



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

**“Desarrollo de un modelo basado en técnicas de
Minería de Datos para clasificar zonas
climatológicamente similares en el estado de
Michoacán”**

T E S I S

QUE PARA OBTENER EL GRADO DE:

**MAESTRO EN INGENIERÍA
(COMPUTACIÓN)**

P R E S E N T A:

GUILLERMO GILBERTO MOLERO CASTILLO

**DIRECTOR DE TESIS:
DR. NICOLÁS KEMPER VALVERDE**

**CO-DIRECTOR DE TESIS:
DRA. JUDITH GUADALUPE RAMOS HERNÁNDEZ**

México, D. F.

2008.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*A Dios y mi familia.
A Lenn Fabricio.*

AGRADECIMIENTO

Más que un agradecimiento, deseo manifestar mi profundo respeto y admiración a las siguientes personas e instituciones, quienes colaboraron de una u otra manera en este proceso, entre ellas:

Al Dr. Nicolás Kemper Valverde, amigo y maestro, por su disposición y apoyo permanente en la realización de esta tesis.

A la Dra. Judith Guadalupe Ramos Hernández por su apoyo, amistad y dedicación para hacer posible el desarrollo y culminación de este trabajo. Siempre le estaré agradecido por todas sus enseñanzas.

A la Mstra. Ildiko Pelczer por su asesoría en algunos temas de la tesis, como el uso de indicadores de validación y la comparación de técnicas de agrupamiento empleadas en la minería de datos.

A mi recordada madre, Raquel, por haber compartido una vida de amor, lucha y sacrificio, dejando como ejemplo una fuente inapreciable de perseverancia y determinación. A mi papá Guillermo por su apoyo, cariño y ejemplo. A ambos, como humilde compensación les dedico este trabajo.

A mis hermanas Mercedes, Rosario y Ruth por compartir momentos importantes en mi vida. A mis hermanos José y Julio gracias por su apoyo en todo momento, siendo una motivación muy importante en mi desarrollo profesional.

A mis compañeros y amigos con quienes compartí momentos gratos y que han contribuido al buen logro de esta meta. A todos mi mayor reconocimiento y gratitud.

Mi profundo agradecimiento a todo el personal del Posgrado de Ciencia de Ingeniería de la Computación, por su innegable apoyo humano y académico, en especial al Coordinador Dr. Boris Escalante Ramírez. No olvido el apoyo incondicional de Cecilia y Lulú, gracias de todo corazón.

A la Universidad Nacional Autónoma de México por permitirme realizar los estudios de maestría.

Índice

Resumen

Capítulo I. Introducción	1
1.1 Descripción del problema	4
1.2 Objetivo General	5
1.3 Hipótesis	6
1.4 Alcances y limitaciones	6
1.5 Organización de la tesis	7
Capítulo II. Minería de Datos	10
2.1 Definición	11
2.2 Proceso de Descubrimiento de Conocimiento en Base de Datos	16
2.3 Modelos y tareas de minería de datos	18
2.3.1 <i>Modelo descriptivo</i>	19
2.3.2 <i>Modelo predictivo</i>	19
2.3.3 <i>Tareas de minería de datos</i>	19
2.3.3.1 Agrupamiento (Clustering)	20
2.3.3.2 Reglas de asociación	21
2.3.3.3 Correlaciones	21
2.3.3.4 Clasificación	22
2.3.3.5 Regresión	22
2.4 Técnicas de minería de datos	23
2.4.1 <i>Técnicas bayesianas</i>	23
2.4.2 <i>Técnicas algebraicas y estadísticas</i>	24
2.4.3 <i>Técnicas basadas en redes neuronales artificiales</i>	24
2.4.4 <i>Técnicas basadas en árboles de decisión y sistema de reglas</i>	24
2.4.5 <i>Técnicas basadas en distancia o densidad</i>	25
2.4.6 <i>Técnicas basadas en núcleo y máquinas de soporte vectorial</i>	25
2.4.7 <i>Técnicas estocásticas y difusas</i>	26
2.5 Metodología para el desarrollo de un proyecto de minería de datos	27
2.5.1 <i>Herramienta de desarrollo</i>	31
2.6 Requerimientos para regionalización en hidrología	33
2.7 Síntesis	37
Capítulo III. Análisis de Datos	38
3.1 Análisis de las series de tiempo	39
3.2 Fuente de la serie de datos climatológicos	40
3.2.1 <i>Estaciones climatológicas tradicionales (CLICOM)</i>	40
3.2.2 <i>Observatorios</i>	42
3.3 Análisis de las bases de datos	47
3.3.1 <i>CLICOM</i>	47
3.3.2 <i>Tratamiento de las bases de datos</i>	47
3.3.3 <i>Construcción y ejecución del modelo de análisis</i>	48

3.3.4	<i>Resultados del análisis de las series temporales</i>	54
3.3.4.1	CLICOM	55
3.3.4.2	Observatorios	59
3.4	Análisis de variabilidad y distribución de los principales elementos climatológicos	60
3.4.1	<i>Estaciones climatológicas del sistema CLICOM</i>	60
3.4.1.1	Precipitación (P)	61
3.4.1.2	Temperatura Observada (Tob)	65
3.4.1.3	Temperatura Máxima (Tmx)	66
3.4.1.4	Temperatura Mínima (Tmn)	67
3.4.2	<i>Observatorio de Morelia</i>	68
3.4.2.1	Precipitación	69
3.4.2.2	Presión en la estación	70
3.4.2.3	Temperatura de bulbo seco	70
3.4.2.4	Temperatura de bulbo húmedo y humedad relativa	71
3.5	Síntesis	73
Capítulo IV. Zona de Estudio		74
4.1	Aspectos Generales	75
4.2	Fisiografía y Geografía	76
4.2.1	<i>Sierra Madre del Sur</i>	78
4.2.2	<i>Eje Neovolcánico</i>	79
4.3	Climatología	82
4.4	Hidrología	86
4.4.1	<i>Aguas Superficiales</i>	86
4.4.2	<i>Aguas Subterráneas</i>	89
4.5	Vegetación	91
4.5.1	<i>Agricultura</i>	91
4.5.2	<i>Vegetación transformada</i>	91
4.5.3	<i>Vegetación natural</i>	92
4.6	Necesidades de regionalización del estado de Michoacán	93
Capítulo V. Metodología		97
5.1	Análisis del problema	98
5.2	Análisis de los datos	98
5.3	Preparación de datos	99
5.3.1	<i>Integración de datos</i>	100
5.3.2	<i>Transformación de datos</i>	101
5.3.3	<i>Normalización de datos</i>	102
5.4	Construcción del modelo de aplicación	102
5.4.1	<i>Bietápico (TwoStep)</i>	103
5.4.2	<i>K-Medias</i>	105
5.4.3	<i>Redes de Kohonen (Mapa autoorganizativo)</i>	106
5.4.4	<i>Integración del modelo</i>	109
5.4.5	<i>Ejecución de las técnicas de minería de datos</i>	110
5.4.6	<i>Análisis de resultados</i>	113
5.4.6.1	Resultados de la comparación	115

5.5 Validación	121
Capítulo VI. Conclusiones, recomendaciones y trabajo futuro	125
6.1 Conclusiones	125
6.2 Conclusiones particulares	129
6.3 Recomendaciones	131
6.4 Líneas de trabajo futuro	132
Referencias Bibliográficas	134
Anexo A. Total de estaciones climatológicas listadas en el Sistema CLICOM	140
Anexo B. Estaciones CLICOM seleccionadas para el análisis de la calidad de las series temporales	144
Anexo C. Descripción de los nodos empelados en el Clementine para el análisis de la calidad de la serie de datos	146
Anexo D. Resultados de las estimaciones mensuales de las variables significativas para el proceso de regionalización	154
Anexo E. Resultados de las estimaciones mensuales de las variables atmosféricas medidas para el observatorio de Morelia	160
Anexo F. Gráficas de los grupos de estaciones CLICOM formados por la técnica K-medias	162

Índice de tablas

2.1	Factores climatológicos mediante los cuales se infiere si se debe o no jugar tenis	15
2.2	Relación existente entre las tareas y técnicas de minería de datos	26
3.1	Variables climatológicas registradas en el Sistema CLICOM	41
3.2	Variables climatológicas registradas en la base de datos de los observatorios	45
3.3	Matriz con resultados porcentuales y numéricos de registros validos por año y variable. Estación climatológica St16052	52
3.4	Matriz con resultados porcentuales y numéricos de registros validos por año y variable, en la cual se muestra una considerable escasez de datos. Estación climatológica St16040	53
3.5	Estaciones climatológicas consideradas significativas para el proceso de clasificación de zonas con climas similares	56
4.1	Acuíferos con problemas de sobreexplotación	94
5.1	Campos disponibles para la preparación de datos	100
5.2	Resultados obtenidos por las técnicas de agrupamiento Bietápico, K-medias y redes de Kohonen considerando 4 grupos como valor óptimo para K-medias y redes de Kohonen	111
5.3	Determinación del número deseado de grupos para K-medias mediante los índices de validación Silueta, Dunn y Davies-Bouldin	124

Índice de figuras

2.1	Asociación de las diversas disciplinas relacionadas a la minería de datos	14
2.2	Proceso de descubrimiento de conocimiento en bases de datos (KDD)	16
2.3	Representación general de los modelos y tareas de minería de datos	20
2.4	Proceso de extracción de patrones mediante el uso de técnicas de minería de datos	23
2.5	Esquema de los cuatro niveles de abstracción de la metodología CRISP-DM	28
2.6	Fases del proceso de minería de datos, basado en la metodología CRISP-DM	29
2.7	Interfaz gráfica de Clementine	32
3.1	Estaciones climatológicas de la base de datos CLICOM para el estado de Michoacán	43
3.2	Ubicación de los Observatorios de Morelia y Zamora en el estado de Michoacán	44
3.3	Modelo de análisis de las series de tiempo elaborado en el Clementine	49
3.4	Identificación de las 52 estaciones climatológicas en el estado de Michoacán	58
3.5	Variación mensual de la precipitación en las 52 estaciones climatológicas	61
3.6	Localización de las estaciones con mayor (a) y menor (b) precipitación	63
3.7	Precipitación media anual de las 52 estaciones climatológicas	64
3.8	Variación estacional de la temperatura del aire observada en las 52 estaciones climatológicas	65
3.9	Variación estacional de la temperatura máxima en las 52 estaciones climatológicas	67
3.10	Variación estacional de la temperatura mínima en las 52 estaciones climatológicas	68
3.11	Suma mensual de precipitación en el observatorio de Morelia	69
3.12	Variación mensual de presión en la estación en el observatorio de Morelia	70
3.13	Variación mensual de la temperatura de bulbo seco en el observatorio de Morelia	71
3.14	Variabilidad mensual de temperatura de bulbo seco y humedad relativa en el observatorio de Morelia	72
4.1	Ubicación geográfica del estado de Michoacán	75
4.2	Fisiográfica del estado de Michoacán	77
4.3	Tipos de climas del estado de Michoacán	85
4.4	Regiones y Cuencas Hidrológicas	88

4.5	Acuíferos con problemas de sobreexplotación	90
4.6	Ubicación de los acuíferos sobreexplotados en Michoacán y sus estados vecinos	94
5.1	Estructura de una red de Kohonen (Mapa Autoorganizativo)	107
5.2	Modelo elaborado para obtener los grupos de zonas climatológicamente homogéneas en el estado de Michoacán	109
5.3	Determinación del número deseado de grupos mediante el algoritmo Bietápico	110
5.4	Carta de Climas del estado de Michoacán	114
5.5	Regiones climáticas formadas por el algoritmo Bietápico	115
5.6	Regiones climáticas formadas por el K-medias	116
5.7	Regiones climáticas formadas por las redes de Kohonen	116
6.1	Mapa de regionalización climática del estado de Michoacán	128

Resumen

Hoy en día, el uso de técnicas de *minería de datos* da lugar al descubrimiento de información y a la adquisición de conocimiento útil a partir de bases de datos, incluso cuando éstas están incompletas y no son continuas. De manera tal, que en la actualidad la minería de datos es un área importante de investigación y desarrollo. En este trabajo de tesis se emplean este tipo de técnicas para obtener zonas o regiones climatológicamente similares en el estado de Michoacán. Hidrológicamente, el disponer de regiones homogéneas permite contar con datos necesarios para establecer un balance de agua confiable que ayude en la toma de decisiones enfocadas a proteger y aprovechar de manera racional el recurso hídrico.

Las series de datos utilizadas corresponden a los registros climatológicos disponibles en las estaciones tradicionales (CLICOM) y observatorios, establecidos en el territorio michoacano. El análisis de la variabilidad climática de la región con base en estas series de datos es importante, puesto que frecuentemente sólo se hacen análisis locales extrapolando datos, lo cual no siempre es adecuado dadas las diferencias climatológicas, topográficas, vegetativas y socio-económicas de las zonas que forman una región, en este caso el estado de Michoacán.

Una de las tareas de la minería de datos para construir modelos de regionalización es el *agrupamiento (clustering)*, cuyo objetivo es evaluar similitudes entre los datos para representarlos en un pequeño número de regiones, es decir, divide una población heterogénea en un número de subgrupos homogéneos de acuerdo a las similitudes de sus registros. Debido a que existen diversas técnicas de agrupamiento, se seleccionaron tres tipos de técnicas de agrupamiento: Bietápico (jerárquico), K-medias (particional) y redes de Kohonen, esto con la finalidad de comparar los resultados obtenidos por cada una de ellas. Los resultados obtenidos fueron validados mediante *índices de validación* para evaluar la certeza en la

formación de los grupos. Los índices empleados fueron: Silueta, Dunn y Davies-Bouldin.

Los registros de las bases de datos se trabajaron para establecer instancias únicas mediante promedios mensuales de las tres variables de temperatura –observada, máxima, y mínima- y la suma mensual de precipitación. Así, se manejaron dos vistas minables: (a) vista establecida por los promedios y sumas integrados en campos mensuales de enero a diciembre, y (b) vista formada por los promedios y sumas integrados en campos mensuales más la variable *altitud*; se adicionó esta última variable con la finalidad de mejorar la calidad de la nueva vista minable.

Para aplicar las tres técnicas de *agrupamiento* a las vistas minables, éstas fueron modeladas en el software *Clementine* bajo la metodología CRISP-MD. Los resultados mostraron como número óptimo de grupos 4, aunque el número de grupos sugerido por los índices de validación fue 2. Las regiones formadas estuvieron fuertemente influenciadas por el accidentado relieve del estado debido a la presencia de las provincias fisiográficas Sierra Madre del Sur y Eje Neovolcánico. Estas a su vez dan lugar a una variedad de climas.

Finalmente, se concluye que la regionalización obtenida por el Bietápico y K-medias, ambos con iguales resultados, permite separar el estado de Michoacán en dos regiones claramente divididas. Las otras dos regiones pese a mostrar una clara división, al centro-norte y este del estado presentan una confusión al encontrarse algunas estaciones en ambas zonas climáticas. Así para lograr mejores resultados se sugiere en trabajos posteriores utilizar otro tipo de combinaciones de las variables atmosféricas, así como emplear un indicador de heterogeneidad basado en L-momentos.

Capítulo I.

Introducción

Una realidad actual de la *minería de datos* es su papel como una tecnología de apoyo que ha venido a resolver dos grandes retos:

- Trabajar con grandes volúmenes de datos para extraer y descubrir información de interés.
- Usar técnicas adecuadas para explorar, analizar, comprender y aplicar el conocimiento adquirido de grandes conjuntos de datos, así como para identificar tendencias y comportamientos que faciliten una mejor comprensión de los fenómenos que nos rodean y sirvan de ayuda en la toma de decisiones.

Debido a la creciente importancia que se ha dado a la minería de datos, instituciones como el Instituto Tecnológico de Massachussets (MIT, 2001) la consideran como “*una de las diez tecnologías emergentes más importantes del siglo 21 que cambiará el sentido de investigación en el mundo*”. De manera tal, que en la actualidad la minería de datos es un área importante de investigación y desarrollo. El impulso al empleo de minería de datos se asocia a una serie de factores como (Larose, 2005):

- El crecimiento exponencial de la recolección de datos y la evolución del poder de cómputo.
- El almacenamiento de los datos con la finalidad de tener acceso a una base de datos actualizada y confiable.
- La creciente disponibilidad de información en Internet.
- La presión competitiva del mercado en una economía globalizada.

- El desarrollo de herramientas comerciales para llevar a cabo la minería de datos.

Estos factores han llevado a emplear la minería de datos en un campo interdisciplinario amplio con aplicaciones en: finanzas, análisis de mercado, procesos industriales, medicina, biología, química, telecomunicaciones, biometría, medio ambiente, hidrología, entre otros. Es precisamente en hidrología donde la minería de datos ha venido a resolver problemas de clasificación de variables atmosféricas en grupos homogéneos, esto es, delimitar zonas que reúnan características similares con el fin de profundizar en la comprensión de los fenómenos ocurridos en la región (Xu y Yao, 2001; Sumathi y Sivanandam, 2006).

La homogeneidad hidrológica se logra mediante la regionalización, esto es dividir la zona de estudio en regiones cuyas características climáticas, topográficas, orográficas, etc. sean similares entre sí (Kyselý *et al*, 2007). La regionalización ha sido utilizada de forma exitosa en análisis regionales por ejemplo, para determinar la tendencia y frecuencia de inundaciones, sequías y el efecto de otras variables (Hall y Minns, 1999). Por tal motivo, la regionalización hidrológica representa una herramienta metodológica útil para la planeación del cuidado y aprovechamiento racional del recurso hídrico.

El análisis del impacto de las condiciones climáticas en zonas extensas es así determinante para establecer métodos de manejo adecuado del agua o de cualquiera de los recursos naturales de una región. En este caso, el comportamiento espacial y temporal de variables atmosféricas como temperatura, precipitación, evaporación, humedad, etc. permitirá conocer la condición real de las fuentes de agua, así como ayudará a establecer los planes de manejo para zonas sobreexplotadas. Sin embargo, la falta de datos es un problema común en diversas regiones del país entre ellas se cuenta el estado de Michoacán, siendo necesaria la optimización de la información disponible de una determinada región geográfica (Pelczar *et al.*, 2007). La práctica común es el realizar análisis locales mediante la extrapolación de datos, lo cual no siempre es adecuado dadas las diferencias

climatológicas, topográficas, vegetativas y socio-económicas de las zonas que forman una región. Para ello, se sugiere el análisis de series temporales, y en particular enfocado a series de datos con registros faltantes.

Entre los métodos de regionalización cabe destacar aquellos enfocados al *agrupamiento (clustering)* de datos, el cual permite formar regiones homogéneas con características similares, esto es, divide una población heterogénea en un número de subgrupos homogéneos de acuerdo a la similitud de sus registros. Gong y Richman (1995) en Muñoz y Rodrigo (2002) señalan que el *agrupamiento* ha sido reconocido como un método eficaz para tratar problemas de clasificación de estaciones en regiones con climas similares

En el *agrupamiento* existen dos tipos principales (Larose, 2005): el *agrupamiento jerárquico* que se caracteriza por el desarrollo recursivo de una estructura en forma de árbol, y el *agrupamiento particional* que organiza los registros dentro de k grupos. Los métodos particionales tienen ventajas en aplicaciones que involucran gran cantidad de datos y los jerárquicos por lo general se utilizan para identificar el número deseado de grupos. Otro tipo de agrupamiento son las redes de Kohonen (mapa autoorganizativo) que se utilizan para agrupar un conjunto de datos en distintas regiones, especialmente cuando no se tiene información de los mismos.

En este documento se presenta el desarrollo y los resultados del trabajo de investigación realizado para clasificar zonas o regiones climatológicamente similares en el estado de Michoacán basado en técnicas de minería de datos. Las técnicas de agrupamiento utilizadas fueron: Bietápico (jerárquico), K-medias (particional) y redes de Kohonen, las cuales fueron aplicadas a las vistas minables generadas a partir de los registros disponibles de las variables de temperatura - observada, máxima y mínima- y precipitación (variables medidas en las estaciones CLICOM).

1.1 Descripción del problema

El estado de Michoacán presenta un rápido crecimiento poblacional, industrial y agrícola, lo que ha llevado a tener problemas por sobreexplotación y contaminación de acuíferos. De acuerdo a CONAGUA (2005), de los 22 acuíferos identificados en el Estado, 9 están sujetos a sobreexplotación: Pastor Ortiz-La Piedad, Ciénega de Chapala, Zamora, Briseñas-Yurécuaro, Ciudad Hidalgo-Tuxpan, Uruapan, Nueva Italia, Apatzingán y Cotija. Ambas situaciones, sobreexplotación y contaminación, han puesto en riesgo la disponibilidad de agua subterránea, la cual ya en el año 2005 era sujeta a desperdicio (Juárez, 2005) en diversas regiones.

Aunado a lo anterior, el estado de Michoacán presenta una gran complejidad fisiográfica, formada por montañas, mesetas, planicies, cuencas, valles y costas, situación que origina la existencia de una gran variedad de climas, vegetaciones y suelos, colocándolo a nivel nacional como uno de los estados con mayor variedad de ecosistemas y recursos naturales.

Varias han sido las acciones tomadas para evitar los problemas de sobreexplotación y contaminación, entre las cuales destacan la veda de los acuíferos para su extracción a fin de que se recuperen los mantos acuíferos en cantidad y calidad del agua. Sin embargo, el crecimiento poblacional e industrial demanda agua, por lo que la veda no es una verdadera solución. Ante esta situación el Gobierno del estado de Michoacán consideró el desarrollo de modelos de manejo que permitan la utilización, protección y el cuidado de los principales mantos acuíferos del estado; siendo prioritario el estudio de los 9 acuíferos con problemas de sobreexplotación.

Debido a que los acuíferos con problemas de sobreexplotación están localizados en una región extensa al oeste del estado y también al este del mismo, resulta difícil contar con datos meteorológicos que permitan obtener un balance de agua real para cada acuífero. Esto último debido a que la mayoría de las estaciones tradicionales (Sistema CLICOM) y observatorios de Zamora y Morelia muestran

series incompletas y no continuas. Además, la dispersión de estaciones con series completas y continuas, ubica una mayor cantidad de estaciones en la zona norte del estado, mientras que un grupo menor se ubica en el centro y aún menor al sur. Por lo que, disponer de regiones homogéneas permitirá contar con datos hidrológicos necesarios para establecer un balance de agua confiable que ayude en la toma de decisiones para proteger y aprovechar de manera racional el recurso hídrico.

1.2 Objetivo General

El principal objetivo de este trabajo de tesis es utilizar técnicas de minería de datos para clasificar zonas o regiones climatológicamente similares en el estado de Michoacán, empleando series de datos temporales no continuas e incompletas, para obtener un balance de agua subterránea que refleje la situación de los acuíferos catalogados bajo problemas de sobreexplotación.

1.2.1 *Objetivos Específicos*

- Determinar la calidad de la serie de datos a emplearse en el proceso de regionalización climática del estado de Michoacán.
- Analizar la variabilidad y distribución de las principales variables atmosféricas medidas en las estaciones disponibles.
- Determinar qué tareas de minería de datos (agrupamiento, correlación y reglas de asociación) permiten la formación de grupos con características similares, en este caso aplicado a bases de datos de estaciones meteorológicas.
- Comparar los resultados con el mapa climatológico del INEGI.
- Validar los resultados obtenidos mediante indicadores de validación, en este caso probar si los índices para métodos de agrupamiento participial y jerárquico ofrecen una alternativa que garanticen la formación de las regiones climáticas.
- Obtener un mapa de regionalización climática.

1.3 Hipótesis

Al aplicar técnicas de agrupamiento de minería de datos para clasificar zonas o regiones climatológicamente similares en el estado de Michoacán, es posible obtener regiones homogéneas con características climáticas similares a partir de las series de datos disponibles y establecer si los métodos de validación son los adecuados para determinar si la homogeneidad obtenida es la más apropiada para los estudios de hidrología.

1.4 Alcances y limitaciones

1.4.1 Alcances

- Probar que la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*) ofrece una aproximación sistemática para estructurar un proyecto de minería de datos, donde a través de una sucesión de pasos se pueda llegar a una regionalización con base en series de datos temporales.
- Determinar si dentro de las tareas de minería de datos, el *agrupamiento* puede emplearse para que a partir del conjunto de datos climatológicos, se construya un modelo basado en la identificación de zonas o regiones climatológicamente homogéneas.
- Determinar cual de los tipos de agrupamiento: jerárquico (Bietápico), particional (K-medias) o redes de Kohonen (Mapa Autoorganizativo), ofrece un mejor resultado en la formación de las regiones homogéneas.
- Establecer si los indicadores de validación son los adecuados para determinar si la formación de las regiones fue la idónea.

1.4.2 Limitaciones

Las principales limitaciones observadas para el desarrollo del trabajo de tesis son:

- La cantidad de estaciones CLICOM que disponen de series datos continuas y completas en un periodo de 20 años.
- La dispersión de las estaciones seleccionadas para el proceso de regionalización, puesto que una mayor cantidad de ellas se encuentran ubicadas en zona norte del estado, mientras que un grupo menor se ubica en el centro y menos aún al sur del mismo.
- La construcción de la vista minable con un número reducido de parámetros, esto debido a la falta de registros para otro tipo de variables a la Precipitación y Temperatura.
- La falta de indicadores de validación diseñados para el método Bietápico y para las redes de Kohonen.

1.5 Organización de la tesis

El documento esta organizado de la siguiente manera:

En el *Capítulo II* se presentan los conceptos de *minería de datos* y sus principales fundamentos. Se revisan los modelos, tareas y técnicas más comúnmente aplicados en la minería de datos. Asimismo, se describe el proceso metodológico que se sigue en un proyecto de minería de datos, en particular, la metodología *CRISP-DM*, el cual permite organizar el desarrollo del proyecto a través de una sucesión de pasos, conocido como desarrollo sistemático. Una de las herramientas que hace uso del desarrollo sistemático es el software *Clementine*, el cual ofrece un método estratégico para encontrar relaciones útiles en grandes conjuntos de datos. Además, se incluye una descripción de los métodos más empleados para la regionalización de zonas homogéneas en hidrología, en particular se mencionan los métodos de agrupamiento porque permiten clasificar conjuntos de datos en grupos homogéneos.

El *Capítulo III* describe las fuentes de datos climatológicos disponibles para el estudio: datos provenientes de las estaciones tradicionales (Sistema CLICOM) y

datos obtenidos de los Observatorios de Zamora y Morelia. Asimismo, describe el proceso de análisis empleado con estas bases de datos para determinar la calidad de la serie de datos y el número de estaciones a utilizar en el proceso de regionalización climática en el estado de Michoacán. Ambos análisis fueron realizados empleando el software Clementine, el cual dio por resultado un total de 52 estaciones CLICOM para la regionalización. Finalmente, se presenta el análisis de variabilidad y distribución en el tiempo de las principales variables registradas en las estaciones disponibles: precipitación y temperatura (observada, máxima y mínima).

En el *Capítulo IV* se describen las principales características del estado de Michoacán, el cual sobresale por su accidentado relieve formado por dos grandes provincias fisiográficas, Sierra Madre del Sur y Eje Neovolcánico. Estas provincias son referencia para la caracterización geológica, climatológica, hidrológica, y vegetativa del estado. A su vez se incluye una descripción de la necesidad de regionalizar zonas similares en el estado de Michoacán para efectos de aprovechamiento de sus recursos hídricos disponibles.

El *Capítulo V* presenta el empleo de la metodología CRISP-DM para obtener las regiones con climas similares en el estado de Michoacán. Las fases del CRISP-DM aplicado a través del software Clementine comprenden las siguientes etapas: *análisis del problema y de los datos* desarrollado en los Capítulos III y IV; *preparación de datos* donde para organizar la vista minable se integró, transformó y normalizó la serie de datos de las 52 estaciones significativas para la regionalización; *construcción del modelo* que constituye el proceso principal de la minería de datos, donde, mediante técnicas de agrupamiento jerárquico (Bietápico), particional (K-medias) y redes de Kohonen (mapa autoorganizativo) se permitió la clasificación de las estaciones en zonas climatológicamente homogéneas. Se muestran los resultados al aplicar los métodos Bietápico, K-medias y redes de Kohonen; así como la *validación de resultados* donde mediante indicadores de validación como el Ancho de la Silueta, Dunn y Davies Bouldin, se validó la regionalización obtenida.

En el *Capítulo VI* se presentan las conclusiones del trabajo realizado y se mencionan recomendaciones, así como las futuras líneas de investigación que se pueden establecer con base en los resultados obtenidos en este trabajo.

Finalmente, se presentan una serie de anexos donde se muestran el total de estaciones CLICOM establecidas en el territorio michoacano (*Anexo A*); las estaciones CLICOM que tienen periodos de datos con al menos 20 años a partir de 1981 (*Anexo B*); la descripción ampliada de la configuración de los nodos de la herramienta Clementine utilizados para el análisis de datos (*Anexo C*); los resultados de las estimaciones mensuales de las variables significativas para la regionalización (*Anexo D*); el resultado de sumas y promedios mensuales de las variables atmosféricas medidas para el observatorio de Morelia (*Anexo E*); y las gráficas de los grupos de estaciones CLICOM formados por la técnica K-medias (*Anexo F*).

Capítulo II.

Minería de datos

Gracias al avance de la tecnología es posible almacenar grandes volúmenes de datos de manera automática en una base continua, que puede ser estructurada en forma de bases de datos, o no automática en forma textual. Una de las ventajas de contar con bases de datos es la posibilidad de descubrir información de interés y adquirir conocimiento mediante el análisis de los datos. Sin embargo, el volumen mismo de las bases de datos es con frecuencia una limitante para análisis manuales por lo que se han desarrollado tecnologías especializadas que faciliten su manejo. Una de estas tecnologías es la *minería de datos*, por medio de la cual se cuenta con mecanismos útiles para la extracción de información a partir de series extensas de datos. Esto ha hecho a la minería de datos una herramienta útil, estableciendo un área activa de investigación y desarrollo.

En este capítulo se define la minería de datos y sus principales fundamentos. Un aspecto importante, es el papel de la minería de datos como parte del proceso de descubrimiento de conocimiento presente en las bases de datos (*Knowledge Discovery in Databases o KDD*). Asimismo, se revisan los modelos, tareas y técnicas más comúnmente aplicados en la minería de datos. Se presenta el proceso metodológico que se sigue en un proyecto de minería de datos, en particular, la metodología *CRISP-DM* (Cross-Industry Standard Process for Data Mining). Finalmente, se presentan los métodos más empleados para la regionalización de zonas homogéneas en hidrología.

2.1 Definición

Debido a la creciente importancia que se ha dado a la minería de datos, ésta ha sido señalada por el Instituto Tecnológico de Massachussets (MIT, 2001) como “*una de las diez tecnologías emergentes más importantes del siglo 21 que cambiará el sentido de investigación en el mundo*”. Asimismo, la revista de tecnología online *ZDNet News* (ZDNet, 2001), considera que la minería de datos “*será uno de los desarrollos más revolucionarios de la siguiente década*”. En efecto, una realidad actual de la minería de datos es su papel como una tecnología de apoyo para explorar, analizar, comprender y aplicar el conocimiento adquirido de grandes volúmenes de datos, así como la identificación de tendencias y comportamientos que faciliten una mejor comprensión de los fenómenos que nos rodean y sirvan de ayuda en la toma de decisiones.

En la literatura se cuenta con varias definiciones para la minería de datos, algunas de ellas son:

Witten y Frank (2005), la minería de datos es el proceso de descubrir patrones en los datos. Los datos se presentan en grandes cantidades. Los patrones descubiertos deben ser significativos de manera que se permitan ventajas, por lo general, de tipo económicas.

Sumathi y Sivanandam (2006), la minería de datos es el proceso eficiente, no trivial, de extraer información valiosa (patrones y tendencias) de una gran colección de datos.

Hand et al. (2001), la minería de datos es el análisis de conjuntos de datos observados, a menudo extensos, para encontrar relaciones insospechadas y resumir los datos en forma comprensible y útil para el usuario de la información.

Berry y Linoff (2004), la minería de datos, es la exploración y análisis de grandes cantidades de datos para descubrir reglas y patrones significativos.

Estas cuatro definiciones tienen en común la extracción de información potencialmente útil (patrones, asociaciones o relaciones entre los datos) para los usuarios finales. Algunas otras definiciones consideran además del proceso de descubrir patrones válidos, la exploración y análisis de series extensas de datos. Debido a las diversas formas de definir la *minería de datos*, Hernández *et al.* (2004) resume el objetivo de la minería de datos como el proceso de convertir datos extraídos de grandes volúmenes de datos en conocimiento útil.

Para Witten y Frank (2005), el establecimiento de patrones es una práctica útil, puesto que permite explicar “algo” con relación a los datos que se tienen. De la misma manera, autores como Vazirgiannis *et al.* (2003) y Geng y Hamilton (2007) consideran que la minería de datos facilita el relevar patrones existentes en el conjunto de datos, los cuales generan información válida, novedosa, útil, comprensible y explotable.

En la definición de Sumathi y Sivanandam (2006), el énfasis está dado en la enorme cantidad de datos existentes como resultado del crecimiento explosivo de la tecnología en base de datos, la cual es representada como un *almacén de datos* (Data Warehouse¹). Un almacén de datos está definido como un proceso de gestión centralizada de datos. Por tanto, Sumathi y Sivanandam (2006) y Berry y Linoff (2004) consideran que un almacén de datos provee datos históricos, mientras que la minería de datos provee inteligencia.

De acuerdo a Hand *et al.* (2001), la mayor diferencia entre la *minería de datos* y la *estadística* radica en que la primera sólo trabaja con datos que ya fueron colectados

¹ *Data Warehouse* o almacén de datos, es definido como un proceso de gestión centralizada de datos, en donde se aplican herramientas que permiten resumir, describir y analizar los datos históricos para dar soporte a la toma de decisiones (Sumathi y Sivanandam, 2006).

(observados no experimentales), mientras que la segunda usa datos experimentales.

En general, el impulso al empleo de minería de datos se asocia a una serie de factores como (Larose, 2005):

- El crecimiento exponencial de la recolección de datos y la evolución del poder de cómputo.
- El almacenamiento de los datos en *data warehouses*, con la finalidad de tener acceso a una base de datos actualizada y confiable.
- La creciente disponibilidad de información en Internet.
- La presión competitiva del mercado en una economía globalizada.
- El desarrollo de herramientas comerciales para llevar a cabo la minería de datos.

Estos factores han llevado a considerar a la minería de datos como un campo interdisciplinario, el cual es desarrollado de manera conjunta o como apoyo de otras disciplinas como: tecnología de bases de datos, sistemas para la toma de decisiones, estadística, inteligencia artificial, aprendizaje automático, redes neuronales, reconocimiento de patrones, visualización de datos, computación paralela, entre otras (Hand *et al.*, 2001; Hernández *et al.*, 2004; Xu y Wunsch, 2005; Sumathi y Sivanandam, 2006). En la *Figura 2.1*, se observa la asociación de las diversas disciplinas relacionadas a la minería de datos.

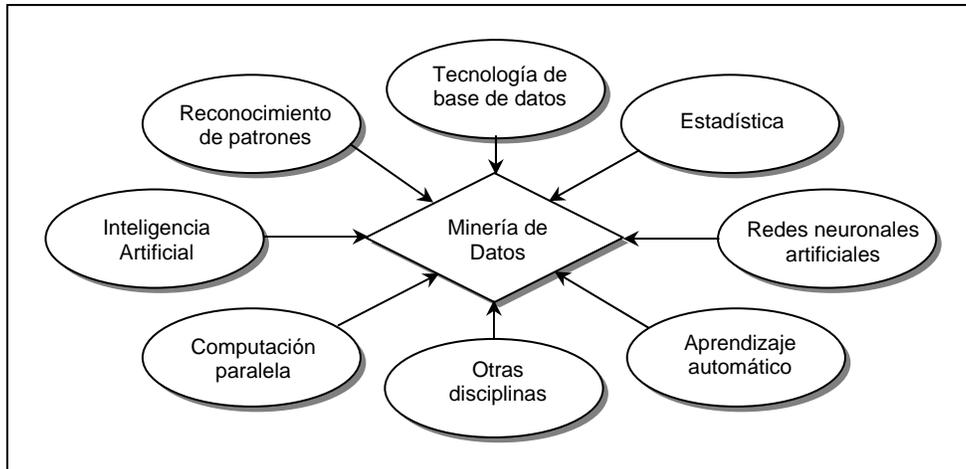


Figura 2.1 Asociación de las diversas disciplinas relacionadas a la minería de datos. Adaptado de Hernández *et al.* (2004)

En la *Figura 2.1* son claras las diferentes disciplinas que se relacionan con la minería de datos y, por tanto, es evidente la gran cantidad de áreas en donde pueden ser aplicadas: finanzas, análisis de mercado, procesos industriales, medicina, biología, química, telecomunicaciones, análisis científico, biometría, análisis ambiental, análisis hidrológicos, entre otros.

Un ejemplo tomado Witten y Frank (2005) muestra la aplicación de la minería de datos para establecer una actividad deportiva con base en variables ambientales. Los autores consideran un conjunto de datos con valores de algunos elementos climatológicos como: clima predominante, temperatura, humedad y viento, mediante los cuales se infiere si se debe o no jugar cierta actividad deportiva (tenis). En la *Tabla 2.1* se muestra a estos elementos con sus respectivos valores, así el clima puede ser *soleado, nublado o lluvioso*; la temperatura *alta, normal o fría*; la humedad *alta o normal*; y el viento *verdadero o falso*.

Tabla 2.1. Factores climatológicos mediante los cuales se infiere si se debe o no jugar tenis.

Clima	Temperatura	Humedad	Viento	Jugar Tenis
Soleado	Alta	Alta	Falso	No
Soleado	Alta	Alta	Verdadero	No
Nublado	Alta	Alta	Falso	Si
Lluvioso	Normal	Alta	Falso	Si
Lluvioso	Fría	Normal	Falso	Si
Lluvioso	Fría	Normal	Verdadero	No
Nublado	Fría	Normal	Verdadero	Si
Soleado	Normal	Alta	Falso	No
Soleado	Fría	Normal	Falso	Si
Lluvioso	Normal	Normal	Falso	Si
Soleado	Normal	Normal	Verdadero	Si
Nublado	Normal	Alta	Verdadero	Si
Nublado	Alta	Normal	Falso	Si
Lluvioso	Normal	Alta	Verdadero	No

Fuente: Witten y Frank (2005)

Para el manejo de la información disponible se deben establecer algunas reglas como:

- Si el clima = soleado y la humedad = alta, *entonces* jugar = no
- Si el clima = lluvioso y el viento = verdadero, *entonces* jugar = no
- Si el clima = nublado, *entonces* jugar = si
- Si la humedad = normal y el viento = falso, *entonces* jugar = si
- Si la humedad = normal, *entonces* jugar = si

Estas reglas se conocen como *reglas de asociación*, las cuales predicen la clasificación del ejemplo en términos de sí se debe o no jugar tenis. Estas reglas de asociación deben ser interpretadas secuencialmente, esto es, se inicia desde la primera regla, pasando por la segunda y, así sucesivamente, hasta llegar a la última.

2.2 Proceso de Descubrimiento de Conocimiento en Base de Datos

La minería de datos es un paso esencial de un proceso más amplio cuyo objetivo es el descubrimiento de conocimiento en bases de datos (*Knowledge Discovery in Databases o KDD*). Fayyad *et al.* (1996) en Klösgen y Zytkow (2002) definen al KDD como “*el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia comprensibles a partir de los datos*”. Con base en la definición de Fayyad *et al.* (1996), Vazirgiannis *et al.* (2003) consideran que en el KDD los *datos* vienen a ser el conjunto de casos (registros) que pertenecen a la base de datos y un *patrón* es un subconjunto de datos que comparten propiedades similares.

Para Hernández *et al.* (2004), el KDD es un proceso general de descubrir conocimiento desde bases de datos, mientras que la minería de datos viene a ser la aplicación de los métodos de aprendizaje y estadísticos. Asimismo, Fayyad *et al.* (1996) consideran que el KDD es un proceso iterativo e interactivo dividido en una secuencia de pasos (Figura 2.2), de los cuales la *minería de datos* es considerada como uno de los pasos más importantes en todo el proceso, esto es, el corazón del proceso KDD mismo (Wong y Leung, 2002).

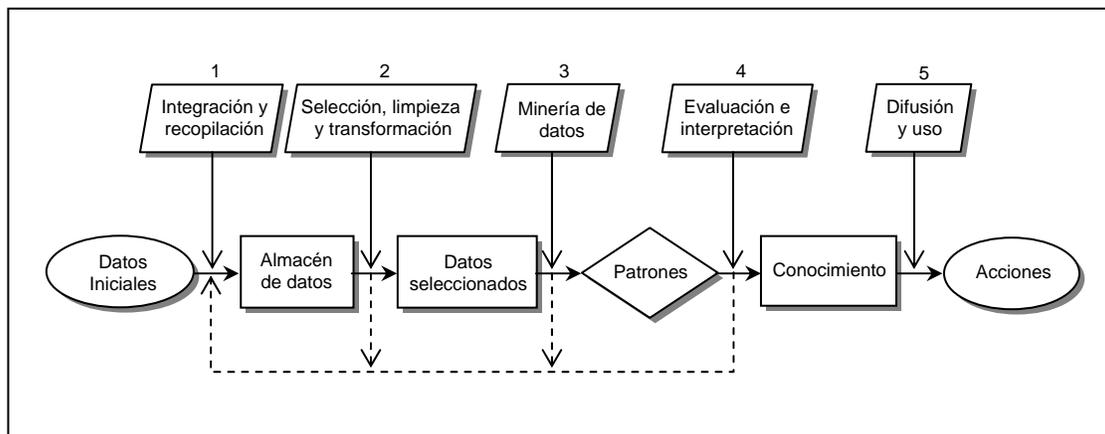


Figura 2.2. Proceso de descubrimiento de conocimiento en bases de datos (KDD). Adaptado de Vazirgiannis *et al.* (2003) y Hernández *et al.* (2004).

En el desarrollo del proceso de descubrimiento de conocimiento, tanto la iteración y la interacción son dos aspectos completamente necesarios, puesto que permiten volver a empezar en alguno de los pasos anteriores para poder obtener un conocimiento de calidad.

En términos generales, el *KDD* es un proceso que incluye el pre-procesamiento de datos, búsqueda de patrones y evaluación del conocimiento extraído. Estas etapas se presentan en una serie de pasos que a continuación se resumen; *para mayor información se recomienda al lector revisar Klösgen y Zytkow (2002), Vazirgiannis et al. (2003) y Hernández et al. (2004).*

1. *Integración y recopilación de datos.* En esta fase inicial se determinan las fuentes de información a utilizarse para la investigación. En ocasiones, este paso es uno de los que más tiempo requiere, puesto que comúnmente se recopila gran cantidad de datos, los cuales pueden ser obtenidos de bases de datos y/o almacenes de datos. Estos datos pueden ser tan diversos como: bancarios, eléctricos, académicos, demográficos, meteorológicos, etc.
2. *Selección, limpieza y transformación.* En ocasiones, dado que los datos provienen de diferentes fuentes, estas por lo general contienen ruido o valores atípicos. Conocidos los objetivos y requerimientos del proyecto, en este paso se selecciona el número efectivo de variables, se eliminan redundancias en los datos y, de ser necesario, se filtran aquellos que son considerados relevantes para el proceso de la minería de datos. Asimismo, se transforman los datos para la aplicación de alguna técnica en particular.
3. *Minería de datos.* En este paso se decide la técnica a utilizar, la cual debe ser cotejada con los objetivos del proyecto. Posterior a la selección y aplicación de la técnica sobre los datos previamente preparados, éste debe ser validado de acuerdo al conocimiento y dominio existente. Puesto que los resultados obtenidos en este paso dependen considerablemente de los anteriores, existe la posibilidad de volver a iniciar desde alguno de los pasos precedentes,

debido a la necesidad de requerir nuevos datos o incluso en caso de que se modifique la definición del problema.

4. *Evaluación e interpretación.* Los patrones extraídos en el paso anterior son interpretados y evaluados de acuerdo a mediciones que garanticen la identificación de patrones verdaderamente significativos (patrones que representen conocimiento). Algunas de estas mediciones podrían ser: *validación cruzada*, la cual divide los datos en dos conjuntos (entrenamiento y prueba); *matrices de confusión* (empleadas generalmente en problemas de clasificación), la cual indica cuantas clasificaciones se han hecho correctamente para cada tipo; entre otras.
5. *Difusión y uso del conocimiento.* En esta parte, el conocimiento extraído es incorporado en algún sistema o procedimiento para su difusión y uso de los usuarios finales. Este resultado debe ser examinado para descartar posibles errores en la representación del conocimiento. Por lo general, en los modelos predictivos se deben comprobar continuamente las prestaciones del mismo, puesto que los patrones podrían cambiar con el tiempo, de ser así, se tendría que reevaluar, reentrenar y posiblemente reconstruir completamente el modelo.

2.3 Modelos y tareas de minería de datos

Al ser la minería de datos un método para extraer conocimiento útil mediante el análisis de los datos, ésta recurre a modelos que permitan encontrar relaciones, patrones o reglas inferidas previamente desconocidas (Hernández *et al.*, 2004). Los modelos empleados en la minería de datos son el *descriptivo* y el *predictivo*.

2.3.1 *Modelo descriptivo*

En el modelo descriptivo se identifican patrones que describen los datos mediante tareas, ej. *agrupamiento (clustering)* y *reglas de asociación* (Vazirgiannis *et al.*, 2003). Hernández *et al.* (2004) destacan que mediante este modelo se identifican patrones que explican o resumen el conjunto de datos, siendo estos útiles para explorar las propiedades de los datos examinados. Los modelos descriptivos siguen un tipo de *aprendizaje no supervisado*, que consiste en adquirir conocimiento desde los datos disponibles, sin requerir influencia externa que indique un comportamiento deseado al sistema (Sumathi y Sivanandam, 2006).

2.3.2 *Modelo predictivo*

Este modelo se emplea para estimar valores futuros de variables de interés. El proceso se basa en la información histórica de los datos, mediante las cuales se predice un comportamiento de los datos, ya sea mediante clasificaciones, categorizaciones o regresiones (Hernández *et al.*, 2004; Sumathi y Sivanandam, 2006). Los modelos predictivos siguen un *aprendizaje supervisado*, que consiste en aprender mediante el control de un supervisor o maestro que determina la respuesta que se desea generar del sistema (Sumathi y Sivanandam, 2006). El atributo a predecir se conoce como variable dependiente u objetivo, mientras que los atributos utilizados para realizar la predicción se llaman variables independientes o de exploración (Hernández *et al.*, 2004).

2.3.3 *Tareas de minería de datos*

Dentro de los modelos descriptivos y predictivos se encuentran diferentes tareas específicas como: *agrupamiento*, *reglas de asociación*, *clasificación*, *regresión*, entre otras. Estas tareas corresponden a un tipo de problema específico en el proceso de minería de datos. En la *Figura 2.3*, se muestra una representación general de los modelos y tareas hallados en el proceso de minería de datos, siendo abordadas aquí brevemente cada una de ellas.

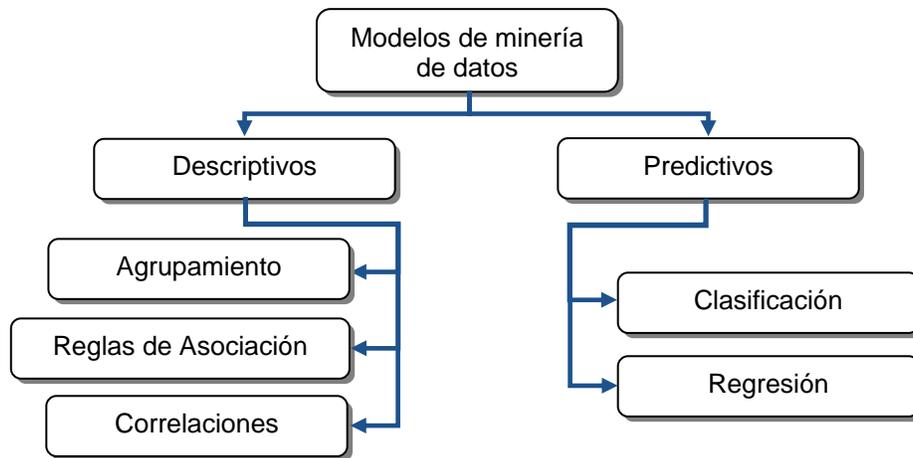


Figura 2.3. Representación general de los modelos y tareas de minería de datos

2.3.3.1 Agrupamiento (Clustering)

En esta tarea se evalúan similitudes entre los datos para construir modelos descriptivos, analizar correlaciones entre las variables o representar un conjunto de datos en un pequeño número de regiones (Sumathi y Sivanandam, 2006). Berry y Linoff (2004) y Sumathi y Sivanandam (2006) consideran al agrupamiento como la tarea de dividir una población heterogénea en un número de subgrupos homogéneos de acuerdo a las similitudes de sus registros. Dentro de esta tarea existen dos tipos principales de agrupamiento (Larose, 2005): el *jerárquico* que se caracteriza por el desarrollo recursivo de una estructura en forma de árbol, y el *particional* que organiza los registros dentro de k grupos. Los métodos particionales tienen ventajas en aplicaciones que involucran gran cantidad de datos para los cuales la construcción de un árbol resulta complicada. Una característica de este tipo de agrupamiento es el establecer *a priori* el número de grupos de entrada (k), por lo que en la práctica es necesario repetir la prueba estableciendo diferentes números de grupos, eligiendo la solución que mejor se adapte al objetivo del problema (Jain *et al.*, 1999). Un método sugerido por Milligan (1980, 1985) y Hair *et al.* (1995) para determinar el número de grupos de entrada (k) es usar el resultado obtenido por algún algoritmo jerárquico, mediante el cual se obtiene el número deseado de grupos, posteriormente se aplica algún algoritmo particional.

2.3.3.2 Reglas de asociación

Mediante esta tarea se identifican afinidades entre la colección de los registros examinados, buscando relaciones o asociaciones entre ellos. Las afinidades son expresadas como reglas de la forma: “Si X entonces Y”, donde X y Y son los registros de una transacción (Sumathi y Sivanandam, 2006). El interés por esta tarea se debe principalmente a que las reglas proporcionan una forma concisa de declarar la información potencialmente útil (Vazirgiannis *et al.*, 2003). Las reglas se evalúan usando dos parámetros: precisión y cobertura. La *cobertura* es el número de instancias o datos hallados correctamente, mientras que la *precisión* es el porcentaje de instancias halladas correctamente (Witten y Frank, 2004). Las ventajas más frecuentes en las reglas de asociación son el *descubrimiento de asociación* y *de secuencia* (Hernández *et al.*, 2004). El descubrimiento de asociación encuentra relaciones que aparecen conjuntamente a un acontecimiento y la secuencia la asocia al tiempo.

2.3.3.3 Correlaciones

Las correlaciones son una tarea descriptiva que se usan para determinar el grado de similitud de los valores de dos variables numéricas. Un mecanismo estándar para medir la correlación es el coeficiente de correlación, para este caso llamado R , el cual es un valor real comprendido entre -1 y 1. Así, si R es 1 las variables están totalmente correlacionadas; si R es -1 las variables están correlacionadas negativamente; y si R es cero no existe correlación. Por consiguiente, cuando R es positivo, las variables tienen un comportamiento similar y cuando R es negativo una variable crece y la otra decrece (Hernández *et al.*, 2004). Una manera de visualizar la posible correlación entre las observaciones de dos variables (X e Y), es a través de un diagrama de dispersión, en el cual los valores que toman estas variables son representados por puntos. Su principal desventaja es que no puede ser usada para hacer predicciones, puesto que no es clara la forma que toma la relación.

2.3.3.4 Clasificación

Es una de las principales tareas en el proceso de minería de datos que se emplea para asignar datos a un conjunto predefinido de variables (Vazirgiannis *et al.*, 2003). El objetivo de la clasificación es encontrar algún tipo de relación entre los atributos de entrada y los registros de salida para comprender el comportamiento de los datos, así mediante el conocimiento extraído se puede predecir el valor de un registro desconocido (Sumathi y Sivanandam, 2006). Sin embargo, el mayor problema de la clasificación es que muchas veces no es representativo y no proporciona un conocimiento detallado, sólo otorga predicciones. Algunos algoritmos comprendidos en esta tarea son: clasificación bayesiana, árboles de decisión, redes neuronales artificiales, entre otros (Hand *et al.*, 2001; Vazirgiannis *et al.*, 2003).

2.3.3.5 Regresión

La regresión es el aprendizaje de una función cuyo objetivo es predecir valores de una variable continua a partir de la evolución de otra variable también continua, la cual por lo general es el tiempo (Sumathi y Sivanandam, 2006). En la regresión, la información de salida es un valor numérico continuo o un vector con valores no discretos (Vazirgiannis *et al.*, 2003). Ésta es la principal diferencia respecto a la clasificación donde el valor a predecir es numérico. Si sólo se dispone de una variable definida se trata de un problema de *regresión simple*, mientras que si se dispone de varias variables se trata de un problema de *regresión múltiple*. A esta tarea también se le conoce como: *interpolación*, cuando el valor o valores predichos están en medio de otros; o *estimación*, cuando se predice valores futuros (Hernández *et al.*, 2004).

2.4 Técnicas de minería de datos

Como se mencionó en la sección anterior, la minería de datos utiliza tareas las cuales requieren ser resueltas a través de técnicas o algoritmos específicos, que de acuerdo con Hernández *et al.* (2004) son un filtro donde al introducir los datos se produce una serie de patrones relevantes que representan el nuevo conocimiento útil. Esta postura se ilustra en la *Figura 2.4*.

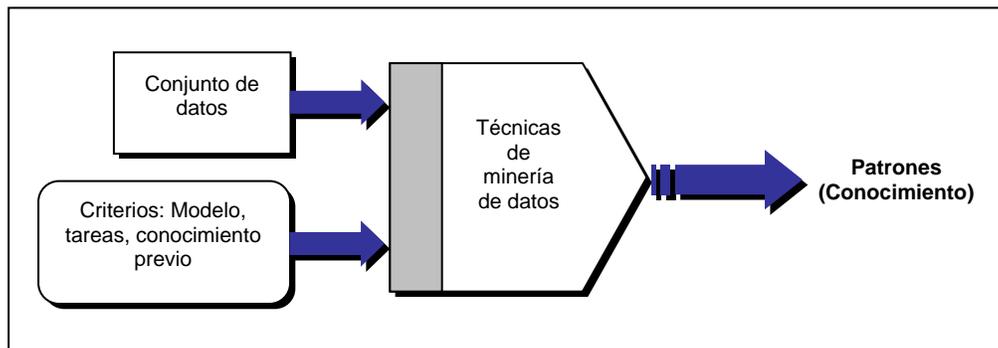


Figura 2.4. Proceso de extracción de patrones mediante el uso de técnicas de minería de datos. Fuente: Adaptado de Hernández *et al.* (2004)

A continuación se presentan brevemente algunos tipos de técnicas disponibles en la minería de datos; *si se requiere conocer más del tema se sugiere al lector revisar los siguientes autores: Klösgen y Zytkow, 2002; Vazirgiannis et al., 2003; Hernández et al., 2004 y Berry y Linoff, 2004.*

2.4.1 Técnicas bayesianas

Mediante las técnicas bayesianas se permiten construir modelos descriptivos y predictivos. Estas técnicas se basan en estimar la probabilidad de pertenencia (a un grupo o clase), mediante la estimación de probabilidades, utilizando para ello el teorema de Bayes. El teorema de Bayes parte de una situación en la que es posible conocer las probabilidades de que ocurran una serie de sucesos. Una de las principales ventajas de las técnicas bayesianas, sobre las otras técnicas de minería de datos, es el uso explícito de la teoría de probabilidad para cuantificar la incertidumbre de datos atípicos.

2.4.2 *Técnicas algebraicas y estadísticas*

El objetivo de estas técnicas consiste en explicar el comportamiento de una variable a partir del comportamiento de otras. Generalmente, se basan en expresar modelos y patrones mediante fórmulas algebraicas, funciones lineales, funciones no lineales y distribuciones estadísticas (medias, varianzas, correlaciones, etc.). Algunos de los algoritmos más conocidos dentro de este grupo de técnicas son: regresión lineal, regresión logística y regresión logarítmica. Estas técnicas son eficientes y en muchos casos comprensibles, siendo una de las primeras opciones a utilizar ante problemas de regresión y clasificación, especialmente cuando los atributos originales son cuantitativos (numéricos).

2.4.3 *Técnicas basadas en redes neuronales artificiales*

Las redes neuronales artificiales son sistemas de cálculo que se asemejan a las redes neuronales biológicas al utilizar nodos (neuronas) interconectados. Estos nodos reciben la información, realizan operaciones sobre los datos y transmiten sus resultados a otros nodos, hasta proporcionar una respuesta de activación de los nodos de salida. De acuerdo a su aprendizaje, las redes neuronales se dividen en *supervisadas* (por ej. perceptrón multicapa) aplicables a tareas de clasificación y regresión; y *no supervisadas* (por ej. redes de Kohonen) utilizadas principalmente para el agrupamiento y reducción de dimensionalidad. Su principal ventaja radica en que cuando están bien ajustadas se obtienen precisiones muy altas. Sin embargo, estas técnicas requieren de cierta experiencia para la elección precisa de los algoritmos, las topologías y los parámetros.

2.4.4 *Técnicas basadas en árboles de decisión y sistema de reglas*

Son técnicas comúnmente usadas en la minería de datos, puesto que son fáciles de usar y admiten valores tanto numéricos y discretos (nominales). Estas técnicas son tolerantes al ruido y a valores faltantes; aportan claridad porque permiten desplegar los resultados en profundidad de mayor a menor detalle. Los algoritmos

más conocidos en este grupo de técnicas son: *divide y vencerás*, como el ID3 (Quinlan, 1983), el C4.5 (Quinlan, 1993), o el CART (Breiman *et al.*, 1984), y *separa y vencerás*, como el CN2 (Clark y Niblett, 1989; Clark y Boswell, 1991). Su principal desventaja es que no son tan precisos como las redes neuronales y son dependientes de la muestra o conjunto de datos inicial.

2.4.5 *Técnicas basadas en distancia o densidad*

Son técnicas que se basan en distancias, como los vecinos más próximos, y de manera más sofisticada, mediante la estimación de funciones de densidad. Mediante estas técnicas se obtienen resultados con bastante expresividad, esto es, muestra la existencia de mejores relaciones entre los datos. Pero su principal desventaja radica en que no son muy eficientes si el número de registros (elementos) es excesivamente grande. Las técnicas más comunes para el agrupamiento son: redes de Kohonen (Kohonen, 1982), K-medias (MacQueen, 1967) y Two-step (Zhang *et al.*, 1996). Las técnicas más usadas para la clasificación son: Estimación bayesiana, K-vecinos (Fix y Hodges, 1951) y LQV (Redes de Cuantización Vectorial) (Kohonen, 1986).

2.4.6 *Técnicas basadas en núcleo y máquinas de soporte vectorial*

Son técnicas que tratan de maximizar el margen entre los grupos y las clases formadas. Para ello, se basan en transformaciones que aumentan la dimensionalidad de los grupos. Estas transformaciones son conocidas como núcleos (*kernels*). Este tipo de técnicas son muy eficientes y precisas por trabajar con datos de alta dimensionalidad. Se conoce como datos de alta dimensionalidad a la presencia de un gran número variables de entrada. El inconveniente al usar estas técnicas es que a veces es necesario elegir una buena función de separación para obtener buenos resultados. Los campos donde las máquinas de soporte vectorial (SVM) han sido aplicadas con éxito son: visión por computadora, bioinformática, recuperación de la información, procesamiento del lenguaje natural y análisis de series temporales (Hernández *et al.*, 2004).

2.4.7 Técnicas estocásticas y difusas

En este grupo de técnicas se encuentran la lógica difusa, el razonamiento probabilístico y los algoritmos genéticos, que junto a las redes neuronales, forman lo que se denomina *soft computing* (computación flexible). *Soft computing* consiste en el uso cooperativo de las técnicas mencionadas para dar solución a diversos problemas de clasificación, predicción, optimización, agrupamiento, etc. Estas técnicas tienen como ventaja la flexibilidad de cambiar la heurística (estrategia de búsqueda) sin cambiar el algoritmo y la expresividad debido al uso de fronteras difusas, en vez de fronteras discretas. Sin embargo, tienen como desventaja el costo computacional (tiempo y memoria) que requieren para llegar a una solución.

En la *Tabla 2.2*, se presenta una relación de técnicas mencionadas, las cuales ayudan a descubrir patrones o propiedades particulares del conjunto de datos para obtener información útil.

Tabla 2.2. Relación existente entre las tareas y técnicas de minería de datos.

Nombre	Predictivo		Descriptivo		
	Clasificación	Regresión	Agrupamiento	Reglas de Asociación	Correlaciones
Redes neuronales	✓	✓	✓		
Árboles de decisión ID3, C4.5, C5.0	✓				
Árboles de decisión CART	✓	✓			
Otros árboles de decisión	✓	✓	✓	✓	
Redes de Kohonen			✓		
Regresión lineal y logarítmica		✓			✓
Regresión logística	✓			✓	
K-medias			✓		
A priori				✓	
Naive Bayes	✓				
Vecinos más próximos	✓	✓	✓		
Análisis de comp. principales					✓
Two-step, Cobweb			✓		
Algoritmos genéticos y evolutivos	✓	✓	✓	✓	✓
Maquinas de soporte vectorial	✓	✓	✓		
Reglas CN2 (cobertura)	✓			✓	
Análisis discriminante multivariente	✓				

Fuente: Hernández *et al.* (2004)

Como se mencionó cada una de éstas técnicas está enfocada a tareas específicas y, por tanto, su empleo en la mayoría de los casos, depende del tipo de dato que se tiene y el modelo hacia el cual está enfocado. Además, en la *Tabla 2.2* se observa que todas las tareas pueden ser resueltas por más de una técnica, y que algunas técnicas pueden resolver más de una tarea, como es el caso de las redes neuronales, los árboles de decisión, las regresiones (lineal, logarítmica y logística) y los algoritmos genéticos, entre otros.

2.5 Metodología para el desarrollo de un proyecto de minería de datos

A pesar de la diversidad de técnicas y herramientas disponibles en la minería de datos, existe la necesidad de disponer de una aproximación sistemática (metodología) para el desarrollo de un proyecto. Por esta razón, diversas empresas y consultorías en el mundo han desarrollado metodologías de trabajo para guiar al usuario para una aplicación exitosa de las técnicas de minería de datos a través de una sucesión de pasos (Sumathi y Sivanandam, 2006).

En la actualidad, las principales metodologías utilizadas en la realización de proyectos de minería de datos son:

- CRISP-DM (*Cross Industry Standard Process for Data Mining*), (Chapman *et al.*, 2000) desarrollada en 1996 por el consorcio de empresas europeas, DaimlerChrysler, SPSS (Statistical Product and Service Solutions) y NCR Systems Engineering Copenhagen.
- SEMMA (*Sampling, Exploration, Modification, Modeled, Assessment*) desarrollado por la empresa *Statistical Analysis Systems Inc.* (SAS Institute), quien la define como: “*el proceso de selección, exploración y modelado de grandes cantidades de datos para descubrir patrones de negocio desconocidos*”.

En términos generales, ambas metodologías comparten la misma filosofía, estructurando el proyecto de minería de datos en diversas fases relacionadas entre sí, con la finalidad de conseguir un proceso repetitivo e interactivo. Sin embargo, para Gondar (2004) una de las diferencias más significativas radica en que la metodología SEMMA se centra en las características técnicas del proyecto, mientras que CRISP-DM mantiene una perspectiva más próxima a los objetivos reales del proyecto. Esta diferencia se aprecia desde la primera fase del proyecto, donde la metodología SEMMA comienza realizando un muestreo de datos y CRISP-DM inicia con un análisis del problema. Otra diferencia a considerar es que SEMMA está ligada a productos comerciales del SAS Institute como: *Enterprise Miner* y *Text Miner*; mientras que CRISP-DM mantiene una posición neutra con respecto a las herramientas utilizadas.

CRISP-DM (Chapman *et al.*, 2000) consta de cuatro niveles, organizados de forma jerárquica en tareas que van desde el nivel más general hasta los casos más específicos. En el nivel general, el proceso está organizado en seis fases: análisis del problema, análisis de datos, preparación de datos, modelado, evaluación e implementación. A su vez, cada fase consta de tareas generales de un segundo nivel (Figura 2.5). Estas tareas generales se proyectan a tareas específicas, donde se describen las acciones que se deben desarrollar para cada situación específica.

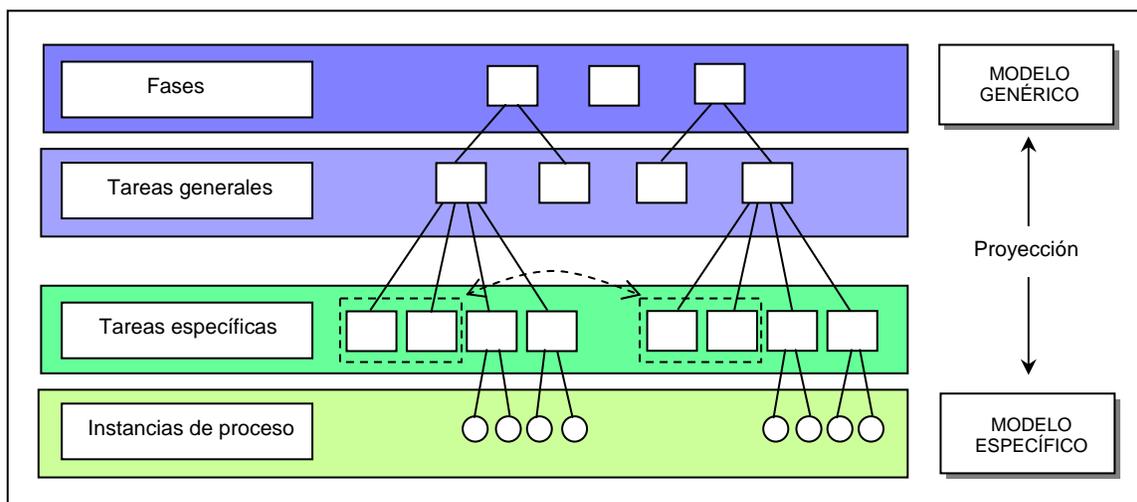


Figura 2.5. Esquema de los cuatro niveles de abstracción de la metodología CRISP-DM. Fuente: Chapman *et al.*, 2000.

Por ejemplo, si la tarea general es “limpieza de datos”, en el tercer nivel se especificarían tareas concretas como: limpieza de datos numéricos y/o limpieza de datos categóricos. Finalmente, en el cuarto nivel se determina el conjunto de acciones y decisiones a realizar para resolver las tareas establecidas en el nivel anterior (Sumathi y Sivanandam, 2006).

Las fases en la metodología CRISP-DM se pueden representar como un ciclo, como se muestra en la *Figura 2.6*. Así, cada una de las fases representa el análisis del problema, el análisis de datos, la preparación de datos, el modelado, la evaluación, y la implementación de estos. La naturaleza iterativa de la metodología CRISP-DM se simboliza por el círculo exterior.

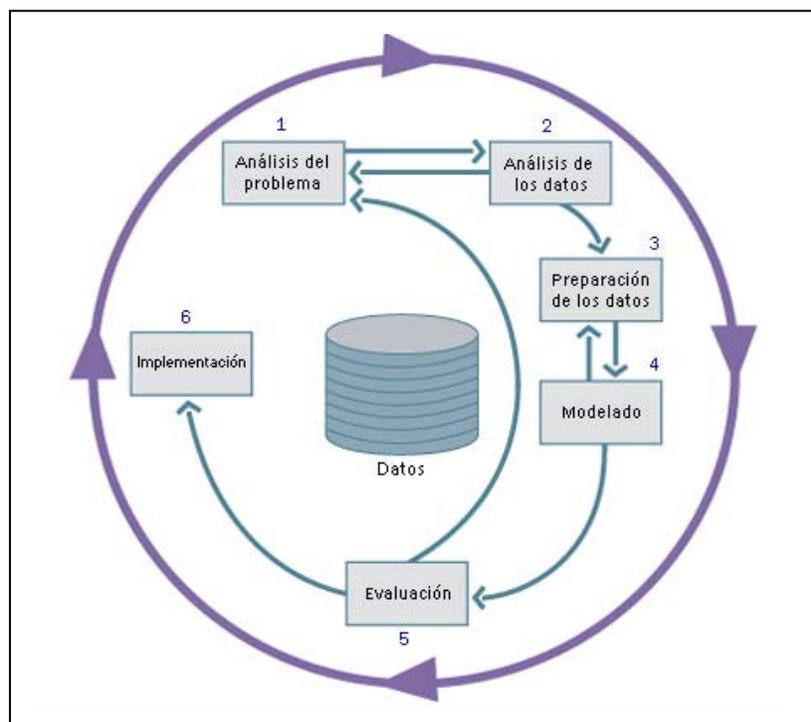


Figura 2.6. Fases del proceso de minería de datos, basado en la metodología CRISP-DM. Fuente: Chapman *et al.*, 2000.

La secuencia de las fases no es estricta, puesto que éstas pueden interactuar entre sí durante el desarrollo del proyecto. De esta manera, la siguiente fase en la secuencia, a menudo depende de los resultados asociados con la fase precedente. Las flechas indican las dependencias significativas entre las diferentes fases, por

ejemplo, si se está en la fase de *modelación*, es posible volver a la fase de *preparación de datos* para refinar los datos, antes de seguir con la fase de *evaluación* (Larose, 2005).

El proceso no concluye cuando se encuentra una solución, puesto que esta solución puede acarrear nuevas investigaciones a explorar (Chapman *et al.*, 2000). A continuación se describe cada una de las fases:

1. *Análisis del problema.* En esta fase inicial se debe comprender con claridad los objetivos y requerimientos del proyecto, con la finalidad de elaborar una buena planeación en el desarrollo.
2. *Análisis de los datos.* Se establece el contacto directo con el problema. Las actividades a realizar son: la recolección inicial de datos, la identificación de la calidad de los datos y el establecimiento de posibles relaciones más evidentes que permitan obtener las primeras hipótesis.
3. *Preparación de los datos.* Aquí se realiza la selección de datos a los que se va a aplicar la técnica de modelado (variables y muestras), la limpieza de los datos, la generación de variables adicionales, la integración de diferentes conjuntos de datos y cambios de formato. Esta fase laboriosa, está directamente relacionada con la fase de *modelado*, puesto que en función de la técnica a utilizar, los datos necesitan ser procesados en diferentes formas.
4. *Modelado.* Aquí se seleccionan las técnicas apropiadas para el desarrollo del proyecto. La técnica a emplearse en esta fase debe ser seleccionada en función a los siguientes criterios: Ser apropiada al problema, disponer de datos adecuados, cumplir los requerimientos del problema, y el conocimiento de la técnica.
5. *Evaluación.* En esta fase se evalúa el modelo, no desde el punto de vista de los datos, sino del cumplimiento a los requerimientos iniciales. Antes de

proceder a su implantación para su uso habitual, se debe revisar todo el proceso teniendo en cuenta los resultados obtenidos, e identificando posibles errores que llevan a repetir algún proceso anterior.

6. *Implementación.* Si el modelo generado es válido, desde el punto de vista de cumplimiento a los requerimientos iniciales, se procede a su implementación y explotación. Normalmente los proyectos de minería de datos no terminan en la implementación del modelo, sino se deben documentar y presentar los resultados de manera comprensible para alcanzar un mejor entendimiento del conocimiento.

2.5.1 *Herramienta de desarrollo*

Metodologías como CRISP-DM que permiten una perspectiva más próxima a los objetivos reales del proyecto bajo un desarrollo sistemático de éste son fundamentales. Para llevar a cabo este desarrollo sistemático, se hace uso de software que siga el procedimiento estándar de la metodología CRISP-DM facilitando la obtención de resultados útiles y confiables. Así, el software *Clementine* ofrece un método estratégico para encontrar relaciones útiles entre grandes conjuntos de datos, el cual incluye el acceso de datos, transformaciones, modelados, evaluación y despliegue de resultados.

Clementine fue originalmente desarrollado por Integral Solutions Ltd. (ISL), posteriormente fue adquirido por SPSS Inc. (*Statistical Product and Service Solutions*) (Sumathi y Sivanandam, 2006). Esta herramienta permite realizar modelos predictivos y descriptivos trabajando directamente sobre bases de datos, tablas Excel, archivos SPSS, o sobre archivos de texto plano tipo txt y archivos SPSS. Además, es un sistema multiplataforma, disponible para sistemas operativos Windows, Solaris, HP-UX, AIX y OS/400 de IBM (Hernández *et al.*, 2004)

Entre las principales características de Clementine esta su capacidad de procesar grandes volúmenes de datos (miles de registros), su disponibilidad de una amplia variedad de técnicas y su ambiente gráfico integrado de desarrollo (Figura 2.7).

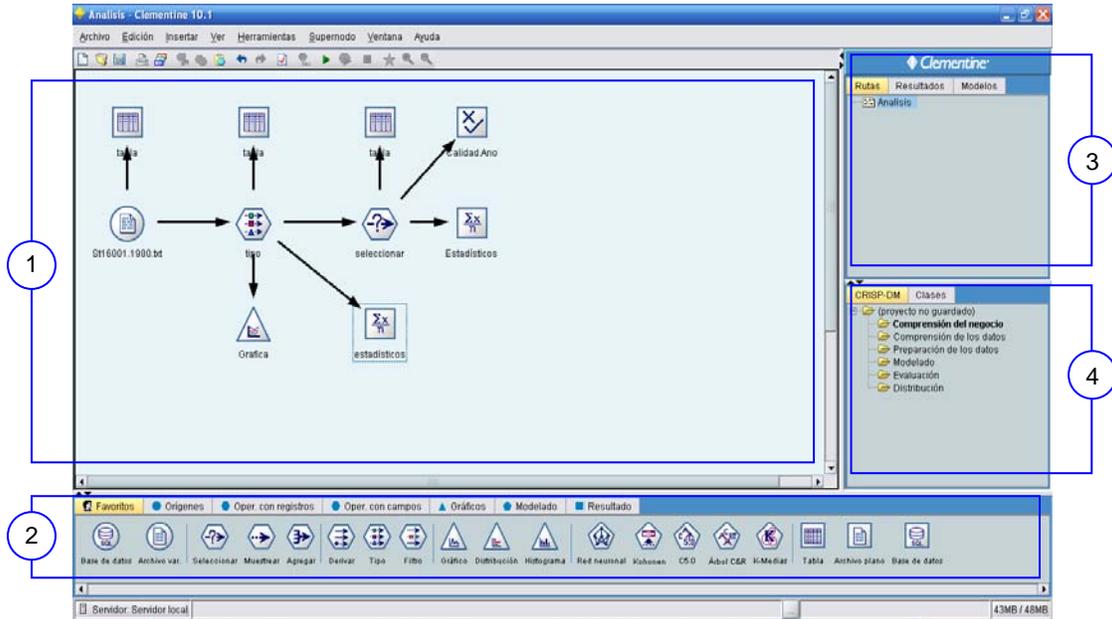


Figura 2.7. Interfaz gráfica de Clementine. Fuente: Clementine, 2006.

A través de la interfaz visual de Clementine se generan rutas, esto es, mapas visuales que representan las distintas etapas del proceso analítico de la minería de datos. Esta secuencia se denomina ruta de datos porque los datos fluyen registro por registro desde el origen, pasando por cada manipulación y, finalmente, hasta llegar al destino, que es el modelo generado o un nodo de resultados.

Una breve descripción de los componentes de este software con base en la distribución de la *Figura 2.7*, es el siguiente (Clementine, 2006c):

- (1) es el espacio o área de trabajo que se denomina “zona de rutas”, en donde cada operación se representa con un nodo y estos nodos son vinculados entre sí para generar una ruta que representa el flujo de datos.

- (2) es el área denominado “paletas”, en donde se encuentran las diversas técnicas analíticas de la minería de datos. Cada paleta contiene un grupo de nodos que representan gráficamente las tareas de minería de datos, como el acceso y el filtrado de datos, la creación de gráficos y la generación de modelos.
- (3) es la zona denominada “administración de objetos y resultados”. Las opciones Rutas, Resultados y Modelos se utilizan para mostrar y administrar los tipos de objetos correspondientes al proyecto en ejecución.
- (4) es la zona denominada “proyectos”, que se utiliza para administrar los proyectos de minería de datos. Mediante esta opción se permite disponer de una ubicación central para el almacenamiento y seguimiento de los requerimientos asociados con el proyecto.

Clementine es flexible, adaptable a cualquier tipo de proyecto por lo que su uso representa ventajas por su fácil acceso a la información, uso de técnicas para su análisis y el despliegue de la información estadística obtenida, así como la reducción en el tiempo de ejecución y, por tanto, reducción de costos.

2.6 Requerimientos para regionalización en hidrología

En Hidrología, la regionalización se apoya en el concepto de homogeneidad hidrológica de una región, es decir, dividir la zona de estudio en regiones cuyas características topográficas, climáticas, orográficas, etc. sean similares (Kyselý *et al*, 2007). La necesidad de analizar cómo se relacionan las variables climáticas que varían en tiempo y espacio son esenciales para el estudio de áreas extensas. El papel de las variables como precipitación, temperatura, evaporación, radiación, etc. es crucial para entender la variación natural a lo largo de una cuenca. Esto permite identificar si un evento es extraño o es parte de los movimientos de la zona. Sin embargo, la escasez de datos es un problema común en hidrología, tornándose

necesaria la optimización de la información disponible de una determinada región geográfica (Pelczer *et al.*, 2007).

Los métodos de regionalización han sido utilizados de manera exitosa para el análisis regional, por ejemplo para determinar la tendencia y frecuencia de inundaciones, sequías y el efecto de otras variables. Entre estos métodos cabe destacar aquellos enfocados a la clasificación de conjuntos de datos, para el cual existe una colección variada de herramientas de informática, como las redes neuronales artificiales (redes de Kohonen) y la lógica difusa, que permiten formar regiones homogéneas con características similares (Hall y Minns, 1999).

Uno de los métodos de regionalización frecuentemente referenciado en la literatura, es la aproximación de *regiones de influencia* (ROI), introducido por Burn (1990). El ROI es un método alternativo de aproximación para el análisis regional. La idea principal de ROI es que no hay necesidad de establecer límites fijos entre las regiones propuestas, más bien las regiones son definidas en forma flexible. El método ROI ha sido empleado para el análisis de la frecuencia de posibles inundaciones que puede ocurrir en las regiones homogéneas establecidas. Otras aplicaciones del ROI son (Gaál *et al.*, 2006): en el del Reino Unido lo utilizaron para la regionalización de características de sequía y en Italia lo emplearon como antecedente para la comparación de similitudes hidrológicas medidas en el norte del país. Otros métodos aplicados en el análisis de la frecuencia de inundaciones son: el método residual (MOR), el análisis de correlación canónica (CCA), la aproximación jerárquica, la extensión del método ROI y el análisis de grupos (CA) (Rao y Srinivas, 2006).

En la investigación climatológica se necesita con frecuencia clasificar las variables en grupos homogéneos e identificar sus características con la finalidad de profundizar en la comprensión de los fenómenos ocurridos (Muñoz y Rodrigo, 2002). El *análisis de grupos* es uno de los métodos que se emplean para obtener tales clasificaciones, el cual permite trabajar con archivos de datos de diferentes tamaños. Así, el *agrupamiento* es reconocido como un método eficaz para tratar

problemas de agrupamiento de estaciones en regiones climatológicamente homogéneas (Gong y Richman, 1995 en Muñoz y Rodrigo, 2002) o el agrupamiento de periodos de tiempo (días, años, etc.) en grupos que reflejan la ocurrencia de determinados sucesos o patrones meteorológicos (Ramos, 2001 en Muñoz y Rodrigo, 2002).

Un grupo consta de uno o más vectores de datos, a su vez estos vectores comprenden varios atributos (variables). En el área de hidrología los atributos utilizados para el análisis de la frecuencia de inundaciones fueron (Rao y Srinivas, 2006):

- características fisiográficas (área de drenaje, extensión de la corriente, densidad de la corriente, etc.),
- características geográficas (latitud, longitud y altitud),
- medida de tiempo de respuesta de la cuenca (periodo de densidad de drenaje y entrada de agua, vía precipitación o aportes subterráneos, medidos en tiempos pico del hidrograma),
- factores meteorológicos (promedio anual de lluvias, intensidad de precipitación, etc.), y
- estadísticas de inundación del lugar.

El objetivo final de cualquier algoritmo de *agrupamiento* es encontrar subgrupos de datos dentro de un conjunto mayor de datos, maximizando la similitud de los elementos dentro de los grupos. Autores como Hand *et al.* (2001), Larose (2005), Witten y Frank (2004) y Berry y Linoff (2004) consideran dos tipos de agrupamiento: jerárquico y no jerárquico (particional). A diferencia de Jain *et al.* (1999), Vazirgiannis *et al.* (2003), Xu y Wunsch (2005) y Berkhin (2006) quienes consideran que debido a la gran cantidad de algoritmos de agrupamiento disponibles, los cuales varían de acuerdo a su estructura, los grupos se clasifican en: jerárquico, no jerárquico, basado en densidad, basado en Grid y difuso, entre otros.

En este documento se aborda el agrupamiento jerárquico y no jerárquico. *Para información relacionada con los otros tipos de agrupamiento se recomienda al lector revisar Jain et al. (1999), Vazirgiannis et al. (2003), Xu y Wunsch (2005) y Berkhin (2006).*

El *agrupamiento jerárquico* se caracteriza por el desarrollo recursivo de una jerarquía o estructura en forma de árbol (*dendograma*). Este tipo de agrupamiento se divide en *aglomerativo* o *divisivo*. El *método aglomerativo* empieza con cada registro formando un grupo independiente. En los pasos subsiguientes, los dos grupos más cercanos son agregados a un nuevo grupo cada vez más grande; de este modo, el proceso continúa hasta que todos los registros formen parte de un sólo grupo. El *método divisivo* considera todos los registros agrupados en un sólo conjunto, y a medida que se va iterando, se dividen en subconjuntos independientes cada vez más pequeños (Larose, 2005). Algunos algoritmos de agrupamiento jerárquico son: Bietápico, BIRCH (Balanced Iterative Reducing and Clustering using Hierarchical), CURE (Clustering Using Representatives), ROCK (RObust Clustering algorithm using linKs) y CHAMELEON, entre otros.

El *agrupamiento no jerárquico o particional* organiza los registros dentro de k grupos. Es decir, determina el número de particiones mediante un procedimiento iterativo que optimiza la estructura local o global de los datos agrupados (Vazirgiannis *et al.*, 2003). Los métodos particionales tienen ventajas en aplicaciones que involucran gran cantidad de datos para los cuales la construcción de un árbol resulta complicado (Witten y Frank, 2005). El problema de los algoritmos particionales es la decisión del número deseado de grupos de salida, por lo que en la práctica es necesario repetir la prueba considerando un número diferente de grupos, eligiendo la solución que mejor se adapte al objetivo del problema. Algunos algoritmos dentro de este tipo de agrupamiento son: K-medias, K-medianas, K-mode, PAM (*Partitioning Around Medoids*), CLARA (*Clustering Large Applications*) y CLARAMS (*Clustering Large Applications based on Randomized Search*), entre otros.

2.7 Síntesis

Hoy en día, la recopilación de información es una tarea común siendo fácil de digitalizar los datos, así como almacenarlos a un bajo costo. La minería de datos ha venido a resolver dos grandes retos:

- Trabajar con grandes volúmenes de datos procedentes de sistemas de información (datos coleccionados para algún propósito, por ejemplo el registro diario de las observaciones climatológicas).
- Usar técnicas adecuadas para identificar tendencias y comportamientos en el conjunto de datos, mediante la extracción de información y el descubrimiento de conocimiento útil.

Para la extracción de la información, los modelos empleados en la minería de datos son el *descriptivo* y el *predictivo*. Su empleo depende del tipo de dato que se tiene y el modelo hacia el cual está enfocado. Una de las tareas para construir modelos descriptivos basados en la regionalización es el agrupamiento cuyo objetivo es evaluar similitudes entre los datos para representarlos en un pequeño número de regiones, es decir, se divide una población heterogénea en un número de subgrupos homogéneos de acuerdo a las similitudes de sus registros. Por esta razón, el agrupamiento es la tarea que se ajusta al planteamiento inicial del problema que es la regionalización de zonas climatológicamente homogéneas en el estado de Michoacán.

Las técnicas de agrupamiento Bietápico (jerárquico), K-medias (particional) y redes de Kohonen serán empleadas para obtener dichas zonas homogéneas mediante la aplicación de la metodología CRISP-DM, debido a las ventajas que ésta ofrece, utilizando el software *Clementine* para evaluar los datos, construir el modelo y establecer los grupos en el proceso de regionalización.

Capítulo III.

Análisis de Datos

En el capítulo II se hizo hincapié que debido al avance de la tecnología es posible disponer de grandes volúmenes de datos y, mediante su análisis usando técnicas como minería de datos, descubrir conocimiento útil o información de interés como la identificación de tendencias y comportamientos del conjunto de datos. Para lograr un proceso efectivo de minería de datos, una de las actividades fundamentales es el análisis de los datos disponibles. En este estudio se empleó el análisis de series de tiempo debido a las características de la fuente de datos.

En este capítulo se describen las fuentes de datos climatológicos disponibles, así como el proceso de análisis empleado con estas bases de datos para determinar la calidad de la serie a emplearse en el proceso de regionalización climatológica en el estado de Michoacán. La principal consideración fue determinar cuántas y cuáles son las estaciones climáticas apropiadas para el estudio. Asimismo, se analizó la variabilidad y distribución de las principales variables climatológicas registradas en las estaciones disponibles.

3.1 Análisis de las series de tiempo

En la última década se ha observado un incremento en la aplicación de minería de datos a problemas relacionados con series de tiempo (Keogh *et al.*, 2003). Dichos trabajos han sido orientados principalmente a la agrupación (*ej.* análisis de regiones homogéneas), clasificación (*ej.* predicción del consumo eléctrico), detección de anomalías (*ej.* análisis de bitácoras para la identificación de intrusos en la red), síntesis y descubrimiento de reglas (*ej.* identificación de patologías).

Las series de tiempo son un caso particular de patrones secuenciales, su análisis ofrece una valoración de la estacionalidad de la serie, describiendo las oscilaciones de los datos en relación a un valor promedio e identificando la presencia de posibles tendencias (Kessler, 2003).

Puerto y Paz (2001) definen una serie de tiempo como “*todo conjunto de valores ordenados cronológicamente que permiten predecir y describir el comportamiento de una o más variables en un determinado periodo*”. Algunas veces estas series pueden ser muy extensas llegando a contener billones de observaciones (Chiu *et al.*, 2003). En las series temporales se identifican cuatro tipos de patrones: tendencia y variaciones estacional, accidental y cíclica.

- Tendencia (T). Refleja la evolución de la serie durante un determinado periodo. Este periodo varía según la naturaleza de la serie, el cual puede ser estacionaria o constante, lineal, exponencial, u otras.
- Variación estacional (S). Es el comportamiento que agrupa las oscilaciones repetitivas en periodos de tiempo. Estos periodos pueden ser estaciones del año, días, meses, años, bimestres, trimestres, semestres, etc.
- Variación accidental (A): Este patrón corresponde a las fluctuaciones accidentales que se dan por la ocurrencia de fenómenos imprevisibles, como

la presencia de huracanes, que afectan a la variable en estudio de manera esporádica y no permanente. También es conocido como variación irregular.

- Variación cíclica (C). Se presenta cuando los datos reflejan oscilaciones periódicas no regulares, ocasionadas por asumir periodos no establecidos. Generalmente aparecen en series de datos climatológicos, por ejemplo en ciclos de sequía.

3.2 Fuente de la serie de datos climatológicos

Las fuentes de datos a partir de las cuales se llevó a cabo el proceso de regionalización fueron: datos provenientes de las estaciones tradicionales (Sistema CLICOM) y datos obtenidos de los Observatorios de Zamora y Morelia. Ambas series de datos se adquirieron a través del Servicio Meteorológico Nacional (SMN). La principal diferencia entre las series de datos empleadas son el tipo de variables registradas en las estaciones climatológicas, así como el tipo y precisión de los sensores usados (Elías y Castellví, 2001).

3.2.1 Estaciones climatológicas tradicionales (CLICOM)

Los registros de las estaciones climatológicas tradicionales se almacenan en un sistema de manejo de bases de datos conocido como CLICOM (*Climatología Computarizada* o *Climat Computing*), el cual permite el acceso rápido y eficiente a la información de las estaciones del Servicio Meteorológico Nacional. El Sistema CLICOM trabaja con aplicaciones ejecutables en lenguaje Fortran¹ versión 5.0, teniendo archivos controladores de tipo Batch². En la actualidad, la versión del CLICOM es la 3.1 y el manejador de base de datos es el DataEasy versión 4.0 (Garza *et al.*, 2003). El periodo de tiempo base es el día para las variables registradas. La

¹ Fortran, lenguaje de programación utilizado principalmente en aplicaciones científicas y análisis numérico.

² Batch, se trata de archivos de texto guardados con extensión *.bat que contienen un conjunto de comandos DOS.

Tabla 3.1 describe brevemente las variables climatológicas almacenadas en la base diaria del Sistema CLICOM (SMN, 2007a; 2007b).

Tabla 3.1. Variables climatológicas registradas en el Sistema CLICOM.

Nombre e Identificador	Instrumento de Medición y U. de Medida	Descripción de la variable
Precipitación (P)	Pluviómetro / Pluviógrafo (mm)	Es el agua en forma líquida o sólida procedente de la atmósfera, que se deposita sobre la superficie de la Tierra, incluye: lluvia, llovizna, rocío, granizo, escarcha, nieve, etc.
Temperatura Observada (Tob)	Termómetro (° C)	Es la medida diaria de calor en un instante de tiempo. También se conoce como el promedio de las temperaturas máxima y mínima registradas en el día.
Temperatura Máxima (Tmx)	Termómetro (° C)	Es la medición de la temperatura más alta alcanzada en el transcurso del día. Siendo medido, por lo general, entre las 14:00 y 16:00 horas.
Temperatura Mínima (Tmn)	Termómetro (° C)	Es la medida de la temperatura más baja alcanzada en el día. Siendo medido, por lo general, entre las 06:00 y 08:00 horas.
Evaporación (Evp)	Evaporímetro (mm)	Es el proceso por el cual se transfiere agua, en la forma de vapor, desde el suelo hacia la atmósfera. La evaporación de la vegetación se conoce como transpiración. La pérdida total combinada de agua de ambos se denomina evapotranspiración.
Cobertura de nubosidad (Cobnub)	Observaciones (octas ³)	La cobertura de las nubes se determina por su distribución en el espacio. Es decir se hace una estimación de la cantidad de nubes suspendidas en la atmósfera. Se mide mediante octas, que corresponde a la octava parte del cielo.
Velocidad del viento (VelVi)	Anemómetro (m/s), (km/h)	Es la intensidad del viento, cuya medida es una cantidad vectorial. La intensidad está sujeta a variaciones, tanto en tiempo como en espacio. Esto se debe a que el flujo de aire no es constante.
Dirección del viento (DirVi)	Veleta (grados)	Es la trayectoria del viento que se expresa en grados, medidos en el sentido de las agujas del reloj. Esta medición indica el ángulo desde donde sopla el viento (0° = norte, 90° = este, 180° = sur, 270° = oeste).

Es importante mencionar, que hoy en día muchos países están dejando de usar el sistema CLICOM, optando por el *Sistema de Manejo de Bases de Datos Climáticos* (CDMS), el cual mejora la accesibilidad y seguridad de los datos (CIIFEN-OMM,

³ Se entiende **octas** como una región y no como una distancia.

2003). La característica principal de los CDMS es la riqueza expresiva de los medios ofrecidos, como la interfaz gráfica y las líneas de comando, para la recuperación y el análisis de los datos.

En el estado de Michoacán se cuenta con un total de 239 estaciones, las cuales se muestran en la *Figura 3.1* y se listan en el *Anexo A* indicando su nombre oficial, ubicación geográfica, año de inicio y año final de los registros disponibles por estación.

3.2.2 Observatorios

El Servicio Meteorológico Nacional cuenta con una red de 80 observatorios en todo el país. Esta red captura los datos de manera tradicional en formatos donde se registran las horas muestreadas para cada día, y en el caso de la velocidad y dirección del viento cada 15 minutos de la hora muestreada. Asimismo, la red de observatorios captura los datos a partir de 2001 a la fecha, en bases de datos sinópticas lo que ha permitido enviar información a organismos internacionales como la Organización Meteorológica Mundial (OMM⁴). Los observatorios miden elementos (variables) climatológicos y meteorológicos, los cuales se realizan por instrumentos especializados en forma constante y a horas establecidas durante el día; por lo general, cada 3 horas en tiempo meridiano de Greenwich, GMT.

Michoacán cuenta con dos observatorios (*Figura 3.2*), el de Morelia (76665) con coordenadas 19°42' Norte/101°11' Oeste a una elevación de 1913 msnm y el de Zamora (76662) con coordenadas 19°51' Norte/102°19' Oeste a una elevación de 1562 msnm (SMN, 2006). Los registros de las variables en los observatorios de Zamora y Morelia, capturados de forma tradicional y digitalizada, se tienen a partir del 1° de enero de 1981 y terminan el 31 de julio y 31 de agosto de 2004, respectivamente. Estos datos se complementan con la base sinóptica, cuyos registros se tienen de 2001 a la fecha.

⁴ OMM - Organización Meteorológica Mundial establecida en 1950 (OMM, 2007).

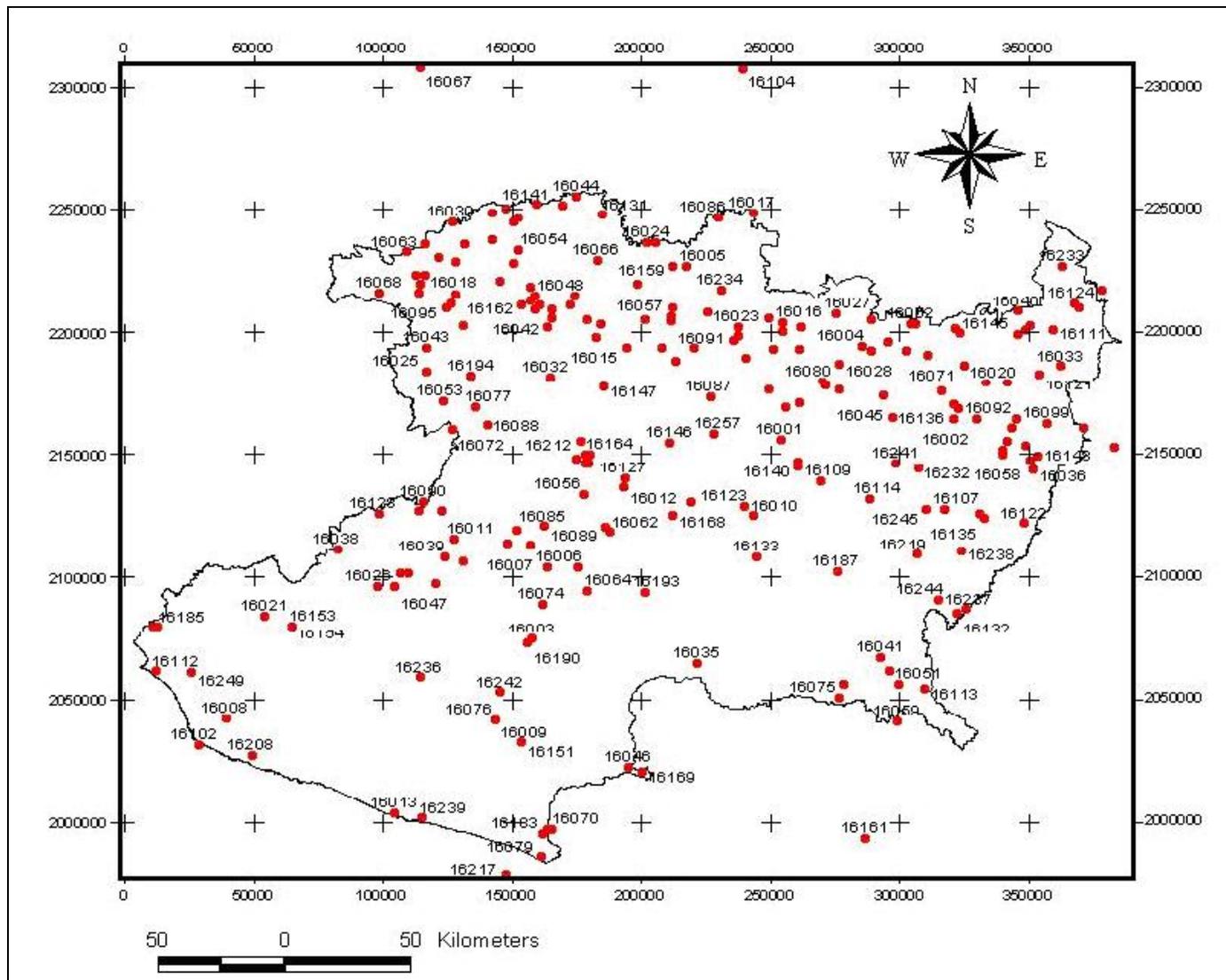


Figura 3.1. Estaciones climatológicas de la base de datos CLICOM para el estado de Michoacán.

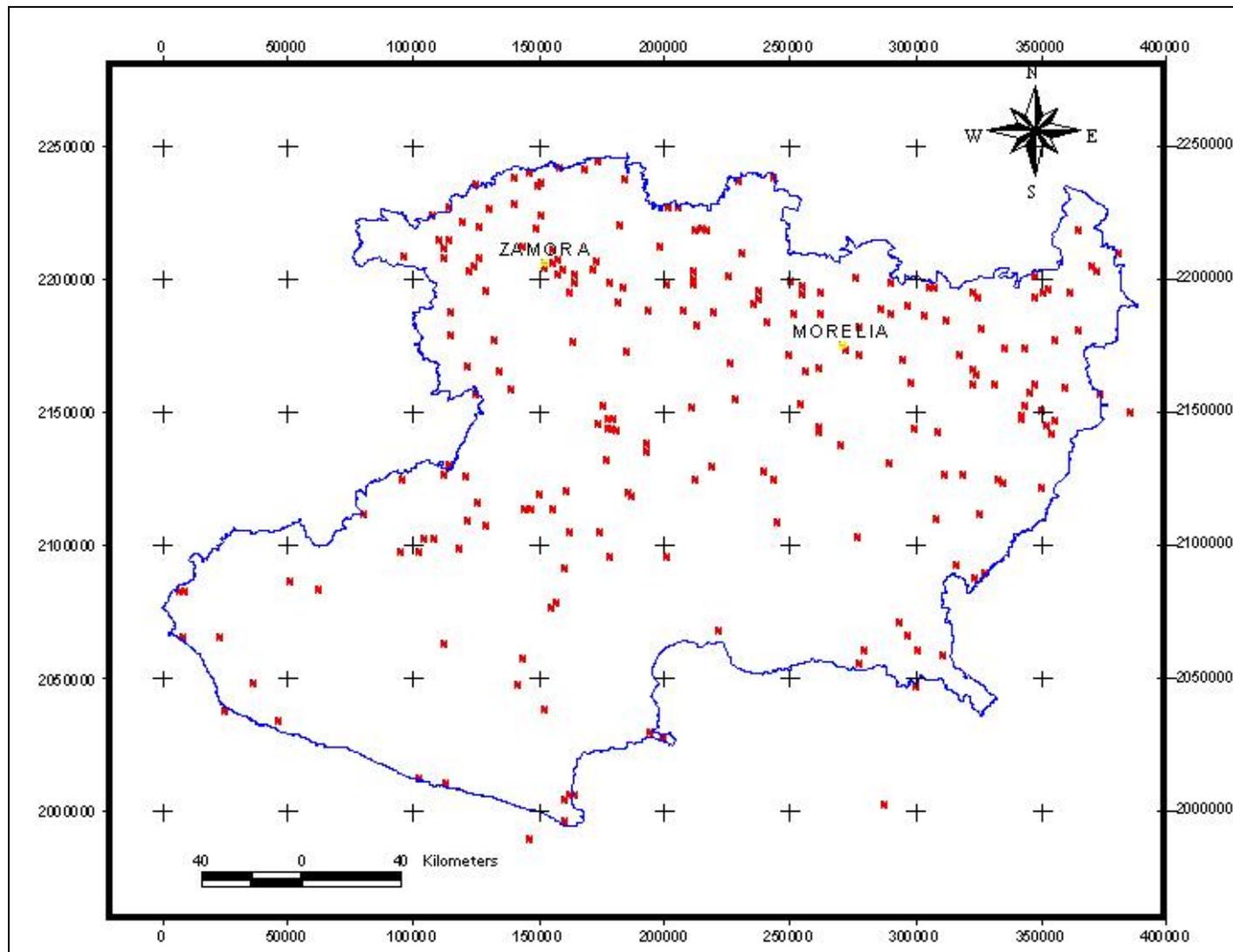


Figura 3.2. Ubicación de los Observatorios de Morelia y Zamora en el estado de Michoacán.

Los observatorios de Zamora y Morelia cuentan con registros horarios variados algunos de siete horas (06:00 a 12:00 horas), otros de 14 horas (06:00 a 19:00 horas), 0 series completas de 24 horas. Una de las causas asociadas a la escasez de datos es la falta de personal para cubrir el día completo (SMN, 2007c). La *Tabla 3.2* describe brevemente las variables medidas en los observatorios (SMN, 2007b).

Tabla 3.2. Variables climatológicas registradas en la base de datos de los observatorios.

Nombre	Instrumento de medición y Unidad de Medida	Descripción de la variable
Temperatura de bulbo seco	Termómetro de bulbo seco (°C)	Es la temperatura observada en un termómetro de bulbo seco expuesto al aire y protegido de la radiación solar directa.
Temperatura de bulbo húmedo	Termómetro de bulbo húmedo (°C)	Es la temperatura observada que registra el termómetro, cuyo bulbo es humedecido con una gasa o algodón, expuesto al aire y protegido de la radiación solar directa.
Humedad relativa	Higrómetro (%)	Es la cantidad de vapor de agua contenido en la atmósfera, representado en porcentajes, donde 0% significa aire seco y 100% aire completamente húmedo.
Humedad relativa media	Higrómetro (%)	Es el promedio diario de la cantidad de vapor de agua contenido en el aire, el cual es obtenido a través de las observaciones horarias.
Presión en la estación	Barómetro (mbar)	Es el registro de la presión o el peso que ejerce la atmósfera en una determinada estación en el transcurso del día.
Presión máxima en la estación	Barómetro (mbar)	Es el registro de la presión máxima que ejerce la atmósfera en una determinada estación en el transcurso del día.
Presión mínima en la estación	Barómetro (mbar)	Es el registro de la presión mínima que ejerce la atmósfera en una determinada estación en el transcurso del día.
Velocidad del viento máximo	Anemómetro (m/s)	Es el registro de la máxima intensidad o fuerza del viento en el transcurso del día.
Dirección del viento máximo	Veleta o brújula (grados)	Es el registro de la máxima trayectoria o dirección del viento en el transcurso del día.
Velocidad media del viento	Anemómetro (m/s)	Es el promedio de la intensidad o fuerza del viento, que se obtiene a través de las observaciones horarias.
Cantidad de nubosidad	Octas	Es el criterio que se establece para determinar la cantidad de nubes suspendidas en la atmósfera expresada en octavas, u octas, de cielo cubierto.

Tabla 3.2. Variables climatológicas registradas en la base de datos de los observatorios.

Precipitación en 24 horas pluviógrafo	Pluviógrafo (mm)	Es la cantidad de agua caída en un periodo de 24 horas, registrándose el inicio, duración e intensidad de la lluvia.
Precipitación horaria total	Pluviómetro (mm)	Es el registro de la cantidad total de agua procedente de la atmósfera, acumulada en una hora.
Insolación -horas y minutos	Heliógrafo (cal/cm ² día o cal/cm ² hora)	Es el registro de la cantidad de energía solar recibida en la superficie horizontal de la estación en horas y/o minutos concretos.
Lluvia total	Pluviómetro (mm)	Es el registro de la precipitación total acuosa en forma de gotas líquidas, caída a la superficie.
Evaporación total	Evaporímetro (mm)	Es el registro del cambio de estado del agua desde la superficie hacia la atmósfera, pasando de ser líquida a gaseosa por absorción de calor.
Ocurrencia de granizo	Probabilidad (%)	Es la probabilidad de ocurrencia de la precipitación de partículas de hielo, en general, de forma esferoidal, cónica o irregular.
Ocurrencia de niebla	Probabilidad (%)	Es la probabilidad de ocurrencia de la existencia de numerosas gotitas de agua (suficientemente pequeñas) que se mantienen suspendidas en el aire.
Ocurrencia de helada	Probabilidad (%)	Es la probabilidad de tener una helada cuando la temperatura desciende por debajo de los 0°C. Es decir, si el cielo es despejado y la temperatura es menor a 3°C, existe la probabilidad de helada.
Ocurrencia de tormenta eléctrica	Probabilidad (%)	Es la probabilidad de ocurrencia de la perturbación violenta de la atmósfera acompañada de fenómenos mecánicos (viento y lluvias) y eléctricos (relámpagos y truenos).
Temperatura máxima extrema	Termómetro (°C)	Es el registro de la temperatura más alta observada durante el día. El cual se presenta entre las 14:00 y las 16:00 horas.
Temperatura mínima extrema	Termómetro (°C)	Es el registro de la temperatura más baja observada durante el día. El cual se presenta entre las 06:00 y las 08:00 horas.
Temperatura mínima a la intemperie	Termómetro (°C)	Es el registro de la temperatura mínima observada en un termómetro expuesto al aire con ventilación adecuada y sin los efectos de la radiación solar directa.

Es importante que estas observaciones se realicen empleando el menor tiempo posible y cubriendo los periodos de 24 horas, dado que la falta de datos se traduce en series incompletas con registros irregulares (SNM, 2007c).

3.3 Análisis de las bases de datos

El análisis de datos es una de las actividades fundamentales en el proceso de minería de datos, mediante el cual se establece un contacto directo con el problema a resolver. El análisis de las bases de datos disponibles se realizó en dos etapas: la *primera* consistió en una revisión preliminar del total de estaciones listadas en el CLICOM para establecer aquellas relevantes en función del periodo de registros en la zona de estudio. Las bases de datos de los observatorios se tienen a partir de 1981 por lo que no se hará el análisis por periodo de registro. En la *segunda* etapa se determinó la calidad de la serie de datos para establecer las estaciones significativas para el proceso de regionalización.

3.3.1 CLICOM

Para realizar el análisis preliminar del total de estaciones listadas en la base de datos del CLICOM se consideraron aquellas estaciones con suficientes registros, esto es, que tuvieran periodos con datos con al menos 20 años a partir de 1981. Esto con la finalidad de tener periodos similares a los reportados por los observatorios. Así, se analizó la frecuencia de registros (fecha de inicio y fecha final de operación de la estación) para el total de 239 estaciones, de las cuales se obtuvieron 111 estaciones que cumplían el criterio establecido (ver Anexo B).

3.3.2 Tratamiento de las bases de datos

La segunda etapa de análisis consistió en el estudio de la calidad de la serie de datos con la finalidad de determinar las estaciones significativas para la regionalización climática de la zona de estudio. En este proceso, se organizaron inicialmente las series de datos en archivos de texto plano (bases de datos textuales independientes) con el propósito de ajustar la fecha de inicio y fin de sus registros. En general, el periodo de registros considerado fue de 20 años. Sin embargo, de las 111 estaciones del CLICOM algunas no contaban con registros en 1981 e incluso en años siguientes, por lo que se tomó como base el año más cercano o próximo con

datos. Este fue el caso para 30 estaciones: St16007, St16023, St16057, St16061, St16080, St16093, St16159, St16192, St16220, St16227, St16228, St16229, St16230, St16231, St16233, St16234, St16235, St16240, St16243, St16245, St16247, St16248, St16250, St16251, St16253, St16254, St16255, St16512, St16513, St16514. En esta parte del proceso se pudo observar que no se disponen de datos de velocidad ni dirección del viento; asimismo para la evaporación y cobertura de nubosidad los datos son escasos, encontrándose un gran porcentaje de valores nulos.

3.3.3 *Construcción y ejecución del modelo de análisis*

Una vez determinado el número de estaciones con base en un periodo teórico de 20 años, se elaboró un procedimiento para evaluar la calidad de los datos en el software *Clementine* (ver Sección 2.5). Este procedimiento consiste en hacer una exploración del conjunto de datos para determinar mediante parámetros de medición (ej. índices de búsqueda) la cantidad de registros válidos, esto es, sin datos nulos y faltantes, por variable y año. La *Figura 3.3* muestra el modelo de análisis en *Clementine*, donde cada proceso es representado por un nodo, el cual es configurado en el espacio de trabajo principal, denominado área o zona de rutas (Sección 2.5).

- b. Nodo *Tabla (2)*, permite crear una matriz a partir de los datos, siendo útil para examinar los valores de los registros y evaluar la cronología de la serie de datos. En este punto se establecieron dos variables de comparación denominadas *fecha* y *hora*. La variable *fecha* se utilizó para la evaluación cronológica de las estaciones del CLICOM, mientras que para los observatorios con registros horarios se emplearon ambas variables. La comparación consistió en verificar la secuencia de datos de las variables *Año, Mes, Día y Hora* según sea el caso, tomando como referencia el calendario del sistema.
- c. Nodo *Tipo (3)*, una vez corroborada la secuencia cronológica, se utiliza este nodo para especificar el tipo de datos en cada una de las variables. Esto permite identificar a los índices de exploración, el cual es un catálogo de búsquedas de registros válidos (sin nulos y/o faltantes). Estos índices son *Nombre, Año, Mes y Día*, que juntos identifican de manera única a los registros de la serie, evitando la redundancia de datos.
- d. Nodo *Seleccionar (4)*. Una vez establecidos los tipos de variables, se seleccionaron subconjuntos de datos de la serie temporal para determinar la cantidad de sus registros válidos por año y variable. Esto es, se filtraron los registros por año, iniciándose por lo general a partir de 1980 en adelante.
- e. Nodo *Calidad (5)*. En este nodo se configuran las condiciones para determinar cuales valores son considerados como no válidos, (ej. nulos - null, cadenas vacías y espacios en blanco). Esta configuración permite que todas las columnas sean analizadas registro por registro, con lo que una vez terminado el proceso de ejecución se emite un reporte con el nombre de la variable, la cantidad porcentual y numérica de registros válidos y nulos, la cantidad de cadenas vacías y el número de espacios en blanco.

- f.* Nodo *Estadístico* (6) es usado para obtener información estadística del subconjunto de datos. Esta información se obtiene para cada una de las variables y se presenta como el valor medio, mínimo, máximo, varianza, desviación típica, error típico, mediana, moda y recuento.

- g.* Nodo *Gráfico* (7) permite representar el comportamiento de las series de datos de manera gráfica; por ejemplo, la relación de las variables numéricas y el tiempo en años o meses. Estos gráficos son útiles para analizar la tendencia y las fluctuaciones de las variables durante un determinado período.

Después de haber aplicado el modelo de análisis al conjunto de datos, se elaboraron tablas resúmenes basadas en los reportes estadísticos generados por dicho modelo (Tablas 3.3 y 3.4). Estas tablas contienen el número identificador de la estación, los años analizados, la cantidad de días por año y la cantidad porcentual y numérica de registros válidos por variable y año.

Tabla 3.3. Matriz con resultados porcentuales y numéricos de registros válidos por año y variable. Estación climatológica St16052.

Nº	Estación	Años	Nº días	Año	Cant.	Mes	Cant.	Día	Cant.	P	Cant.	Tob	Cant.	Tmx	Cant.	Tmn	Cant.	Evp	Cant.	Cobnub	Cant.
1	St16052	1980	366	91.8%	336	91.8%	336	91.8%	336	91.8%	336	91.8%	336	91.8%	336	91.8%	336	91.8%	336	91.8%	336
2	St16052	1981	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365
3	St16052	1982	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365
4	St16052	1983	365	91.5%	334	91.5%	334	91.5%	334	91.5%	334	91.5%	334	91.5%	334	91.5%	334	57.0%	208	57.0%	208
5	St16052	1984	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	0.0%	0
6	St16052	1985	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365
7	St16052	1986	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	99.2%	362
8	St16052	1987	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	0.0%	0	0.0%	0
9	St16052	1988	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	8.5%	31	100.0%	366
10	St16052	1989	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	80.5%	294	80.5%	294
11	St16052	1990	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	0.0%	0	0.0%	0
12	St16052	1991	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	0.0%	0	0.0%	0
13	St16052	1992	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	0.0%	0	0.0%	0
14	St16052	1993	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	0.0%	0	0.0%	0
15	St16052	1994	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	0.0%	0	0.0%	0
16	St16052	1995	365	91.5%	334	91.5%	334	91.5%	334	91.5%	334	91.5%	334	91.5%	334	91.5%	334	0.0%	0	0.0%	0
17	St16052	1996	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	0.0%	0	0.0%	0
18	St16052	1997	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	0.0%	0	0.0%	0
19	St16052	1998	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	0.0%	0	0.0%	0
20	St16052	1999	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	0.0%	0	0.0%	0

- P = Precipitación
- Tob = Temperatura observada
- Tmx = Temperatura máxima
- Tmn = Temperatura mínima
- Evp = Evaporación
- Cobnub = Cobertura Nubosa

Tabla 3.4. Matriz con resultados porcentuales y numéricos de registros válidos por año y variable, en la cual se muestra una considerable escasez de datos. Estación climatológica St16040.

Nº	Estación	Años	Nº días	Año	Cant.	Mes	Cant.	Día	Cant.	P	Cant.	Tob	Cant.	Tmx	Cant.	Tmn	Cant.	Evp	Cant.	Cobnub	Cant.
1	St16040	1980	366	16.4%	60	16.4%	60	16.4%	60	16.4%	60	16.4%	60	16.4%	60	16.4%	60	16.4%	60	16.4%	60
2	St16040	1981	365	58.9%	215	58.9%	215	58.9%	215	58.9%	215	58.9%	215	58.9%	215	58.9%	215	58.9%	215	58.9%	215
3	St16040	1982	365	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0
4	St16040	1983	365	49.6%	181	49.6%	181	49.6%	181	49.6%	181	49.6%	181	49.6%	181	49.6%	181	49.6%	181	49.6%	181
5	St16040	1984	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366	100.0%	366
6	St16040	1985	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	100.0%	365	91.5%	334
7	St16040	1986	365	58.1%	212	58.1%	212	58.1%	212	58.1%	212	58.1%	212	58.1%	212	58.1%	212	58.1%	212	58.1%	212
8	St16040	1987	365	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0
9	St16040	1988	366	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0
10	St16040	1989	365	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0
11	St16040	1990	365	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0
12	St16040	1991	365	16.7%	61	16.7%	61	16.7%	61	16.7%	61	16.7%	61	16.7%	61	16.7%	61	16.7%	61	16.7%	61
13	St16040	1992	366	16.4%	60	16.4%	60	16.4%	60	16.4%	60	16.4%	60	16.4%	60	16.4%	60	16.4%	60	16.4%	60
14	St16040	1993	365	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0
15	St16040	1994	365	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0
16	St16040	1995	365	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0
17	St16040	1996	366	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0
18	St16040	1997	365	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0
19	St16040	1998	365	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0
20	St16040	1999	365	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0

- P = Precipitación
- Tob = Temperatura observada
- Tmx = Temperatura máxima
- Tmn = Temperatura mínima
- Evp = Evaporación
- Cobnub = Cobertura Nubosa

La *Tabla 3.3* muestra los resultados obtenidos del análisis de datos de las variables disponibles para la estación St16052 (Huingo, Zinapecuaro). El total de años evaluados fueron 20 (1980-1999, algunas estaciones contaban con datos de 1980 por lo que se consideraron); se observa que las variables *P*, *Tob*, *Tmx* y *Tmn* disponen de una mayor cantidad de datos (con bajos o ningunos registros faltantes y/o nulos), con porcentajes de 100 % anual, a excepción de los años 1980, 1982 y 1996 con 91.8 %, 91.5 % y 91.5 %, respectivamente. De acuerdo a los registros de *Evp* y *Cobnub*, la cantidad de datos es escasa llegando a tener periodos sin datos hasta en 10 años consecutivos (1990 a 1999), confirmándose la escasa disponibilidad de datos en ambas variables.

La *Tabla 3.4* muestra la escasa disponibilidad de datos que presentan las variables disponibles para la estación St16040 (El Gigante, Maravatío), donde el porcentaje más alto de datos válidos para *P*, *Tob*, *Tmx* y *Tmn* fue de 100 % en 1984 y 1985; sin embargo para el resto años el porcentaje apenas alcanza los 58.9 % (1981), observándose periodos consecutivos sin datos en 1987-1990 y 1993-1999. En el caso de *Evp* y *Cobnub*, la cantidad de datos es escasa, llegando a tener periodos sin datos de hasta en 13 años (1987-1999).

3.3.4 *Resultados del análisis de las series temporales*

En esta sección se presentan los resultados obtenidos al aplicar el modelo desarrollado a las 111 estaciones del CLICOM y a los observatorios de Zamora y Morelia. Este análisis de resultados permitió identificar las estaciones que disponen de datos continuos en un periodo mínimo de 20 años, esto es, series completas con una baja o nula cantidad de registros faltantes.

Para el análisis de los datos y seleccionar las estaciones a utilizar en la regionalización, tanto para las estaciones tradicionales (CLICOM) como para los observatorios, se establecieron las mismas consideraciones.

3.3.4.1 CLICOM

Las consideraciones para aplicar el modelo a la serie de datos del CLICOM son:

- a. Tener por lo menos el 90 % de registros válidos por variable y año de las variables climatológicas consideradas para el proceso de regionalización: precipitación (P), temperatura observada (Tob), temperatura máxima (Tmx) y temperatura mínima (Tmn). Esto es, para estas variables se acepta la existencia de hasta un 10 % de registros nulos y/o faltantes por año, que aproximadamente equivale a un mes sin datos (36 días).
- b. Establecer un porcentaje de aceptación entre el número de años analizados y el número de años considerados favorables (determinado en el paso anterior). Así, para que una estación sea considerada adecuada, este porcentaje debe ser mayor o igual a 85%, en caso contrario se descarta la estación por tener una considerable cantidad de registros faltantes. Por ejemplo, para la estación St16052 (Tabla 3.3) el número de años analizados es 20 y la cantidad de años considerados favorables es también 20, el cual representa un 100 % de aceptación y, por consiguiente, se considera ésta como una estación adecuada para la regionalización. El caso contrario ocurre con la estación St16040 que dispone de sólo 2 años favorables de un total de 20 años analizados, representado un 10 % de aceptación, siendo así descartada.

Al aplicar las consideraciones anteriores, en el caso del *CLICOM* se observó que 52 de las 111 estaciones cumplen con los criterios establecidos (Tabla 3.5).

Tabla 3.5. Estaciones climatológicas consideradas significativas para el proceso de clasificación de zonas con climas similares.

Nº	Estación	Fecha de Inicio	Fecha Final	Nº de Años	Años Favorables	% Aceptación
1	St16006	01/01/1980	31/12/1999	20	19	95.0
2	St16007	01/01/1985	31/12/2003	19	18	94.7
3	St16014	01/01/1980	31/12/1999	20	20	100.0
4	St16016	01/01/1980	31/12/1999	20	20	100.0
5	St16017	01/01/1980	31/12/1999	20	19	95.0
6	St16020	01/01/1980	31/12/1999	20	19	95.0
7	St16023	01/01/1982	31/12/2000	19	17	89.5
8	St16024	01/01/1980	31/12/1999	20	18	90.0
9	St16027	01/01/1980	31/12/1999	20	19	95.0
10	St16028	01/01/1980	31/12/1999	20	18	90.0
11	St16031	01/01/1980	31/12/1999	20	20	100.0
12	St16033	01/01/1980	31/12/1999	20	18	90.0
13	St16043	01/01/1980	31/12/1999	20	18	90.0
14	St16045	01/01/1980	31/12/1999	20	20	100.0
15	St16048	01/01/1980	31/12/1999	20	18	90.0
16	St16050	01/01/1980	31/12/1999	20	18	90.0
17	St16052	01/01/1980	31/12/1999	20	20	100.0
18	St16057	01/01/1982	31/12/1999	18	17	94.4
19	St16059	01/01/1980	31/12/1999	20	18	90.0
20	St16074	01/01/1980	31/12/1999	20	19	95.0
21	St16075	01/01/1980	31/12/1999	20	18	90.0
22	St16081	01/01/1980	31/12/1999	20	19	95.0
23	St16090	01/01/1980	31/12/1999	20	19	95.0
24	St16091	01/01/1980	31/12/1999	20	18	90.0
25	St16093	01/01/1983	31/12/2002	20	19	95.0
26	St16094	01/01/1980	31/12/1999	20	19	95.0
27	St16095	01/01/1980	31/12/1999	20	20	100.0
28	St16107	01/01/1980	31/12/1999	20	19	95.0
29	St16109	01/01/1980	31/12/1999	20	17	85.0
30	St16118	01/01/1980	31/12/1999	20	18	90.0
31	St16120	01/01/1980	31/12/1999	20	19	95.0
32	St16123	01/01/1980	31/12/1999	20	18	90.0
33	St16124	01/01/1980	31/12/1999	20	20	100.0
34	St16133	01/01/1980	31/12/1999	20	18	90.0
35	St16136	01/01/1980	31/12/1999	20	20	100.0
36	St16141	01/01/1980	31/12/1999	20	19	95.0
37	St16142	01/01/1980	31/12/1999	20	19	95.0
38	St16151	01/01/1980	31/12/1999	20	18	90.0
39	St16159	01/01/1982	31/12/2001	20	18	90.0
40	St16162	01/01/1980	31/12/1999	20	19	95.0
41	St16165	01/01/1980	31/12/1999	20	18	90.0
42	St16188	01/01/1980	31/12/1999	20	19	95.0
43	St16192	01/01/1987	31/12/2004	18	17	94.4
44	St16194	01/01/1980	31/12/1999	20	18	90.0
45	St16208	01/01/1980	31/12/1999	20	18	90.0
46	St16219	01/01/1980	31/12/1999	20	19	95.0
47	St16225	01/01/1980	31/12/1999	20	18	90.0
48	St16228	01/01/1981	31/12/2000	20	20	100.0
49	St16248	01/01/1982	31/12/2001	20	18	90.0
50	St16254	01/01/1982	31/12/2001	20	19	95.0
51	St16255	01/01/1981	31/12/2000	20	19	95.0
52	St16515	01/01/1980	31/12/1999	20	18	90.0

El número de estaciones obtenidas representan el 46.8 % del total de 111, lo que muestra la falta de datos en más de la mitad de estaciones evaluadas. La estación con el menor porcentaje de aceptación (85 %) fue St16109, mientras que el resto de estaciones comparten porcentajes superior a 85, alcanzando en algunos casos una aceptación de hasta 100 % como es el caso de las estaciones St16014, St16016, St16031, St16045, St16052, St16095, St16124, St16136 y St16228. Aún cuando las estaciones St16007, St16023, St16057 y St16192 presentan las series de tiempo con menos años evaluados (19, 19, 18 y 18, respectivamente), éstas se consideraron por presentar series consecutivas y por su ubicación dentro del estado. La *Figura 3.4* muestra la distribución de las 52 estaciones en el estado de Michoacán.

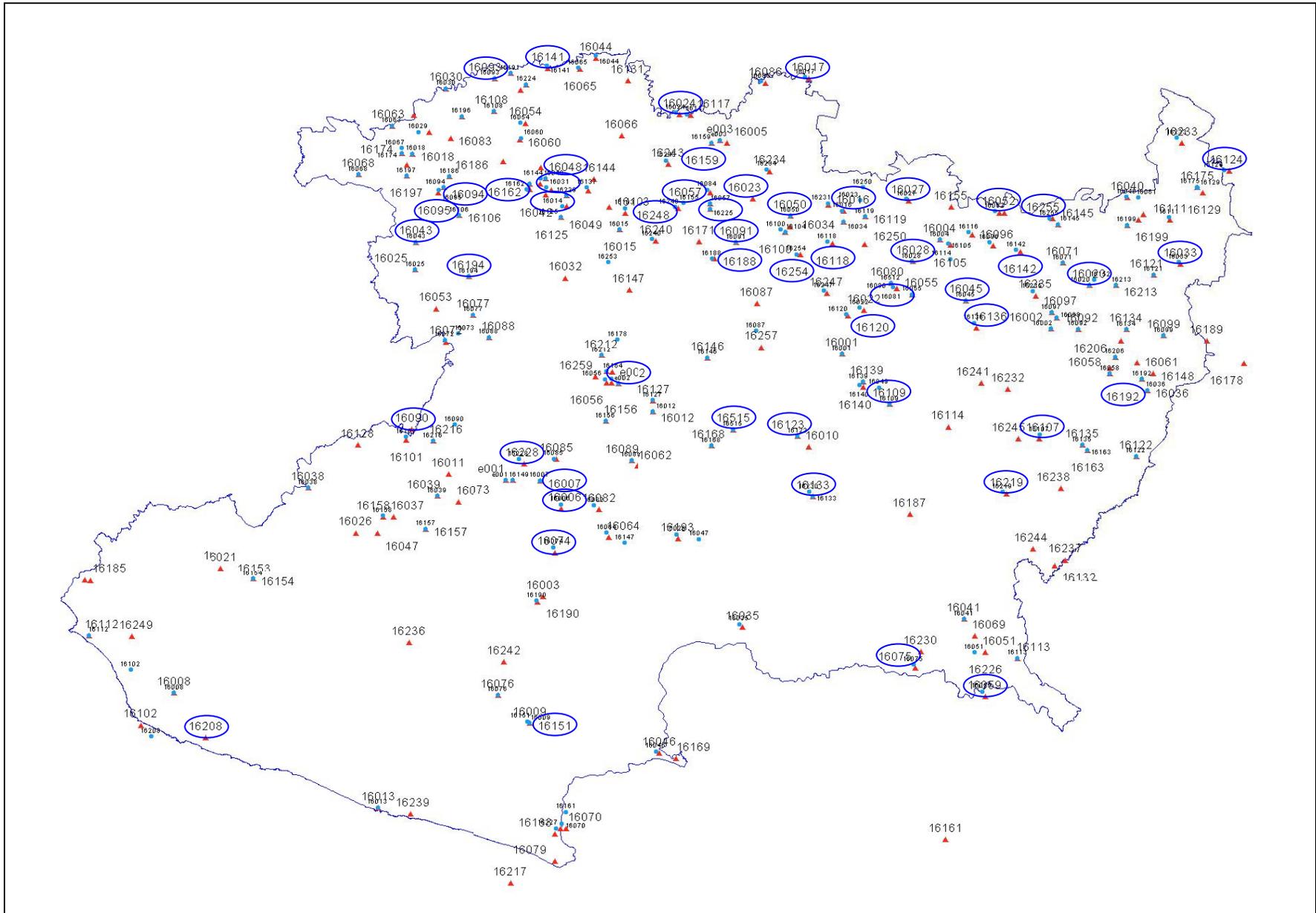


Figura 3.4. Identificación de las 52 estaciones climatológicas (círculo) en el estado de Michoacán.

En la *Figura 3.4*, se observa que la mayor cantidad de las 52 estaciones se encuentran situadas en la zona norte del estado, mientras que un grupo menor se ubica en el centro y al sur del mismo. Asimismo se distingue que del total de estaciones (239) presentadas en la figura, tres de las ellas se encuentran fuera del territorio michoacano (St16161, St16178 y St16217); esto se puede asociar a que dichas estaciones están fuera de operación desde 1985, 1991 y 1985, respectivamente, por lo que sus datos de ubicación geográfica se encuentran desactualizados. Otro aspecto que es de interés se relaciona a que gran parte de las estaciones instaladas se ubican al norte y centro del Estado donde el relieve michoacano se destaca por la zona montañosa del Eje Neovolcánico con picos como Tancítaro (3857 msnm), Patambán (2750 msnm), Cerro de Quinceo (3300 msnm), Tzirate (3605 msnm) y el Volcán de San Andrés (3605 msnm); así como con valles en Zamora, Tarímbaro, Morelia, Queréndaro y Maravatío.

3.3.4.2 Observatorios

Aplicando las consideraciones mostradas en la sección anterior, los resultados obtenidos para los observatorios de Zamora y Morelia fueron:

- a) Observatorio de Zamora (1981-2004), se observó que las variables *temperatura de bulbo seco*, *precipitación horaria*, *humedad relativa* y *presión en la estación*, por lo general poseen registros válidos con porcentajes entre los 16.2 y 99.9 % anual, aunque los años 1994 y 1995 no cuentan con registros. Una menor disponibilidad de datos se da en la variable *temperatura de bulbo húmedo*, cuyo porcentaje más alto fue de 82.2 % en 1981. Realizando el análisis para identificar las estaciones significativas, se observó que la cantidad de años favorables con datos disponibles de este observatorio es 13 de un total de 24 años evaluados, el cual representa un 54.17 % de aceptación. Este valor no alcanza el mínimo porcentaje establecido (85%), por lo que se descarta el uso de esta serie en el proceso de regionalización.

b) Observatorio de Morelia (1981-2004), se observó que las variables *temperatura de bulbo seco, temperatura de bulbo húmedo, precipitación horaria, humedad relativa y presión en la estación*, por lo general disponen de registros válidos que están por arriba del 92 % anual. El año que registra el menor porcentaje de datos disponibles fue 1989 con 0.27 % de registros. En consecuencia, la cantidad de años favorables con datos disponibles de este observatorio es 21 de un total de 24 años evaluados, el cual representa un 87.5 % de aceptación, por lo que esta serie se acepta para su uso en la regionalización.

3.4 Análisis de variabilidad y distribución de los principales elementos climatológicos

Una vez establecidas las estaciones significativas para la regionalización, se procedió al análisis de variabilidad y distribución de las cuatro variables principales del CLICOM: *P, Tob, Tmx* y *Tmn*, en las 52 estaciones; así como también de las variables disponibles en el observatorio de Morelia.

3.4.1 Estaciones climatológicas del sistema CLICOM

Para la precipitación (*P*) se estimó la suma mensual y la media anual, ambas expresadas en milímetros (mm). Para la temperatura observada (*Tob*) se calcularon las medias mensuales, mientras que para la temperatura -máxima (*Tmx*) y mínima (*Tmn*)- se estimaron las medias mensuales de los máximos y mínimos, respectivamente, expresados en grados centígrados °C. En el Anexo D se muestra la tabla general con el resumen de los resultados obtenidos por estación y variable. A continuación se hará un resumen de los resultados obtenidos.

3.4.1.1 Precipitación (P)

La *Figura 3.5* muestra la distribución y variabilidad mensual de la precipitación en las 52 estaciones. Como se aprecia, la temporada de lluvias se presenta de mayo a octubre, mientras que la temporada de sequía es de noviembre a abril. Los meses más lluviosos son julio y agosto con valores promedio de agua precipitada que oscilan entre los 133 mm (St16074) y 320 mm (St16107).

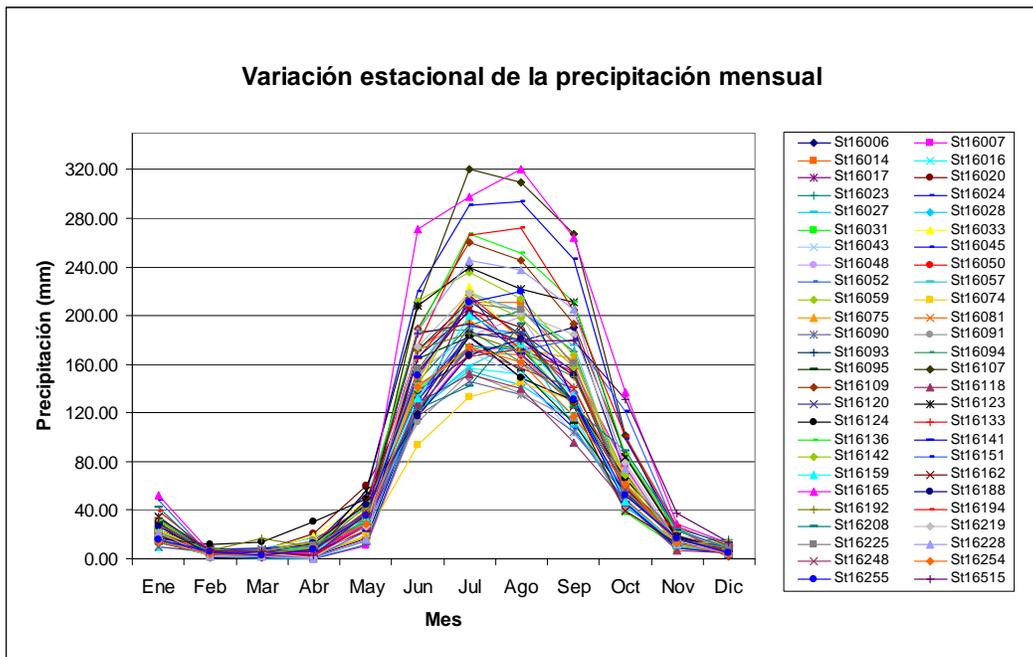


Figura 3.5. Variación mensual de la precipitación en las 52 estaciones climatológicas.

De acuerdo con la *Figura 3.5*, se observa que las precipitaciones mensuales de las 52 estaciones se pueden dividir en 3 grupos: en el *primero*, las estaciones con una mayor incidencia de lluvias: St16045, St16059, St16107, St16109, St16123, St16136, St16165, St16194, y St16228, cuyas precipitaciones en los meses más húmedos (julio y agosto) son mayores de 230 mm; en el segundo grupo se encuentran aquellas estaciones con precipitaciones entre 230 y 150 mm, para finalmente en el tercer grupo agruparse aquellas estaciones con menos de 150 mm de agua precipitada en los meses más lluviosos (St16016, St16027, St16028, St16074, St16090, St16091, St16118 y St16208).

Ubicando geográficamente las estaciones en la *Figura 3.6*, se observa que para el primer grupo (*Figura 3.6a*) las estaciones se encuentran dispersas en la zona noroeste, centro y noreste del estado, con altitudes que van de 305 msnm (St16059) a 2,240 msnm (St16045). En el segundo grupo las estaciones no siguen un patrón teniendo variaciones en la altitud de 305 msnm (St16059) a 2,500 msnm (St16188). Por su parte en el tercer grupo existen cinco estaciones que se ubican en la zona norte del estado, con altitudes que van de 1,384 msnm (St16091) a 2,056 msnm (St16118); mientras que otras se localizan al suroeste del estado, con altitudes entre los 45 msnm (St16208) y 304 msnm (St16090) (*Figura 3.6b*).

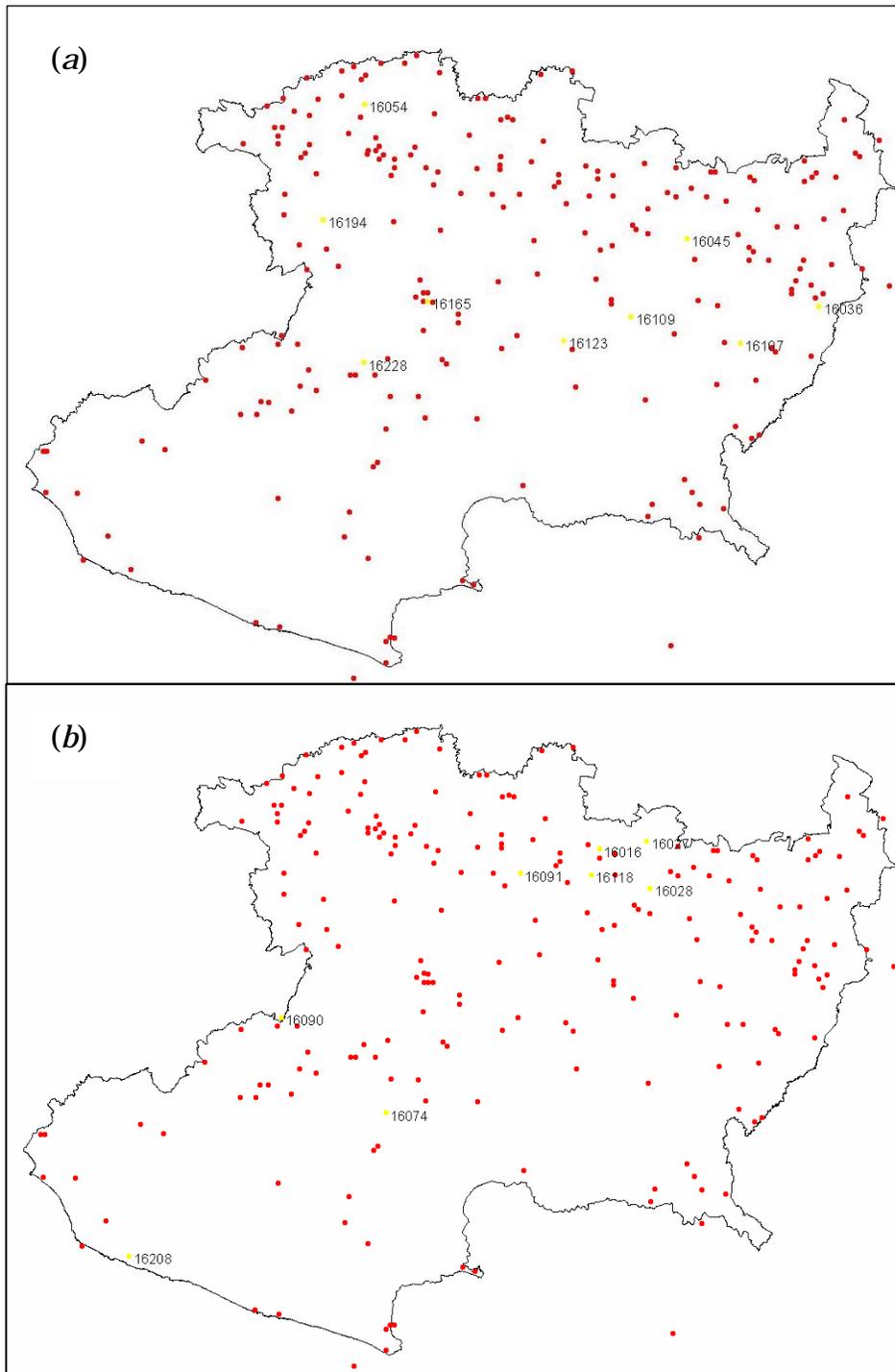


Figura 3.6. Localización de las estaciones con mayor (a) y menor (b) precipitación.

Para analizar la variabilidad de las precipitaciones anuales (Figura 3.7), se establecieron intervalos de confianza calculados a partir de la suma y diferencia de la media anual de las precipitaciones ($P=858.2$ mm) y la desviación estándar ($DE=168.4$ mm). El límite *superior* ($P+DE$, línea azul) identifica aquellas estaciones con una precipitación anual superior a los 1,026 mm, y el límite *inferior* ($P-DE$, línea rosa) a aquellas estaciones con una precipitación anual menor a los 690 mm.

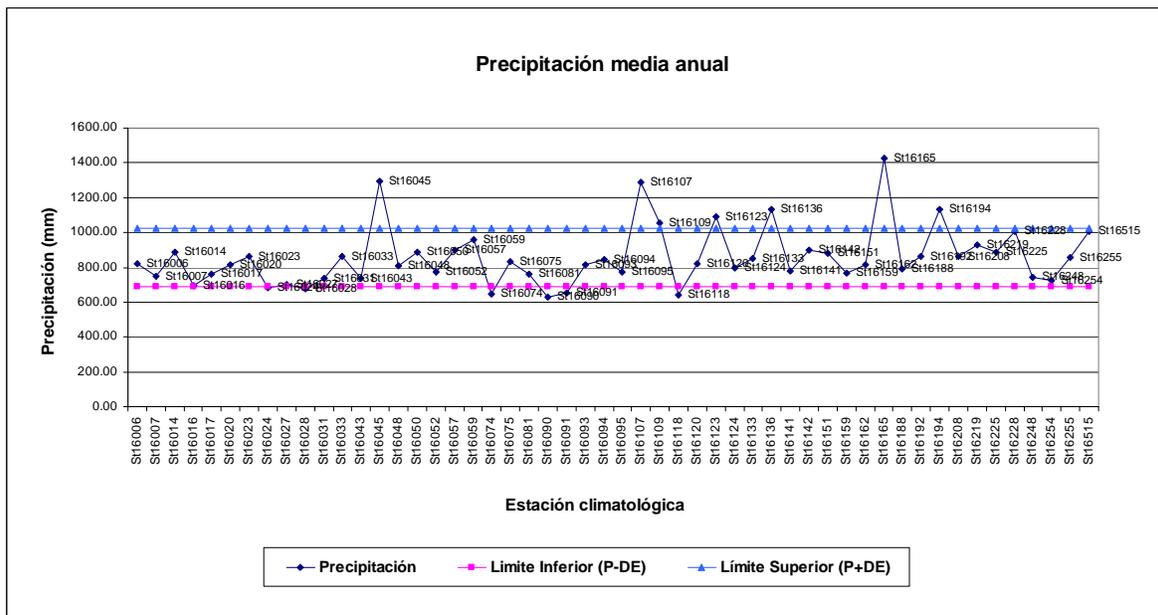


Figura 3.7. Precipitación media anual de las 52 estaciones climatológicas.

En la *Figura 3.7* se observa que las estaciones que rebasan el límite superior fueron: St16045 (1294 mm), St16107 (1286 mm), St16109 (1027 mm), St16123 (1092 mm), St16136 (1134 mm), St16165 (1428 mm) y St16194 (1133 mm), las cuales geográficamente se encuentran ubicadas en una franja que va de este a oeste cubriendo la zona centro del estado. Estas estaciones coinciden con lo observado en el análisis mensual de la precipitación (*Figura 3.5*), confirmando que éstas son las estaciones con una mayor cantidad de lluvia durante el año. Por el contrario, las estaciones que están por debajo del límite inferior fueron: St16024 (685 mm), St16028 (675 mm), St16074 (648 mm), St16090 (630 mm), St16091 (652 mm) y St16118 (642 mm), las cuales geográficamente se encuentran en la zona norte

(St16024, St12028, St16091, St16118) y suroeste (St16074 y St16090) del estado, es decir sin ninguna relación climática aparente salvo que son estaciones con mayor sequía en el estado, tal como se observó en el análisis mensual de precipitaciones (Figura 3.5). El resto de estaciones se encuentran distribuidas entre ambos límites con valores promedio de agua precipitada que oscilan entre los 698 y 1,010 mm anuales en todo el territorio del Estado.

3.4.1.2 Temperatura Observada (Tob)

En la *Figura 3.8* se observa un cambio gradual en la temperatura del aire observada a lo largo del año, con un incremento de febrero a mayo alcanzando las mayores temperaturas en mayo y junio, de julio a septiembre se tiene un descenso moderado de la temperatura y, a partir de septiembre, el descenso es rápido hasta alcanzar las temperaturas más bajas en enero.

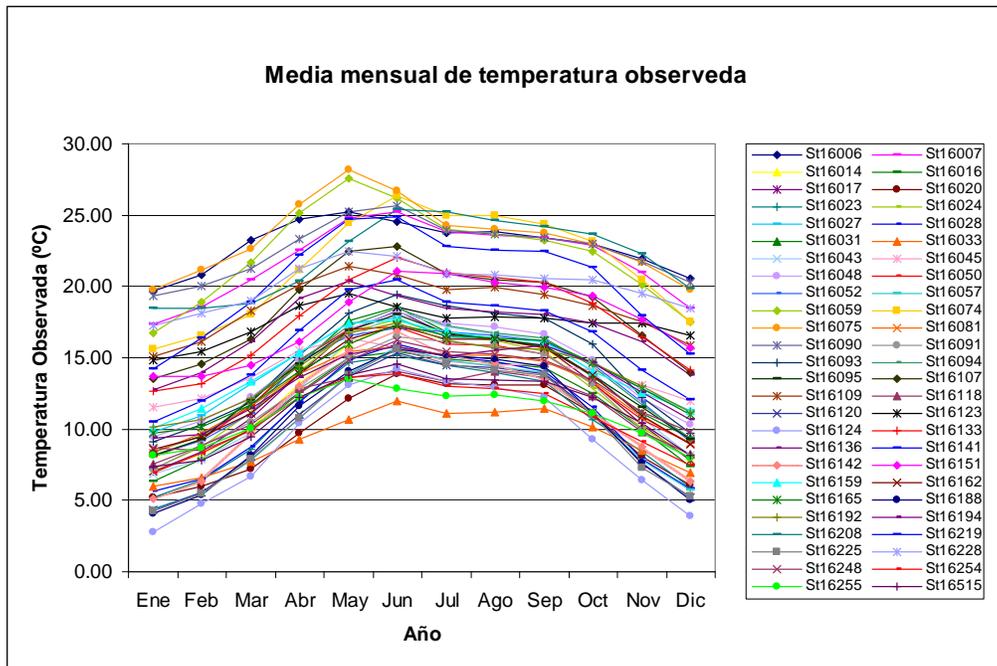


Figura 3.8. Variación estacional de la temperatura del aire observada en las 52 estaciones climatológicas.

Como se mencionó, en todas las estaciones el periodo con la temperatura más alta se presenta en mayo y junio (Figura 3.8). En mayo destacan las estaciones St16075 y St16059 con valores de 28.2 y 27.6 °C, respectivamente; ambas estaciones se localizan al sur del Estado cerca al límite con el estado de Guerrero, situándose en zonas cálidas subhúmedas con lluvias en verano. En junio se concentra la mayor cantidad de temperaturas altas con valores que oscilan entre los 12 y 26.7 °C. Por lo general, la variación de temperatura mensual entre las estaciones muestra una diferencia aproximada de 17.5 °C. Por ejemplo, la estación St16075 en mayo registra un valor de 28.2 °C, mientras que en el mismo mes la estación St16033 registra 10.7 °C. Estas diferencias se deben fundamentalmente a la accidentada orografía del Estado, conformada por diversas elevaciones sobre el nivel del mar, lo que origina variadas temperaturas de un lugar a otro.

3.4.1.3 Temperatura Máxima (Tmx)

La *Figura 3.9* muestra la variabilidad mensual de las temperaturas máximas registradas en las 52 estaciones. De enero a abril se observa un incremento progresivo de la temperatura alcanzando, por lo general, valores más altos en mayo y junio; de julio a diciembre el descenso de la temperatura es moderado hasta alcanzar las temperaturas máximas más bajas en diciembre (25.5 y 39.5 °C) y enero (25 y 39.4 °C).

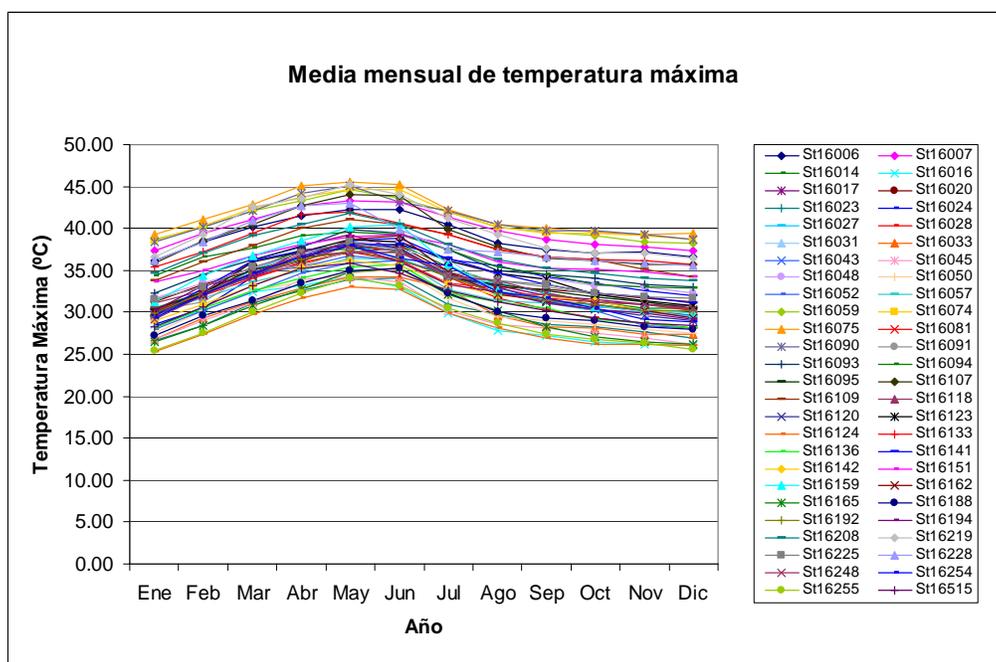


Figura 3.9. Variación estacional de la temperatura máxima en las 52 estaciones climatológicas.

En la *Figura 3.9*, se aprecia que en los meses de mayor temperatura máxima (mayo y junio), los valores oscilan entre los 32.5 y 45.5 °C. La variación mensual de temperatura máxima en las estaciones, por lo general, tiene una diferencia aproximada de hasta 14 °C. La estación más caliente en el año fue St16075 con 41.7 °C, esta estación se ubica al sur del Estado, en una zona cálida con lluvias en verano. Mientras que la estación con la menor temperatura máxima anual fue St16124 con 28.6 °C, ubicada al noreste del estado cerca al límite con Querétaro, en una zona semifría con lluvias en verano.

3.4.1.4 Temperatura Mínima (Tmn)

En cuanto a la variabilidad mensual de las temperaturas mínimas registradas en las 52 estaciones, se observa en la *Figura 3.10* que de enero a mayo la temperatura mínima tiene un incremento constante alcanzando, por lo general, en junio y julio las más altas temperaturas mínimas con oscilaciones entre los 4 y 18 °C. Durante los meses de agosto a diciembre la temperatura desciende notablemente hasta

alcanzar las temperaturas más frías del año en enero, con oscilaciones entre -5 a 11 °C.

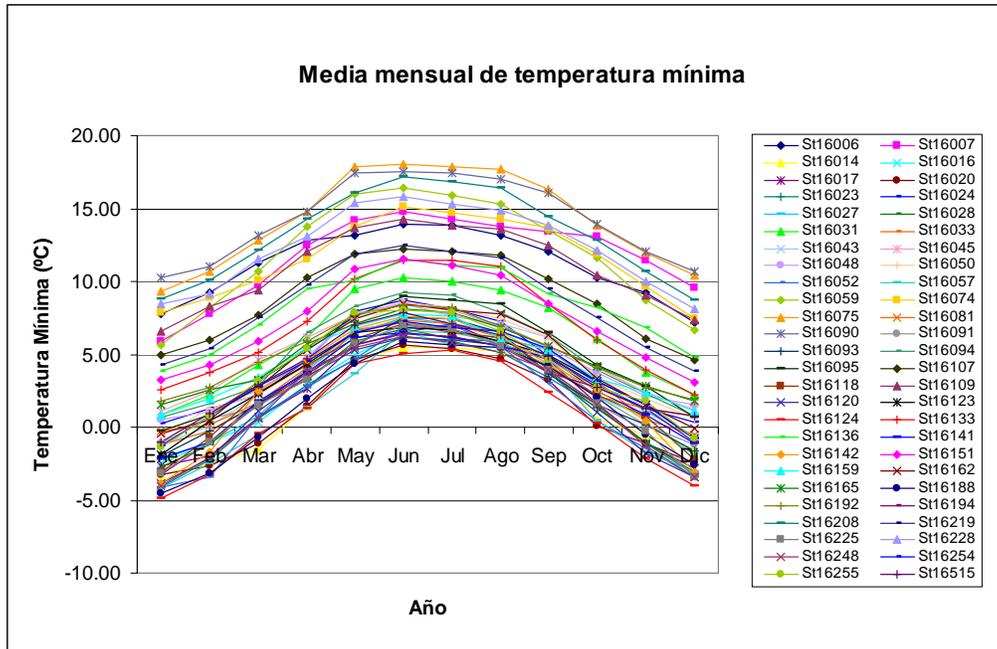


Figura 3.10. Variación estacional de la temperatura mínima en las 52 estaciones climatológicas.

En la *Figura 3.10* se aprecia que la variación mensual de la temperatura mínima en las estaciones tiene una diferencia aproximada de hasta 15 °C. La estación con la mayor temperatura mínima anual fue St16075 con 14.3 °C, mientras que la estación con la menor temperatura mínima anual fue St16124 con 0.7 °C (Anexo D). Ambas estaciones coinciden con las observadas en la *Figura 3.9*, confirmándose de esta manera que St16075 y St16124 son las estaciones más cálida y fría del Estado, respectivamente.

3.4.2 Observatorio de Morelia

Los elementos atmosféricos analizados para el observatorio de Morelia fueron la temperatura de bulbo seco, temperatura de bulbo húmedo, precipitación horaria, humedad relativa y presión en la estación. Tanto para la temperatura de *bulbo seco*,

así como para la de *bulbo húmedo* se estimaron medias mensuales expresadas en grados centígrados °C, mientras que para la *precipitación horaria* se estimó la suma mensual y la media anual, ambas expresadas en milímetros (mm). Para la *humedad relativa* y *presión en la estación* se calcularon las medias mensuales, la primera expresada en porcentaje y la segunda en milibares (mbar). En el Anexo E se muestra la tabla resumen con los resultados obtenidos para este observatorio.

3.4.2.1 Precipitación

La *Figura 3.11* muestra la distribución y variabilidad mensual de la precipitación en el observatorio de Morelia. Se aprecia que la temporada de lluvias se presenta de mayo a octubre, mientras que la temporada de sequía se da de noviembre a abril. Los meses más lluviosos son julio y agosto con una de precipitación de 179.3 y 162.19 mm, respectivamente.

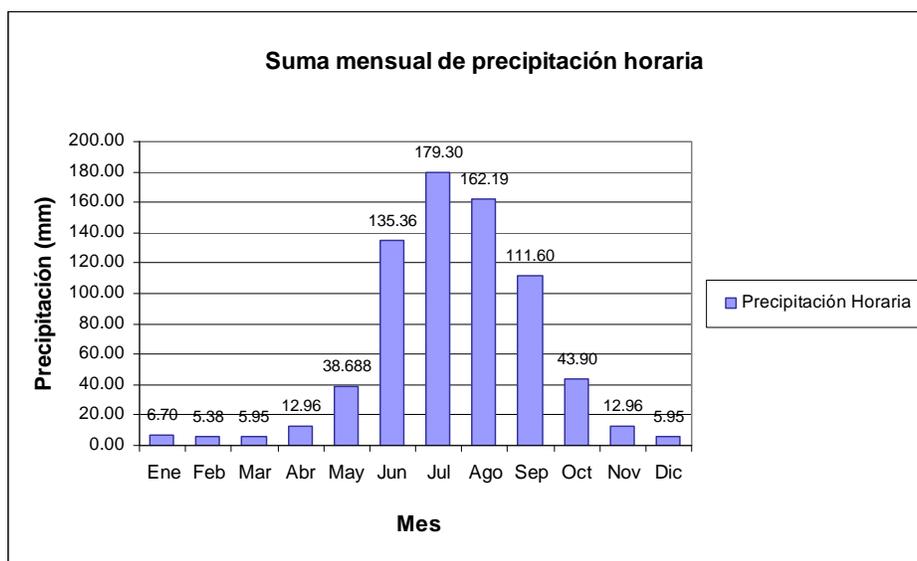


Figura 3.11. Suma mensual de precipitación en el observatorio de Morelia.

La precipitación anual en este observatorio alcanza los 720.9 mm (ver Anexo E); siendo junio, julio, agosto y septiembre los meses que cuentan con una mayor abundancia de precipitaciones, aportando alrededor del 81.6 % del total anual.

3.4.2.2 Presión en la estación

En cuanto a la presión en la estación, en la *Figura 3.12* se observa que ésta se mantiene casi constante a lo largo del año, con valores entre los 766 y 813 mbar. Los meses con la menor cantidad de presión en la estación son: abril con 777.5 mbar, junio con 766.7 mbar y octubre con 767.6 mbar, mientras que en el resto de los meses se tiene una presión entre los 810 y 812.7 mbar.

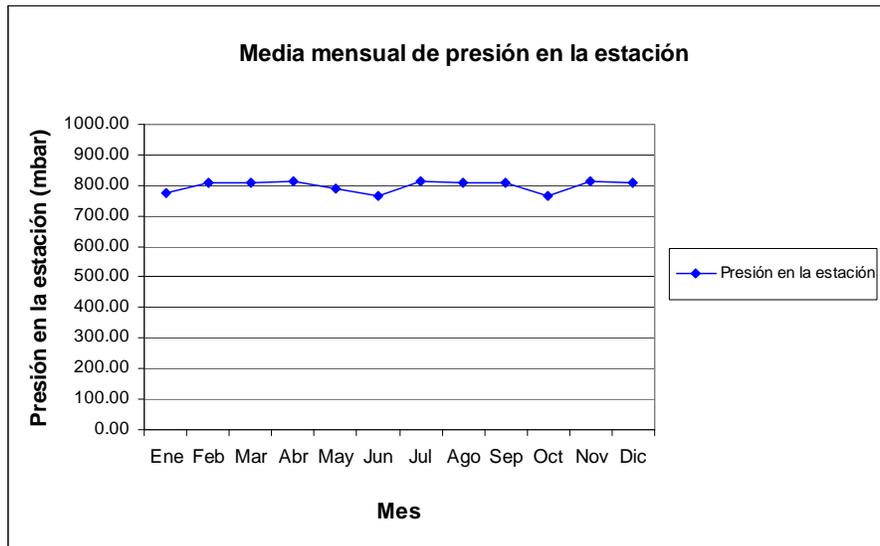


Figura 3.12. Variación mensual de presión en la estación en el observatorio de Morelia.

A su vez, se observa que la media anual de presión en la estación en el observatorio de Morelia tiene un valor aproximado de 799 mbar (Anexo E).

3.4.2.3 Temperatura de bulbo seco

En la *Figura 3.13* se presenta la variabilidad mensual de la temperatura de bulbo seco. Se observa un cambio gradual de calor a través del año, con un incremento de febrero a mayo, alcanzando la mayor temperatura en mayo con un valor de 21.6 °C, de mayo a julio el descenso es rápido y, de julio a septiembre la temperatura se mantiene ligeramente constante en torno a los 18.8 °C, para en octubre descender moderadamente hasta alcanzar las temperaturas más bajas en enero (14.6 °C).

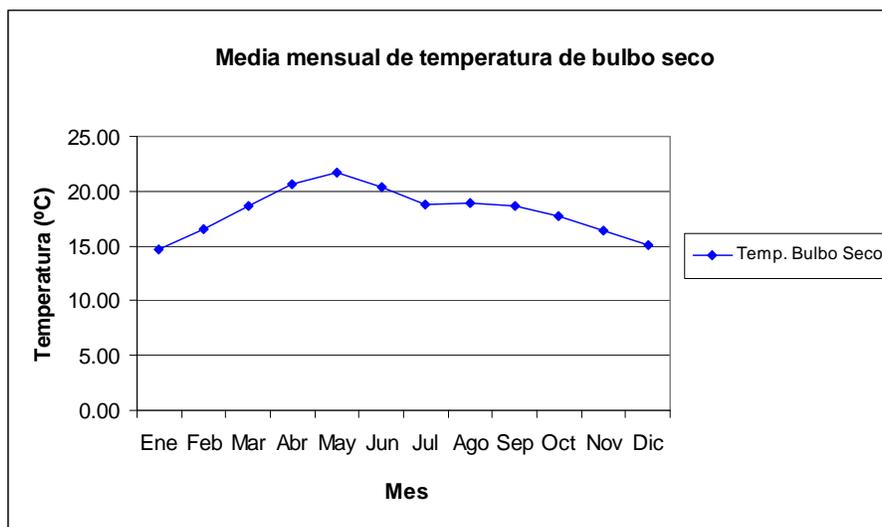


Figura 3.13. Variación mensual de la temperatura de bulbo seco en el observatorio de Morelia.

Por lo general, la media anual de temperatura de bulbo seco en el observatorio de Morelia alcanza los 18.2 °C (Anexo E).

3.4.2.4 Temperatura de bulbo húmedo y humedad relativa

En la *Figura 3.14* se presenta la variabilidad mensual tanto de la temperatura de bulbo húmedo, así como la humedad relativa en el observatorio de Morelia. Para la *temperatura de bulbo húmedo*, se observa que el mes más frío es enero con una temperatura media de 9.8 °C, mientras que de febrero a junio el incremento de calor es constante, alcanzando la mayor temperatura en junio con un valor de 16 °C, por lo general de julio a septiembre la temperatura se mantiene constante en torno a los 15.7 °C.

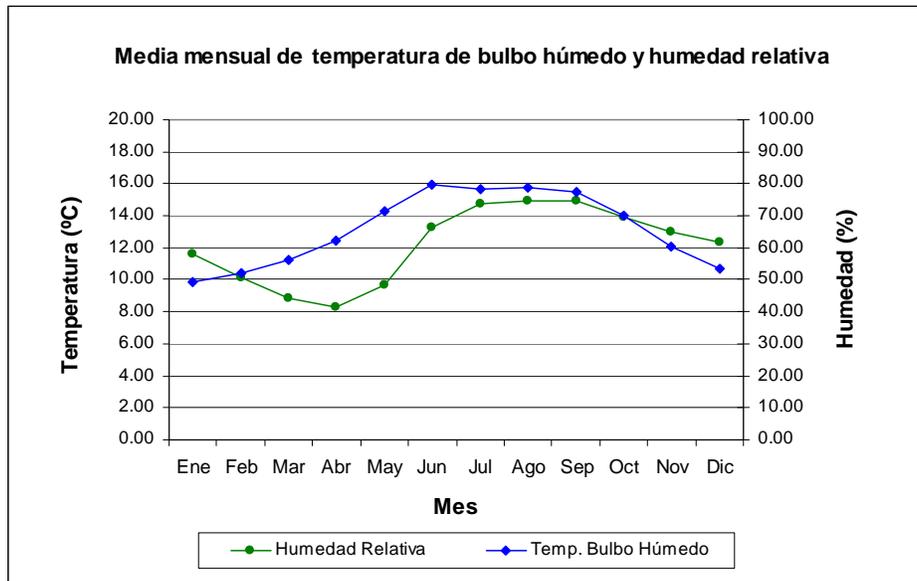


Figura 3.14. Variabilidad mensual de temperatura de bulbo seco y humedad relativa en el observatorio de Morelia.

Por su parte la humedad relativa desciende gradualmente de enero a abril, alcanzando la menor humedad en abril con un valor de 41.5 %, de mayo a junio el porcentaje se incrementa rápidamente hasta alcanzar los 66.4 %, de julio a septiembre la humedad se mantiene ligeramente constante en torno a los 74.7 %, este valor representa el máximo porcentaje de saturación en el observatorio, y a partir de septiembre desciende moderadamente hasta alcanzar en diciembre una humedad de 61.8 %.

La relación principal entre la temperatura de bulbo húmedo y la humedad relativa, se basa en la estimación del contenido de humedad existente en el aire, para lo cual, el bulbo del termómetro es cubierto con una tela o algodón húmedo. La evaporación del bulbo húmedo desde la superficie enfría el bulbo húmedo hasta alcanzar una temperatura estacionaria (punto de saturación), esto es, la existencia de un equilibrio entre el calor perdido por la evaporación y el ganado por la radiación. La media anual de temperatura de bulbo húmedo en el observatorio es de 13.15 °C, mientras que porcentaje anual de humedad relativa es de 60.7 % (Anexo E).

3.5 Síntesis

El análisis de datos es una de las actividades fundamentales en el proceso de minería de datos. Los datos analizados corresponden a las estaciones climatológicas tradicionales (CLICOM) y los Observatorios de Zamora y Morelia. Este análisis se realizó en dos etapas, en la *primera* se hizo una evaluación preliminar de la disponibilidad de datos de todas las estaciones listadas para la zona de estudio (estaciones tradicionales y observatorios) y en la *segunda* etapa se determinó la calidad de la serie de datos para establecer las estaciones significativas para el proceso de regionalización.

De la evaluación quedaron 52 estaciones que pertenecen al grupo de las estaciones tradicionales; mientras que en el caso de los observatorios de Zamora y Morelia, el primero quedó descartado por su falta de datos siendo Morelia la estación a usarse en el desarrollo del modelo de regionalización climática. Posterior a la determinación de las estaciones significativas, se analizó la variabilidad y distribución de las cuatro variables principales del CLICOM (precipitación y temperaturas -observada, máxima y mínima-), así como también de algunas de las variables del observatorio (temperatura de -bulbo seco, bulbo húmedo-, precipitación horaria, humedad relativa y presión en la estación); este análisis proporcionó la identificación de tendencias y comportamientos del conjunto de datos disponibles.

Capítulo IV.

Zona de Estudio

En el capítulo III se describió el proceso de análisis mediante el cual se comprobó la disponibilidad y calidad de la serie de datos para las variables medidas en las estaciones del CLICOM y observatorios. Con esto se determinó el número de estaciones significativas para el proceso de clasificación de zonas climatológicamente similares en el estado de Michoacán. En este capítulo se presentan las características generales de la zona de estudio, sobresaliendo su topografía y geología dado su accidentado relieve, su litología por su variedad de suelos y climatología por su variedad de climas; factores que han determinado la riqueza de sus recursos naturales permitiendo el establecimiento de sectores económicos importantes para la región, como el desarrollo de zonas agrícolas e industriales.

Este capítulo contiene la descripción de las principales características fisiográficas del estado de Michoacán. La fisiografía es muy importante ya que divide en dos grandes provincias al Estado: Sierra Madre del Sur y Eje Neovolcánico, las cuales son referencia para la caracterización geológica, climatológica, hidrológica, y vegetativa del estado. Finalmente, se describe la necesidad de regionalizar zonas similares en el estado de Michoacán para efectos de aprovechamiento de sus recursos disponibles.

La información contenida en este capítulo se basa principalmente en la Síntesis de Información Geográfica del Estado de Michoacán publicada por el Instituto Nacional de Estadística, Geografía e Informática (INEGI) en 1985, México.

4.1 Aspectos Generales

El estado de Michoacán de Ocampo se ubica en la zona centro occidente de la República Mexicana, entre las coordenadas 20°23'44" y 18°09'49" de latitud norte y 100°04'48" y 103°44'20" de longitud oeste, con una superficie de 58,837 km². Limita con los estados de Jalisco y Guanajuato al norte, Querétaro al noroeste, Estado de México al este, Guerrero al sur y sureste, Colima al oeste y con el Océano Pacífico al suroeste (Figura 4.1).



Figura 4.1. Ubicación geográfica del estado de Michoacán.
Fuente: Mapas de México (2006)

De acuerdo a INEGI (1985), Michoacán cuenta con 113 municipios, siendo Morelia su capital. El estado se divide en dos *provincias fisiográficas*¹: Sierra Madre del Sur y Eje Neovolcánico. En la primera provincia los picos de mayor elevación son el Coalcomán (2,985 msnm) y Cantador (2,463 msnm), ubicadas en las localidades de Aguililla y Coalcomán. En la provincia del Eje Neovolcánico destacan el volcán de Tancítaro (3,525 msnm) en el municipio de Tancítaro, los cerros de Patamban (3,857 msnm) en Tangancícuaro, San Andrés (3,605 msnm) en Ciudad Hidalgo y Tzirate (3,340 msnm) en Quiroga. Entre ambas provincias se localiza una zona

¹ Son regiones de características geomorfológicas distintas.

plana denominada Tierra Caliente, en ella se encuentran los valles de Apatzingán, Churumuco, Tepalcatepec, Tuzantla, Tiquicheo y Huetamo.

La provincia más poblada del estado es el Eje Neovolcánico, destacando los municipios de Morelia, Uruapan, Zamora, Zitácuaro y La Piedad. Asimismo esta provincia cuenta con importantes cuerpos de agua, como los lagos de Cuitzeo, Pátzcuaro, Zirahuén y parte del lago de Chapala, que han permitido el desarrollo de la agricultura, ganadería y otras actividades de la región.

4.2 Fisiografía y Geografía

Como se mencionó, el territorio michoacano está dividido en dos grandes provincias, la Sierra Madre del Sur y el Eje Neovolcánico (Figura 4.2), por lo que a continuación se describen sus principales aspectos fisiográficos y geológicos con base en dicha división.

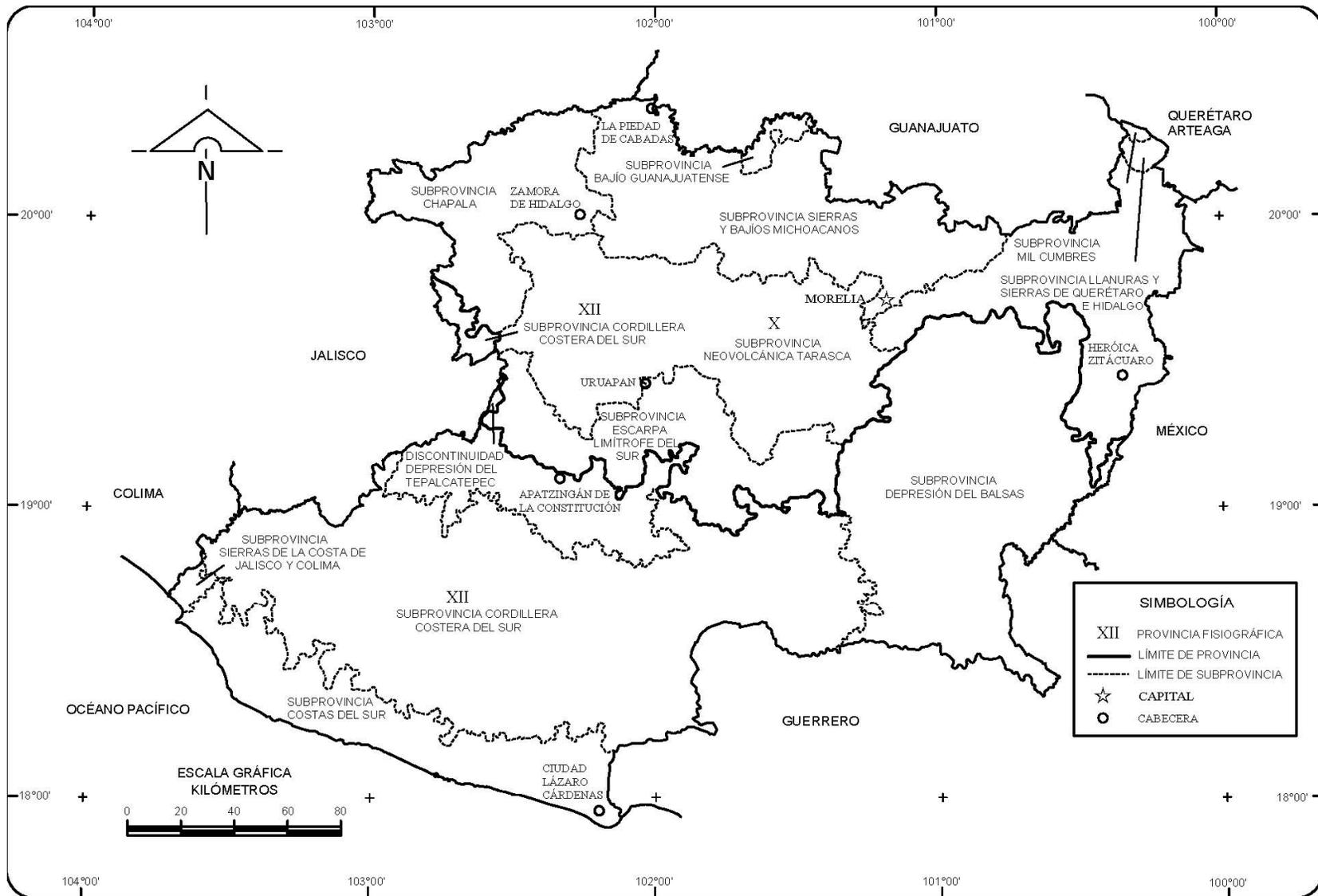


Figura 4.2. Fisiográfica del estado de Michoacán. Fuente: INEGI, 2006. Carta Fisiográfica, 1:1'000,000.

4.2.1 Sierra Madre del Sur

Los principales factores geológicos que caracterizan a esta provincia son el magmatismo y el tectonismo, a excepción de la zona costera donde domina la erosión marina y el tectonismo. Entre los tipos de rocas que conforman esta provincia, por lo general, se encuentran las metamórficas, sedimentarias, intrusivas y extrusivas. Esta provincia a su vez se subdivide en las siguientes subprovincias:

- ***Cordillera Costera del Sur.*** Se extiende a lo largo de las costas del Estado. Se encuentran muchas y diferentes rocas como calizas y calizas-lutitas. Sus tipos de suelos predominantes son: *Luvisoles* (rojos arcillosos), ricos en materia orgánica con un drenaje lento y una baja permeabilidad; *Litsoles* (grises superficiales), moderadamente ácidos con menos de 10 cm de profundidad con un drenaje moderado y una baja permeabilidad; y *Feozem*, (negros de textura franca) moderadamente ricos en materia orgánica con una permeabilidad media y un drenaje moderado, estos suelos se encuentran en menor proporción.
- ***Depresión del Balsas.*** Se extiende por Morelia y Zitácuaro, hasta los márgenes del río Balsas. Por lo general, presenta una dominancia de suelos de poco desarrollo, de colores claros y textura arenosa llamados *Regosoles*. Estos suelos disponen de escasa materia orgánica, cuentan con un drenaje moderado y una permeabilidad media. También en menor proporción se encuentran los *Litsoles* de fertilidad moderada con baja permeabilidad y drenaje moderado.
- ***Discontinuidad Depresión del Tepalcatepec.*** Es un valle rodeado por elevadas montañas. Casi la totalidad de sus suelos son arcillosos y fértiles, motivo por el cual se han desarrollado actividades agrícolas tanto de riego como de temporal. La mitad de sus suelos son profundos mayores a 90 cm, con permeabilidad baja y drenaje lento; y la otra mitad esta formado por suelos medianamente profundos que tienen niveles altos de fertilidad.

- **Costas del Sur.** Su paisaje corresponde a una angosta llanura costera constituida principalmente por sierras bajas de origen sedimentario, volcánico y metamórfico, y algunos valles y llanuras formados con materiales aluviales. Sus suelos son moderadamente fértiles de composición variable, destacando los *Regosoles*, con alta permeabilidad y drenaje rápido; los *Luvisoles*, con baja permeabilidad y drenaje lento; y los *Feozem*, con baja permeabilidad y drenaje moderado.
- **Sierras de la Costa de Jalisco y Colima.** Llanura costera cuyos suelos dominantes son: *Vertisoles* y *Feozem*. Los *Vertisoles* son arcillosos y poco permeables, pero muy fértiles a pesar de ser duros y agrietados en época de caudal mínimo, y lodosos y adhesivos la época lluviosa. Los *Feozem* presentan cantidades moderadas de materia orgánica, siendo fértiles con una permeabilidad media y un drenaje moderado.

4.2.2 Eje Neovolcánico

La actividad magmática dio origen a la formación del paisaje y relieve característico de esta provincia, que básicamente esta constituida por rocas volcánicas, sedimentarias y areniscas, siendo integrada por sierras volcánicas. El volcán más importante es el Tancitaro (3,857 msnm), que es el punto más alto del relieve michoacano. En esta provincia se localiza casi la totalidad de la cuenca del río Lerma, la cual nace al este de Toluca y después de atravesar el Bajío Guanajuatense deposita sus aguas en el lago de Chapala. Esta región a su vez se subdivide en ocho subprovincias:

- **Bajío Guanajuatense.** En esta región se halla un sistema de topofomas denominado *gran llano*. Sus suelos son de origen aluvial y, en parte, residual procedentes de basaltos y tobas. El tipo de suelo dominante son los *Vertisoles*, los cuales son fértiles dedicados a la agricultura de riego y de temporal. Estos

suelos, por lo general, son arcillosos con contenidos de materia orgánica moderada, de permeabilidad baja y drenaje alto.

- ***Llanuras y Sierras de Querétaro e Hidalgo.*** La principal característica de esta subprovincia son los lomeríos de colinas redondeadas con cañadas. Por lo general, sus suelos son de fertilidad moderada, siendo los más importantes: los *Feozem*, de permeabilidad baja con una capacidad de absorción moderada; los *Luvisoles*, moderadamente ricos en materia orgánica con drenaje moderado y baja permeabilidad; y los *Litosoles*, ricos en materia orgánica con una permeabilidad baja y drenaje moderado.

- ***Chapala.*** En esta región se encuentra el lago más grande del país (lago de Chapala), el cual tiene una longitud de este a oeste de 80 km. Por lo general, sus suelos van de muy fértiles a moderadamente fértiles, destacando los *Vertisoles*, arcillosos de baja permeabilidad y drenaje moderado; *Luvisoles*, ricos en materia orgánica con permeabilidad media y drenaje moderado; y *Andosoles*, extremadamente ricos en materia orgánica con una permeabilidad media y drenaje alto.

- ***Sierras y Bajíos Michoacanos.*** Se localizan en el corazón de la cuenca del río Lerma, siendo la principal ocupación de la tierra la agricultura. Las efusiones volcánicas en el oriente generaron, por bloqueo de drenaje, los lagos de Yuriria y Cuitzeo. Sus suelos tienen alta fertilidad, siendo los más importantes: *Vertisoles*, que ocupan el 70% de la región con altos contenidos de arcilla por lo que son poco permeables y de drenaje lento; *Feozem*, ricos en materia orgánica con permeabilidad media y drenaje moderado; y *Luvisoles*, moderadamente ricos en materia orgánica con permeabilidad media y drenaje moderado.

- ***Mil Cumbres.*** Es una región accidentada y compleja por la diversidad de sus geofomas como: sierras volcánicas, mesetas, lomeríos y el valle por el cual el río Lerma se dirige hacia el norte. Sus tipos de suelos dominantes son:

Andosoles, ricos en materia orgánica con permeabilidad alta y drenaje rápido;
Luvisoles, de fertilidad moderada con baja permeabilidad y drenaje lento; y
Regosoles, fértiles con permeabilidad media y drenaje moderado.

- ***Neovolcánica Tarasca.*** Se caracteriza por la presencia de un gran número de volcanes. La mayor parte de su extensión está clasificada como sierra volcánica con llanuras. Sus suelos son fértiles, destacando los siguientes: *Andosoles*, moderadamente ricos en materia orgánica con permeabilidad media y drenaje moderado; *Luvisoles*, arcillosos de baja permeabilidad y drenaje lento; y *Acrisoles*, arcillosos de permeabilidad moderada y drenaje moderado.

- ***Escarpa Limítrofe del Sur.*** Se denomina así por la inclinación que presenta en aproximadamente 30 km, con un descenso que va desde los 2,000 hasta los 300 msnm. Esta región sirve de límite con la provincia Sierra Madre del Sur. En su mayoría está constituida por mesetas lávicas asociadas a lomeríos y sierras. Entre sus tipos de suelos dominantes se encuentran los *Vertisoles*, fértiles con baja permeabilidad y pésimo drenaje, y los *Andosoles*, ricos en materia orgánica con permeabilidad media y drenaje moderado.

- ***Lagos y Volcanes de Anáhuac.*** Abarca una porción reducida (126 km²) en el oriente del Estado. Esta integrada por sierras volcánicas. Sus suelos son de fertilidad moderada, destacando los *Andosoles* y *Acrisoles*. Los *Andosoles* tienen altos contenidos de materia orgánica con baja permeabilidad y drenaje moderado, mientras que los *Acrisoles* son moderadamente ricos en materia orgánica con permeabilidad media y drenaje moderado.

4.3 Climatología

Michoacán presenta una gran variedad de climas que van desde los más cálidos, en la región de Tepalcatepec, hasta los semifríos de las zonas altas de la Meseta Tarasca y Mil Cumbres. El régimen de humedad predominante es el subhúmedo con lluvias en verano y seco en invierno. La frecuencia de *heladas* tiene relación con los diversos tipos de climas, debido a esto, en la zona costera y en la depresión del Balsas y del Tepalcatepec no se registran heladas; en las zonas semicálidas subhúmedas del centro y norte del Estado su presencia es de 0 a 20 veces al año; y en las zonas templadas es de 20 a 80 veces; tiene mayor incidencia en las zonas templadas y semifrías de altitudes superiores a 2,500 msnm, donde alcanzan de 80 a 140 veces al año.

La presencia de *granizadas* también tiene relación con los tipos de climas, de manera que en zonas cálidas como las de la costa y sierra de Coalcomán no se presentan granizadas durante todo el año, en las zonas semicálidas y templadas como la Uruapan y Pátzcuaro se presentan de 2 a 4 veces al año, y en las zonas templadas y semifrías de altitudes superiores a 2,500 msnm su presencia anual es de 4 a 8 veces. La distribución de climas en Michoacán se relaciona a los siguientes factores (INEGI, 1985):

- Las variadas diferencias del relieve debido a las diversas altitudes sobre el nivel del mar.
- La presencia de cadenas montañosas alineadas a la costa, las cuales actúan como barrera orográfica.
- La cercanía al mar, que implica la presencia de vientos húmedos que ascienden por las montañas y provocan precipitaciones. En el verano los vientos soplan con dirección este y noreste, mientras que en invierno predomina el aire seco.

De acuerdo a INEGI (1985) los tipos de climas que se registran en Michoacán, con base en la relación de la altitud y la vegetación, son las siguientes (Figura 4.3):

1. *Seco muy cálido y cálido*. BS(h'). Es un tipo de clima con lluvias en verano y escasas en los meses restantes. Tiene un promedio de precipitación total anual de 500 a 700 mm y una temperatura media anual que oscila entre los 27 y 31 °C. Se localiza en una franja irregular paralela al cauce de los ríos Balsas y Tepalcatepec.
2. *Semiseco muy cálido y cálido*. BS1(h'). Se localiza en los valles y lomeríos de la zona que va del suroeste de Huetamo hasta Tazumbo (límite con Jalisco). Asimismo, ocupa una pequeña franja al suroeste de Apatzingán, Gabriel Zamora, Nueva Italia y Nuevo Churumuco. Tiene un promedio de precipitación total anual de 800 mm y una temperatura media anual que fluctúa entre los 23.8 y 30.3 °C.
3. *Cálido subhúmedo con lluvias en verano*. A(w). Este clima abarca la mayor parte del Estado ocupando toda el área de las Planicies Costeras, las partes bajas de la Sierra Madre del Sur y partes altas de la Cuenca del Balsas. El periodo de lluvias se presenta de junio a octubre, con una mayor incidencia en septiembre (alrededor de 258 mm). El promedio de precipitación total anual es de 989 mm y la temperatura media anual es de 23 °C.
4. *Semicálido subhúmedo con lluvias en verano*. ACw. Este tipo de clima ocupa todo el altiplano michoacano y las partes bajas de la cordillera Neovolcánica. Tiene un promedio de precipitación total anual que fluctúa de 720 a 900 mm y su temperatura media anual alcanza los 20.9 °C.
5. *Semicálido húmedo con abundantes lluvias en verano*. ACm. La zona comprendida por este clima es la más húmeda de Michoacán (inmediaciones de Uruapan, Nuevo Parangaricutiro y Ziracuaretiro). El promedio de precipitación total anual alcanza los 1622 mm, con una mayor incidencia en

septiembre (352 mm), mientras que la temperatura media anual apenas alcanza los 19 °C.

6. *Templado subhúmedo con lluvias en verano*. C(w). Estos climas se encuentran en el Eje Neovolcánico. Su temperatura media anual fluctúa entre los 12.5 y 17.4 °C y el promedio de precipitación total anual va desde los 790 a 1343 mm. Conforme se avanza hacia el norte las condiciones se tornan menos húmedas, pues la precipitación disminuye de 790 a 741 y de 1,343 a 959 mm anuales.
7. *Templado húmedo con abundantes lluvias en verano*. C(m). Se localiza en una pequeña región entre Tancítaro y Uruapan, en ella el promedio de precipitación total anual es de 1,500 mm (lluvia suficiente para que en los meses de invierno no exista sequía). Asimismo la temperatura media anual de esta zona alcanza los 17.5 °C.
8. *Semifrío húmedo con abundantes lluvias en verano*. C(E)(m). Este tipo de clima influye en dos zonas, *la primera* en los alrededores de Tancítaro y *la segunda* por Zinapécuaro, Ciudad Hidalgo y Los Azufres. Tiene un promedio de precipitación total anual de 1,534 mm, siendo agosto el mes con mayor abundancia de lluvias (362 mm). Su temperatura media anual es de 10.6 °C.

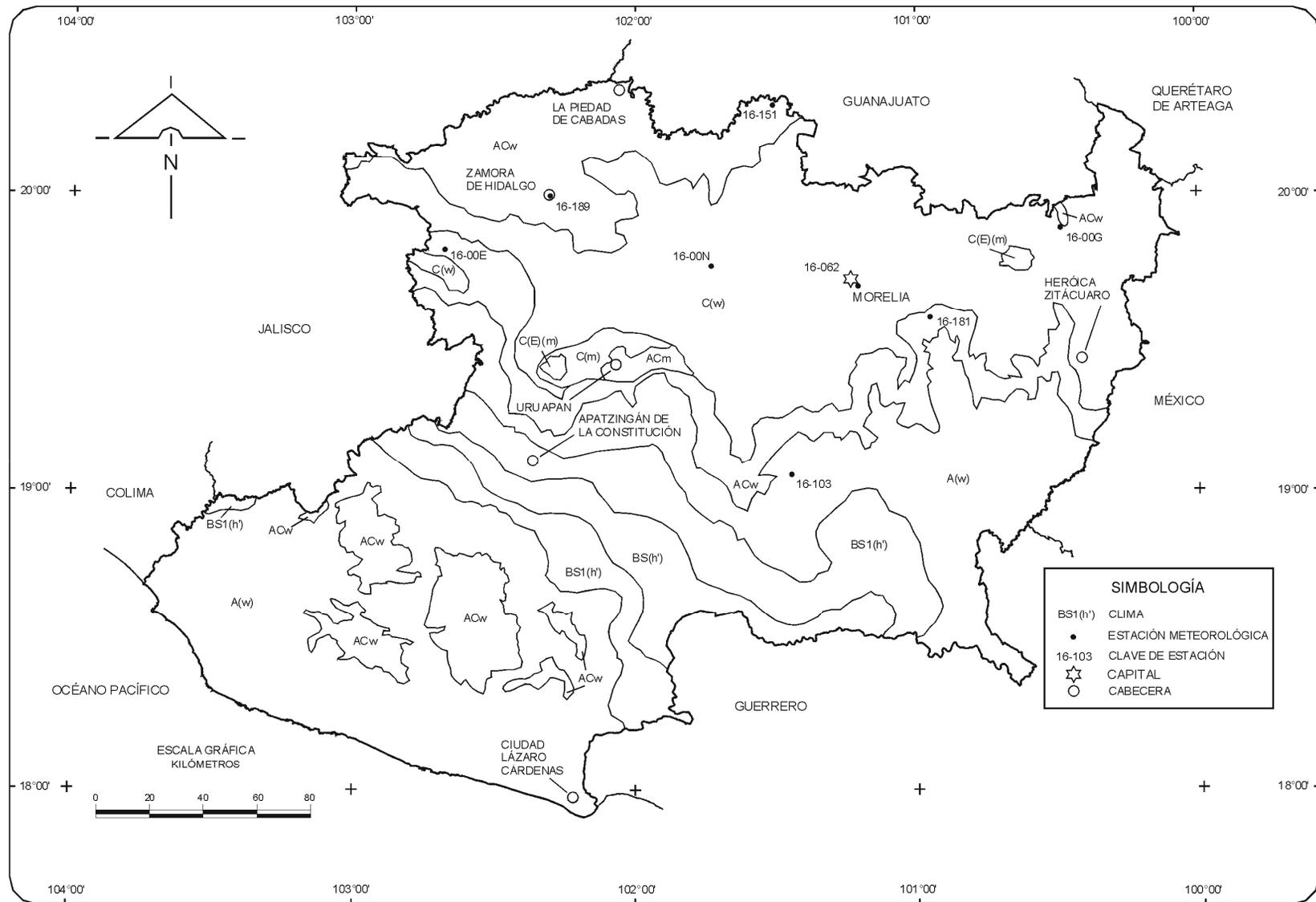


Figura 4.3. Tipos de climas del estado de Michoacán. Fuente: INEGI, 2003. Carta de Climas, 1:1'000,000.

4.4 Hidrología

De acuerdo a la Comisión Nacional del Agua, el estado de Michoacán enfrenta carencias de agua debido a que su distribución no es uniforme. Es así, que en algunas regiones como en la zona de Tierra Caliente el agua se presenta de manera escasa, mientras que en otras como en la Costa el agua es abundante (Juárez, 2006).

En el relieve michoacano se formaron tres vertientes: la del *Norte* que comprende los lagos de Cuitzeo, Pátzcuaro, Chapala y Zirahuén; la del *Centro* que comprende los ríos Tepalcatepec y Balsas; y la del *Sur*, donde los escurrimientos de la Sierra de Coalcomán desembocan al Océano Pacífico (INEGI, 1985).

4.4.1 Aguas Superficiales

De acuerdo a INEGI (1985) son cuatro las regiones hidrológicas situadas en el estado de Michoacán (Figura 4.4). Sin embargo esta entidad enfrenta carencias en la disposición de aguas superficiales, sobretodo en la región de Lerma-Santiago, debido a que las demandas exceden la oferta natural del recurso (Juárez, 2006).

a. Región Hidrológica 12 "Lerma-Chapala-Santiago"

Esta región se caracteriza por tener zonas planas y amplias que cubren una superficie aproximada de 14,818 km². Sus valles están separados por elevaciones con altitudes de 1,600 a 2,000 msnm. Las cuencas hidrológicas que conforman esta región son: Río Lerma-Toluca, Río Lerma-Salamanca, Río Lerma-Chapala, Lago de Chapala, y Lago de Pátzcuaro-Cuitzeo y Laguna de Yuriria. Sus aguas son utilizadas en los Distritos de Riego Morelia-Queréndaro (DR20), Rosario Mezquite (DR87), Zamora (DR61), Zacapu (DR22), Ciénega de Chapala (DR24) y Tzurumutaro (DR21). Asimismo esta región surte de agua potable a la ciudad de Morelia y también el agua es aprovechada para la generación de energía eléctrica.

b. Región Hidrológica 16 "Armería-Coahuayana"

Esta región está formada por las cuencas de los ríos Armería y Coahuayana, pero la mayor parte de su superficie se encuentra en el estado de Jalisco. Sólo una porción reducida de la cuenca del río Coahuayana pertenece a Michoacán (aprox. 1,495 km²), donde sus corrientes drenan hasta desembocar al Océano Pacífico, en la boca de Apiza. Pese a su reducida extensión, ésta región dispone de escurrimientos que representan el 5.2 % del total estatal, pero debido a su accidentada superficie no tiene centros de población de importancia ni zonas industriales.

c. Región Hidrológica 17 "Costa de Michoacán"

Toda la región está ubicada en territorio michoacano cubriendo un área de 8,078 km². La región se forma por las corrientes entre los ríos Coahuayana y Mezcala, las cuales desembocan en el Océano Pacífico. Esta región comprende a dos cuencas principales: el Río Nexpa y otros, y Río Coalcomán y otros; sin embargo, pese a su accidentada superficie no cuenta con centros de población, y las áreas de importancia agrícola sólo se encuentran en parte de la cuenca del Río Nexpa y otros al utilizar sus aguas en el Distrito de Riego José María Morelos (DR98).

d. Región Hidrológica 18 "Río Balsas"

El Río Balsas conocido también como Atoyac o Mezcala es una de las corrientes más importantes del país. La Depresión del Balsas está situada en la zona central de Michoacán con una superficie de 34,294 km² y altitudes menores a los 600 msnm. Esta región comprende las siguientes subcuencas: Río Balsas-Zirándaro, Río Balsas-Infiernillo, Río Cutzamala, Río Tacámbaro, Río Tepalcatepec-Infiernillo y Río Tepalcatepec. Sus aguas son utilizadas en los distritos de riego José María Morelos (DR98), Tuxpan (DR45), Lázaro Cárdenas (DR97) y Quitupan-La Magdalena (DR99). Asimismo, son aprovechadas para la generación de energía eléctrica.

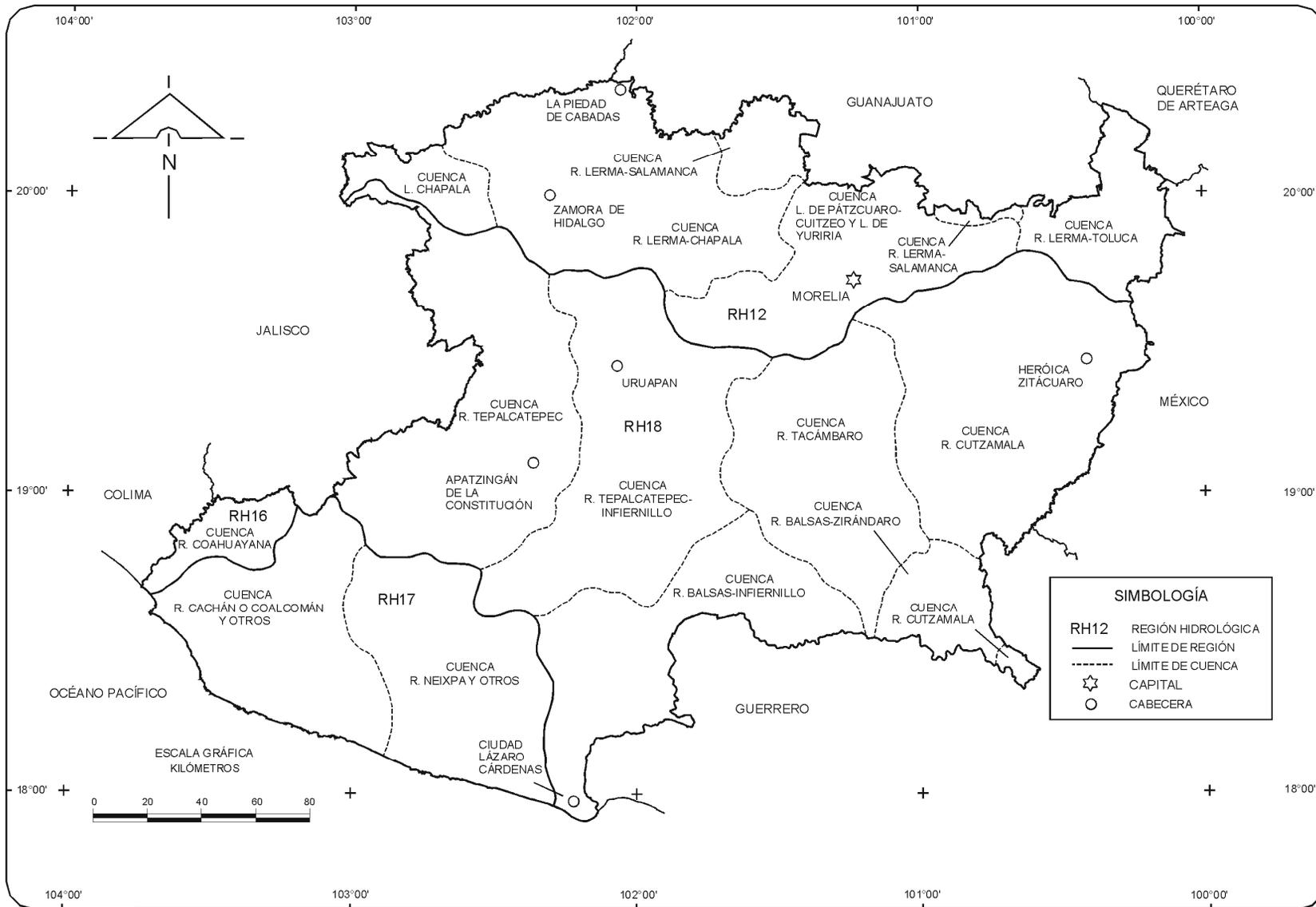


Figura 4.4. Regiones y Cuencas Hidrológicas. Fuente: INEGI, 2003. Carta Hidrológica, 1:1'000,000.

4.4.2 Aguas Subterráneas

De acuerdo a INEGI (1985), debido a su geografía Michoacán presenta dos zonas de aguas subterráneas: la zona norte y la zona sur.

- *Zona Norte.* Forma parte de la provincia fisiográfica del Eje Neovolcánico, encontrándose las mejores condiciones geohidrológicas del estado. Se explotan acuíferos en los valles de Maravatío, Zamora, Zacapu, Morelia y Queréndaro. Una buena producción acuífera se da en los valles de Pastor Ortiz y la Piedad de Cabadas; también en la región de Apatzingán las posibilidades de su explotación son buenas. Otros acuíferos disponibles se encuentran en los flancos del valle de la Ciénega de Chapala.
- *Zona Sur.* Forma parte de la provincia fisiográfica de la Sierra Madre del Sur. Por su formación de estructuras complejas y terrenos abruptos ha dificultado la exploración de posibilidades acuíferas. En Apatzingán y sus zonas aledañas existe sólo un acuífero de importancia, denominado unidad hidrológica de Apatzingán, el cual se aprovecha para uso doméstico y agrícola. El resto de la zona se abastece de agua potable mediante manantiales, ríos y pozos a cielo abierto, como es el caso de la población de Arteaga.

En general, las zonas más importantes de recarga de los acuíferos se localizan en las sierras de Michoacán, donde las rocas que las constituyen tienden a ser más permeables facilitando la infiltración de los escurrimientos (INEGI, 1985). Por su parte, el drenaje de los acuíferos se realiza de manera *artificial* y *natural*. La artificial se da por medio de pozos y norias principalmente en las regiones de Zamora-Ciénega de Chapala, Morelia-Queréndaro, Coahuayana, Maravatío y Pastor Ortiz. El drenaje natural se efectúa por medio de ríos, los cuales funcionan como drenes superficiales y subterráneos (INEGI, 1985).

De acuerdo a CONAGUA (2005) en Michoacán se tienen identificados 22 acuíferos, de los cuales 9 de ellos están en condición de sobreexplotados (Figura 4.5). Debido a esta sobreexplotación, la Comisión Nacional del Agua prohibió en todo el Estado la perforación de pozos para la extracción del agua subterránea (Juárez, 2006). Los acuíferos sobreexplotados son: Pastor Ortiz-La Piedad (1605), Ciénega de Chapala (1607), Zamora (1608), Briseñas-Yurécuaro (1609), Ciudad Hidalgo-Tuxpan (1610), Uruapan (1614), Nueva Italia (1616), Apatzingán (1620) y Cotija (1622).

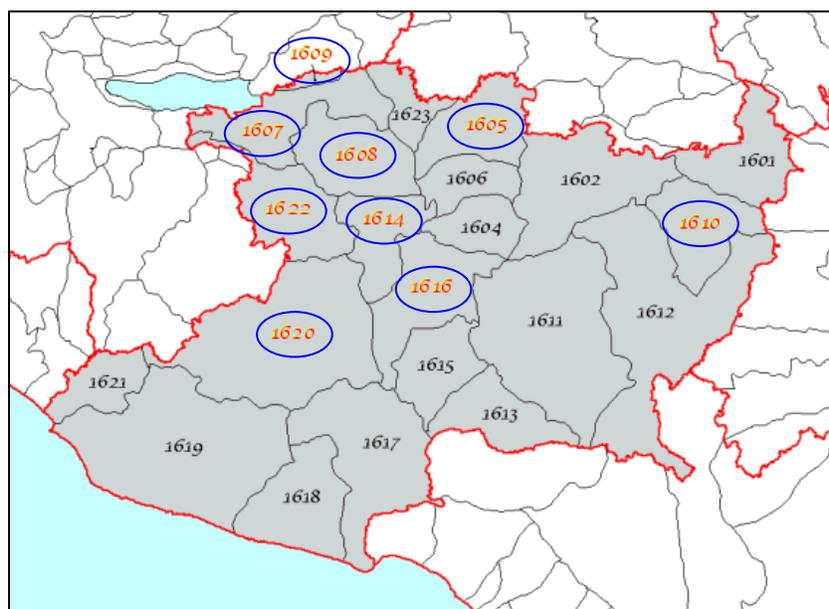


Figura 4.5. Acuíferos con problemas de sobreexplotación (círculo azul).
Fuente: CONAGUA, 2005

Para 2005, Michoacán contaba con 6,212 pozos, de los cuales 3,850 eran de uso agrícola, 1,967 de uso público, 148 de uso doméstico, 125 para hotelería y servicios, 108 destinados a la actividad pecuaria, 50 a la acuicultura y 17 para usos agroindustriales (Juárez, 2005).

4.5 Vegetación

En Michoacán, la presencia de las provincias fisiográficas Sierra Madre del Sur y Eje Neovolcánico diversifica la topografía, influye sobre el clima y da lugar a la variedad de suelos. Ante este escenario, Michoacán presenta una gran diversidad de vegetación, siendo uno de los estados más importantes del país por su producción agrícola y de vegetación natural. Entre los grupos de vegetación se tiene la agricultura, vegetación transformada y vegetación natural (bosques, selvas y matorrales).

4.5.1 Agricultura

De acuerdo al INEGI (2006), la cantidad de superficie utilizada para la agricultura en Michoacán representa el 28.23 % del territorio estatal. Su variedad de suelos y climas ha permitido el desarrollo de zonas agrícolas de gran importancia como: la zona aguacatera en los alrededores de Uruapan, la zona de producción de cultivos básicos al norte del estado y la productora de cultivos tropicales como el mango, plátano y coco en la zona costera. