

# **CORRECCIÓN ESTADÍSTICA (MOS) AL PRONÓSTICO ATMOSFÉRICO DEL MODELO WRF TH1201.1 Informe Final**

SUBCOORDINACIÓN  
HIDROMETEOROLOGÍA

Elaboró:

DR. EFRAÍN MATEOS FARFAN  
DR. JULIO SERGIO SANTANA SEPULVEDA  
DR. JOSÉ ANTONIO SALINAS PRIETO  
MÓNICA ROMANA TABOADA MEDINA  
DR. RICARDO PRIETO GONZALEZ

Con la colaboración de:

ING. AURA IVETTE GALICIA RIVERA  
ING. JORGE ARMANDO IBARRA ROMERO  
ING. LUIS ALBERTO SÁNCHEZ HERNÁNDEZ  
M.C. OBED PÉREZ CORTÉS

# Índice general

<b>1. RESUMEN</b>	<b>5</b>
Resumen	5
<b>2. INTRODUCCIÓN</b>	<b>7</b>
2.1. Incertidumbre sobre la atmósfera . . . . .	7
<b>3. OBJETIVO</b>	<b>9</b>
<b>4. ANTECEDENTES</b>	<b>10</b>
4.1. Modelos de post-procesamiento . . . . .	10
4.2. Perfect-Prog . . . . .	11
4.3. MOS . . . . .	13
<b>5. METODOLOGÍA</b>	<b>14</b>
5.1. Curva de ajuste . . . . .	14
5.2. Regresión . . . . .	15
5.2.1. Método de mínimos cuadrados . . . . .	16
5.2.2. Recta de mínimos cuadrados . . . . .	17
5.2.3. Regresión lineal múltiple . . . . .	17
5.2.4. Selección de variables independientes (predictores) . . . . .	18
5.3. Forward selection . . . . .	18
5.3.1. Regla de paro . . . . .	19

<b>6. RESULTADOS</b>	<b>20</b>
6.1. Puntos geográficos analizados . . . . .	20
6.2. Potenciales variables . . . . .	22
6.3. <i>Forward selection</i> . . . . .	24
6.3.1. Temperatura . . . . .	24
6.3.2. Precipitación . . . . .	30
<b>7. CONCLUSIONES</b>	<b>35</b>
<b>Bibliografía</b>	<b>36</b>
<b>A. Variables y coeficientes de las rectas de regresión</b>	<b>37</b>

# Índice de figuras

5.1. Relación lineal entre las variables $x$ e $y$ . . . . .	15
5.2. Relación no lineal entre las variables $x$ e $y$ . . . . .	15
5.3. Variables $x$ e $y$ no relacionadas. . . . .	16
5.4. Mínimos cuadrados. . . . .	16
6.1. Diagrama de dispersión de 3 variables del modelo WRF correspondientes a la estación INIFAP en 2009. . . . .	23
6.2. Variables incluidas en la ecuación de regresión de la estación INIFAP. . . . .	25
6.3. Coeficiente de correlación de Pearson de la recta de regresión lineal en la estación INIFAP. . . . .	26
6.4. logaritmo natural de MSE de la recta de regresión lineal en la estación INIFAP. . . . .	26
6.5. Perfil temporal de la temperatura observada y predicha por el modelo WRF y la recta de regresión en la estación CEIEPO. . . . .	28
6.6. Perfil temporal de la temperatura observada y predicha por el modelo WRF y la recta de regresión en la estación Tetela del Volcan. . . . .	29
6.7. Variables incluidas en la ecuación de regresión de la estación “El Hospital.” . . . .	30
6.8. Coeficiente de correlación de Pearson de la recta de regresión lineal en la estación “El Hospital.” . . . .	31
6.9. MSE de la recta de regresión lineal en la estación “El Hospital.” . . . .	31

# Índice de tablas

6.1. Nombre y ubicación de las 21 estaciones meteorológicas analizadas . . . . .	21
6.2. Variables físicas del modelo WRF que serán consideradas para conformar las rectas de regresión lineal múltiple. . . . .	22
6.3. Variables del modelo WRF que serán utilizadas para construir las rectas de regresión lineal. . . . .	24
6.4. Variables y coeficientes de las rectas de regresión lineal. . . . .	27
6.5. Coeficiente de correlación de Pearson y error cuadrático medio de las predicciones del modelo WRF y las rectas de regresión lineal. . . . .	32
6.6. Variables y coeficientes de las rectas de regresión lineal. . . . .	33
6.7. Coeficiente de correlación de Pearson de las rectas de regresión lineal de la variable precipitación. . . . .	34
A.1. Variables y coeficientes de las rectas de regresión (continuación). . . . .	38
A.2. Variables y coeficientes de las rectas de regresión (continuación). . . . .	39
A.3. Variables y coeficientes de las rectas de regresión (continuación). . . . .	40
A.4. Variables y coeficientes de las rectas de regresión (continuación). . . . .	41
A.5. Variables y coeficientes de las rectas de regresión (continuación). . . . .	42
A.6. Variables y coeficientes de las rectas de regresión (continuación). . . . .	43
A.7. Variables y coeficientes de las rectas de regresión (fin). . . . .	44

# Capítulo 1

## RESUMEN

Una de las herramientas más importantes en el estudio de la dinámica atmosférica son los modelos numéricos de predicción (MNP). Por medio de éstos es posible, entre otras cosas, predecir el comportamiento de variables meteorológicas de interés como la temperatura, precipitación entre otras.

Los MNP consisten en modelos matemáticos que representan los procesos físicos que ocurren en la atmósfera. Para dar solución al conjunto de ecuaciones planteadas por éstos, deben realizarse aproximaciones tanto físicas como matemáticas. Como resultado de este proceder las predicciones que se obtienen de los MNP difieren de lo que se observa.

En este trabajo se implementó la técnica MOS con la finalidad de mejorar el pronóstico de la temperatura del modelo Weather Research and Forecasting (WRF), en puntos geográficos del estado de Morelos.

Para ser más específicos, se emplearon datos del modelo WRF y datos observados, registrados por 22 estaciones Agro-meteorológicas de la Fundación Produce (COFUPRO) durante el periodo 2009-2011, con los que se construyeron rectas de regresión lineal múltiple para la variable temperatura.

A partir de los parámetros de las ecuaciones de regresión lineal, se estimó el valor de la temperatura para los días 1-8 de agosto de 2012 en los mismos puntos geográficos y se calculó el error cuadrático medio (MSE) de dicha estimación y los valores observados. En la estación CEIEPO el valor encontrado fue 4. Por otro lado se calculó el mismo parámetro, el MSE, con los valores de la temperatura predichos por el modelo WRF y los valores observados. En la estación citada el valor determinado fue 40. Al comparar el MSE del pronóstico

de las rectas de regresión lineal (4) y el pronóstico del modelo WRF (40) se verifica que el primero es menor comparado con el segundo. Además del MSE se calculó el coeficiente de correlación de Pearson (R) para las mismas parejas de datos. Se encontró que el valor de R generado por las rectas de regresión lineal es mayor que el generado por el modelo WRF (0.82 y 0.8 respectivamente).

# Capítulo 2

## INTRODUCCIÓN

Una herramienta matemática ampliamente utilizada en las ciencias atmosféricas es la estadística. En términos coloquiales, y aplicados a las ciencias atmosféricas, se dice que la estadística trata sobre la incertidumbre. Por medio de la estadística se evalúa y cuantifica la incertidumbre. Por otra parte, por medio de la estadística se hacen inferencias y pronósticos a pesar de la incertidumbre.

La estadística se divide en dos ramas, la estadística descriptiva e inferencial. La primera se emplea para organizar y presentar información de manera ordenada y objetiva mientras que la segunda consiste en métodos cuyo objetivo es extraer conclusiones sobre los procesos físicos que generan los datos observados.

### 2.1. Incertidumbre sobre la atmósfera

Una de las herramientas más importantes para el estudio de los procesos atmosféricos son los modelos numéricos de predicción (NMP) los cuales permiten comprender y predecir el comportamiento dinámico de variables como la temperatura, presión, entre otras, como función de la posición (longitud, latitud y altitud). Uno de los modelos más empleados es el denominado WRF (acrónimo de su nombre en inglés Weather research forecasting) [WRF, 2012]. A partir de variables como la temperatura, presión, entre otras, es posible hacer pronósticos sobre el clima de una región. Para que el modelo WRF opere es necesario proporcionarle condiciones iniciales de ciertas variables, estos valores se obtienen de mode-

los climatológicos globales. Los MNP consisten en modelos matemáticos que representan los procesos físicos que ocurren en la atmósfera. De manera que para construir los MNP es necesario conocer los procesos físicos involucrados, contar con las herramientas matemáticas que permitan resolver las ecuaciones planteadas y disponer de la infraestructura computacional para implementar de manera operativa los MNP.

