

Análisis espectral de series temporales de
temperaturas y su aplicación al estudio de su
comportamiento en una red de observatorios
meteorológicos

Francisca Dolores Matlalcuatzi Rugerio

2006

Contenido

| | |
|---|----|
| Agradecimientos | iv |
| Resumen | v |
| Introducción | 1 |
| 1. Área geográfica de estudio e información meteorológica | 4 |
| 1.1. Delimitación del área geográfica de estudio | 4 |
| 1.2. Información Meteorológica | 6 |
| 1.2.1. Instrumentos de medición | 7 |
| 1.2.2. Estaciones meteorológicas | 7 |
| 2. Obtención de datos faltantes | 14 |
| 2.1. Algoritmo para obtener los datos faltantes | 17 |
| 3. Filtrado de las series de temperaturas | 20 |
| 4. Ajuste de las series Filtradas | 29 |
| 4.1. Resultados | 30 |
| 5. Discusión de resultados | 42 |
| Conclusiones | 44 |
| Apéndices. | 45 |
| A. Programa para obtener datos faltantes | 45 |
| B. Programa que calcula el filtrado y el ajuste de las series | 48 |
| Bibliografía | 52 |

Resumen

El objetivo de este trabajo es obtener una función suavizada y manejable que describa de manera aproximada la variación de la temperatura máxima anual y analógicamente otra que describa la temperatura mínima con base en los datos registrados en cada uno de los observatorios meteorológicos que, por su ubicación geográfica, proporcionan información del comportamiento de la temperatura en la cuenca del Valle de Puebla.

Para obtener dichas funciones se usaron los datos diarios de temperaturas máximas y mínimas en el periodo comprendido entre los años 1967 y 2004. Se empleó el método de análisis espectral (reportado en Chatfield, 2003; Fuller, 1976; Gottman, 1984) que resulta directamente del análisis de Fourier. Se aplicó análisis espectral a los datos diarios de temperaturas máximas/mínimas en cada observatorio, para detectar la periodicidad más amplia y eventuales periodicidades ocultas. Se implementó un criterio para seleccionar las frecuencias más representativas asociadas a dichas periodicidades. Empleando dichas frecuencias, se filtraron los datos; se seleccionó una sola serie filtrada por estación y se ajustó por medio de un polinomio trigonométrico, siendo este polinomio la función buscada. Un procedimiento similar se efectuó para temperatura mínima; se procedió igual con los otros observatorios.

Introducción

El agua es el componente principal de la materia viva; constituye del 50 al 90 por ciento de la masa de los organismos vivos; la existencia del agua a lo largo de la historia ha determinado la ubicación de los núcleos de población humana, debido a que tanto la agricultura como la industria están íntimamente relacionadas con la disponibilidad de agua. En la actualidad, son muchas las actividades humanas donde el agua reviste gran importancia, de aquí la necesidad de conocer mejor la dinámica histórica del agua y realizar pronósticos de su evolución para enfrentar la problemática actual de disminución acelerada de volúmenes de agua dulce.

Con la intención de responder a esta problemática, existe un trabajo de grupo, del que éste forma parte, en el que se desea formular un modelo matemático de balance hídrico, cuyo objetivo consiste en efectuar una modelación matemática que describa, de manera dinámica y acoplada, los principales procesos que intervienen en el balance hídrico de una cuenca y aplicarla al desarrollo de una metodología para el manejo sustentable de la cuenca. Dicha metodología deberá traducirse en un sistema interactivo que sirva como herramienta para la toma de decisiones. En la Figura 1 se representa el diagrama empleado para formular el modelo matemático de balance hídrico.

El objetivo de este trabajo es proporcionar funciones suaves que describan la variación temporal de las temperaturas máximas y mínimas en cada uno de los ocho observatorios meteorológicos que estamos considerando (TABLA 1). En un trabajo posterior, estas funciones se emplearán para obtener una función que dependa tanto del tiempo como del espacio, que describa la variación de temperaturas en la cuenca del Valle de Puebla; dicha función será empleada en el cálculo de los coeficientes de evapotranspiración, variables de entrada para un modelo matemático de balance hídrico.

En este trabajo se efectúa el procesamiento de datos de temperaturas máximas y mínimas diarias obtenidas de los registros de ocho estaciones meteorológicas que por su

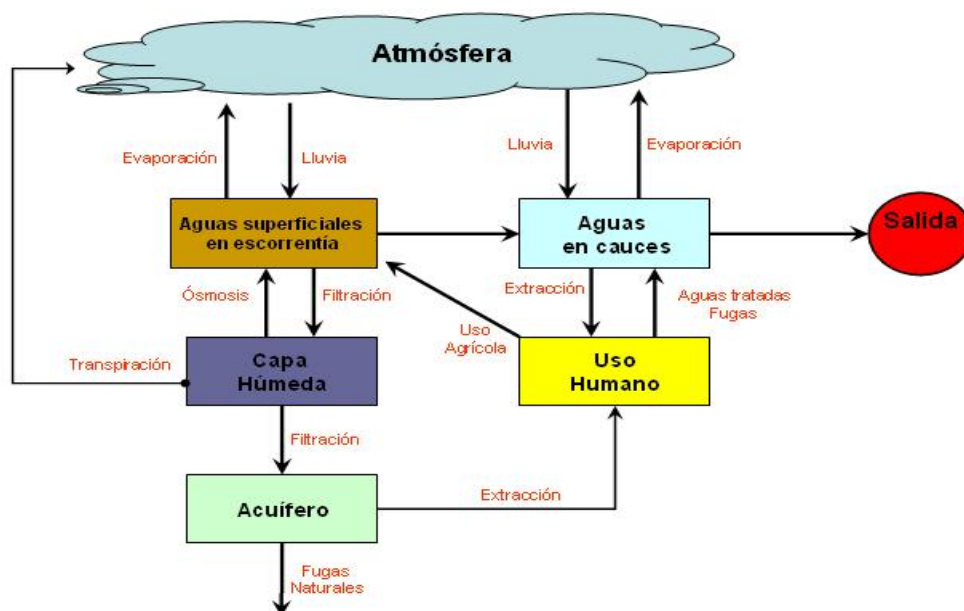


Figura 1. Diagrama para la formulación del modelo matemático de balance hídrico.

ubicación geográfica proporcionan información del comportamiento de la temperatura en la cuenca del Valle de Puebla. Con dicho procesamiento podemos describir analíticamente la variación de temperaturas máximas y mínimas por medio de polinomios trigonométricos.

En general existen dos métodos empleados para describir datos de fenómenos registrados en ciertos instantes de tiempo: los métodos en el dominio del tiempo y los métodos en el dominio de la frecuencia; entre los métodos en el dominio de la frecuencia se encuentran el método de análisis espectral y los métodos para estimar funciones de distribución de probabilidad (como los descritos en Breña, 2005). Entre los métodos en el dominio del tiempo se encuentran los modelos de medias móviles (MA); modelos autorregresivos de orden p (AR (p)); los procesos mixtos (ARMA); los métodos anteriores son descritos en Gottman, 1984; Fuller, 1976; Guerrero, 1987.

En esta tesis se aplica el método de análisis espectral (reportado en Chatfiel, 2003; Fuller, 1976; Gottman, 1984) para obtener de forma determinista las funciones deseadas.

Las razones de aplicar dicho análisis surgen del hecho conocido de que la temperatura tiene periodicidades anuales y trimestrales, que se reflejan en las estaciones del año, y de que el análisis espectral es una buena herramienta para descubrir las periodicidades ocultas de fenómenos físicos que se suponen periódicos, así como para representarlos (léase

comentario de Yule G. U., citado en Gottman, 2003). En este caso se emplean las frecuencias asociadas a las periodicidades ocultas en los datos de temperatura máxima/mínima diaria, para aproximar dichos datos por un polinomio trigonométrico.

Para obtener la curva suavizada se requería obtener: el intervalo de tiempo adecuado para describir las temperaturas; verificar si, como se presuponía, la periodicidad más amplia era la anual, cabe señalar que los resultados obtenidos dependieron del método empleado y de la cantidad de datos que se tenían. Una vez que se determinó que la periodicidad más amplia que se detecta es la anual se convino que la curva suavizada debe describir la variación de temperaturas a lo largo de un año. Para identificar periodicidades mayores a la anual se requería tener datos diarios de temperatura de varios años consecutivos, sin embargo la información no estaba completa, por tal motivo se estimaron los datos faltantes por medio de algunos criterios que se describen en el Capítulo 2. Una vez que se confirmó que la periodicidad más amplia detectada es la anual, se aplicó nuevamente análisis espectral, pero ahora a cada año de datos de temperaturas; para identificar las periodicidades ocultas menores a un año. Los resultados obtenidos se ejemplifican en este trabajo empleando los datos de temperaturas máximas de 1967-2004 de la estación Puebla.

Esta tesis está organizada como sigue: En el Capítulo 1 se delimita y describe el área geográfica de estudio; se detalla la forma en la que se organizó la información recopilada. En el Capítulo 2 se presenta un procedimiento y un algoritmo para obtener datos faltantes de temperaturas máximas/mínimas en cada observatorio meteorológico considerado. En el Capítulo 3 se expone uno de los resultados importantes de este trabajo: una serie filtrada, por cada estación, que describe la variación de la temperatura máxima/mínima. El capítulo se divide en dos partes, en la primera se hace uso de los datos obtenidos en el Capítulo 2 para tratar de detectar periodicidades más amplias a la anual; en la segunda parte se propone un criterio para identificar la frecuencias más representativas de cada estación, que se utilizan para proponer el polinomio trigonométrico. En el Capítulo 4 se presenta el resultado principal de la tesis: los polinomios trigonométricos que describen las temperaturas máximas diarias por estación y los que describen las temperaturas mínimas diarias de cada estación y sus gráficas. En el último capítulo se presenta una breve discusión de los resultados. Finalmente se incluyen dos apéndices, con los programas, realizados en Matlab, empleados para obtener los resultados de esta tesis.

Capítulo 1

Área geográfica de estudio e información meteorológica

En este capítulo se presenta el concepto de cuenca hidrológica, se describe el área de estudio y además se detalla la forma en la que se organizó la información meteorológica disponible.

1.1. Delimitación del área geográfica de estudio

Una cuenca hidrológica (River,2001) se define como una unidad territorial en la cual el agua que cae por precipitación se reúne y escurre o fluye toda al mismo río, lago o mar. También se define como una unidad fisiográfica conformada por la reunión de un sistema de cursos de ríos de agua definidos por el relieve.

Los límites de la cuenca o divisoria de aguas se definen naturalmente, y en forma práctica corresponden a las partes más altas del área que encierra un río. Una cuenca no solamente abarca la superficie, largo y ancho, sino también la profundidad, comprendida desde el extremo superior de la vegetación hasta los estratos geológicos limitantes bajo la tierra. En este área viven seres humanos, animales y plantas, todos ellos relacionados entre sí.

El manejo sustentable de cuencas se define como la gestión que el hombre realiza en forma integral para aprovechar y proteger los recursos naturales con el fin de obtener una producción óptima y sostenida de agua para satisfacer las necesidades presentes y futuras

de todos los actores que conviven en la región. De esta definición se aprecia que el manejo de cuencas integra muchas disciplinas del quehacer humano, cuya importancia relativa dependerá de las condiciones y características de cada cuenca (Martínez, 1999).

El área geográfica de estudio de este trabajo es la cuenca del Valle de Puebla (Figura 2) que pertenece a la región hidrológica No. 18 del Río Balsas (RH18) y está comprendida entre los $18^{\circ} 50'$ y $19^{\circ} 15'$ de latitud norte y los $98^{\circ} 00'$ y $98^{\circ} 20'$ de longitud oeste, abarcando la mayor parte del municipio de Puebla y parte de los municipios de Amozoc, Cuautinchán, Tzicatlacoyan y Ocoyucan.

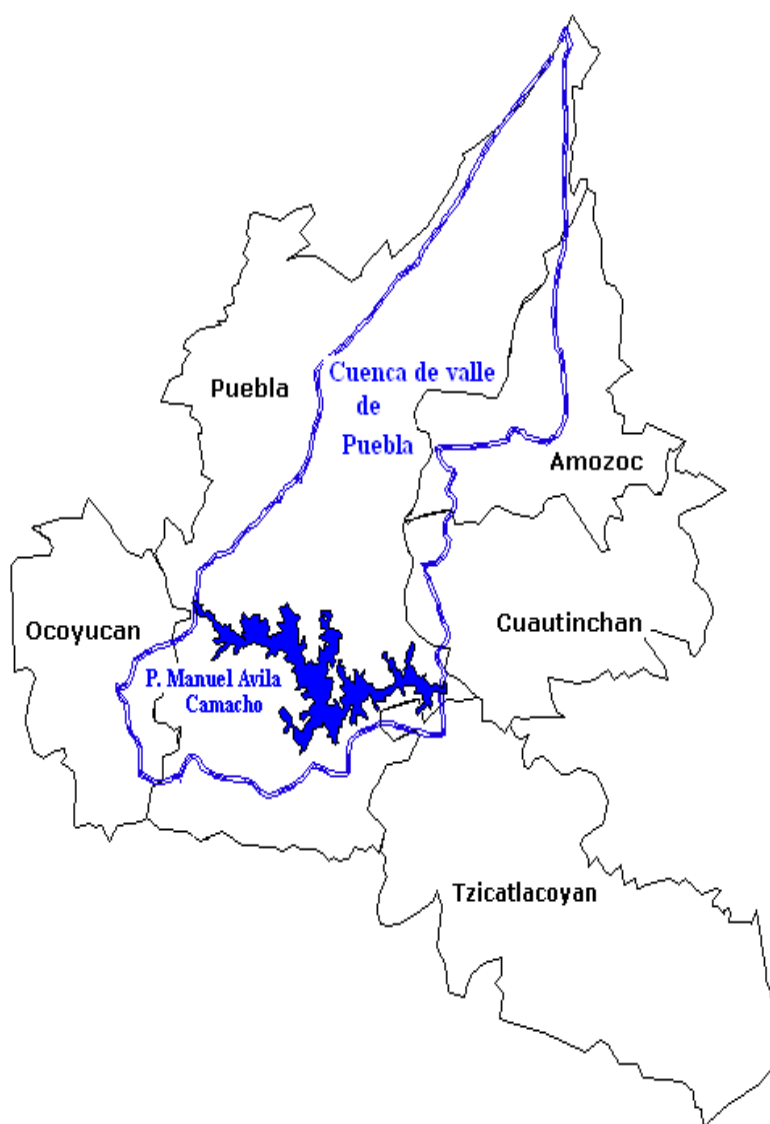


Figura 2. Cuenca del Valle de Puebla.

La cuenca del Valle de Puebla abarca desde la parte suroeste del volcán Malinche hasta la presa Valsequillo. La cuenca está constituida por dos corrientes principales: el río Atoyac, que escurre hasta entrar al vaso de la presa de Valsequillo, y el río Alseseca, con un recorrido de norte a sur dentro de la ciudad de Puebla, que desemboca en el vaso de la presa Valsequillo.

En la cuenca del Valle de Puebla, el clima se clasifica como templado, de una temperatura media anual entre 23°C y 24°C , y la del mes más frío entre 4°C y 7°C ; con lluvias en verano (Romero, 2004).

1.2. Información Meteorológica

A pesar de que los procesos meteorológicos varían continuamente en el tiempo y en el espacio, se suelen medir como muestras puntuales, es decir, se realizan mediciones a través del tiempo en lugares fijos del espacio. Estos lugares fijos se denominan **observatorios** o **estaciones meteorológicas**, que son instalaciones donde se realizan observaciones sensoriales e instrumentales de los elementos meteorológicos (precipitación, temperatura, vientos, heladas, etc.) en un lugar, en un momento y/o durante las 24 horas del día y por tiempo indefinido (INEGI, 2001). La temperatura varía continuamente en el espacio de una cuenca, pero el termómetro mide la temperatura en un punto específico (estación meteorológica) de ella. La información resultante forma una **serie de tiempo**. También se estudian muestras distribuidas a lo largo de un área en el espacio en un momento específico del tiempo. Por ejemplo, consideremos un instante dado, en el cual se miden las temperaturas en las diferentes estaciones meteorológicas, la información obtenida forma una **serie espacial**.

Las mediciones de los datos meteorológicos se efectúan con el fin de obtener información de los procesos meteorológicos. Esta información se utiliza para entender mejor estos procesos y como información de entrada en modelos hidrológicos, para análisis y toma de decisiones.

1.2.1. Instrumentos de medición

Los instrumentos con los que se miden los datos meteorológicos son varios, sin embargo, dado que en este trabajo nos interesa el análisis de temperaturas, sólo se describe el termómetro empleado en las estaciones meteorológicas.

El aparato con el que se miden las temperaturas mínima, máxima y ambiente es el **termómetro tipo Six** (Figura 3), que cuenta con una columna de mercurio y dos índices metálicos sobre una escala en grados Celsius.

Se halla instalado dentro del abrigo de madera para protección contra el sol; cuenta con un imán que sirve para llevar los índices a cero, dejando el termómetro listo para nuevas mediciones.

Para medir la temperatura se procede como sigue:

1. Se hace la lectura de la temperatura máxima en la rama derecha del termómetro, en el extremo del índice más cercano al mercurio. Aunque la temperatura máxima ocurre entre las dos y las cuatro de la tarde del día anterior a la observación, ésta se anotará en el día correspondiente a la observación.
2. Se procede a la lectura de la temperatura mínima, el índice de la rama izquierda, también en su extremo más cercano al mercurio. Este valor corresponde a la mínima de la misma fecha de la observación.
3. La lectura de la temperatura ambiente se realiza en la rama izquierda del termómetro, en el extremo de la columna de mercurio.

1.2.2. Estaciones meteorológicas

Para el estudio de la cuenca del Valle de Puebla se cuenta con datos diarios de temperatura máxima y temperatura mínima de los registros de las estaciones, que se muestran en la TABLA 1, correspondientes a los años de 1943 a 2004, con varios faltantes, en algunos casos incluso años completos.

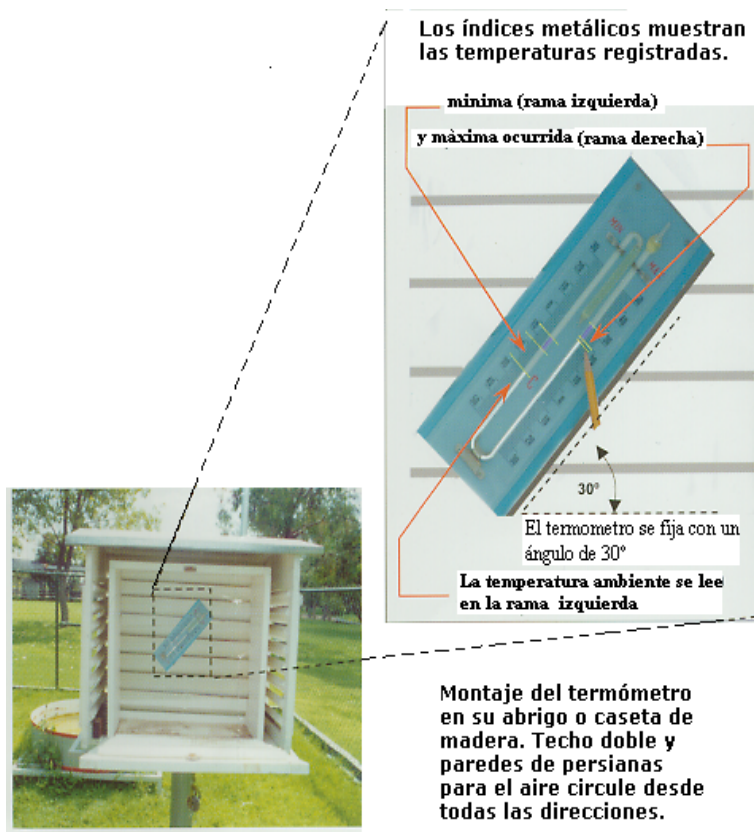


Figura 3. Termómetro.

TABLA 1. Nombre y localización de las estaciones meteorológicas.

| Número | Estación | Municipio | Latitud | Longitud |
|--------|---------------------|------------------------|---------|----------|
| 1 | Acajete | Tepatlatxco de Hidalgo | 19° 07' | 97° 58' |
| 2 | Atlixco | Atlixco | 18° 55' | 98° 25' |
| 3 | Balcón del Diablo | Tzicatlacoyan | 18° 55' | 98° 07' |
| 4 | San Baltazar Tetela | Tzicatlacoyan | 18° 54' | 98° 06' |
| 5 | San Miguel Canoa | Puebla | 19° 09' | 98° 06' |
| 6 | Echeverría | Puebla, Ocoyucan | 18° 59' | 98° 17' |
| 7 | Puebla | Puebla | 18° 59' | 98° 11' |
| 8 | Teolocholco | Teolocholco, Tlax. | 19° 05' | 98° 08' |

La información meteorológica fue proporcionada por la Comisión Nacional del Agua (CNA) del estado de Puebla (discos de 1.44 MB), CNA del estado de Tlaxcala (fotocopias) y CNA desde su coordinación nacional (Archivo de Word). Puesto que el formato de datos

que maneja CNA no es apropiado para nuestro procesamiento, se organizó la información proporcionada por dichas instancias en cuatro tablas de datos por cada estación: una tabla para temperaturas máximas, otra para temperaturas mínimas, una más para precipitación pluvial y finalmente una para evaporación. Cada columna de las tablas contiene la serie de tiempo anual de temperaturas máximas, temperaturas mínimas, precipitación o evaporación, según corresponda, desde 1943 a 2004, aunque en ciertos casos faltan datos en algunos intervalos de tiempo.

En total se cuenta con bases de datos de temperaturas máximas y mínimas diarias, bases de datos de precipitación y evaporación con datos diarios de 34 estaciones meteorológicas localizadas en el estado de Tlaxcala; 33 estaciones localizadas en el estado de Puebla; 6 estaciones del estado de México y una del estado de Morelos. La información antes mencionada se recopiló para contar con la plataforma de datos necesaria para formular un modelo de balance hídrico en la cuenca del Alto Atoyac. Para capturar la información escrita se contó con el apoyo de alumnos de servicio social y algunos becarios. Para organizar la información proporcionada en los discos de 1.44 MB y los demás archivos proporcionados por CNA, se elaboraron dos programas en Matlab.

Como este trabajo está restringido al estudio de la cuenca del Valle de Puebla, sólo consideramos las estaciones que se enumeran en la TABLA 1, siete de las cuales pertenecen al estado de Puebla y una al estado de Tlaxcala. Las estaciones que se están considerando no son las únicas que registran el comportamiento de la temperatura dentro de la cuenca, sin embargo algunas no fueron consideradas porque no poseen suficiente información y otras porque no están incorporadas a CNA.

Se realizó también un programa en Matlab que permitió determinar cuántos datos faltan por año en cada estación, tanto de temperatura máxima (TABLA 2) como de temperatura mínima (TABLA 3); el vacío indica que no hay datos en ese año y el 0 indica que no falta ningún dato.

TABLA 2. Número de datos faltantes de temperatura máxima.

| | Balcón | Acajete | Canoa | Baltazar | Echeverría | Puebla | Teolocholco | Atlixco |
|------|--------|---------|-------|----------|------------|--------|-------------|---------|
| 1943 | 334 | 365 | | | 338 | | | |
| 1944 | 0 | 278 | | | 0 | | | |
| 1945 | 0 | 0 | | | 0 | | | |
| 1946 | 0 | 30 | | | 0 | | | |
| 1947 | 0 | 92 | | | 0 | | | |
| 1948 | 365 | 152 | | | 0 | | | |
| 1949 | 0 | 0 | | | 0 | | | |
| 1950 | 31 | 215 | | | 289 | | | |
| 1951 | 0 | 62 | | | 365 | | | |
| 1952 | 0 | 31 | | | 365 | 365 | | |
| 1953 | 0 | 31 | | | 31 | 304 | | |
| 1954 | 0 | 2 | | 143 | 0 | 31 | | |
| 1955 | 0 | 0 | | 30 | 0 | 0 | | |
| 1956 | 0 | 0 | | 0 | 0 | 0 | | |
| 1957 | 0 | 0 | | 0 | 0 | 0 | | |
| 1958 | 28 | 31 | | 28 | 28 | 28 | | |
| 1959 | 0 | 244 | | 0 | 0 | 0 | | |
| 1960 | 0 | 61 | | 0 | 0 | 0 | | |
| 1961 | 7 | 365 | | 7 | 0 | 7 | | |
| 1962 | 7 | 365 | | 37 | 0 | 7 | | 174 |
| 1963 | 7 | 365 | | 38 | 0 | 7 | | 31 |
| 1964 | 6 | 365 | | 7 | 0 | 6 | | 0 |
| 1965 | 7 | 365 | | 7 | 0 | 20 | | 0 |
| 1966 | 97 | 365 | | 97 | 92 | 98 | | 92 |
| 1967 | 7 | 365 | | 7 | 0 | 7 | 0 | |
| 1968 | 6 | 365 | | 35 | 0 | 6 | | 0 |
| 1969 | 31 | 334 | | 0 | 31 | 0 | | 0 |
| 1970 | 0 | 365 | | 0 | 0 | 0 | | 60 |
| 1971 | 0 | 365 | | 0 | 0 | 0 | | 0 |
| 1972 | 0 | 365 | | 0 | 0 | 0 | | 0 |
| 1973 | 0 | 334 | | 0 | 0 | 0 | | 0 |
| 1974 | 0 | 365 | | 0 | 0 | 0 | | 0 |

CONTINUACIÓN TABLA 2

| | Balcón | Acajete | Canoa | Baltazar | Echeverría | Puebla | Teolocholco | Atlixco |
|------|--------|---------|-------|----------|------------|--------|-------------|---------|
| 1975 | 0 | 0 | | 0 | 0 | 0 | | 0 |
| 1976 | 0 | 53 | | 0 | 0 | 0 | | 0 |
| 1977 | 0 | 365 | 318 | 0 | 0 | 0 | | 30 |
| 1978 | 212 | 63 | 0 | 0 | 0 | 0 | | 0 |
| 1979 | 0 | 244 | 0 | 0 | 0 | 0 | | 0 |
| 1980 | 0 | 96 | 0 | 0 | 0 | 0 | 334 | 0 |
| 1981 | 0 | 304 | 0 | 0 | 0 | 0 | | 0 |
| 1982 | 0 | 365 | 0 | 0 | 0 | 0 | | 0 |
| 1983 | 31 | 7 | 0 | 0 | 0 | 0 | | 0 |
| 1984 | 0 | 123 | 0 | 6 | 6 | 6 | | 0 |
| 1985 | 0 | 306 | 0 | 7 | 7 | 7 | | 73 |
| 1986 | 0 | 1 | 227 | 248 | 7 | 36 | | 352 |
| 1987 | 0 | 0 | 149 | 249 | 7 | 7 | 222 | 365 |
| 1988 | 29 | 59 | 248 | 365 | 6 | 6 | 14 | 365 |
| 1989 | 30 | 31 | 91 | 365 | 7 | 7 | 185 | 365 |
| 1990 | 273 | 365 | 0 | 365 | 45 | 7 | | 182 |
| 1991 | 0 | 0 | 0 | 273 | 24 | 0 | | 304 |
| 1992 | 5 | 1 | 31 | 0 | 43 | 0 | | 184 |
| 1993 | 0 | 1 | 0 | 31 | 0 | 0 | 93 | 61 |
| 1994 | 31 | 152 | 275 | 61 | 0 | 0 | 0 | 365 |
| 1995 | 62 | 365 | 212 | 93 | 365 | 61 | 0 | 365 |
| 1996 | 61 | 365 | 62 | 92 | 220 | 0 | 0 | 365 |
| 1997 | 0 | 365 | 0 | 150 | 6 | 0 | 0 | 365 |
| 1998 | 12 | 365 | 245 | 1 | 365 | 0 | 0 | 305 |
| 1999 | 0 | 365 | 6 | 66 | 365 | 0 | 0 | 154 |
| 2000 | 17 | 120 | 103 | 34 | 365 | 0 | 92 | 30 |
| 2001 | 21 | 2 | 36 | 158 | 365 | 0 | 0 | 4 |
| 2002 | 1 | 0 | 1 | | 209 | 0 | 0 | 1 |
| 2003 | 13 | 3 | 0 | | 0 | 0 | 0 | |
| 2004 | 0 | 7 | 0 | | 150 | 0 | 0 | |

TABLA 3. Número de datos faltantes de temperatura mínima.

| | Puebla | Balcón | Acajete | Canoa | Echeverría | Baltazar | Teolocholco | Atlixco |
|------|--------|--------|---------|-------|------------|----------|-------------|---------|
| 1943 | | 334 | 365 | | 338 | | | |
| 1944 | | 0 | 279 | | 0 | | | |
| 1945 | | 0 | 0 | | 0 | | | |
| 1946 | | 1 | 30 | | 0 | | | |
| 1947 | | 0 | 92 | | 0 | | | |
| 1948 | | 365 | 152 | | 0 | | | |
| 1949 | | 0 | 0 | | 0 | | | |
| 1950 | | 31 | 215 | | 288 | | | |
| 1951 | | 0 | 62 | | 365 | | | |
| 1952 | 365 | 0 | 30 | | 3656 | | | |
| 1953 | 304 | 0 | 31 | | 31 | | | |
| 1954 | 31 | 0 | 0 | | 0 | 144 | | |
| 1955 | 0 | 0 | 0 | | 0 | 30 | | |
| 1956 | 0 | 0 | 0 | | 0 | 0 | | |
| 1957 | 0 | 0 | 0 | | 0 | 0 | | |
| 1958 | 28 | 28 | 30 | | 28 | 28 | | |
| 1959 | 0 | 0 | 242 | | 0 | 0 | | |
| 1960 | 0 | 0 | 61 | | 0 | 0 | | |
| 1961 | 0 | 0 | 365 | | 0 | 0 | | |
| 1962 | 0 | 0 | 365 | | 0 | 31 | | 174 |
| 1963 | 0 | 0 | 365 | | 0 | | 31 | |
| 1964 | 0 | 0 | 365 | | 0 | 0 | | 0 |
| 1965 | 14 | 0 | 365 | | 0 | 0 | | 0 |
| 1966 | 92 | 92 | 365 | | 92 | 92 | | 92 |
| 1967 | 0 | 0 | 365 | | 0 | 0 | | 0 |
| 1968 | 0 | 0 | 365 | | 0 | 30 | | |
| 1969 | 0 | 32 | 334 | 31 | 1 | | | |
| 1970 | 0 | 0 | 365 | | 0 | 0 | | 60 |
| 1971 | 0 | 0 | 365 | | 0 | 0 | | 0 |
| 1972 | 0 | 0 | 365 | | 0 | 0 | | 0 |
| 1973 | 0 | 0 | 334 | | 0 | 0 | | 0 |
| 1974 | 0 | 0 | 365 | | 0 | 0 | | 0 |

CONTINUACIÓN TABLA 3

| | Puebla | Balcón | Acajete | Canoa | Echeverría | Baltazar | Teolocholco | Atlixco |
|------|--------|--------|---------|-------|------------|----------|-------------|---------|
| 1975 | 0 | 0 | | 0 | 0 | 0 | | 0 |
| 1976 | 0 | 0 | 53 | | 0 | 0 | | 0 |
| 1977 | 0 | 0 | 365 | 318 | 0 | 0 | | 30 |
| 1978 | 0 | 212 | 63 | 0 | 0 | 0 | | 0 |
| 1979 | 0 | 0 | 244 | 0 | 0 | 0 | | 0 |
| 1980 | 0 | 0 | 96 | 0 | 0 | 0 | 334 | 0 |
| 1981 | 0 | 0 | 304 | 0 | 0 | 0 | | 0 |
| 1982 | 0 | 0 | 365 | 0 | 0 | 0 | | 0 |
| 1983 | 0 | 31 | 7 | 0 | 0 | 0 | | 0 |
| 1984 | 0 | 0 | 122 | 0 | 0 | 0 | | 0 |
| 1985 | 0 | 0 | 306 | 0 | 0 | 0 | | 73 |
| 1986 | 30 | 0 | 0 | 227 | 0 | 245 | | 352 |
| 1987 | 0 | 0 | 0 | 148 | 0 | 245 | 221 | 365 |
| 1988 | 0 | 29 | 59 | 248 | 0 | 365 | 14 | 365 |
| 1989 | 0 | 30 | 31 | 91 | 0 | 365 | 184 | 365 |
| 1990 | 0 | 273 | 365 | 0 | 38 | 365 | | 182 |
| 1991 | 0 | 0 | 0 | 91 | 24 | 273 | | 304 |
| 1992 | 0 | 5 | 0 | 31 | 44 | 0 | | 184 |
| 1993 | 0 | 0 | 1 | 0 | 0 | 31 | 93 | 61 |
| 1994 | 0 | 31 | 152 | 275 | 0 | 61 | 0 | 365 |
| 1995 | 61 | 62 | 365 | 212 | 365 | 93 | 0 | 365 |
| 1996 | 0 | 61 | 365 | 62 | 220 | 92 | 0 | 365 |
| 1997 | 0 | 0 | 365 | 0 | 6 | 150 | 0 | 365 |
| 1998 | 0 | 12 | 365 | 245 | 365 | 0 | 0 | 304 |
| 1999 | 0 | 0 | 365 | 6 | 365 | 62 | 0 | 154 |
| 2000 | 0 | 17 | 120 | 104 | 365 | 32 | 92 | 30 |
| 2001 | 0 | 21 | 2 | 36 | 365 | 157 | 0 | 3 |
| 2002 | 0 | 1 | 0 | 1 | 210 | | 0 | 1 |
| 2003 | 0 | 0 | 0 | 0 | | | 0 | |
| 2004 | 0 | 0 | 7 | 0 | | | 0 | |

Capítulo 2

Obtención de datos faltantes

En vista de que a las ocho estaciones consideradas (TABLA 1) les faltan datos de temperatura máxima (TABLA 2) y temperatura mínima (TABLA 3) en este capítulo se describe un procedimiento para obtener dichos datos faltantes. Las nuevas tablas de datos generadas se emplean en el capítulo siguiente, principalmente para revisar si se detectan periodicidades importantes mayores a la anual.

