230	0.8424	129.0990	9.8335
RN	1029.4234	32.0846	18.9554	0.3613	389.0094	27.8361
TW	135.6686	11.6477	7.3445	0.8011	141.2221	10.3032
UWS5	77.0813	8.7796	4.9359	0.8851	106.4480	7.7184

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
UWS4	81.9285	9.0514	5.1401	0.8811	109.7440	8.1111
UWS3	89.8332	9.4780	5.3134	0.8700	114.9162	8.5169
UWS2	111.1646	10.5435	5.8067	0.8469	127.8340	9.7822
UWS1	131.2548	11.4566	5.5856	0.8470	138.9059	11.2383

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 75. Análisis de regresión lineal simple para precipitación y 25% de datos faltantes para la región climática muy seco semicálido

Modelo	m	b	r ²
RD	0.7021	3.1052	0.5196
RN	0.3323	18.8970	0.0293
TW	0.6305	4.9322	0.4428
UWS5	0.7001	2.8837	0.6358
UWS4	0.7108	2.8821	0.6197
UWS3	0.6999	2.9481	0.5890
UWS2	0.7131	3.0914	0.5300
UWS1	0.8440	2.9343	0.5448

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

c. Evaporación

Cuadro 76. Resultados de indicadores y parámetros estadísticos para la variable climática evaporación y 25% de datos faltantes para la región climática muy seco semicálido

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	532.8360	23.0832	17.6160	0.9601	13.1226	20.6901
RN	1831.2621	42.7932	30.1275	0.8426	24.3275	33.6639
TW	513.1754	22.6534	17.8922	0.9644	12.8782	22.0333
UWS5	353.4803	18.8011	13.8893	0.9739	10.6882	16.4755
UWS4	389.6859	19.7405	14.9127	0.9710	11.2222	16.8653
UWS3	563.1402	23.7306	17.0433	0.9568	13.4906	19.1594
UWS2	488.9812	22.1129	15.3287	0.9635	12.5709	18.5387
UWS1	908.5040	30.1414	20.0167	0.9285	17.1350	19.7131

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 77. Análisis de regresión lineal simple para evaporación y 25% de datos faltantes para la región climática muy seco semicálido

Modelo	m	b	r ²
RD	0.8393	23.7530	0.8666
RN	0.5805	81.8120	0.5400
TW	0.9097	19.4590	0.8706
UWS5	0.8636	19.5130	0.9156
UWS4	0.8543	19.8290	0.9101
UWS3	0.8158	23.5310	0.8774
UWS2	0.8439	19.4650	0.8911
UWS1	0.7699	22.2810	0.8576

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

C. Templado subhúmedo

a. Temperatura

Cuadro 78. Resultados de indicadores y parámetros estadísticos para la variable climática temperatura y 25% de datos faltantes para la región climática templado subhúmedo

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	7.5209	2.7424	2.1621	0.8206	26.3152	2.0799
RN	7.2840	2.6989	2.0290	0.8442	25.8974	2.2919
TW	4.0726	2.0181	1.3865	0.9173	19.3645	1.7332
UWS5	6.7264	2.5935	2.0632	0.8438	24.8865	1.8509
UWS4	6.7490	2.5979	2.0437	0.8468	24.9282	1.8854
UWS3	6.9233	2.6312	2.0094	0.8389	25.2480	1.9434
UWS2	7.3469	2.7105	2.1116	0.8324	26.0090	2.1664
UWS1	9.9958	3.1616	2.4674	0.7911	30.3374	2.7412

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 79. Análisis de regresión lineal simple para temperatura y 25% de datos faltantes para la región climática templado subhúmedo

Modelo	m	b	r ²
RD	0.5431	5.1816	0.5009
RN	0.6278	4.0925	0.5248
TW	0.7502	3.0263	0.7339
UWS5	0.5880	5.2132	0.5977
UWS4	0.6026	5.0968	0.6006
UWS3	0.5776	5.1552	0.5653

Modelo	m	b	r ²
UWS2	0.5789	4.7004	0.5125
UWS1	0.5836	4.2746	0.4002

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

b. Precipitación

Cuadro 80. Resultados de indicadores y parámetros estadísticos para la variable climática precipitación y 25% de datos faltantes para la región climática templado subhúmedo

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	1552.8916	39.4067	22.7558	0.8996	73.3443	38.9684
RN	3868.7802	62.1995	44.7532	0.5580	115.7665	39.0911
TW	1165.2820	34.1362	22.4539	0.9196	63.5347	33.6884
UWS5	1177.9259	34.3209	21.0232	0.9205	63.8785	33.7555
UWS4	1192.4618	34.5320	21.2117	0.9192	64.2714	34.0050
UWS3	1128.9617	33.6000	20.4700	0.9221	62.5368	33.0050
UWS2	1611.3338	40.1414	22.8556	0.8960	74.7117	39.7138
UWS1	2833.0089	53.2260	28.8653	0.8445	99.0649	51.9986

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 81. Análisis de regresión lineal simple para precipitación y 15% de datos faltantes para la región climática templado subhúmedo

Modelo	m	b	r ²
RD	0.9379	9.1434	0.6817
RN	0.1988	41.4840	0.0872
TW	0.8961	6.5526	0.7234
UWS5	0.9210	9.3194	0.7334
UWS4	0.9170	9.0604	0.7289
UWS3	0.9032	9.0715	0.7346
UWS2	0.9359	9.1989	0.6724
UWS1	1.0378	10.1300	0.5955

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

c. Evaporación

Cuadro 82. Resultados de indicadores y parámetros estadísticos para la variable climática evaporación y 25% de datos faltantes para la región climática templado subhúmedo

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RD	1141.0637	33.7796	24.4155	0.8539	22.8161	31.4673

Modelo	MSE	RMSE	MAE	IW	DR%	SSE
RN	1273.0183	35.6794	26.7599	0.7975	24.0992	29.4245
TW	882.1714	29.7014	22.1239	0.8892	20.0614	28.7407
UWS2	1136.5002	33.7120	24.3429	0.8546	22.7704	31.6623
UWS1	1029.1999	32.0811	24.6730	0.8733	21.6688	31.1271

RD= Razón de distancia; RN= Redes neuronales; TW= Transformada de wavelet; UWS5=US Weather Service 5 estaciones cercanas; MSE= Error cuadrático medio; RMSE= Raíz cuadrada del cuadrado medio del error; MAE= Error absoluto medio; IW= Índice de Wilmott; DR%= Diferencia relativa; SSE= Error estándar de estimación

Cuadro 83. Análisis de regresión lineal simple para evaporación y 25% de datos faltantes para la región climática templado subhúmedo

Modelo	m	b	r ²
RD	0.7823	24.4230	0.5576
RN	0.5517	64.0370	0.4176
TW	0.8466	26.4120	0.6389
UWS2	0.7853	24.9290	0.5565
UWS1	0.8465	18.3930	0.6013

Dónde: m= pendiente de regresión; b= ordenada al origen; r²= coeficiente de determinación

9.3. Coeficientes de Correlación

Cuadro 84. Coeficientes de correlación simple para evaporación y 15 % de datos faltantes

EVAPORACIÓN	TEMPERATURA	PRECIPITACIÓN
0 mss e	0.8007	-0.0194
1 mss e	0.8716	0.0189
1 ts e	0.5325	0.0697
2 ts e	0.2874	0.0412
3 ts e	0.2121	-0.2026
4 ts e	0.3238	-0.1690
5 ts e	0.0764	-0.4091
6 ts e	0.5174	0.0306
7 ts e	0.7035	0.1716
0 cs e	0.5470	-0.1889
1 cs e	0.7298	0.2247
2 cs e	0.4506	0.1655
3 cs e	0.5375	0.0741

Correlación arriba de 0.7

Correlación menor de 0.7

Cuadro 85. Coeficientes de correlación simple para precipitación y 15% de datos faltantes

PRECIPITACIÓN	TEMPERATURA	EVAPORACIÓN
0p ts tp	0.6683	0.2028
2p ts tp	0.5446	0.0266
4p ts tp	0.3738	NA
5p ts tp	0.5843	-0.0391
6p ts tp	0.3439	-0.2517
0p mss tp	0.0990	NA
1p mss tp	0.4060	0.1860
0p cs tp	0.2254	-0.1863
1p cs tp	0.5155	-0.0599
2p cs tp	0.6331	NA
3p cs tp	0.5367	-0.2427
4p cs tp	0.4568	0.2239
5p cs tp	0.3366	0.1677
6p cs tp	0.4757	-0.0196

Correlación arriba de 0.7

Correlación menor de 0.7

Cuadro 86. Coeficientes de correlación simple para temperatura y 15% de datos faltantes

TEMPERATURA	PRECIPITACIÓN	EVAPORACIÓN
0 ts tp	0.6617	0.2817
2 ts tp	0.5363	0.1339
3 ts tp	0.4175	0.5547
4 ts tp	0.3383	NA

TEMPERATURA	PRECIPITACIÓN	EVAPORACIÓN
5 ts tp	0.5962	0.5330
6 ts tp	0.3447	0.4934
0 mss tp	0.1009	NA
1 mss tp	0.4090	0.8825
0 cs tp	0.2252	0.5405
1 cs tp	0.5212	0.6388
2 cs tp	0.6222	NA
3 cs tp	0.5327	0.3089
4 cs tp	0.4169	0.7215
5 cs tp	0.3513	0.4969
6 cs tp	0.4750	0.7190

Correlación arriba de 0.7

Correlación menor de 0.7

Cuadro 87. Coeficientes de correlación simple para evaporación y 25 % de datos faltantes

EVAPORACIÓN	TEMPERATURA	PRECIPITACIÓN
0 mss e	0.7950	-0.0125
1 mss e	0.8843	0.0410
1 ts e	0.5170	0.0242
2 ts e	0.2500	0.0718
3 ts e	0.2383	-0.2008
4 ts e	0.2803	-0.2138
5 ts e	0.0431	-0.4081
6 ts e	0.5328	0.0806
7 ts e	0.6930	0.1551
0 cs e	0.5362	-0.1853
1 cs e	0.7260	0.1947
2 cs e	0.4556	0.1928
3 cs e	0.5773	0.0882

Correlación arriba de 0.7

Correlación menor de 0.7

Cuadro 88. Coeficientes de correlación simple para precipitación y 25% de datos faltantes

PRECIPITACIÓN	TEMPERATURA	EVAPORACIÓN
0p ts tp	0.6648	0.1627
2p ts tp	0.5150	0.0614
4p ts tp	0.3645	NA
5p ts tp	0.6136	0.0068
6p ts tp	0.3987	-0.2084
0p mss tp	0.1379	NA
1p mss tp	0.3739	0.1485
0p cs tp	0.2351	-0.1927
1p cs tp	0.4897	-0.0640

PRECIPITACIÓN	TEMPERATURA	EVAPORACIÓN
2p cs tp	0.6484	NA
3p cs tp	0.5614	-0.2649
4p cs tp	0.4348	0.2065
5p cs tp	0.3497	0.1757
6p cs tp	0.4753	-0.0493

Correlación arriba de 0.7

Correlación menor de 0.7

Cuadro 89. Coeficientes de correlación simple para temperatura y 25% de datos faltantes

TEMPERATURA	PRECIPITACIÓN	EVAPORACIÓN
0 ts tp	0.6594	0.2619
2 ts tp	0.5309	0.1009
3 ts tp	0.3954	0.5511
4 ts tp	0.4083	NA
5 ts tp	0.5560	0.5632
6 ts tp	0.3420	0.5181
0 mss tp	-0.0154	NA
1 mss tp	0.4153	0.8844
0 cs tp	0.2262	0.5788
1 cs tp	0.4897	0.6734
2 cs tp	0.6740	NA
3 cs tp	0.5053	0.4174
4 cs tp	0.4324	0.7224
5 cs tp	0.3803	0.4482
6 cs tp	0.4827	0.7320

Cuadro 90. Coeficientes de correlación múltiple para cada variable climática para 15% de datos faltantes

EVAPORACIÓN	MÚLTIPLE
0 mss e	0.8587
1 mss e	0.8964
1 ts e	0.6465
2 ts e	0.3148
3 ts e	0.5415
4 ts e	0.5198
5 ts e	0.6196
6 ts e	0.6287
7 ts e	0.7469
0 cs e	0.6288
1 cs e	0.7384
2 cs e	0.4513
3 cs e	0.5581

PRECIPITACIÓN	MÚLTIPLE
0p ts tp	0.6243
2p ts tp	0.5435
4p ts tp	NA
5p ts tp	0.7214
6p ts tp	0.6000
0p mss tp	NA
1p mss tp	0.4980
0p cs tp	0.4163
1p cs tp	0.7184
2p cs tp	NA
3p cs tp	0.6765
4p cs tp	0.4804
5p cs tp	0.3368
6p cs tp	0.6783

TEMPERATURA	MÚLTIPLE
0 ts tp	0.6459
2 ts tp	0.5449
3 ts tp	0.8002
4 ts tp	NA
5 ts tp	0.8102
6 ts tp	0.6958
0 mss tp	NA
1 mss tp	0.9108
0 cs tp	0.6275
1 cs tp	0.8486
2 cs tp	NA
3 cs tp	0.6761
4 cs tp	0.7738
5 cs tp	0.5647
6 cs tp	0.8651

Cuadro 91. Coeficientes de correlación múltiple para cada variable climática para 25% de datos faltantes

EVAPORACIÓN	MÚLTIPLE
0 mss e	0.8473
1 mss e	0.9055
1 ts e	0.6506
2 ts e	0.2608
3 ts e	0.5657
4 ts e	0.5132
5 ts e	0.5834
6 ts e	0.6014
7 ts e	0.7372
0 cs e	0.6134
1 cs e	0.7344
2 cs e	0.4575
3 cs e	0.5923

PRECIPITACIÓN	MÚLTIPLE
0p ts tp	0.6240
2p ts tp	0.5090
4p ts tp	NA
5p ts tp	0.7241
6p ts tp	0.6002
0p mss tp	NA
1p mss tp	0.4875
0p cs tp	0.4288
1p cs tp	0.7171
2p cs tp	NA
3p cs tp	0.6775
4p cs tp	0.4616
5p cs tp	0.3503
6p cs tp	0.6879

TEMPERATURA	MÚLTIPLE
0 ts tp	0.6295
2 ts tp	0.5323
3 ts tp	0.8040
4 ts tp	NA
5 ts tp	0.8213
6 ts tp	0.7140
0 mss tp	NA
1 mss tp	0.9119
0 cs tp	0.6637
1 cs tp	0.8564
2 cs tp	NA
3 cs tp	0.7182
4 cs tp	0.7786
5 cs tp	0.5396
6 cs tp	0.8686



9.4. Código MATLAB para el modelo de redes neuronales

```
function ENTRENAR_Y_PROBAR_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton1 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
cd('C:\Users\OMAR\Documents\MATLAB\RN\Evaporación');
file = uigetfile('*.prn');
if ~isequal(file, 0)
A=load(file);
end
A=A';
file = uigetfile('*.prn');
if ~isequal(file, 0)
B=load(file);
end
B=B';
Ne=str2num(get(handles.edit1, 'string'));
p=A;
t=B;
[pn,minp,maxp,tn,mint,maxt]=premnmx(p,t);    %normalizar para tansig
[R,Q]=size(pn);
iitst=2:10:Q;
iival=10:10:Q;
iitr=[1:10:Q 3:10:Q 4:10:Q 5:10:Q 6:10:Q 7:10:Q 8:10:Q 9:10:Q];
val.P=pn(:,iival);
val.T=tn(:,iival);
test.P=pn(:,iitst);
test.T=tn(:,iitst);
ptr=pn(:,iitr);
ttr=tn(:,iitr);
apoyo=0;
coeficiente=0.90;
R2=0;
while R2<coeficiente
net=newff(minmax(ptr),[Ne,1,1],{'tansig','purelin','purelin'},'trainlm');
[net,tr]=train(net,ptr,ttr,[],[],val,test);
an=sim(net,pn);
sal=postmnmx(an,mint,maxt);
R=corrcoef(sal,B)
R2=R(2,1)
apoyo=apoyo+1;
if apoyo==30
coeficiente=coeficiente-0.5;
end
```



```
if apoyo==60
coeficiente=coeficiente-0.5;
end
if apoyo==90
coeficiente=coeficiente-0.5;
end
if apoyo==120
coeficiente=coeficiente-0.5;
end
end
set(handles.text3,'string',R2);
file = uigetfile('*.prn');
if ~isequal(file, 0)
C=load(file);
end
C=C';
pnewn=trammx(C,minp,maxp); % PARA TANSIG
ANEW=sim(net,pnewn);
SALIDA=postmnmx(ANEW, mint, maxt); %salida tansig
SALIDA=SALIDA';
xlswrite('SALIDA.xls', SALIDA);
cd('C:\Users\OMAR\Documents\MATLAB\RN');
```