En principio los MNP proveen pronósticos sin incertidumbre, pero no lo hacen debido a dos razones. La atmósfera terrestre es catalogada como un sistema complejo, esto significa dos cosas: primero, no se conocen expresiones analíticas que describan todas las interacciones que experimentan los componentes atmosféricos, segundo, la evolución temporal de las variables de interés está determinado por las condiciones iniciales de las mismas de manera no lineal. El primer punto mencionado significa que el modelo realiza una descripción aproximada de la dinámica atmosférica, y el segundo significa que si las condiciones iniciales no son las adecuadas entonces el MNP empleado puede proporcionar datos que no se apegan a la realidad; a este último fenómeno se conoce como caos dinámico. El caos dinámico implica que aun cuando se contara con un modelo basado en un formalismo teórico exacto las predicciones hechas por el modelo, a largo plazo, pudieran ser completamente erróneas.

Con la finalidad de mejorar los pronósticos de los MNP se han desarrollado procedimientos estadísticos denominados *modelos de post-procesamiento* que emplean los datos de salida de algún MNP. Estos modelos incluyen todos los formalismos que utilizan los datos de salida de algún MNP y se agrupan en dos categorías: perfect prog y MOS. Los modelos de post-procesamiento Las variables que proporciona el MNP pueden emplearse directamente, o bien se puede utilizar en combinación con datos de variables meteorológicas de otras fuentes.

Los *modelos de post-procesamiento* utilizan los datos provistos por algún MNP para generar pronósticos de variables meteorológicas de interés (como temperatura, precipitación, etc) en lugares específicos. En conjunto, todos los formalismos estadísticos forman parte importante en el proceso de pronóstico. Algunos de estos procedimientos estadísticos pueden reducir los sesgos inherentes a algún MNP.

# Capítulo 3

## OBJETIVO

Implementar una metodología estadística (*Model Output Statistics-MOS*) a los pronósticos meteorológicos del modelo WRF (*Weather Research and Forecasting*) a tres días en las variables de temperatura y precipitación, con la finalidad de mejorar sus predicciones.

Para alcanzar el objetivo general se deben cumplir los siguientes objetivos específicos:

- Plantear el formalismo matemático que permitirá construir las rectas de regresión lineal.
- Obtener datos históricos de la temperatura observada y datos de salida del modelo WRF.
- Validar las ecuaciones de regresión lineal empleando datos de un periodo diferente al empleado para construir las primeras.

# Capítulo 4

## ANTECEDENTES

### 4.1. Modelos de post-procesamiento

De manera general, algunas variables del tiempo meteorológico, como la visibilidad y tormentas eléctricas, no son predichas por los modelos numéricos de predicción (MNP) y no se pueden derivar directamente de los datos proporcionados por éstos, no obstante a partir de estos últimos pueden generarse pronósticos de las primeras. De la otra parte debido a la incertidumbre intrínseca de los resultados de los MNP se han propuesto metodologías cuyo objetivo es mejorar sus pronósticos.

De manera que para para coadyuvar a mejorar los pronósticos y dar cuenta de variables que no proporcionan los MNP se proponen ecuaciones que relacionan una variable de interés (variable independiente) con otras variables proporcionadas por el MNP (variables independientes).

Los formalismos que se emplean para el pronóstico estadístico del tiempo meteorológico se agrupan en dos categorías:

- Prosnóstico perfecto (perfect-prog).
- Modelo estadístico de salida (MOS)

Ambos enfoques son análogos a lo que es un pronosticador experimentado hace en la práctica. Un pronosticador establece relaciones entre las observaciones y los datos de algún

MNP, para luego interpretar los resultados del MNP con la finalidad de mejorar sus estimaciones. Por medio de los formalismos estadísticos se realiza la labor descrita empleando criterios matemáticos.

## 4.2. Perfect-Prog

Los pronósticos realizados por métodos estadísticos clásicos para escalas de tiempo de días por lo general no se emplean debido a que los modelos dinámicos ofrecen mayor precisión en sus estimaciones. Sin embargo, se emplean dos metodologías estadísticas que mejoran algunos aspectos de los pronósticos dinámicos, éstos emplean los resultados de los modelos numéricos. Ambas metodologías emplean un conjunto amplio de ecuaciones de regresión de manera semejante a las metodologías clásicas.

Las diferencias ente dichos enfoques y la estadística clásica radica en el rango de variables predictoras (variables independientes) disponibles. De manera adicional a los predictores convencionales tales como por ejemplo las observaciones meteorológicas, las predictores pueden ser las salidas de los modelos dinámicos.

Existen tres razones por las que las interpretaciones estadísticas de los datos de salida de los modelos dinámicos son útiles en el pronóstico de variables meteorológicas:

1. Existen diferencias importantes entre el mundo real y su representación en los modelos dinámicos las cuales tienen implicaciones importantes en los resultados que los modelos dinámicos generan. Los modelos dinámicos simplifican y consideran homogéneas condiciones superficiales al considerar al mundo con un arreglo de puntos de malla a los cuales pertenecen las salidas que estos generan. También, lugares y variables que son necesarios para hacer un pronóstico pueden no representarse de manera explícita. Sin embargo pueden establecerse relaciones estadísticas entre la información provista por los modelos y la cantidad de interés a fin de mitigar los efectos negativos de todas esas aproximaciones e “inexactitudes”.
2. Los modelos dinámicos no son completas ni verdaderas representaciones de la atmósfera, de manera particular a escalas pequeñas de tiempo y espacio y los modelos “inician” en estados que difieren del estado real de la atmósfera. Debido a estas razones su pronósticos son sujetos a errores. Si los errores del modelo son sistemáticos,

el pos-procesamiento puede compensarlos y corregir los sesgos.

3. Los modelos dinámicos son deterministas. Esto significa que aunque de antemano se sabe que el estado futuro del tiempo meteorológico es incierto, dadas las condiciones iniciales los modelos numéricos proporcionan un valor específico para todas las variables que considera. Al emplear la información generada por los modelos dinámicos junto con metodologías estadísticas es posible estimar la incertidumbre inherente de las predicciones de los primeros.

El primer enfoque estadístico desarrollado para aprovechar las ventajas de los modelos numéricos de predicción (MNP) se conoce como *perfect-prog* (abreviatura de *perfect prognosis*). Tal y como el nombre lo sugiere, la técnica *perfect-prog* no intenta corregir los errores del MNP sino que emplea los pronósticos generados por estos para estimar valores futuro de alguna otra variable meteorológica asumiendo que los pronósticos del modelo son perfectos.

Las ecuaciones de regresión de la técnica *perfect-prog* son semejantes a las que se desarrollan en los métodos estadísticos clásicos de predicción del tiempo meteorológico, esto es, se emplean ciertas variables independientes para determinar cierta variable dependiente. *Perfect-prog* emplea datos históricos para generar pronósticos. La principal diferencia entre *perfect-prog* y los métodos estadísticos clásicos de predicción del tiempo meteorológico es el tiempo de retardo. En las ecuaciones clásicas se incorpora el retardo temporal relacionando variables independientes disponibles antes de que el pronóstico deba ser emitido con aquellas variables dependientes que serán observadas en un futuro. Por su parte *perfect-prog* no incorpora tiempo de retardo sino que de manera simultánea emplea valores de la variable independiente junto con los de las variables independientes para construir las ecuaciones de regresión.

Pareciera ser poco útil la técnica *perfect-prog* para hacer pronósticos. Posiblemente la presión atmosférica del día de mañana sea una excelente variable independiente para predecir la temperatura pero este valor se conocerá mañana. Sin embargo, en la técnica *perfect-prog* los pronósticos del MNP se emplean como variables independientes para estimar el valor de una variable futura. De manera que el tiempo de retardo está implícito en el MNP. Variables no predichas por el modelo no pueden ser empleadas como variables independientes a menos que se conozcan de antemano.

## 4.3. MOS

Este es el segundo enfoque que incorpora los datos de los MNP a las metodologías estadísticas clásicas de predicción del tiempo; *MOS* es al acrónimo de *Model output statistics*. Esta técnica es preferida sobre *perfect-prog* porque tiene la capacidad de incluir de manera directa en las ecuaciones de regresión la influencia de características bien específicas de un MNP particular con diferentes tiempo de retardo a futuro.

Si bien *MOS* y *perfect-prog* emplean como variables independientes los resultados de los MNP, ambos enfoques aplican la información de manera diferente. *Perfect-prog* utiliza los resultados de los MNP sólo para hacer pronósticos mientras que *MOS* utiliza los resultados de los MNP para desarrollar y hacer los pronósticos. Suponga que hoy se requiere hacer un pronóstico para el día de mañana. Las ecuaciones generadas por *MOS* para la variable independiente de mañana emplea las variables independientes del MNP del día mañana. Los valores “reales” de dichas variables independientes aun son desconocidas pero el MNP puede estimar sus valores hoy. Para desarrollar las ecuaciones en la técnica *MOS* es necesario contar con datos históricos (del mismo periodo) observados y provistos por un MNP.

# Capítulo 5

## METODOLOGÍA

Se utilizó la técnica MOS, descrita en el capítulo anterior, para la corrección estadística del modelo WRF. Este método consiste en usar ecuaciones de regresión lineal múltiples para obtener una ecuación lineal que corrija el pronóstico de la temperatura del WRF en sitios específicos.

En las siguientes secciones se plantea el formalismo matemático que permite construir las rectas de regresión lineal múltiple.

### 5.1. Curva de ajuste

Muy a menudo, en la práctica, se encuentra que existe una relación entre dos (o más) variables, y se desea expresar esta relación en forma matemática determinando una ecuación que conecte las variables [3]. Un primer paso es la colección de datos indicando los valores correspondientes de las variables. Por ejemplo, si  $x$  e  $y$  denotan la estatura y peso de un adulto, respectivamente, entonces una muestra de  $n$  individuos resultaría en las estaturas  $x_1, x_2, \dots, x_n$  y los pesos correspondientes  $y_1, y_2, \dots, y_n$ . El paso siguiente es dibujar los puntos  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  en un sistema de coordenadas cartesianas. El conjunto resultante de puntos se denomina diagrama de dispersión.

Del diagrama de dispersión es posible visualizar una curva que se aproxime a los datos. Dicha curva se denomina *curva de aproximación*. En la figura 5.1, por ejemplo, se observa que los datos se aproximan bien por una recta y entonces se dice que existe una relación lineal entre las variables.

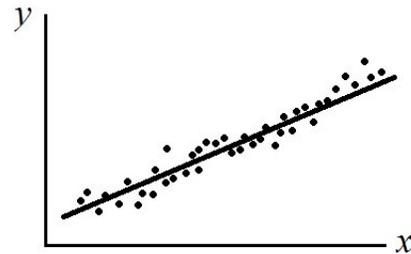


Figura 5.1: Relación lineal entre las variables  $x$  e  $y$ .

Sin embargo, en la figura 5.2 aunque existe una relación entre las variables ésta no se ajusta a una recta por lo que se denomina *relación no lineal*. Finalmente, en la figura 5.3 parece que no hay ninguna relación entre las variables.



Figura 5.2: Relación no lineal entre las variables  $x$  e  $y$ .

## 5.2. Regresión

Uno de los propósitos principales de la curva de ajuste es estimar una de las variables (la variable dependiente) de la otra (la variable independiente). El proceso de estimación se conoce como *regresión* [2].

Si la variable  $y$  se va a estimar a partir de la variable  $x$  por medio de alguna ecuación  $y = ax + b$  a ésta se denomina *ecuación de regresión de  $y$  sobre  $x$*  y a la curva correspondiente *curva de regresión de  $y$  sobre  $x$* .

El problema general de hallar ecuaciones de curvas de aproximación que se ajusten a conjuntos de datos se denomina *curva de ajuste*. En la práctica el tipo de ecuación se deduce frecuentemente del diagrama de dispersión. Así para la figura 5.1 podría emplearse una recta.

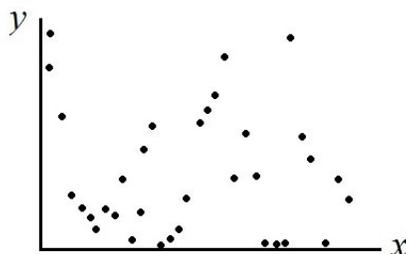


Figura 5.3: Variables  $x$  e  $y$  no relacionadas.

Generalmente, más de una curva de un tipo dado parece ajustar un conjunto de datos. Para evitar el juicio individual en la construcción de rectas, parábolas, u otras curvas de aproximación, es necesario obtener una definición de la “mejor recta de ajuste”; “mejor parábola de ajuste”, etc.

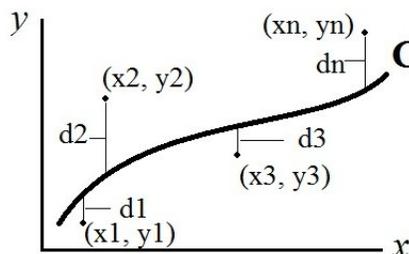


Figura 5.4: Mínimos cuadrados.

### 5.2.1. Método de mínimos cuadrados

Para establecer una posible definición de “mejor recta de ajuste” considérese la figura 5.4 en la cual los puntos de datos son  $(x_1, y_1), (x_2, y_2), \dots (x_n, y_n)$ . Para un valor dado de  $x$ , por ejemplo  $x_1$  habrá una diferencia entre el valor de  $y_1$  y el valor correspondiente determinado de la curva  $C$ . Denotamos esta diferencia por  $d_1$  (se dice que  $d_1$  es la desviación, error, o residuo). De manera análoga en correspondencia a los valores  $x_2, x_3, \dots x_n$  se obtienen las desviaciones  $d_2, d_3, \dots d_n$ .

Una medida que cuantifica “qué tan buena” es la curva  $C$  al conjunto de datos la suministra la cantidad  $d_1 + d_2 + \dots + d_n = S$ . Si la suma es pequeña el ajuste “es bueno” si es grande el ajuste “es malo”. Por tanto se adopta la siguiente definición: *de todas las curvas de aproximación de un conjunto de puntos de datos dados, la curva para la cual  $S$  sea mínimo*

es la mejor curva de ajuste.

Una curva con esta propiedad se dice que ajusta los datos en el sentido de mínimos cuadrados y se llama *curva de regresión de mínimos cuadrados*. Por lo tanto una recta con esta propiedad se llama *recta de mínimos cuadrados*, una parábola con esta propiedad se llama *parábola de mínimos cuadrados*, etc. Se acostumbra emplear la definición anterior cuando  $x$  es la variable independiente e  $y$  es la variable dependiente.

### 5.2.2. Recta de mínimos cuadrados

Empleando las definiciones anteriores es posible demostrar que la recta de mínimos cuadrados de aproximación al conjunto de datos  $(x_1, y_1), (x_2, y_2), \dots (x_3, y_3)$  tiene la expresión analítica análoga a la ecuación de la recta  $y = ax + b$  donde las constantes  $a$  y  $b$  se denominan *pendiente* e *intercepto*.

### 5.2.3. Regresión lineal múltiple

La regresión lineal múltiple es el caso más general de la regresión lineal [3]. Al igual que en el caso de la regresión lineal simple, existe una variable dependiente pero existen más de una variable independiente.

Sea  $K$  el número de variables independientes, entonces, la ecuación de regresión lineal se escribe como [2]:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots b_x.x_k \quad (5.1)$$

Cada variable independiente tiene un coeficiente análogo a la pendiente en el caso de la regresión lineal simple. Note que si  $k=1$  entonces se recupera la ecuación de la regresión lineal simple.

En el caso de la regresión múltiple, la ecuación de residuos ( $e_i$ ) se define de la misma manera que en el caso de la regresión simple [3]:

$$e_i = y_i - y(x_1, x_2 \dots x_k) \quad (5.2)$$

Los parámetros SSE (*sum of squared differences*), SST (*sum of squared total*), MSE (*mean squared error*) se calculan de manera idéntica al caso de la regresión lineal simple.

Por otro lado el coeficiente de determinación  $R^2$  se calcula de la manera usual pero ya no es el cuadrado del coeficiente de correlación de Pearson [3].

Note que si bien la ecuación de regresión lineal múltiple representa una recta en un espacio k-dimensional esto no significa que su perfil corresponda a una recta en un espacio bidimensional.

#### 5.2.4. Selección de variables independientes (predictores)

Por lo general, existen más potenciales variables independientes de las que pueden ser utilizadas en un procedimiento de predicción estadístico. El proceso de selección de las variables adecuadas no es trivial. Aparentemente el procedimiento consistiría en agregar variables independientes hasta que se consiga un buen acuerdo entre los datos predichos con los “reales”, sin embargo, no es así.

En general  $K=n-1$  variables independientes producirán un ajuste perfecto para cualquier variable para la cual existan  $n$  valores observados (aun cuando no exista ninguna relación entre las variables independientes y la dependiente).

Entonces para seleccionar de manera adecuada a las potenciales variables independientes se sugiere: Elegir como potenciales variables independientes aquellas que tienen significado físico razonable y reservar 1/4, 1/3 o bien 1/2 de los datos correspondiente a la variable dependiente a fin de verificar las predicciones de la recta de regresión múltiple tentativa.

### 5.3. Forward selection

Se asume que se cuenta con un conjunto de potenciales variables independientes para construir una recta de regresión lineal múltiple. Además dicho conjunto incluye solamente aquellas variables con relevancia física.

Se excluyen las variables independientes que están correlacionadas para evitar información redundante. Incluir variables que presenten correlación alta conlleva a estimaciones pobres de la variable dependiente.

En el argot de los pronósticos climatológicos, la selección de las variables independientes adecuadas se denomina *screening regresion* [3]. Uno de los procedimientos más comunes

se denomina *forward selection* y se describe a continuación:

1. Sean M potenciales variables independientes. El proceso inicia suponiendo  $K=0$ , es decir  $y = b_0$  donde  $b_0$  es, necesariamente, la media de la variable dependiente.
2. Sea ahora  $K=1$ . Esto significa que  $y = b_0 + b_1x_i$ . El uso del subíndice  $i$  en la variable independiente significa que se deberá realizar el análisis de ajuste lineal simple con cada una de las potenciales variables independientes; aquella que presente el mejor ajuste es seleccionada. En general  $b_0$  será diferente que aquella encontrada en el paso anterior.
3. Sea  $K=2$ . Esto significa que  $y = b_0 + b_1x_1 + b_2x_2$ ; igual que el paso anterior, se determina la mejor curva de ajuste al considerar el resto de las M-1 potenciales variables independientes.
4. Repetir el paso anterior.

Un procedimiento alternativo al *forward selection* se denomina *backward selection*, el procedimiento es análogo pero en orden inverso.

### 5.3.1. Regla de paro

El proceso *forward selection* se suspenderá cuando ninguno de las potenciales variables que restan por considerarse disminuyan el coeficiente de correlación de Pearson en 0.05 %. De manera paralela se revisará el comportamiento del MSE correspondiente, y deberá disminuir conforme el coeficiente de correlación aumente.

# Capítulo 6

## RESULTADOS

En este capítulo se mostrarán las rectas de regresión lineal de la temperatura de cada una de las estaciones y su validación. Los datos empleados para construir las rectas son aquellos generados por el modelo WRF (todas las variables que proporciona este MNP) y observados por estaciones meteorológicas (temperatura) correspondientes al periodo enero 2009 — enero 2011.

El primer paso en la construcción de las rectas de regresión lineal consiste en definir los puntos geográficos que se analizarán. Para ello se escogieron 21 puntos geográficos del estado de Morelos en los que se ubican las estaciones meteorológicas del estado en cuestión. Como segundo paso se deben elegir de entre el universo de variables proporcionadas por el modelo WRF aquellas con significado físico relevante. El tercer paso consiste en elegir del conjunto de variables previamente seleccionadas aquellas que no presenten alto coeficiente de correlación; al conjunto resultante se le denominara *potenciales variables*. Finalmente se aplica el procedimiento *forward selection* empleando las *potenciales variables*. Las ecuaciones resultantes representan la “estimación corregida” de la temperatura en un punto geográfico.

### 6.1. Puntos geográficos analizados

Para el análisis estadístico se extrajeron las variables del modelo correspondiente a los puntos de malla más cercanos a cada estación meteorológica. En la tabla 6.1 en la que se

listan los nombres de las 21 estaciones junto con su ubicación geográfica y los índices de malla correspondientes.

Tablas 6.1: Nombre y ubicación de las 21 estaciones meteorológicas analizadas

	Estación	Ubicación	Punto de malla
1	INIFAP	-99.2008,18.6531	42,80
2	El Calvario	-99.3744,18.7373	42,79
3	Rancho el potrerillo	-99.3338,18.593	41,80
4	CEIEPO	-99.2363,19.0439	43, 80
5	Tlaquiltenango	-99.0851,18.532	41,80
6	Ocuituco	-98.8194,18.8716	42,81
7	El Hospital	-98.9878, 18.8321	42,81
8	Ayala	-99.0464,18.7391	42,81
9	Axochiapan	-98.7387,18.5341	41,82
10	Jonacatepec	-98.8317,18.6971	42,81
11	Puente de Ixtla	-99.3449,18.6476	42,79
12	Coatetelco	-99.293,18.7218	42,80
13	Tepalcingo	-98.8997, 18.6336	42,81
14	Emiliano Zapata	-99.1691,18.8411	42,80
15	Tepoztlan	-99.0891, 18.9383	43,80
16	Tlayacapan	-98.9885,18.9523	43,81
17	Tlaltizapan	-99.1342, 18.6912	42,80
18	Tetela del Volcan	-98.7108,18.8828	43,82
19	Tehuixtla	-99.2617,18.5417	41,80
20	Huazulco	-98.8099, 18.7548	42,81
21	Tlalnepantla	-98.9968,19.0115	43,81

## 6.2. Potenciales variables

Después de elegir los puntos geográficos con los que se trabajará, se deben elegir variables que tengan significado físico relevante, por ejemplo: temperatura superficial, presión flujo de calor, etc. Entonces, se eligieron veintisiete variables que cumplen el requisito anterior y se listan en la tabla 6.2.

Tablas 6.2: Variables físicas del modelo WRF que serán consideradas para conformar las rectas de regresión lineal múltiple.

Variables		Variables		Variables		Variables		Variables	
1	T2	7	Q2	13	QVAPOR*	19	LH	25	RAINNC
2	PH	8	TH2	14	QCLOUD	20	QFX	26	I_RAINC
3	P	9	PSFC	15	QRAIN	21	HFX	27	I_RAINNC
4	PB	10	V10	16	TMN	22	ALBEDO		
5	SR	11	U10	17	RAINC	23	SMOIS		
6	POTEVP	12	QVAPOR	18	PBLH	24	TSLB		

Tal como es requerido por el procedimiento *forward selection*, las variables a emplear no deben estar correlacionadas. Del diagrama de dispersión de dichas variables se puede inferir cuáles presentan alto coeficiente de correlación y cuáles no. De aquél par que lo estén se debe conservar una y descartar la otra. A manera de ejemplo, de la figura 6.1 se puede inferir que las variables T2 y P no guardan relación por lo que ambas se conservan (esta figura corresponde a los datos de la estación INIFAP del año 2009). Sin embargo, un análisis visual representa un procedimiento subjetivo, por lo que se opta por uno objetivo: calcular el coeficiente de correlación de Pearson.

Para establecer qué variables están correlacionadas se calcula el coeficiente de correlación de Pearson, se dirá que dos variables están correlacionadas si éste es mayor que 0.8. De las parejas que presenten alto coeficiente de correlación se descarta una de ellas. Al conjunto de variables “no correlacionadas” se les denomina *potenciales variables*.

De manera que para seleccionar las potenciales variables de las 21 estaciones del SMN se analizó la matriz de correlaciones de las variables de interés, esta labor se realiza de manera automática por medio de un código escrito en R. El algoritmo que se siguió se lista

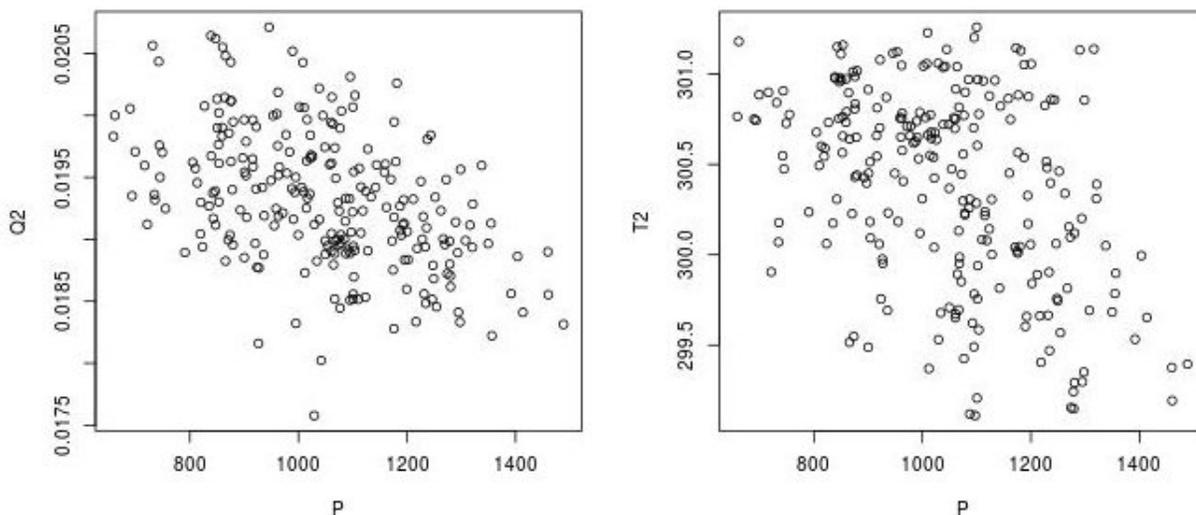


Figura 6.1: Diagrama de dispersión de 3 variables del modelo WRF correspondientes a la estación INIFAP en 2009.

a continuación y el código escrito se proporcionará como archivo adjunto a este reporte (“numpv”):

1. Leer la matriz de las pre-potenciales variables (en lo posterior denominada *file* formada por 2920 filas X 32 columnas).
2. Calcular el coeficiente de correlación de cada una de las variables con todas las demás. En este paso se genera la matriz de correlación. La matriz de correlación consiste en una matriz de 27X27. El elemento  $c_{m,n}$  de la matriz no es otra cosa sino el coeficiente de correlación de Pearson de la variable ubicada en la columna m de la matriz *file* con la variable ubicada en la columna n de la misma matriz.
3. Se identifican aquellas variables cuya desviación estándar es pequeña. Éstas se distinguen por generar el valor lógico NA como coeficiente de correlación de Pearson. Todos estos valores deberán ser reemplazados por un valor numérico a fin de que se

despliegue en la gráfica de la matriz de correlación cuando sea necesario. El valor numérico es arbitrario y se eligió 0.31415.

4. Se identifican todas aquellas variables cuyo coeficiente de correlación es mayor que 0.8.
5. Se genera un vector que contenga el nombre de las variables que resultan de aplicar los dos pasos anteriores. Este vector se denomina *allpv* y es el vector que contiene el nombre de las *potenciales variables*.

De manera que al implementar los pasos anteriores se obtienen las potenciales variables de un punto geográfico particular. A manera de ejemplo considere la tabla 6.3 en la que se listan las potenciales variables del punto de malla correspondiente a la estación INIFAP.

Tablas 6.3: Variables del modelo WRF que serán utilizadas para construir las rectas de regresión lineal.

1	T2	2	P	3	Q2	4	V10
5	U10	6	QVAPOR	7	QVAPOR*	8	QRAIN
9	TMN	10	RAINC	11	PBLH	12	ALBEDO
13	RAINNC						

## 6.3. Forward selection

Toda vez que se determinaron cuáles son las potenciales variables en cada uno de los puntos geográficos considerados, se aplica el procedimiento *forward selection* cuyo algoritmo se expuso en la sección 5.3. Para implementarlo de manera automática se escribió un código en R el cual se entrega de manera adjunta a este reporte (“forwardselectionpy”).

### 6.3.1. Temperatura

Con la finalidad de exponer de manera gráfica el procedimiento *forward selection* aplicado a los datos de la estación INIFAP, considere la figura 6.2. En esta imagen se muestra una

speudo-gráfica con un eje que representa el número de pasos del procedimiento en cuestión. Además en el recuadro inferior se muestran las potenciales variables de la estación analizada. Ahora bien, en cada uno de los pasos se eligió una variable que forma parte de la ecuación de regresión final y se muestra el nombre de la variable. Por ejemplo, la primer variable incorporada fue T2 y la última Q2. Compare los datos mostrados en la figura en cuestión con los proporcionados en la tabla 6.4.



Figura 6.2: Variables incluidas en la ecuación de regresión de la estación INIFAP.

El orden en que aparecen las variables puede interpretarse como un grado de “importancia” esto es, la variable T2 es más *importante* que cualquiera otra variable. Con *variable importante* se quiere decir que la variable en cuestión es aquella que al incorporarse en la recta de regresión conlleva a reducir el MSE y aumentar el coeficiente de correlación de Pearson, de manera más significativa que otra.

Para demostrar que T2 es la variable más importante considere la figura 6.3 y 6.4 que muestran el valor del coeficiente de correlación de Pearson y el logaritmo natural de MSE de la recta de regresión lineal en función del número de variables incluidas en la recta de regresión lineal. Las figuras en cuestión muestran que la primer variable incluida en el ajuste, T2, deriva en el menor MSE y mayor coeficiente de correlación entre la recta y los datos observados para la temperatura (primer variable, primer paso). Al incluir la variable P en el segundo paso disminuye el MSE y el coeficiente de correlación se mantiene. Al continuar aplicando el procedimiento de manera progresiva se incrementa el coeficiente de correlación y disminuye el MSE tal y como lo indica el procedimiento *forward selection*.

De manera que el resultado de aplicar el procedimiento *forward selection* a todas las

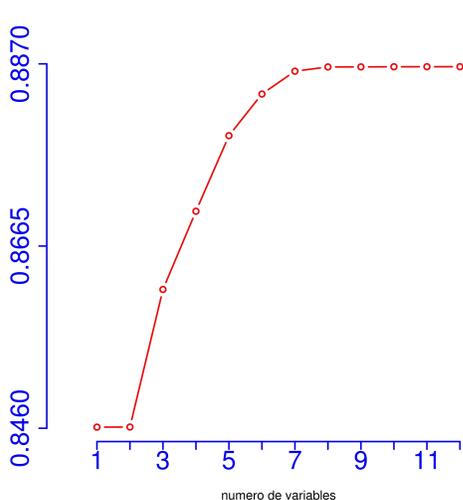


Figura 6.3: Coeficiente de correlación de Pearson de la recta de regresión lineal en la estación INIFAP.

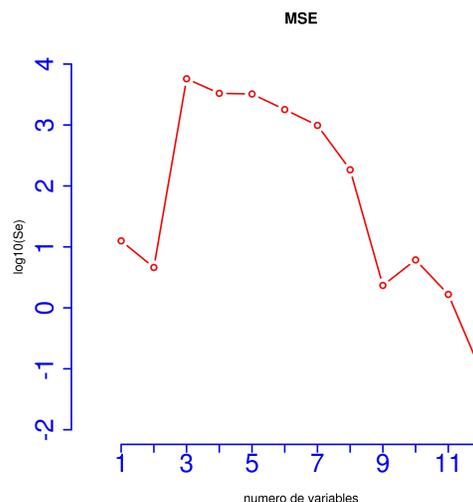


Figura 6.4: logaritmo natural de MSE de la recta de regresión lineal en la estación INIFAP.

estaciones son 21 ecuaciones. Las variables seleccionadas y su coeficiente respectivo serán utilizadas para hacer las correcciones al modelo WRF. A manera de ejemplo en la tabla 6.4 se muestran las variables y coeficientes de tres estaciones.

En el anexo A se listan los nombres de las variables y sus respectivos coeficientes de las 21 estaciones.

Las rectas de regresión lineal fueron empleadas para hacer estimaciones de la variable de interés, temperatura. Para ello se emplearon registros de la temperatura de las 21 estaciones de interés cuyo periodo comprende del 1 de agosto a 8 de agosto de 2012. Esto permitirá hacer una validación parcial de los datos obtenidos. Una validación completa exige implementar este análisis a datos que abarquen periodos mayores de tiempo en diferentes épocas del año; el análisis también tendría que aplicarse a datos que el modelo WRF genera día a día.

Para validar, de manera parcial, la mejoría se calcula el error cuadrático medio (MSE) de las variables meteorológicas medidas por las estaciones meteorológicas y aquellas pronosticadas tanto por las rectas de regresión, por el modelo WRF y un ensamble del las rectas de regresión y el modelo (Av.).

Tablas 6.4: Variables y coeficientes de las rectas de regresión lineal.

INIFAP		El Calvario		Rancho el potrerrillo	
Variable	Coeficiente	Variable	Coeficiente	Variable	Coeficiente
b0	-447.4907	b0	-380.1113	b0	-486.0974
T2	1.4274	T2	1.0568	T2	1.3459
P	0.0015	P	0.002	P	0.0017
U10	0.4824	U10	0.511	ALBEDO	81.3997
ALBEDO	61.8215	ALBEDO	65.5976	PBLH	-0.0024
PBLH	-0.0028	QVAPOR*	-1092.55	QVAPOR*	-695.84
V10	0.4047	PBLH	-0.0017	U10	0.4038
QVAPOR2	-617.1625	V10	0.3065	TMN	0.3401
TMN	0.1395	TMN	0.2843	V10	0.3133
RAINC	-0.0397	QVAPOR	-2511.72	Q2	-175.1461
QRAIN	1461.4862	RAINC	-0.1423	QRAIN	4416.3907
QVAPOR	415.8542	QRAIN	5979.80	RAINC	-0.1409
		RAINC	-0.0831		

Los resultados de aplicar el procedimiento *forward selection* se resumen en la tabla 6.5. Cada una de las columnas contiene el valor del coeficiente de correlación de Pearson (R) y el error cuadrático medio (MSE) de la variable temperatura predicha por el modelo WRF (T2), la estimada por las rectas de regresión lineal (RRL) y el ensamble de las rectas de regresión y el modelo (Av).

Tal como puede verificarse en la tabla 6.5 el modelo WRF proporciona buenos pronósticos, de hecho en 10 estaciones (el 47 %, en Rancho potrerrillo, El Hospital, Ayala, Axochiapan, Jonacatepec, Tepalcingo, Tlaltizapan, Tetela del Volcan, Tehuixtla, Huazulco) el modelo WRF proporcionó mejores resultados que el ajuste lineal. No obstante en el resto de las estaciones (53 %) el ajuste lineal o el denominada *ensamble* mejoraron los pronósticos del modelo. Para ser más específicos, las rectas de ajuste lineal mejoran las predicciones del modelo WRF en 4 estaciones: CEIEPO, Ocuituci, Puente de Ixtla, Tlalnepantla. En el resto, 11 estaciones, el ensamble permite mejorar las predicciones del modelo.

Continuando con la exposición de los resultados de la validación parcial de las rectas de regresión, considere la figura 6.5 que muestra el perfil de la temperatura registrada por la estación CEIEPO (Estación), los valores predichos por la recta de regresión (Ajuste lineal) y el ensamble (Av.) en el mismo punto geográfico; todos estos datos correspondientes al periodo de tiempo analizado. En la figura en cuestión se muestran los datos de 3 días solamente (1-3 de agosto) en intervalos de 3 horas. Tal y como puede apreciarse en esta figura los datos generados por la recta de regresión se ajustan mejor que los predichos por el modelo WRF. Este “mejor ajuste” de los puntos mostrados en la gráfica es congruente con el valor de MSE mostrado en la tabla 6.5 para la estación en cuestión. Se eligieron los datos de la estación CEIEPO por tratarse de un punto geográfico en el que la recta de ajuste lineal construida en efecto mejora las predicciones del modelo WRF.

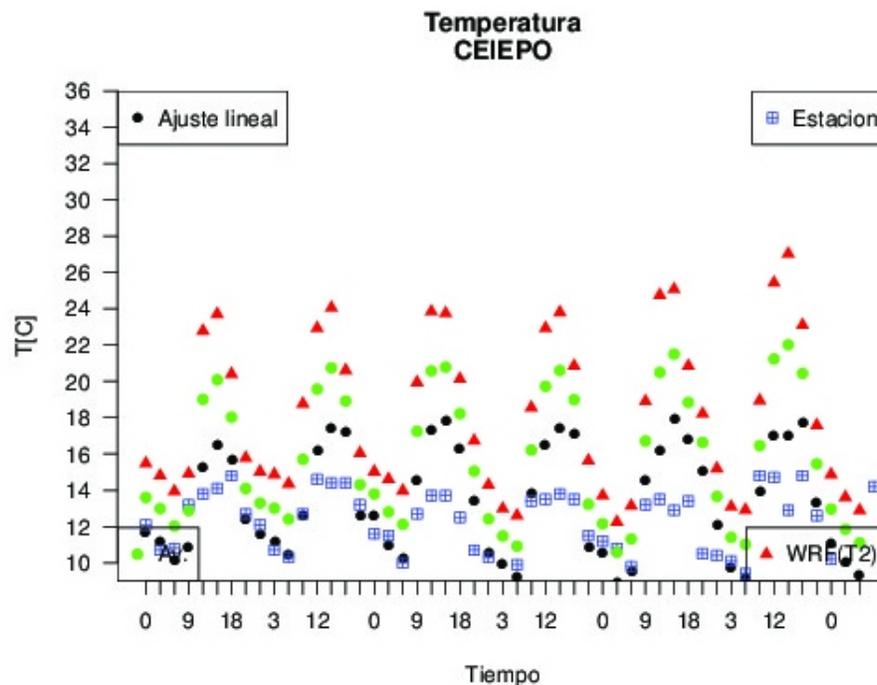


Figura 6.5: Perfil temporal de la temperatura observada y predicha por el modelo WRF y la recta de regresión en la estación CEIEPO.

Ahora bien, durante el mismo periodo considerado, en otros puntos geográficos, por ejemplo en Tetela del Volcan, el modelo WRF proporciona mejores resultados que el ajuste lineal construido. Esta afirmación se concluye al revisar los datos correspondientes a la estación en cuestión en la tabla 6.5 y se verifica de manera visual en la figura 6.3.1

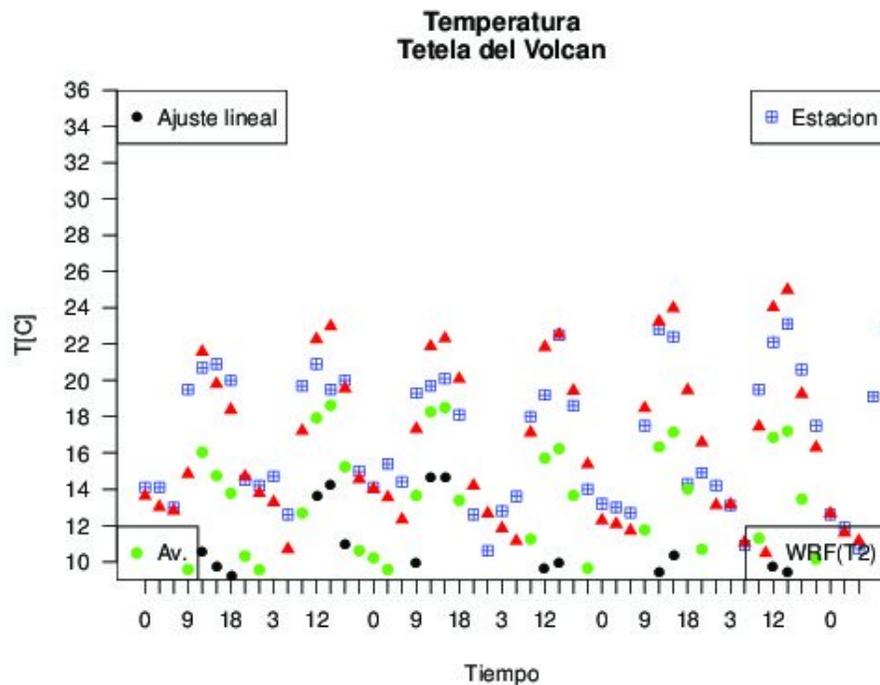


Figura 6.6: Perfil temporal de la temperatura observada y predicha por el modelo WRF y la recta de regresión en la estación Tetela del Volcan.

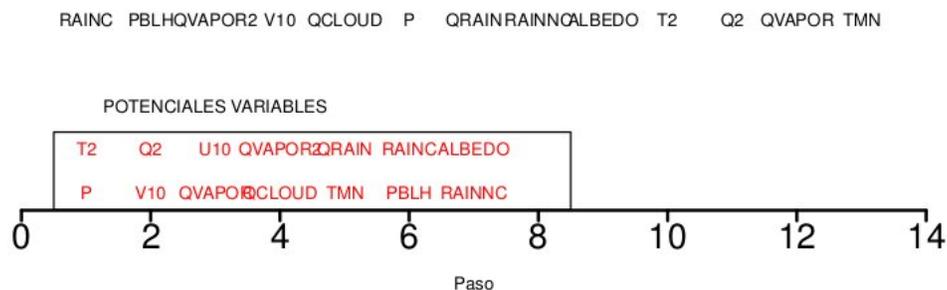


Figura 6.7: Variables incluidas en la ecuación de regresión de la estación “El Hospital.”

### 6.3.2. Precipitación

En el caso de la precipitación, el procedimiento es exactamente el mismo que el aplicado a la temperatura. Las potenciales variables son las mismas y los códigos empleados también. De manera que considere la figura 6.7 que representa de manera esquemática el procedimiento *forward selection* con los datos de la estación “El Hospital”. Esta pseudográfica muestra las variables que forman parte de la ecuación de regresión lineal de esta estación, también muestra el orden en que las variables fueron seleccionadas y exhibe las potenciales variables.

Considere ahora la figura 6.8 y 6.9 que muestran el valor del coeficiente de correlación de Pearson y el logaritmo natural de MSE de la recta de regresión lineal en función del número de variables incluidas en la recta de regresión.

De manera similar al caso de la temperatura, el resultado de aplicar el procedimiento *forward selection* a todas las estaciones son 21 ecuaciones de regresión lineal. Las variables seleccionadas y su coeficiente respectivo serán utilizadas para hacer las correcciones al modelo WRF. A manera de ejemplo en la tabla 6.6 se muestran las variables y coeficientes de tres estaciones. Note que el número de variables incluidas en el ajuste es menor que el correspondiente a las rectas de ajuste de la temperatura. Esto se debe a que durante el proceso de construcción de estas rectas ninguna variable que fuese incluida incrementaba el coeficiente de correlación de Pearson ni disminuía el MSE.

A diferencia de la variable temperatura, el procedimiento *forward selection* no arrojó bue-

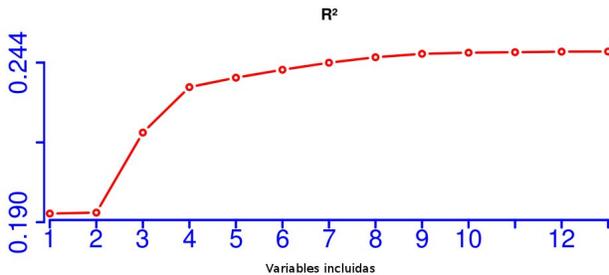


Figura 6.8: Coeficiente de correlación de Pearson de la recta de regresión lineal en la estación “El Hospital.”

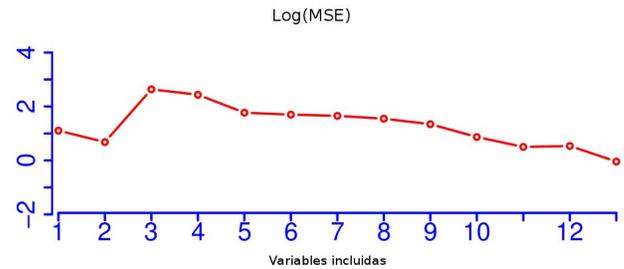


Figura 6.9: MSE de la recta de regresión lineal en la estación “El Hospital.”

nos resultados. Esto significa que las rectas de regresión construidas no proporcionaban valores que se ajustaran a los datos observados de esta variable. De hecho los correspondientes coeficientes de correlación de Pearson fueron muy bajos como se demuestra en la tabla 6.7. La columna etiquetada como “R-in” representa el coeficiente de Pearson de los datos observados y la recta de regresión lineal al inicio del procedimiento *forward selection*; al final del procedimiento se determinó el mismo parámetro y es el que se exhibe en la misma tabla bajo la etiqueta “R-fin”. Entonces, dado que durante el proceso de generar las rectas de ajuste no se obtuvieron rectas que se ajustaran a los datos observados, no se procedió a la etapa de verificación con datos de otra época.

Tablas 6.5: Coeficiente de correlación de Pearson y error cuadrático medio de las predicciones del modelo WRF y las rectas de regresión lineal.

	Estación	R-RRL	MSE-RRL	R-T2	MSE-T2	R-Av	MSE-Av
1	INIFAP	0.8911	6.2409	0.9175	3.4585	0.9167	3.3633
2	El Calvario	0.8873	3.5827	0.8971	5.7839	0.9167	3.5525
3	Rancho el potrerrillo	0.8492	7.1369	0.9324	3.9542	0.9094	3.9857
4	CEIEPO	0.8233	4.3442	0.806	40.20	0.8224	16.6667
5	Tlaquiltenango	0.8829	5.0795	0.9177	4.104	0.9197	3.3626
6	Ocuituco	0.8297	4.6264	0.9082	8.8072	0.8825	5.38
7	El Hospital	0.8563	6.2246	0.9239	2.3904	0.8994	3.5837
8	Ayala	0.815	6.9086	0.8991	3.2997	0.8807	3.9215
9	Axochiapan	0.898	6.2963	0.9364	2.4362	0.9306	3.1934
10	Jonacatepec	0.7938	9.1241	0.9217	2.4037	0.8805	4.1648
11	Puente de Ixtla	0.8956	4.1091	0.9018	7.846	0.9215	4.5807
12	Coatetelco	0.8721	5.1364	0.8844	3.5097	0.9012	3.28
13	Tepalcingo	0.7523	11.614	0.9025	3.2892	0.8592	5.249
14	Emiliano Zapata	0.8948	4.0342	0.9225	3.2778	0.9187	3.0831
15	Tepoztlan	0.8968	4.5063	0.9198	11.450	0.9184	3.8562
16	Tlayacapan	0.8178	5.2588	0.8947	5.472	0.8715	3.4285
17	Tlaltizapan	0.8774	6.7896	0.9356	2.8843	0.9171	3.7912
18	Tetela del volcan	0.8466	116.17	0.9115	2.7725	0.9039	30.7622
19	Tehuixtla	0.902	6.2514	0.9488	2.1426	0.9368	2.7933
20	Huazulco	0.7877	7.8034	0.906	3.5615	0.8634	4.4668
21	Tlalnepantla	0.7988	4.2416	0.8578	5.3033	0.8426	4.2702

Tablas 6.6: Variables y coeficientes de las rectas de regresión lineal.

El Hospital		Ayala		Axochiapan	
Variable	Coeficiente	Variable	Coeficiente	Variable	Coeficiente
b0	-381.2759	b0	-243.8402	b0	16.3214
QVAPOR2	-601.3075	RAINC	-0.2849	Q2	599.8357
T2	1.3488	T2	0.9122	U10	0.7946
Q2	-123.4916				
RAINC	-0.0315				
QCLOUD	-5507.9962				
PBLH	-0.0026				
RAINNC	-0.002				
V10	0.3103				
QRAIN	1417.6487				
QVAPOR	308.957				
P	0.0012				
U10	0.3104				
ALBEDO	53.955				

Tablas 6.7: Coeficiente de correlación de Pearson de las rectas de regresión lineal de la variable precipitación.

	Estación	R-in	R-fin
1	INIFAP	0.2063	0.2603
2	El Calvario	0.1723	0.2281
3	Rancho el potrerrillo	0.1734	0.2419
4	CEIEPO	0.264	0.3464
5	Tlaquiltenango	0.1514	0.2358
6	Ocuituco	0.2169	0.2634
7	El Hospital	0.1896	0.2444
8	Ayala	0.2151	0.2704
9	Axochiapan	0.1869	0.2569
10	Jonacatepec	0.1934	0.2598
11	Puente de Ixtla	0.2378	0.2861
12	Coatetelco	0.1776	0.2323
13	Tepalcingo	0.1936	0.2585
14	Emiliano Zapata	0.1294	0.2021
15	Tepoztlan	0.2051	0.2818
16	Tlayacapan	0.2078	0.2657
17	Tlaltizapan	0.1723	0.2188
18	Tetela del volcan	0.2366	0.296
19	Tehuixtla	0.2129	0.2711
20	Huazulco	0.1643	0.225
21	Tlalnepantla	0.172	0.2298

# Capítulo 7

## CONCLUSIONES

En el presente estudio, se mostró la implementación de la técnica MOS con el propósito de mejorar el pronóstico de la temperatura superficial y precipitación, dada por el modelo WRF, en localidades específicas del Estado de Morelos.

Con relación a la temperatura, en términos generales y de manera parcial (en el sentido especificado en la sección 6.3.1), se encontró que en las localidades donde el WRF tiene un error cuadrático medio superior a 5, la corrección estadística (MOS) fue mejor, por ejemplo en la estación CEIEPO. El pronóstico por el MOS fue mejor que solo el WRF en 6 localidades. Por otro lado el ensamble (WRF-MOS), mostró una mejoría en el pronóstico en 11 localidades. Es probable que la mejora en ciertas estaciones se deba principalmente a las diferencias en la topografía.

Por otra parte, la técnica que se empleó en este trabajo no permitió mejorar los pronósticos de la variable precipitación. Esto se concluye al revisar el coeficiente de correlación de Pearson de los datos observados y los datos predichos por las rectas de regresión en cada una de las estaciones después de aplicar el procedimiento *forward selection*.

## Bibliografía

- [1] Lowry D.A. Glahn, H.R. The use of model output statistics (mos) in objective weather forecasting. *Journal of applied meteorology*, 11:1203–1211, 1972.
- [2] Downie N.M. *Métodos estadísticos aplicados*. Harla, México, primera edition, 1973.
- [3] Wilks Daniels S. *Statistical Methods in the Atmospheric Sciences*. Elsevier, Amsterdam, second edition, 2006.

## **Anexo A**

### **Variables y coeficientes de las rectas de regresión**

Tablas A.1: Variables y coeficientes de las rectas de regresión (continuación).

CEIEPO		Tlaquiltenango		OCUITUCO	
Variable	Coeficiente	Variable	Coeficiente	Variable	Coeficiente
b0	-313.1947	b0	-472.1032	b0	-374.6298
T2	1.0393	T2	1.2591	T2	1.0876
QVAPOR	32.1961	P	0.0016	QCLOUD	-4505.1793
PBLH	-0.002	QVAPOR2	-851.0981	ALBEDO	55.708
ALBEDO	97.5594	ALBEDO	85.4186	QVAPOR2	-746.9421
P	0.0017	PBLH	-0.0022	PBLH	-0.0021
QVAPOR2	-395.3771	U10	0.3832	U10	0.2492
QRAIN	1452.0201	TMN	0.3792	P	0.0018
U10	0.0538	Q2	-216.2141	V10	0.2085
TMN	0.0353	V10	0.3429	Q2	-227.4148
QCLOUD	988.461	QRAIN	4209.2969	TMN	0.2305
Q2	-17.6061	RAINNC	-0.1025	QRAIN	2100.0905
V10	0.0077	RAINNC	-0.0129	RAINNC	-0.0257
RAINNC	0.0119			RAINNC	-0.0085

Tablas A.2: Variables y coeficientes de las rectas de regresión (continuación).

El Hospital		Ayala		Axochiapan	
Variable	Coeficiente	Variable	Coeficiente	Variable	Coeficiente
b0	-383.0195	b0	-387.9123	b0	-537.4806
T2	1.3476	T2	1.1424	T2	1.563
QCLOUD	-5512.5136	P	6e-04	TMN	0.3007
ALBEDO	54.4281	ALBEDO	71.0757	ALBEDO	72.5629
PBLH	-0.0026	PBLH	-0.0026	HFX	-0.0123
U10	0.311	QVAPOR2	-1031.5665	Q2	-307.8385
QVAPOR2	-597.5834	V10	0.3919	PBLH	-0.0017
V10	0.3111	U10	0.3028	V10	0.4029
Q2	-125.2272	TMN	0.2286	QVAPOR2	-495.6864
P	0.0012	Q2	-142.4297	P	0.0031
QRAIN	1372.0149	QVAPOR	-1658.7366	U10	0.2795
RAINC	-0.0323	QCLOUD	-11539.6714	QRAIN	4571.2978
QVAPOR	325.0948	RAINC	-0.0513	RAINC	-0.1199
TMN	0.0069	QRAIN	1152.0777	RAINNC	0.0328

Tablas A.3: Variables y coeficientes de las rectas de regresión (continuación).

Jonacatepec		Puente de Ixtla		Coatetelco	
Variable	Coeficiente	Variable	Coeficiente	Variable	Coeficiente
b0	-404.0951	b0	-389.0788	b0 -	418.1691
T2	1.3481	T2	1.1377	T2	1.2138
QCLOUD	-9189.9064	P	0.0022	P	7e-04
ALBEDO	56.4441	U10	0.5528	ALBEDO	71.9215
PBLH	-0.0031	ALBEDO	65.4776	PBLH	-0.003
QVAPOR2	-760.3335	QVAPOR2	-1063.2519	QVAPOR2	-950.2962
V10	0.3341	PBLH	-0.0017	V10	0.393
U10	0.223	V10	0.3178	U10	0.4532
P	0.0014	TMN	0.2324	TMN	0.2552
QVAPOR	-2753.9233	QVAPOR	-2937.4466	Q2	-55.1823
TMN	0.0766	RAINC	-0.1054	RAINC	-0.0509
Q2	-62.9535	Q2	63.0932	QVAPOR	-1355.074
RAINC	-0.0708	QRAIN	6394.4403	RAINNC	-0.009
QRAIN	1927.1112	RAINNC	-0.1396		

Tablas A.4: Variables y coeficientes de las rectas de regresión (continuación).

Tepalcingo		Emiliano Zapata		Tepoztlán	
Variable	Coeficiente	Variable	Coeficiente	Variable	Coeficiente
b0	-409.7173	b0	-390.9653	b0	-431.1725
T2	1.2343	T2	1.1865	T2	1.3121
QCLOUD	-10495.0904	P	0.0011	QCLOUD	-4833.7926
ALBEDO	70.4933	ALBEDO	61.6746	ALBEDO	193.4207
PBLH	-0.0029	U10	0.3492	PBLH	-0.0019
QVAPOR2	-975.9141	PBLH	-0.0024	V10	0.2473
V10	0.4187	V10	0.4048	QVAPOR2	-876.5545
U10	0.2748	QVAPOR2	-708.831	P	0.0025
TMN	0.2065	TMN	0.1879	U10	0.2029
P	0.0014	Q2	-114.8357	TMN	0.1594
Q2	-107.6693	RAINC	-0.0398	Q2	-87.9437
QVAPOR	-2349.9154	RAINNC	0.0085	QVAPOR	-3721.8115
RAINC	-0.0922	QRAIN	-207.9698	RAINNC	-0.0832
QRAIN	1724.0328			QRAIN	2989.7924

Tablas A.5: Variables y coeficientes de las rectas de regresión (continuación).

Tlayacapan		Tlaltizapan		Tetela del Monte	
Variable	Coficiente	Variable	Coficiente	Variable	Coficiente
b0	-290.7972	b0	-449.6092	b0	-352.4424
T2	1.0017	T2	1.3345	T2	0.9248
Q2	-61.1189	P	0.002	U10	0.0371
U10	0.2735	ALBEDO	66.8116	QVAPOR2	-661.9745
QVAPOR2	-964.068	U10	0.443	ALBEDO	55.4726
P	0.0027	PBLH	-0.0021	P	0.0024
PBLH	-0.0011	V10	0.4438	TMN	0.295
V10	0.136	QVAPOR2	-551.7637	HFX	0.0091
QCLOUD	-3803.2612	TMN	0.2317	PBLH	-0.0016
QVAPOR	-3835.3564	RAINC	-0.092	Q2	-211.5587
TMN	0.0646	Q2	-29.3966	V10	0.1803
RAINC	0.0672	QRAIN	1265.8712	QCLOUD	14906.6831
RAINNC	-0.0412	QVAPOR	574.4069	QVAPOR	1277.9514
				RAINC	-0.0984
				RAINNC	0.0662

Tablas A.6: Variables y coeficientes de las rectas de regresión (continuación).

Tetela del Volcan		Tehuixtla		Huazulco	
Variable	Coeficiente	Variable	Coeficiente	Variable	Coeficiente
b0	-321.9492	b0	-527.7195	b0	-392.1416
T2	1.0545	T2	1.5005	T2	1.2157
PBLH	-6e-04	P	0.0025	QCLOUD	-8187.332
ALBEDO	152.7283	ALBEDO	84.6369	ALBEDO	63.0138
QVAPOR2	-618.6578	U10	0.4512	PBLH	-0.0025
V10	0.1312	QVAPOR2	-619.0974	QVAPOR2	-819.5322
P	0.0028	PBLH	-0.002	V10	0.3346
QCLOUD	-3811.291	TMN	0.3216	U10	0.2493
U10	0.0993	Q2	-205.2838	P	0.0015
Q2	-117.9411	V10	0.2869	Q2	-162.3443
RAINC	0.1272	QRAIN	4777.7929	TMN	0.1637
TMN	0.0404	RAINC	-0.136	QVAPOR	-1125.1126
QVAPOR	-1937.1622	QVAPOR	919.3058	RAINC	-0.0608
RAINNC	-0.0299			QRAIN	1948.2037

Tablas A.7: Variables y coeficientes de las rectas de regresión (fin).

Talnepantla	
Variable	Coefficiente
b0	-244.2057
T2	0.9629
PBLH	-0.0012
QVAPOR2	-1107.3459
U10	0.2165
P	0.0023
TMN	-0.0593
V10	0.1038
Q2	-158.7803
QCLOUD	-2816.4174
QVAPOR	-4046.6478
RAINC	0.0531
RAINNC	-0.0308