La idea central del procedimiento para obtener datos faltantes es aprovechar que no hay grandes cambios climáticos en la región donde se ubican las estaciones y, en consecuencia, la semejanza que tienen los datos de temperaturas máximas/mínimas del mismo año de estaciones cercanas, geográficamente hablando. Se propone que si se desconoce la temperatura máxima en cierto día pero se conoce la del mismo día en una estación con la que se correlaciona altamente (mayor o igual a 0.75), entonces se emplee el dato conocido para estimar el faltante. Para validar esta propuesta se ubicaron gráficamente las estaciones meteorológicas en la cuenca del Valle de Puebla, empleando el sistema de información geográfica ArcView GIS version 3.1 (Figura 4); posteriormente se calcularon los coeficientes de correlación entre series de tiempo del mismo año de estaciones distintas y cercanas, geográficamente hablando; también se calcularon los coeficientes de autocorrelación entre los datos de una misma serie anual.

Para calcular la correlación y la autocorrelación entre las series de temperatura se crearon dos programas en Matlab que calculan los coeficientes antes mencionados, en dicho cálculo se empleó la siguiente relación de los coeficientes de correlación R con la

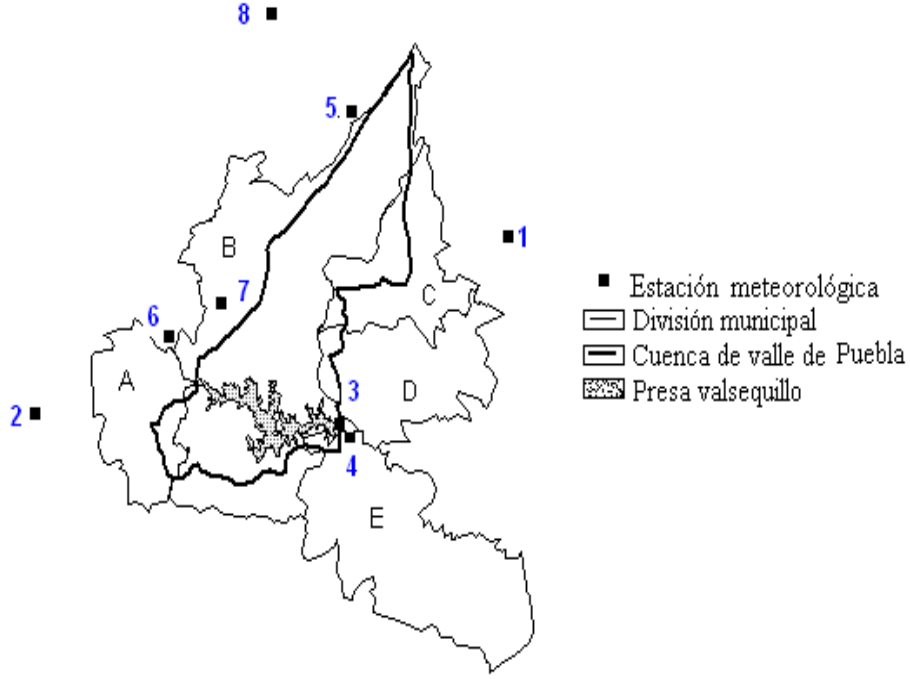


Figura 4. Para identificar las estaciones ver TABLA 1.

matriz de covarianza C (Johnson, 1976; Mendenhall, 2002) de una matriz M que contiene como columnas los datos de temperaturas de años completos de estaciones cercanas.

$$R(i, j) = \frac{C(i, j)}{\sqrt{C(i, i)C(j, j)}}. \quad (2.1)$$

Para estimar la autocorrelación de longitud k , Matlab emplea el estimador Yule-Walker dado por:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}, \quad (2.2)$$

para el caso en el que $k = 1$ se tiene

$$\hat{\rho}_1 = \frac{\sum_{t=1}^{n-1} (x_t - x_{t+1})^2}{\sum_{t=1}^n (x_t - \bar{x})^2}. \quad (2.3)$$

En el caso anterior se estima la correlación entre los elementos de $\{x_{1+k}, x_{2+k}, \dots, x_n\}$ y $\{x_1, x_2, \dots, x_{n-k}\}$.

La siguiente tabla muestra los coeficientes de correlación entre la estación Puebla y las dos estaciones con las que se correlaciona altamente:

TABLA 4. Coeficientes de correlación.

| año | Echeverría | Balcón |
|------|------------|--------|
| 1955 | 0.885 | 0.881 |
| 1956 | 0.846 | 0.857 |
| 1957 | 0.864 | 0.775 |
| 1959 | 0.835 | 0.765 |
| 1960 | 0.837 | 0.835 |
| 1970 | 0.77 | 0.787 |
| 1971 | 0.763 | 0.835 |
| 1972 | 0.841 | 0.823 |
| 1980 | 0.817 | 0.854 |
| 1982 | 0.86 | 0.87 |
| 1993 | 0.848 | 0.858 |
| 2003 | 0.8 | 0.750 |
| 2004 | 0.75 | 0.8 |

Recordemos que la correlación es la medida de la fuerza de la relación lineal entre dos variables (Johnson, 1976); en vista de los resultados de la tabla anterior, podemos emplear datos existentes de temperatura máxima en dos de las estaciones para estimar los faltantes en la otra. De aquí que se emplearan los datos de las estaciones Echeverría y Balcón para estimar los faltantes en la estación Puebla. Sin embargo algunos años seguían teniendo faltantes (de 1 a 8), así que se calculó la autocorrelación en las series anuales para estimar por medio de los datos de temperatura anteriores y posteriores a los faltantes. La siguiente tabla muestra la autocorrelación en algunos años de la estación Puebla:

TABLA 5. Coeficientes de autocorrelación.

| longitud k | 2004 | 2000 | 1995 | 1990 |
|------------|------|--------|-------|-------|
| 1 | 0.85 | 0.65 | 0.81 | 0.79 |
| 2 | 0.8 | 0.511 | 0.8 | 0.78 |
| 3 | 0.74 | 0.4238 | 0.795 | 0.777 |
| 4 | 0.65 | 0.3613 | 0.78 | 0.76 |
| 5 | 0.6 | 0.330 | 0.77 | 0.758 |
| 6 | 0.59 | 0.28 | 0.765 | 0.741 |
| 7 | 0.55 | 0.25 | 0.742 | 0.721 |
| 8 | 0.5 | 0.23 | 0.69 | 0.710 |

La autocorrelación es una medida que indica cómo se relacionan las variables x en el tiempo con respecto a observaciones pasadas de x . Se obtuvieron tablas similares con los datos de las otras estaciones, tanto para temperatura máxima como para temperatura mínima; la TABLA 6 proporciona el nombre de las estaciones empleadas para obtener datos faltantes tanto de temperaturas máximas como de temperaturas mínimas de cada estación; de dichos resultados se propone el siguiente algoritmo para obtener datos faltantes.

2.1. Algoritmo para obtener los datos faltantes

Basados en los resultados obtenidos, se creó un programa en Matlab para obtener los datos faltantes de las tablas de temperaturas (máximas o mínimas) de cualquiera de las estaciones de la cuenca, este programa consta de dos partes esenciales:

- Obtención de datos faltantes de una estación empleando datos de las dos estaciones con las que más se correlaciona.
- Autollenado.

A continuación se detalla el algoritmo que se programó en Matlab para obtener datos faltantes.

Algoritmo

1. Se tienen como variables de entrada 3 bases de datos: la base de datos 1 es la tabla de temperaturas máximas/mínimas de la estación en la que se desea obtener datos faltantes; las bases de datos 2 y 3 son las tablas de temperaturas máximas/mínimas de las dos estaciones con las que más se correlaciona la estación de interés.
2. (Parte 1)
 - a) Localizar todos los datos faltantes de la base de datos 1, reemplazar por los datos de la misma fecha, si es que existen, de la base de datos 2, de este modo se genera una nueva base de datos 4.
 - b) Localizar todos los datos faltantes de la base de datos 4, reemplazarlos por los datos de la misma fecha, si es que existen, de la base de datos 3, así se genera una nueva base de datos 5, y se pasa a la Parte 2.
3. (Parte 2) Si en algún año de la base de datos 5 siguen faltando datos (menos de ocho), entonces se llenan estos vacíos con el siguiente criterio: si existen datos anterior y posterior, el dato anterior se emplea para obtener la mitad de los faltantes y el dato posterior se emplea para obtener la otra mitad. En caso de no haber dato anterior o posterior, entonces se inserta el único dato existente en lugar de los datos faltantes.

Programando en Matlab el algoritmo antes descrito (véase Apéndice A) y aplicando el procedimiento en cada una de las ocho estaciones, se obtuvieron los resultados reportados en la TABLA 6. Nótese que la estación Zacatelco no pertenece a la cuenca, pero por su cercanía con Teolocholco se consideró para aportar los datos faltantes en dicha estación. En la columna DATOS OBTENIDOS se muestran los años de datos completos posteriores a 1966 obtenidos después de completar mediante el procedimiento descrito.

TABLA 6. Datos obtenidos.

| ESTACIÓN | ESTACIONES EMPLEADAS | DATOS OBTENIDOS |
|-------------|----------------------|-----------------------|
| Acajete | Puebla-Teolocholco | 1967-2004 |
| Atlixco | Echeverría-Puebla | 1967-1994 y 1996-2004 |
| Balcón | Baltazar-Puebla | 1967-2004 |
| Baltazar | Balcón-Puebla | 1967-2004 |
| Canoa | Puebla | 1964-1985 y 1987-2004 |
| Echeverría | Puebla-Balcón | 1967-2004 |
| Puebla | Echeverría-Balcón | 1967-2004 |
| Teolocholco | Zacatelco | 1994-2004 |

Capítulo 3

Filtrado de las series de temperaturas

En este capítulo se ejemplifica el análisis espectral empleando los datos de temperatura máxima de la estación Puebla; se realiza un análisis similar en las otras estaciones, tanto para temperatura máxima como para temperatura mínima. El análisis espectral esencialmente sirve para descubrir las eventuales periodicidades ocultas que posee una serie de datos, para representar dicha serie por un polinomio trigonométrico. Claramente se intuye que existe una periodicidad anual y trimestral en la temperatura, que se observa en las estaciones del año: primavera, verano, otoño e invierno, a causa de las variaciones climáticas que sufre la tierra, debido a la inclinación del eje terrestre. Dividimos el trabajo en dos partes: En la primera se aplica el análisis espectral a la cadena de datos de temperaturas máximas/mínimas de la estación Puebla sin faltantes de 1967-2004, obtenidos en el capítulo anterior, para ejemplificar, con los datos de esta estación, que la periodicidad detectable (con el periodo de tiempo de 1967 al 2004) más amplia es la anual. Una vez detectada la existencia de periodicidades anuales en las series de temperaturas de cada estación, se convino que la curva suavizada debe representar la variación de la temperatura a lo largo de un año. En la segunda parte se descubren las frecuencias representativas de cada serie anual y posteriormente se seleccionan las frecuencias que se emplean en el siguiente capítulo para representar la variación de temperaturas por un polinomio trigonométrico. Finalmente se presenta una serie de datos filtrada que describe

las series de temperaturas máximas y otra que describe las temperaturas mínimas de cada estación. Para obtener dicha serie de datos empleamos el análisis espectral (Chatfield, 2003).

El análisis espectral para funciones discretas consiste en calcular una función denominada periodograma (Gottman, 1984), la cual se interpreta como la energía total de la función en la frecuencia ω , también se denomina el espectro (Chatfield, 2003; Bracewell, 2000; Brigham, 1988).

Para comprender el concepto de espectro consideremos que una serie x_t , con $t = 1, \dots, N$; se aproxima por un modelo simple

$$x_t = a \cos(2\pi ft) + b \sin(2\pi ft) \quad (3.1)$$

empleando mínimos cuadrados tenemos

$$\hat{a} = 2 \frac{\left[\sum_0^N x_t \cos(2\pi ft) \right]}{N} \quad (3.2)$$

$$\hat{b} = 2 \frac{\left[\sum_0^N x_t \sin(2\pi ft) \right]}{N}. \quad (3.3)$$

El periodograma para este modelo se define por (Chatfield, 2003)

$$Sp(f) = \frac{N}{4\pi} \frac{R_p^2}{2},$$

donde $R_p = \sqrt{a^2 + b^2}$, de aquí que el periodograma proporcione información acerca de la contribución que tiene la frecuencia $\omega = 2\pi f$ en la descripción de x_t . A una gráfica de $\frac{R_p^2}{2}$ contra ω se la conoce con el nombre de espectro de línea (Chatfiel, 2003); así, un pico en la gráfica del espectro indica una contribución importante de la frecuencia en la serie x_t . Realizando algunos cálculos se obtiene una expresión del periodograma en términos de la norma cuadrada de la transformada discreta de Fourier de x_t (Fuller, 1976, Brigham, 1988):

$$Sp(f) = \frac{N}{4\pi} \frac{\left\| \sum_{t=1}^N x_t \exp^{i\omega t} \right\|^2}{2}. \quad (3.4)$$

De manera natural se obtienen resultados similares cuando $x(t)$, para $t = 1, \dots, N$; se aproxima por un polinomio trigonométrico de grado $m > 1$ dado por

$$x(t) = C_1 + \sum_{j=1}^m C_{2j} \cos[f(j)(2\pi/365)t] + C_{2j+1} \sin[f(j)(2\pi/365)t.] \quad (3.5)$$

donde C_i , para $i = 1, \dots, 2m + 1$, son los coeficientes de la serie, y $f(j)$, para $j = 1, \dots, m$, son las frecuencias representativas de la serie. A la expresión (3.5) se la conoce como representación en serie de Fourier de $x(t)$. De manera análoga, se encuentra una expresión del espectro Sp en términos de la transformada discreta de Fourier, excepto por una constante que en nuestros cálculos no consideramos:

$$Sp(\omega) = \left(\frac{2}{N}\right)^2 \left[\sum_{t=1}^N x_t \exp^{i\omega t} \right] \left[\sum_{t=1}^N x_t \exp^{-i\omega t} \right] \quad (3.6)$$

donde $\omega = \frac{2\pi}{365}f(j)$. Nuevamente $Sp(\omega)$ representa una cantidad que se interpreta como la contribución que tiene la frecuencia $f(j)$ en la descripción de $x(t)$. La relación que se presenta en (3.6) y la interpretación que hemos dado al periodograma en las líneas anteriores son fundamentales en el desarrollo de nuestro trabajo: nos permiten realizar un análisis espectral para descubrir frecuencias $f(j)$ importantes en los datos de temperaturas máximas/mínimas, empleando la transformada de Fourier, que se sustituyen en el modelo (3.5) para aproximar a los datos de temperaturas.

A continuación se describe la manera en la que se realizó el análisis espectral para verificar que la periodicidad oculta más amplia detectable en los datos de temperaturas máximas/mínimas de que disponemos es la anual. Se elaboró un programa en Matlab que permitió graficar todos los años de temperaturas máximas como una sola serie de tiempo de años consecutivos por estación, para realizar un análisis espectral en la serie concatenada de los últimos años de cada una de las tablas de datos, con el fin de observar eventuales cambios considerables en los datos de temperatura y obtener, si existen, periodos más amplios que el anual que puedan influir en los resultados que se desean presentar; sin embargo se observó que la periodicidad relevante es de un año.

Como se mencionó al principio del capítulo, con el fin de ejemplificar el análisis espectral, se eligieron aleatoriamente los datos diarios de temperatura máxima de la estación

Puebla. En la Figura 5 se muestra la gráfica de años consecutivos 1967 – 2004 de temperaturas máximas de la estación Puebla. Denotando por x_t el valor de la temperatura en el día t y considerando que el 1 de enero de 1967 es el día uno y el 31 de diciembre de 2004 es el día trece mil ochocientos setenta, calculamos la trasformada discreta de Fourier X (Brigham, 1988) de la serie concatenada x , empleando un programa en Matlab, en el cual se aplica la función $\text{fft}(x)$ (Bracewell, 2000; Brigham, 1988) para un vector de longitud N :

$$X(k) = \sum_{j=1}^N x(j) w_N^{(j-1)(k-1)}, \quad (3.7)$$

donde

$$w_N = \exp^{-2\pi i/N}.$$

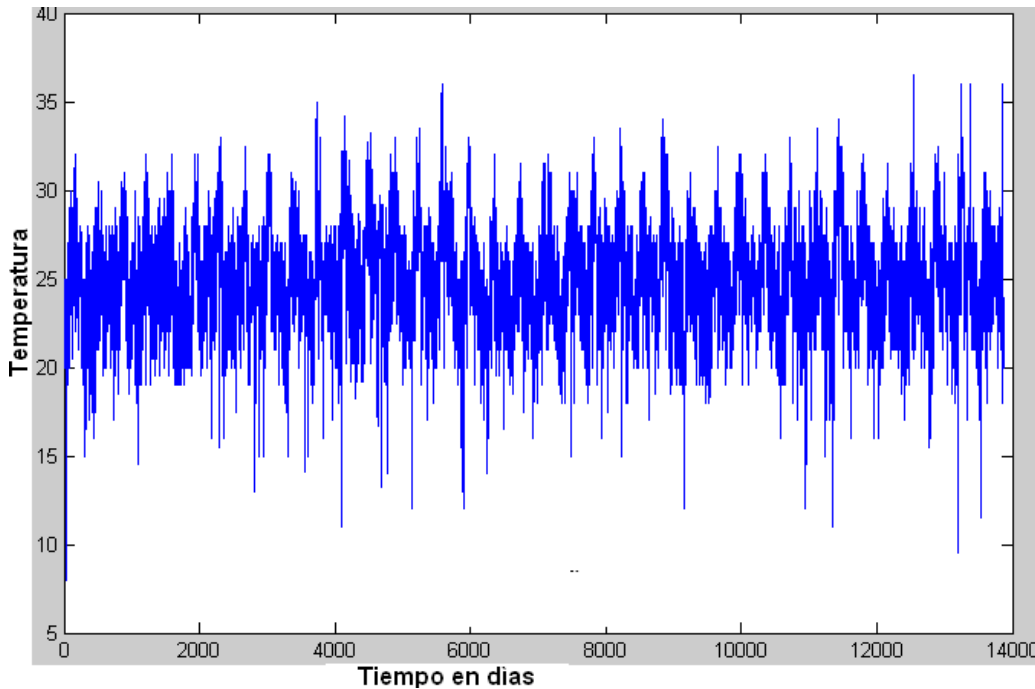


Figura 5. Temperaturas máximas de 1967-2004.

Posteriormente, se calculó el espectro de potencias $Sp(k) = X(k)X^*(k)$ (según (3.6)). Es importante recordar que la interpretación física del espectro representa la contribución a la variación de las componentes de las frecuencias (Chatfield, 2003). En otras palabras, al graficar Sp , un pico en la frecuencia ω_0 del espectro Sp indica que ω_0 contribuye de manera importante en la representación en series de Fourier de la función de temperaturas.

En la Figura 6 se muestra la gráfica del espectro de la serie concatenada, en realidad el espectro es un vector de tantas posiciones como datos se tengan, sin embargo graficamos sólo los primeros ciento cincuenta y dos valores del espectro para identificar mejor las posiciones ω en las que hay picos en Sp . Obsérvese que el pico más alto se localiza en la posición treinta y ocho, por la relación inversa de la frecuencia con el periodo se tiene $P = \frac{N * 365}{\omega}$, donde N es el número de años, ω es la frecuencia, en este caso $\omega = 38$, por lo que $P = 365$, de este modo se determina que la periodicidad más amplia es la anual, de manera análoga se procede para identificar las periodicidades menores a un año: semestrales y cuatrimestrales. En algunas otras estaciones se observó también una periodicidad trimestral.

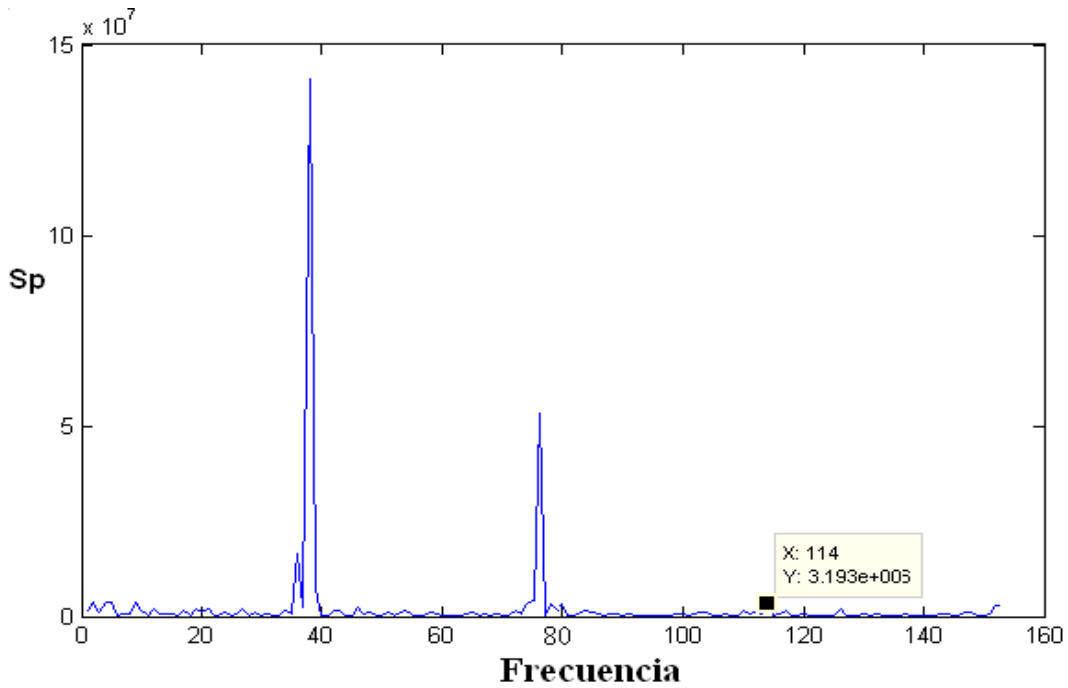


Figura 6. Espectro de la serie de temperaturas máximas de 1967-2004.

A continuación se ejemplifica el análisis espectral realizado para obtener una serie filtrada que representa la variación de temperatura de la estación Puebla. Se calculó la transformada discreta de Fourier (mediante 3.7) de la serie de tiempo de cada año; análogamente se calculó y graficó el espectro de potencias de cada serie anual (la Figura 7a y 7b muestra sólo dos de las treinta y ocho gráficas del espectro) y se identificaron los quince

valores de las frecuencias para las cuales se tiene un valor mayor del espectro. En total se obtuvieron setenta y dos frecuencias distintas que en cierto año son uno de los quince valores más grandes que se obtuvieron; de dichos valores se seleccionaron los presentes en la mayoría de los años de la estación Puebla, tal como se muestra en la TABLA 7; para fines prácticos se muestran sólo los más representativos, pues las demás frecuencias se presentaron en menos de diecisiete años de los treinta y ocho años considerados.

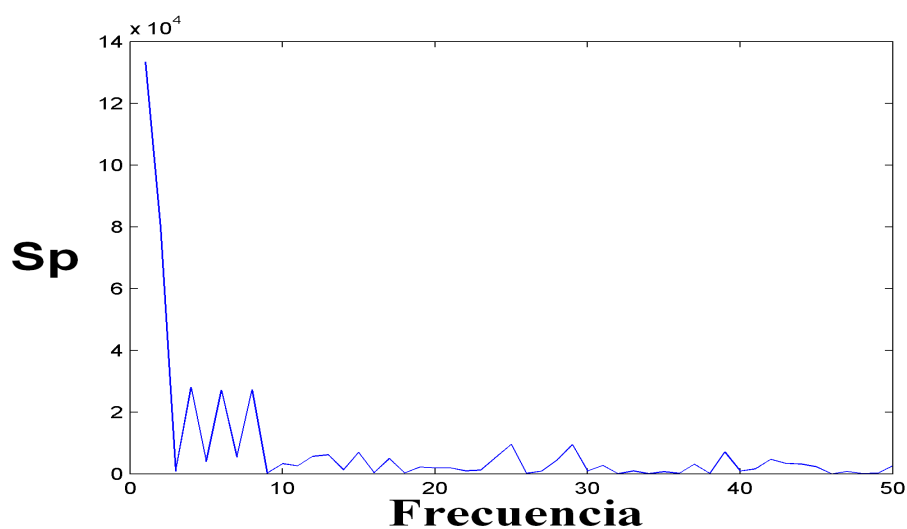


Figura 7a. Espectro del año 1976.

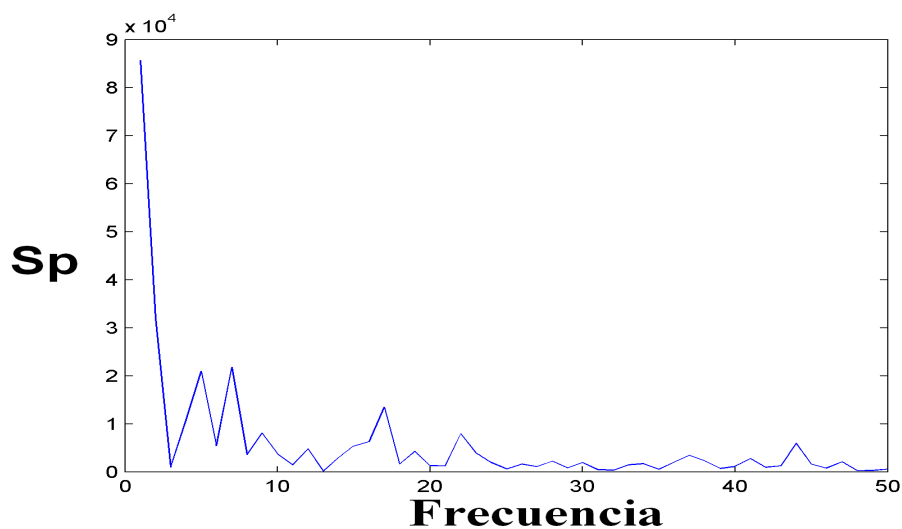


Figura 7b. Espectro del año 1989.

TABLA 7. Frecuencias representativas.

| frecuencia $f(j)$ | número de estaciones en las que está presente |
|----------------------|--|
| 1 | 38 |
| 2 | 38 |
| 3 | 37 |
| 4 | 24 |
| 5 | 32 |
| 6 | 20 |
| 7 | 18 |
| 8 | 22 |

El criterio que propusimos para seleccionar las frecuencias que se consideran para recuperar la serie es: Sea N el número de años considerados, si el número de estaciones en las que está presente la frecuencia $f(j)$ es mayor o igual que la parte entera de $N/2$ entonces dicha frecuencia se emplea para recuperar la serie.

Empleando el criterio anterior, los resultados de la TABLA 7, y la relación inversa que existe entre el periodo y la frecuencia, se deduce que las periodicidades a considerar se presentan cada año, medio año, cuatro meses, tres meses, $\frac{12}{5} \approx$ dos meses y trece días, dos meses y cada $\frac{12}{8} \approx$ un mes y medio.

Aparecen también otras frecuencias altas, sin embargo dichas frecuencias no son comunes a todos los años. Se creó un vector con las frecuencias asociadas a las periodicidades antes identificadas

$$f = \{1, 2, 3, 4, 5, 6, 8\} \quad (3.8)$$

asociado a este vector un filtro para cada una de las 38 series de tiempo. Dicho filtro es un vector de trescientas sesenta y cinco entradas definido de la siguiente manera: es un vector de ceros excepto en las posiciones $\{1, 2, 3, 4, 5, 6, 8, 358, 360, 361, 362, 363, 364, 365\}$ donde

contiene 1; como se puede observar dichas posiciones se relacionan con las frecuencias f antes identificadas.

Aplicando este filtro se recuperaron las series anuales de temperaturas máximas de la estación Puebla.

Con el fin de mantener claro el procedimiento, sean x_i las series de datos anuales, F_i con $i = 1, \dots, 38$ las series de datos filtradas obtenidas mediante el procedimiento antes descrito, de modo que $x_i(j)$ denota el valor real de la serie del año i en el día j , considerando que el año $i = 1$ es 1967 y el año $i = 38$ es 2004, el día $j = 1$ es el 1 de enero del año i y $j = 365$ es el 31 de Diciembre del año i . Deseamos encontrar una sola serie F_i que represente a todas las x_i , es decir, elegir una sola de las series filtradas que represente el comportamiento de la temperatura en ese punto del espacio. Para lograr nuestro objetivo resolvimos el siguiente problema.

Seleccionar la serie $F = F_i$ tal que

$$\text{Min}_i d_i, \quad (3.9)$$

donde

$$d_i = \sum_{k=1}^{k=38} \sqrt{\sum_{j=1}^{365} (x_k(j) - F_i(j))^2} \quad \text{para cada } i = 1, \dots, n,$$

dicha serie será elegida tomando en cuenta también el hecho de que no sea la que representa un año que tenía más de trece datos faltantes, pues recordemos que en el Capítulo 2 se obtuvieron los datos faltantes y no es apropiado tomar como modelo un año al que le faltaban muchos datos. Después de resolver el problema de minimización (3.9), que para el caso particular de los datos de temperaturas máximas de la estación Puebla resultó ser la serie filtrada de los datos del año 1993, se procede a realizar el ajuste de la serie $F = F_{27}$ por medio de mínimos cuadrados, para encontrar una expresión en términos de una suma alternada de senos y cosenos que represente la temperatura máxima diaria en todo el año en la estación Puebla, tal como se muestra en el siguiente capítulo.

Es importante notar que con el análisis anterior se determinó el número de términos que tendrá el polinomio trigonométrico, ya que éste es dos veces el número de frecuencias

que se identificaron más uno, el término constante.

Un análisis similar se efectuó en las otras estaciones, tanto para temperaturas máximas como para temperaturas mínimas. En la TABLA 8 se muestra el año de la serie filtrada que resultó de resolver (3.9), para cada una de las estaciones.

TABLA 8. Series filtradas que representan la variación de temperaturas por estación.

| Nombre | T. Máxima F | No. de datos faltantes (TABLA 2) | T. Mínima F | No. de datos faltantes (TABLA 3) |
|-------------|----------------|--|----------------|--|
| Acajete | 2003 | 3 | 2001 | 2 |
| Atlixco | 2001 | 4 | 2002 | 1 |
| Balcón | 2003 | 13 | 1993 | 0 |
| Baltazar | 1998 | 1 | 1985 | 0 |
| Canoa | 1997 | 0 | 2004 | 0 |
| Echeverría | 1993 | 0 | 1997 | 6 |
| Puebla | 1993 | 0 | 1992 | 0 |
| Teolocholco | 2003 | 0 | 1994 | 0 |

De la tabla anterior, es importante destacar que el año que se está tomando como representativo, por cada estación, fue seleccionado considerando que la serie de temperaturas de dicho año debe aproximar adecuadamente años completos o con pocos faltantes y que en sí misma fuera una serie asociada a un año con pocos faltantes.

En el capítulo discusión de resultados se presentan los valores de la desviación cuadrática media (5.1), mismos que permiten confirmar que efectivamente los polinomios trigonométricos que ajustan las series de la TABLA 8 son buenas aproximaciones, en el sentido de que describen la variación máxima/mínima en cada estación meteorológica.

Capítulo 4

Ajuste de las series Filtradas

Una vez que tenemos una serie filtrada F que representa la temperatura máxima (mínima) por estación (TABLA 8), empleamos el método de mínimos cuadrados para ajustar la serie F por medio de un polinomio trigonométrico de la forma:

$$X_F(t) = C_1 + \sum_{j=1}^9 C_{2j} \cos[f(j)(2\pi/365)t] + C_{2j+1} \sin[f(j)(2\pi/365)t] \quad (4.1)$$

para $t = 1, \dots, 365$. Donde C_j y $f(j)$ se muestran en la TABLA 9, para el caso de temperaturas máximas, y en la TABLA 10, para temperaturas mínimas. Para identificar las frecuencias $f(j)$ representativas de cada estación se procedió de la misma manera que para proponer (3.8).

Es importante mencionar que en algunas estaciones no se consideran todas las frecuencias del 1 al 9, por lo que en la suma anterior son cero los coeficientes asociados a dicha frecuencia.

Retomemos el caso de temperaturas máximas de la estación Puebla. De la TABLA 8 concluimos que la serie filtrada a ajustar corresponde a la del año 1993, y de (3.8) se determinaron las frecuencias para recuperar las series de datos; dichas frecuencias influyen en los valores de los coeficientes de (4.1), que se muestran en la novena columna de la TABLA 9. Observemos que, al parecer, el coeficiente C_1 de la serie es el más representativo, sin embargo es importante notar que se trata la media de los datos (24,9507) antes del filtrado.

4.1. Resultados

El análisis ejemplificado con la estación Puebla en la sección anterior se realizó en cada una de las ocho estaciones, tanto para temperaturas máximas como para temperaturas mínimas.

Recordemos que después de realizar el filtrado de cada una de las series anuales por estación se eligió una serie representativa de cada observatorio de datos de temperaturas máximas y una para datos de temperaturas mínimas (TABLA 8), posteriormente dicha serie filtrada se ajustó por medio de la expresión trigonométrica (4.1) empleando mínimos cuadrados. En la TABLA 9 se muestran los coeficientes para los ajustes de temperatura máxima de cada estación.

TABLA 9. Coeficientes para los ajustes de temperatura máxima.

| subíndice j | Frecuencia f(j) | Acajete | Atlixco | Balcón | Baltazar |
|-------------|-----------------|----------|-----------|-----------|-----------|
| 1 | | 24,332 | 26,434 | 23,878 | 24,655 |
| 2 | 1 | -1,39 | -0,73545 | -1,3255 | -3,5743 |
| 3 | 1 | 1,4107 | 1,1695 | 1,4479 | 1,6221 |
| 4 | 2 | -1,3193 | -1,056 | -0,48123 | -0,58562 |
| 5 | 2 | -0,12111 | -0,31889 | -0,87196 | -1,6155 |
| 6 | 3 | 0,34526 | 0,077833 | 0,40001 | 0,14613 |
| 7 | 3 | -0,43002 | -0,45108 | -0,14607 | -0,62712 |
| 8 | 4 | -0,54674 | -0,10708 | -0,77662 | -0,15396 |
| 9 | 4 | 0,28069 | 0,068003 | -0,099618 | -0,25762 |
| 10 | 5 | 0,10307 | -0,098228 | 0,68933 | 0,07979 |
| 11 | 5 | -0,17455 | 0,12264 | 0,13565 | 0,35972 |
| 12 | 6 | 0 | 0 | -0,29713 | -0,25587 |
| 13 | 6 | 0 | 0 | -0,075081 | -0,1829 |
| 14 | 7 | 0,036524 | -0,11438 | 0 | 0 |
| 15 | 7 | 0,2697 | -0,22002 | 0 | 0 |
| 16 | 8 | 0 | 0 | -0,39584 | 0 |
| 17 | 8 | 0 | 0 | 0,42828 | 0 |
| 18 | 9 | 0 | 0 | 0 | -0,085136 |
| 19 | 9 | 0 | 0 | 0 | 0,027514 |

CONTINUACIÓN TABLA 9

| subíndice j | Frecuencia f(j) | Canoa | Echeverría | Puebla | Teolocholco |
|-------------|-----------------|----------|------------|----------|-------------|
| 1 | | 21,827 | 23,29 | 24.95 | 24,205 |
| 2 | 1 | -1,2744 | -0,96613 | -1.7018 | -1,1722 |
| 3 | 1 | 1,2315 | 0,95528 | 0.81916 | 1,8907 |
| 4 | 2 | -0,91128 | -0,41653 | -0.47621 | -1,2223 |
| 5 | 2 | 0,10662 | -0,90895 | -0.69082 | -0,65286 |
| 6 | 3 | -0,10851 | 0,15705 | 0.02515 | 0,41834 |
| 7 | 3 | -0,22926 | -0,18527 | -0.19466 | -0,34228 |
| 8 | 4 | -0,65865 | -0,5555 | -0.1445 | -1,0947 |
| 9 | 4 | -0,30366 | -0,1291 | 0.10796 | 0,49325 |
| 10 | 5 | 0,18971 | 0,46337 | 0.59249 | 0,13462 |
| 11 | 5 | 0,1049 | 0,01238 | 0.07265 | -0,31226 |
| 12 | 6 | 0 | -0,46212 | -0.17915 | 0 |
| 13 | 6 | 0 | -0,14452 | 0.86947 | 0 |
| 14 | 7 | 0,22344 | 0,60232 | 0 | 0,033999 |
| 15 | 7 | 0,086835 | -0,10777 | 0 | 0,47118 |
| 16 | 8 | -0,20552 | 0 | -0.1457 | 0,067962 |
| 17 | 8 | -0,45726 | 0 | -0,2921 | -0,13206 |
| 18 | 9 | 0 | 0 | 0 | 0 |
| 19 | 9 | 0 | 0 | 0 | 0 |

En la TABLA 10 se muestran los coeficientes para el ajuste de la forma (4.1) para temperatura mínima de las estaciones consideradas.

TABLA 10. Coeficientes para los ajustes de temperatura mínima.

| subíndice j | Frecuencia f(j) | Acajete | Atlixco | Balcón | Baltazar |
|-------------|-----------------|--------------|----------|----------|----------|
| 1 | | 6,1671 | 11,647 | 8,3644 | 9,0466 |
| 2 | 1 | -3,641 | -2,4226 | -3,7465 | -4,2516 |
| 3 | 1 | -1,0729 | -0,36152 | -0,96533 | -0,21392 |
| 4 | 2 | -0,809351125 | -1,1675 | -0,67401 | -1,3175 |
| 5 | 2 | -0,73031 | -0,12411 | -0,3923 | -0,46922 |
| 6 | 3 | 0,16317 | -0,16242 | -0,26106 | -0,26557 |
| 7 | 3 | 0,96821 | 0,57321 | 0,68183 | 0,48186 |

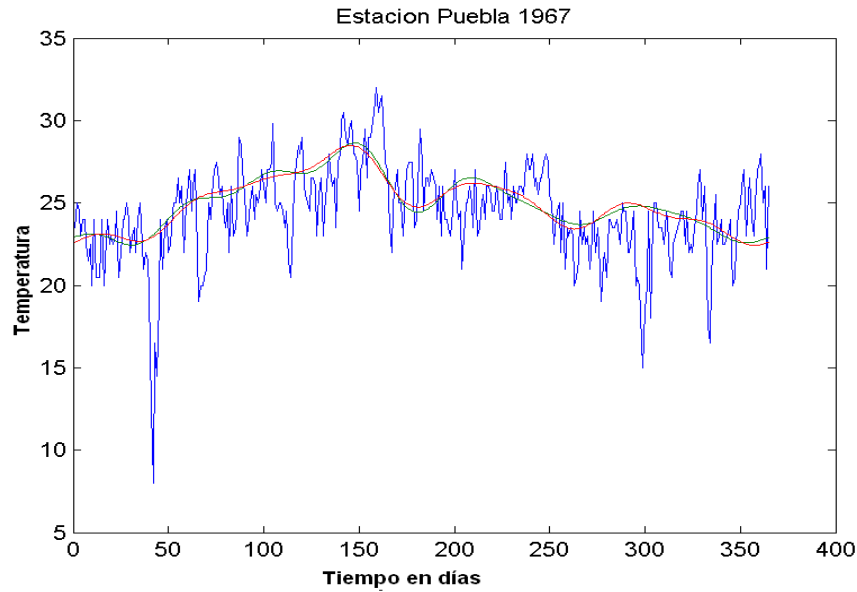
CONTINUACIÓN TABLA 10

| subíndice j | Frecuencia f(j) | Acajete | Atlixco | Balcón | Baltazar |
|-------------|-----------------|-----------|-----------|------------|-----------|
| 8 | 4 | -0,052749 | 0,31726 | 0,16781 | 0,6458 |
| 9 | 4 | -0,20477 | 0,19592 | -0,2757 | 0,10961 |
| 10 | 5 | 0,44743 | -0,081334 | -0,16456 | 0,53507 |
| 11 | 5 | 0,051256 | 0,18357 | 0,18395 | 0,11201 |
| 12 | 6 | 0,33703 | 0,16247 | 0 | 0,046843 |
| 13 | 6 | -0,21439 | 0,010414 | 0 | -0,28431 |
| 14 | 7 | 0,69688 | 0 | -0,0090168 | 0,5169 |
| 15 | 7 | 0,099689 | 0 | 0,19619 | -0,10377 |
| 16 | 8 | -0,14072 | -0,04559 | 0 | -0,049797 |
| 17 | 8 | -0,050527 | -0,26777 | 0 | -0,094874 |
| 18 | 9 | 0 | 0 | 0 | 0 |
| 19 | 9 | 0 | 0 | 0 | 0 |

| subíndice j | Frecuencia f(j) | Canoa | Echeverría | Puebla | Teolocholco |
|-------------|-----------------|------------|------------|-----------|-------------|
| 1 | | 6,6614 | 7,3507 | 9,5685 | 7,3973 |
| 2 | 1 | -2,4703 | -3,8773 | -3,2252 | -2,7338 |
| 3 | 1 | -0,17735 | -1,2681 | -0,534 | -0,44216 |
| 4 | 2 | -0,58022 | -1,5709 | -0,87984 | -0,96392 |
| 5 | 2 | -0,49213 | -0,59408 | -0,30115 | -0,20133 |
| 6 | 3 | 0,045139 | -0,65517 | -0,33775 | -0,042522 |
| 7 | 3 | 0,54073 | -0,064987 | 0,010745 | 0,47311 |
| 8 | 4 | 0,11114 | 0,62006 | 0,25016 | -0,066866 |
| 9 | 4 | -0,1249 | -0,76531 | -0,11489 | -0,45301 |
| 10 | 5 | 0,1417 | -0,39937 | 0,19643 | 0,17912 |
| 11 | 5 | 0,024116 | 0,12177 | 0,6349 | 0,036692 |
| 12 | 6 | 0 | -0,34559 | -0,18642 | 0 |
| 13 | 6 | 0 | -0,27361 | 0,31751 | 0 |
| 14 | 7 | 0,00050777 | 0,49488 | 0,23594 | 0 |
| 15 | 7 | 0,08708 | -0,34442 | 0,01805 | 0 |
| 16 | 8 | 0 | 0,65404 | 0,056363 | 0 |
| 17 | 8 | 0 | -0,13163 | -0,045723 | 0 |
| 18 | 9 | 0 | 0 | 0 | 0 |
| 19 | 9 | 0 | 0 | 0 | 0 |

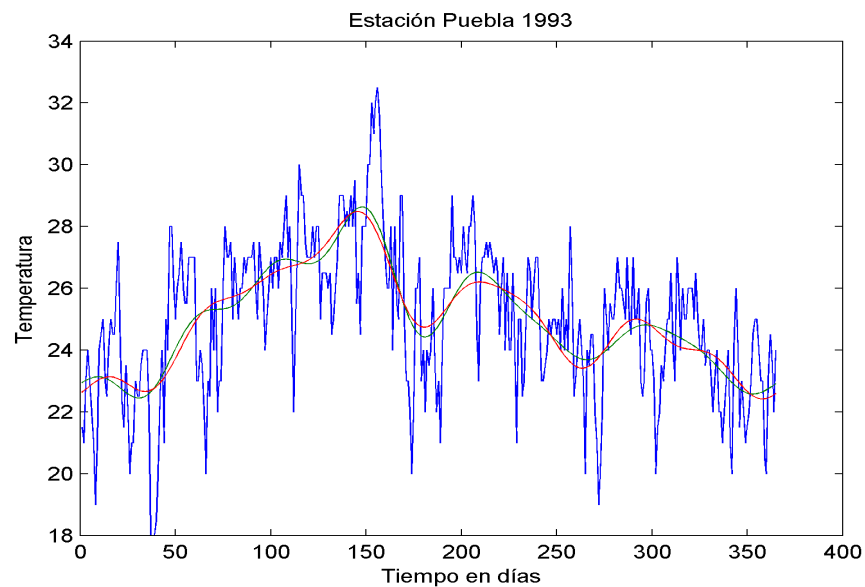
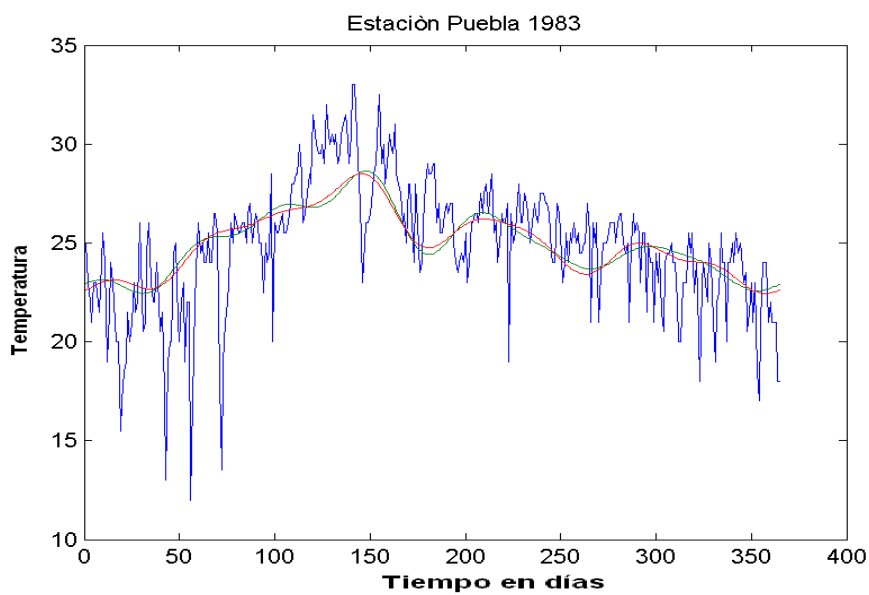
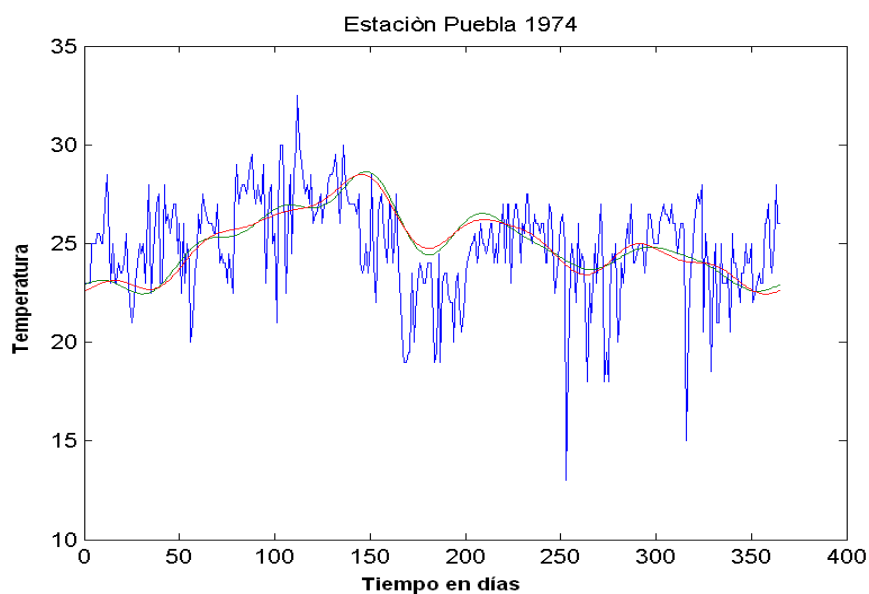
Esto muestra que los polinomios trigonométricos tienen como máximo 17 términos. En otras palabras, se recupera la serie temporal de temperaturas con menos de nueve frecuencias, dichas frecuencias no son las mismas en todos los observatorios, más aún, no coinciden las frecuencias a pesar de tratarse de la misma estación, pues no se emplean las mismas frecuencias para recuperar temperaturas máximas que temperaturas mínimas y viceversa.

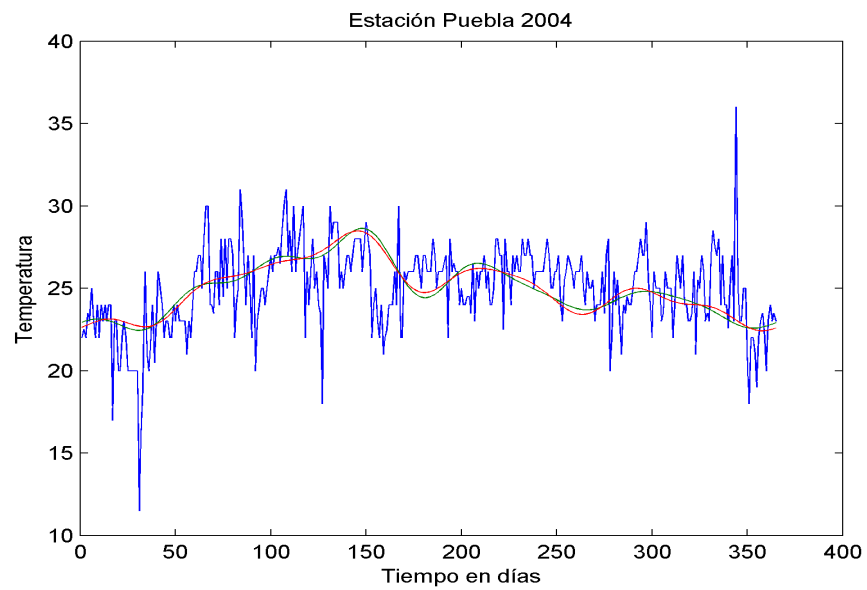
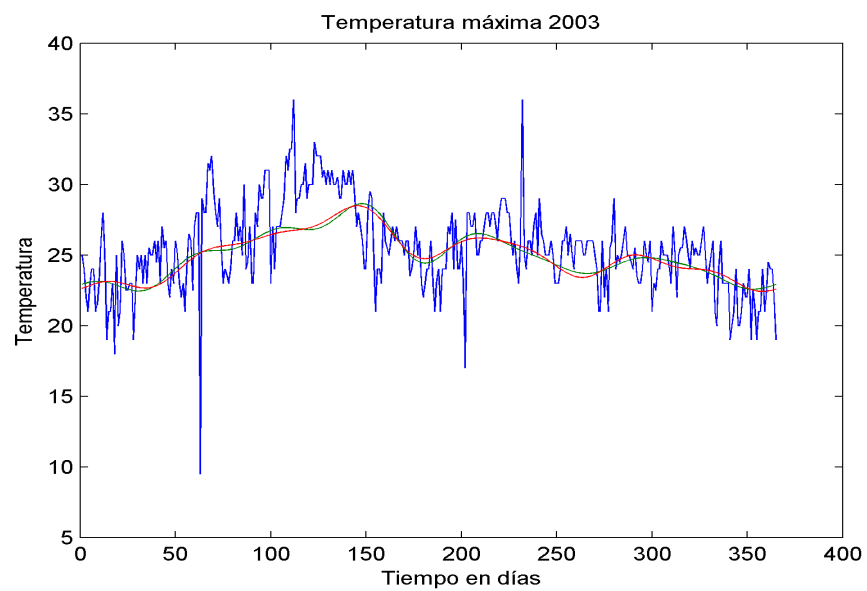
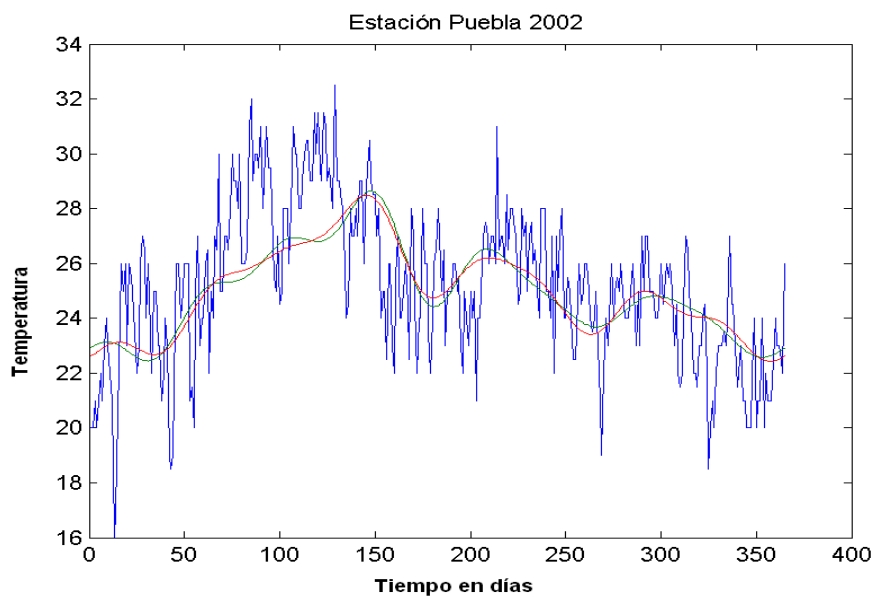
Las siguientes figuras muestran las gráficas de las series de datos de temperaturas máximas de la estación de Puebla de algunos años. En cada figura se muestra en azul la



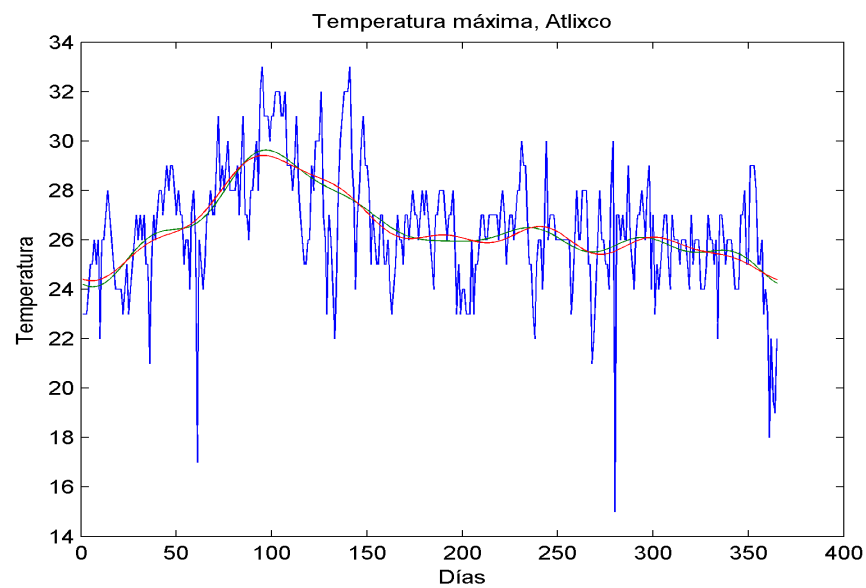
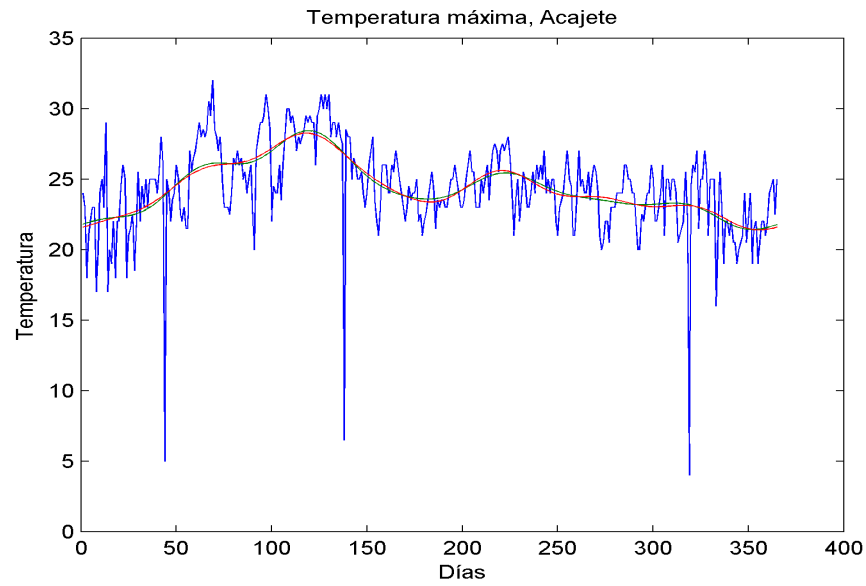
serie de datos obtenida de las mediciones, en color verde la serie F obtenida de resolver el problema (3.9), que para el caso de temperaturas máximas de la estación Puebla corresponde al filtrado del año 1993 (TABLA 8), y la roja, su respectiva serie ajustada, con coeficientes dados en la TABLA 9.

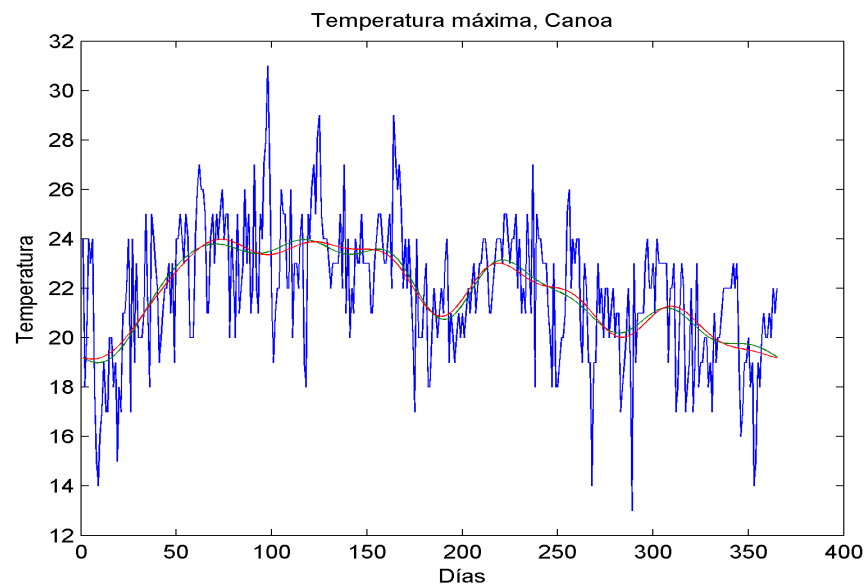
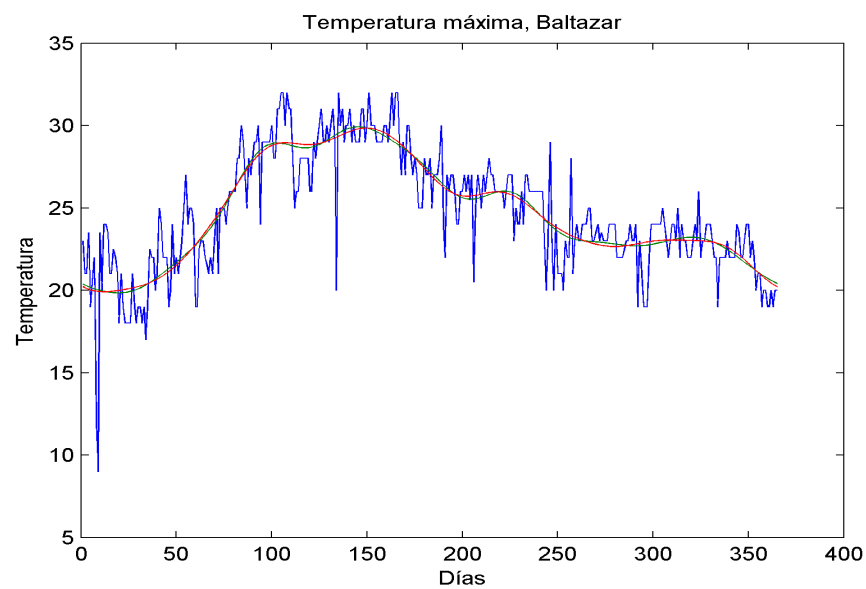
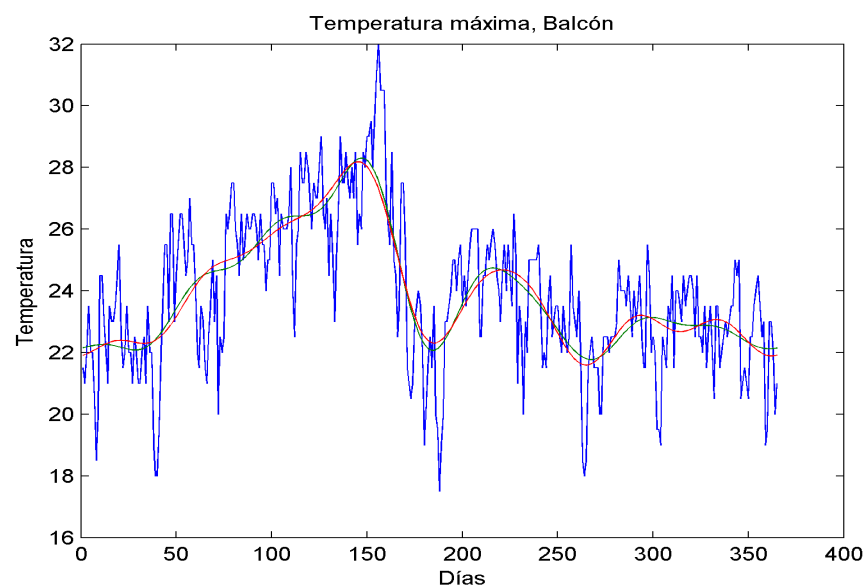
NOTA: Las figuras corresponden a datos de temperaturas asociados a **años completos** originalmente, tal como puede verificarse en la TABLA 2; además la gráfica que se muestra en rojo es la de la función suavizada para la estación Puebla que se deseaba (se obtuvo con los datos del año 1993), observese que efectivamente es un buen ajuste para todos los años considerados.

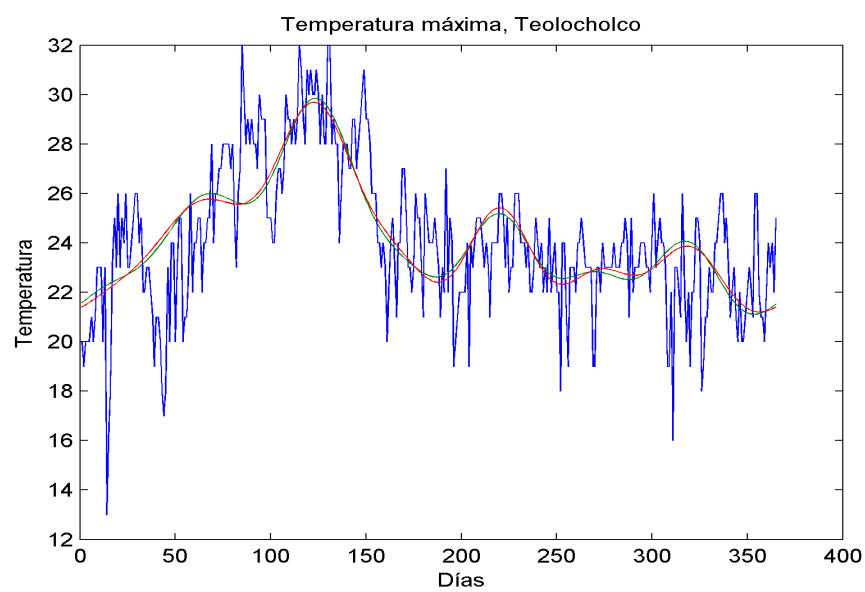
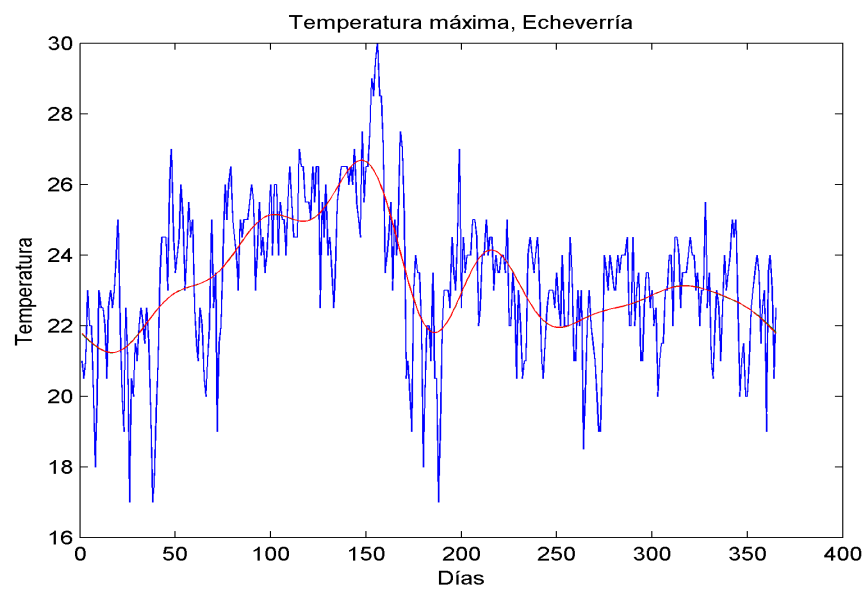




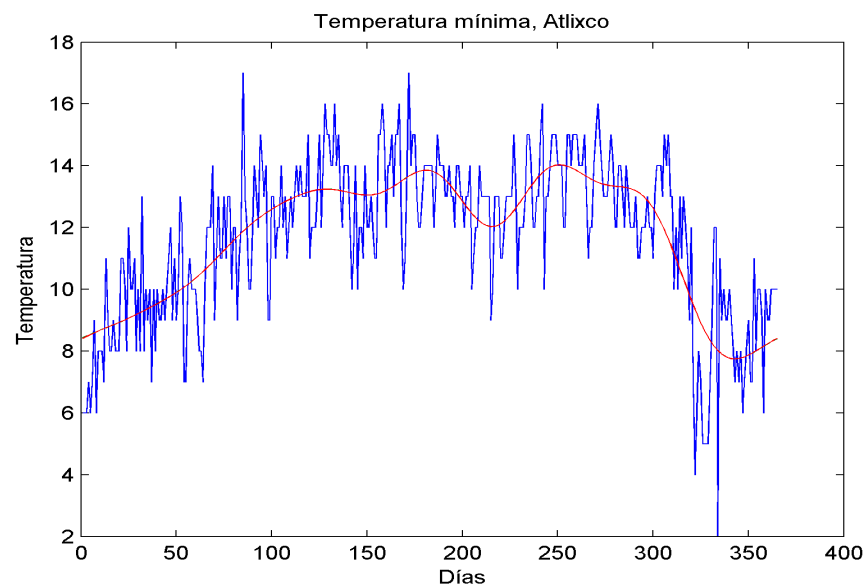
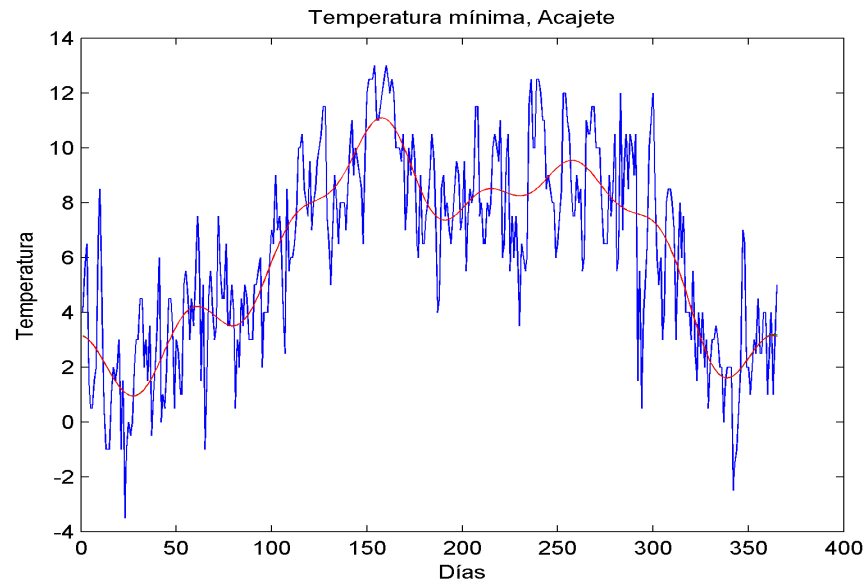
A continuación se muestran las gráficas de los polinomios trigonométricos (TABLA 9) que representan la variación de temperaturas máximas por estación.

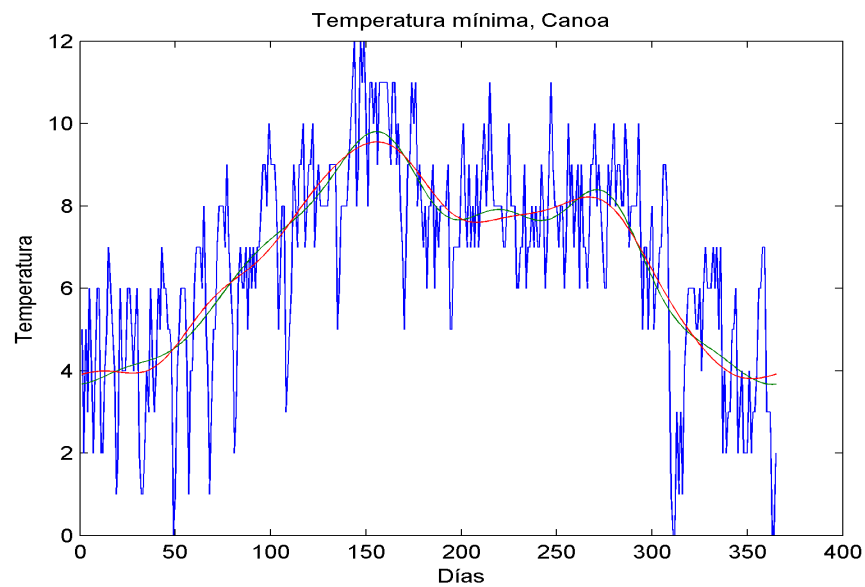
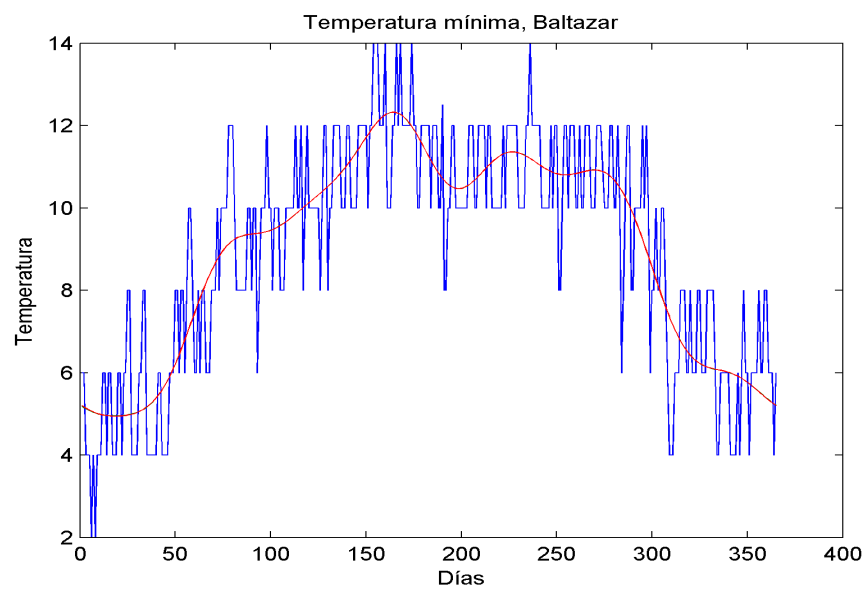
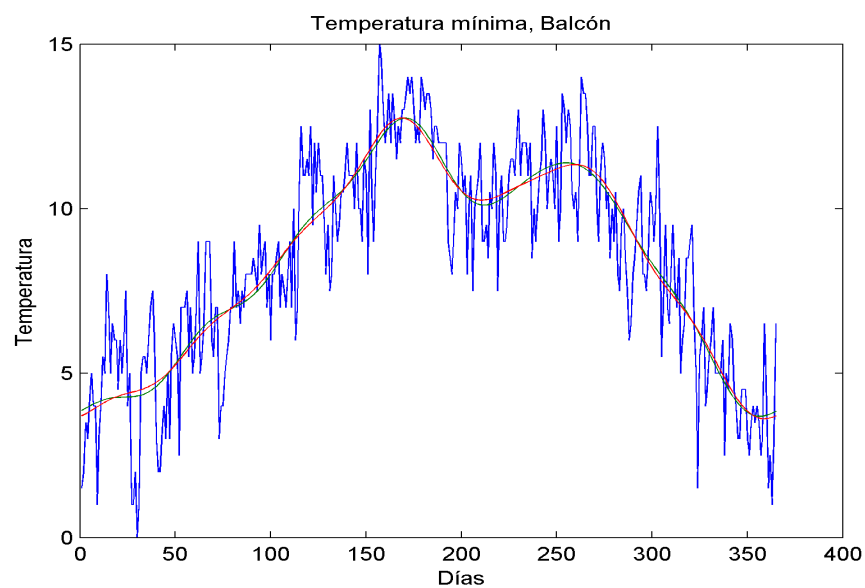


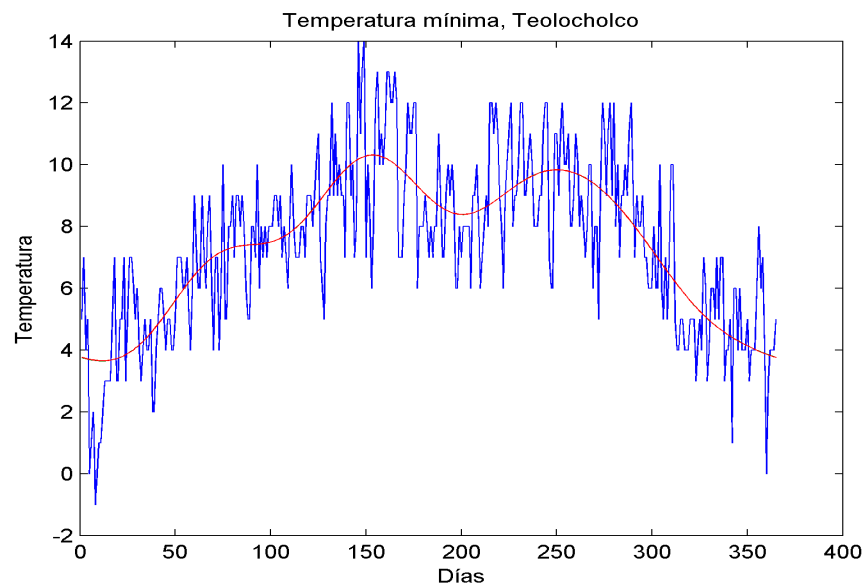
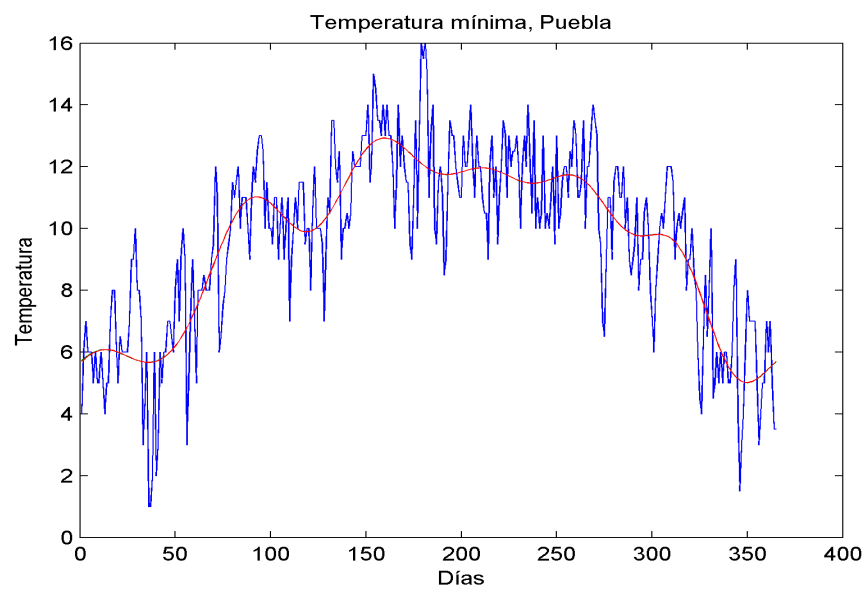
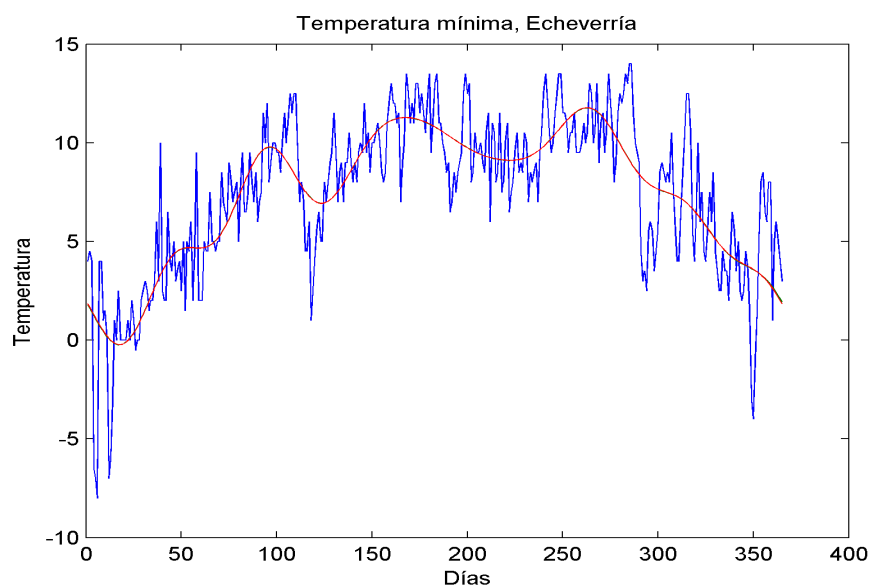




A continuación se muestran las gráficas de los polinomios trigonométricos (TABLA 10) que representan la variación de temperaturas mínimas por estación.







Capítulo 5

Discusión de resultados

El problema de datos faltantes es una situación muy común cuando se trabaja con datos reales; desde el punto de vista estadístico hay diferentes métodos para estimarlos. En este trabajo se propuso otra manera de estimarlos; consideramos que es conveniente verificar la efectividad de esta propuesta y compararla con los métodos ya conocidos.

En el desarrollo de este trabajo se empleó la estimación de datos faltantes principalmente para explorar si podían detectarse periodicidades ocultas mayores o iguales a un año, (lo cual se hizo con todas las estaciones); se obtuvo que la periodicidad mas importante detectada es la anual, lo cual nos permitió continuar el análisis con ciclos anuales. Cabe señalar que debido a la longitud del intervalo de tiempo más largo de que disponemos no se puede garantizar que no existan periodicidades más amplias que la anual, sin embargo, sí se garantiza que la anual es la más importante dentro de este intervalo de tiempo.

Como resultado principal se presentan los polinomios trigonométricos (4.1) que describen la variación de temperaturas máximas/mínimas con coeficientes dados en las TABLAS 9 y 10. Dichos polinomios se obtuvieron de años completos y aunque no hubiera sido así, los polinomios trigonométricos propuestos se compararon con series de datos que originalmente estaban completas, como se ejemplificó con los datos de temperatura máxima de la estación Puebla, (véase la nota del final de la página 34), por lo que la estimación de datos faltantes no afecta la validez de los resultados.

Al final del Capítulo 4 se muestran, en rojo, las gráficas de los polinomios trigonométricos. La teoría del análisis de Fourier garantiza que los polinomios obtenidos aproximan adecuadamente a los datos de temperaturas que se poseen. Más aun, los polinomios trigonométricos son una buena aproximación en el sentido de que la desviación cuadrática media (5.1) , entre los datos reales y los descritos por los polinomios trigonométricos, es pequeña

$$E = \frac{\sqrt{\sum_{t=1}^{365} (X_a(t) - X_F(t))^2}}{365} \quad (5.1)$$

donde $a = 1, \dots, 38$, (considerando que $X_{a=1}$ está asociado al año 1967 y $X_{a=38}$ está asociado a 2004), X_F es el polinomio trigonométrico que describe la variación de la temperatura máxima de la estación en cuestión. La desviación cuadrática media es de 0,12827325 para el caso de la estación Puebla.

La tabla siguiente muestra el promedio de la desviación cuadrática media en cada estación.

TABLA 11. Desviación cuadrática media por estación.

| ESTACIÓN | T. Máxima E | T. Mínima E |
|-------------|----------------|----------------|
| ACAJETE | 0.14448969 | 0.169200216 |
| ATLIXCO | 0.151845053 | 0.159229833 |
| BALCÓN | 0.124922422 | 0.116098558 |
| BALTAZAR | 0.152427974 | 0.113175342 |
| CANOA | 0.169339654 | 0.15001031 |
| ECHEVERRÍA | 0.125992502 | 0.157361211 |
| PUEBLA | 0.128273252 | 0.112700288 |
| TEOLOCHOLCO | 0.145089166 | 0.110010212 |

De los resultados reportados en la tabla anterior concluimos que los polinomios propuestos son una buena aproximación a los datos de temperaturas máximas/mínimas según corresponda.

Conclusiones

El desarrollo del trabajo se basó fundamentalmente en el análisis espectral de series de tiempo para describir de forma determinista, a través de polinomios trigonométricos, el comportamiento de una serie temporal de datos de temperaturas máximas y mínimas registradas en ocho observatorios.

Como primer paso se determinó que el ciclo anual es el más importante, esto permitió que los polinomios trigonométricos describieran la variación de temperatura máxima/mínima a lo largo de un año, en cada observatorio meteorológico. Dicho polinomio describe la temperatura para lo que denominamos un año típico.

El resultado principal de este trabajo son los polinomios trigonométricos que representan la variación de temperaturas máximas y mínimas, dichos polinomios son adecuados en el sentido de que minimizan la desviación media cuadrática entre los datos reales y los descritos por los polinomios.

Los resultados obtenidos proporcionan información para desarrollar un modelo de la variación de temperatura en la cuenca del Valle de Puebla que dependa tanto del espacio como del tiempo y que será utilizado para calcular los coeficientes de evapotranspiración, variables de entrada para el modelo matemático de balance hídrico.

Apéndice A

Programa para obtener datos faltantes

En este programa se implementó el algoritmo descrito en el Capítulo 2.

Este programa proporciona los datos que faltan en una estación con datos de las dos estaciones con las que está más correlacionada.

```
clear all; clc
```

```
[Total1,C]=xlsread('E:\AATESIS\Estaciones\Teolocholco.xls','Minima');  
[Total3]=U(Total1);  
[Total2,C]=xlsread('E:\AATESIS\Estaciones\Zacatelcoteolo.xls','Minima');  
[Totals]=Ullenacon(Total3,Total2); Totalo=Totals; l=size(Totals);  
[Total22,C]=xlsread('E:\AATESIS\Estaciones\Zacatelcoteolo.xls','Minima');  
[Totals]=Ullenacon(Totalo(1:368,:),Total22);  
Totals=[Totals(1:368,:); Totalo(369:l(1),:); Totals(370,:)];  
[Total]=Uauto(Totals(1:368,:));  
l=size(Totals);  
Total=[Total(1:368,:); Totals(369:l(1),:); Total(370,:)];
```

La matriz Total contiene la nueva tabla de datos después del llenado.

Sub-rutina Ullenacon.m

Este programa inserta datos de la base de datos Total1 en datos faltantes de la base Total2.

```
function [Total2]=Ullenacon(Total2,Total1)  
%Total2 es la tabla a llenar  
%Total1 es la tabla con la que se llena  
[ndf2]=fndfalta(Total2);  
[ndf1]=fndfalta(Total1);  
Total1=[Total1;  
ndf1(2,:)]; Total2=[Total2; ndf2(2,:)]; Total22=Total2;
```



```

l1=size(ndf1); l2=size(ndf2); for i=1:l2(2)
    if (Total2(369,i) ~= 0) && (Total2(369,i)<=365)
        for j=1:l1(2)
            if Total1(367,j)==Total2(367,i)
                for k=1:365
                    if isnan(Total2(k,i))==1
                        Total2(k,i)=Total1(k,j);
                    end
                end
                break;
            end
            if Total1(367,j)>Total2(367,i)
                break;
            end
        end
    end
end [ndf2]=fndfalta(Total2); Total2=[Total2; ndf2(2,:)];

```

Sub-rutina Uauto.m

Este programa se emplea en el programa anterior y realiza el llenado de hasta siete datos faltantes consecutivos, con datos anteriores o posteriores a dichos datos faltantes.

```

function [Total2]=Uauto(Total2)

[ndf1]=fndfalta(Total2); Total2=[Total2; ndf1(2,:); ndf1(2,:)];
Total222=Total2; l2=size(ndf1); l3=l2(2); vacio=0; jtas=[]; for
i=1:l3
    if (Total2(370,i) ~= 0) && (Total2(370,i)<59)
        for j=1:365
            if isnan(Total222(j,i))==1
                vacio=vacio+1; jtas=[jtas j];
            end
            if isnan(Total222(j,i))==0 | j==365
                if (vacio==1)
                    Total2(j-1,i)=Total222(j,i);
                end
                if (vacio>1) && (vacio<=8)
                    l=floor((max(jtas)-min(jtas))/2); ml=min(jtas)+1;
                    if (min(jtas)~=1) && (max(jtas)~=365)
                        Total2(min(jtas):ml,i)=Total222(min(jtas)-1,i).*ones(l+1,1);
                        z=max(jtas)-ml;
                        Total2(ml+1:max(jtas),i)=Total222(max(jtas)+1,i).*ones(z,1);
                    else
                        if min(jtas)==1
                            Total2(min(jtas):max(jtas),i)=Total222(max(jtas)+1,i).*ones(vacio,1);

```

```

        else
            Total2(min(jtas):max(jtas),i)=Total222(min(jtas)-1,i).*ones(vacio,1);
        end
    end
end
    vacio=0;
    jtas=[];
end
end
end

if vacio==1
    Total2(365,i)=Total222(364,i);
end
vacio=0;    jtas=[];
end
[ndf2]=fndfalta(Total2); Total2(370,:)= ndf2(2,:);

```

Sub-rutina fndfalta.m

Determina cuántos datos faltan en cada columna de una matriz A.

```

function [ndf]=fndfalta(A)
tbd=size(A); ndf=[];
fori=1:tbd(2)
    x=A(1:365,i);
    nan0=isnan(x);
    n = nnz(nan0);
    colum=[A(367,i); n];
    ndf=[ndf colum];
end

```

Apéndice B

Programa que calcula el filtrado y el ajuste de las series

En este programa se filtran las series temporales anuales, se secciona la que representa las temperaturas por estación y se ajusta dicha serie filtrada por mínimos cuadrados.

```
nmax=15; % posiciones del espectro a considerar
[lista]=facomple(Total);
aan=lista(367,:); y=lista(1:365,:); lgy=size(y); x=[];
for j=1:lgy(2)
    x=[x; y(:,j)]; %contiene la serie concatenada
end [sub]=grfacons(lista,x); l=size(sub);
s=sub(:,4)-sub(:,3)+ones(l(1),1); c=find(s==(max(s)));
% ANALISIS ANUAL
[sp]=ffiltxao1(lista(:,sub(c,3):sub(c,4))); %calcula el espectro de anual
[maxsp,ff]=fmax(sp,nmax);
%maxsp da las posiciones en donde el espectro tiene valores altos
maxsp=[maxsp; lista(367,sub(c,3):sub(c,4))];
%ff da las posiciones del espectro que mas se repiten en las primeras dos
%columnas ordenados.
l=size(ff);
pos=[1 2 3 4 5 6 8]; % se escriben las frecuencias que son representativas
[xf,valor]=ffiltxao(lista(:,sub(c,3):sub(c,4)),pos); valor1=valor;
xf1=xf; pos1=pos; for i=10:l(1)
    if ff(i,4)>=16
        pos=[pos ff(i,3)];
        %coordenadas para el filtro se prueba con todas las posibilidades
        [xf,valor]=ffiltxao(lista(:,sub(c,3):sub(c,4)),pos);
        if (valor1(1)>valor(1) && (valor(2)~=9) && (valor(2)~=4))% && (valor(2)~=24))
            valor1=valor; pos1=pos; xf1=xf;
        end
    end
end
end
```

```

aaaa=lista(367,sub(c,3):sub(c,4));
posi=valor1(2);

    lista1=lista(:,sub(c,3):sub(c,4));  y=lista1(1:365,:);  l=size(y);  t=1:365;
% se calcula el AJUSTE ANUAL
[xa,b]=fajuste(xf1(:,posi),pos1); mm=size(y);
    dst=zeros(3,mm(2));
    dst(1,:)=1996:2004;
    for i=1:l(2)
        dst(3,i)=sqrt(sum((y(:,i)-xa).*(y(:,i)-xa)));
        dst(2,i)=sqrt(sum((y(:,i)-xf1(:,posi)).*(y(:,i)-xf1(:,posi))));
    end

for i=1:l(2)
    figure;
    plot(t,y(:,i),t,xf1(:,posi),t,xa);
    title(['Filtrado representativo',int2str(lista1(367,posi)),
'-Estacin ',int2str(lista1(366,1)), 'an-o', int2str(lista1(367,i)),
' -dato ', int2str(lista1(368,1))]);
end

```

Sub-rutina grfacons.m

Se grafican los años consecutivos.

```

function [sub]=grfacons(lista,x) lgy=size(lista);
a=lista(367,1); i=1;  aaa=[];  m=lgy(2)*365;  sub=[]; for
j=1:lgy(2)
    if a==lista(367,j)
        a=a+1;          k=j;
    else
        k=k*365;  w=k-i+1;  t=1:w;  a1=floor(i/365)+1;  a2=floor(k/365);
        sub=[sub;[i,k,a1,a2]];
        figure;  plot(t,x(i:k,1));
        title(['Original -Estacion ',int2str(lista(366,1)),' -dato ',
int2str(lista(368,1)),'del',int2str(lista(367,a1)),'al',int2str(lista(367,a2))]);
        a=lista(367,j)+1;  i=k+1;  k=j;
    end
end
end

    k=k*365;  w=k-i+1;  t=1:w;  a1=floor(i/365)+1;  a2=floor(k/365);
    sub=[sub;[i,k,a1,a2]];
    figure;
    plot(t,x(i:k,1));
    title(['Original -Estacion ',int2str(lista(366,1)),' -dato ',
int2str(lista(368,1)),'del',int2str(lista(367,a1)),'al',int2str(lista(367,a2))]);

\bigskip

```

Sub-rutina fmax.m

```

function[msp,ffo]=fmax(sp,n)
l=size(sp);  v=[]; msp=[];
for i=1:l(2)
    s=sp(1:184,i);
    for j=1:n
        c=max(s);      p=find(s>=c);    s(p)=-1;      v=[v; p(1)];
    end
    msp=[msp v];      v=[];
end
ff=[];  l=size(msp);  ff=zeros(1,2);  ff(1,1)=msp(1,1);
ff(1,2)=0;  cc=0;
for i=1:l(1)
    for j=1:l(2)
        l2=size(ff);
        for k=1:l2(1)
            if ff(k,1)==msp(i,j)
                ff(k,2)=ff(k,2)+1;
                cc=1;
            end
            if k==l2(1) && cc==0
                ff=[ff; [msp(i,j) 1]];
            end
        end
    end
    cc=0;
end
end
%ordena ff
fft=ff;  l1=size(ff);  ffo=[]; for i=1:l1(1) c=min(fft(:,1));
p=find(fft(:,1)==c);  ffo=[ffo; ff(p,:)]; fft(p,:)=[1000,1000];
end
ffo=[ffo ff];

```

Sub-rutina ffiltxao.m

Cálculo de la serie filtrada por año.

```

function[sf,valor]=ffiltxao(lista,pos)
y=lista(1:365,:); l=size(y);  sf=[];
sp=[];  t=1:365;
for i=1:l(2)
    s=y(:,i);      [sf1]=ffiltrar(s,pos);      sf=[sf sf1];
end ds=[];

for i1=1:l(2)
    dst=zeros(1,365);

```

```

for i=1:l(2)
    if lista(367,i)>=1989
sumita=sum((y(:,i)-sf(:,i1)).*(y(:,i)-sf(:,i1))); dst(i)=sqrt(sumita);
    end
end
dstt=sum(dst); ds=[ds; dstt];
end v=min(ds); posi=find(ds==v); valor=[v, posi];

```

Sub-rutina filtrar.m

Esta función filtra una serie s (vertical) y retorna sf la serie filtrada, sp el espectro, el filtro empleado y el vector $posif$ de posiciones donde filtro tiene componentes no nulas

```

function[sf]=ffiltrar(s,pos)
z=fft(s); %se calcula la transformada de Fourier
sp=z.*conj(z); %calcula el espectro de potencias
l=length(sp); filtro=zeros(l,1);

%crea un filtro poniendo 1 en las posiciones dadas en posif
ll=length(pos); posif=pos; for i=ll:-1:1
    p=366-posif(i); posif=[posif p];
end l=length(posif); for i=1:l
    filtro(posif(i))=1;
end zf=filtro.*z; sf=real(ifft(zf));

```

Sub-rutina fajuste.m

```

function[sa,b]=fajuste(sf,pos)
% devuelve sa la serie ajustada, b contiene los coeficientes del ajuste,
% r son los residuos del ajuste.
ll=length(pos); posif=pos; for i=ll:-1:1
    p=366-posif(i);
posif=[posif p]; end ll=length(sf); t=[]; t=1:1:ll; t=t';
w=(2*pi)/365; longi=floor(length(posif)/2);
X=[ones(ll,1)]; coeficientes=[];
for k=1:longi
    f=[]; f=posif(k);
    X=[X cos((f*w).*t) sin((f*w).*t)];
end
y1=sf-mean(sf).*ones(size(sf)); %se le resta la media a cada dato de x
[b,bint,r,rint,stats]=regress(y1,X,0.05);
%b contiene los coeficientes del ajuste.
sa=X*b+mean(sf).*ones(size(sf));
media=mean(sf)

```

Bibliografía

Bracewell, Ronald. 2000. *The Fourier Transform and Its Applications*. McGraw-Hill, 3a Ed. –ISBN 0-0730-3938-1. .

Breña, Agustín Felipe; Breña, José Agustín. 2005. *Frecuencia de valores extremos en Hidrología*. Universidad Autónoma Metropolitana. ISBN: 970-692-173-7.

Brigham, Oran. 1988. *The Fast Fourier Transform and its applicvations*. Prentice Hall. ISBN 0-13-307505-2.

Chatfield, Chris. 2003. *The Analysis of Time Series*. Chapman Hall/CRC, sixth edition. ISBN 1584883170.

Fuller, Wayne. 1976. *Introduction to Statistical Time Series*. John Wiley-Sons. ISBN 0-471287156.

Gilat, Amos. 2004. *Matlab: an Introduction with Applications*. Hoboken NJ: Wiley. – ISBN 0471439975.

Gottman, John. 1984. *Time Series Analysis; A Comprehensive Introduction for Social —Scientists*. Cambrige University Press. ISBN 0 521 23597 9.

Guerrero, Víctor. 1987. *Modelos Estadísticos para series de Tiempo Univariadas*. Centro de Investigación y Estudios Avanzados del IPN.

Instituto Nacional de Estadística Geografía e Informática (INEGI). 2001. *Diccionario de datos Climáticos*. ISBN 970-13-3263-6.

Johnson, Robert. 1976. *Estadística elemental*. Editorial Trillas.

Lantada, Nieves; Núñez, M. Amparo. 2002. *Sistemas de Información Geográfica: Prácticas con ArcView*. Barcelona : Edicions UPC. ISBN 8483016125.

Levine, David; Ramsey, Patricia; Smidt, Robert. 2001. *Applied statistics for engineers and scientists : using Microsoft Excel and Minitab* Prentice Hall. ISBN 0134888014.

Martínez, M. 1999. *Manejo integral de cuencas. Pasado, presente y futuro*. IX Congreso Nacional de Irrigación, Culiacán, Sinaloa, México.

Mendenhall, William. Wackerly, D. 2002. *Estadística Matemática con Aplicaciones*. Thomson. ISBN 970-686-194-7.

Rivera P., N. 2001. *Introducción al manejo de cuencas hidrográficas y su importancia. El deterioro ambiental del río Atoyac en la ciudad de Puebla: acciones para su restauración*. Universidad Iberoamericana Campus Puebla. Puebla, Puebla, México.

Romero Flores Granados, Arturo. 2004. *Manejo del agua en la cuenca del valle de Puebla*. (Tesis de Maestría) Posgrado en Ciencias Ambientales. Benemérita Universidad Autónoma de Puebla (BUAP).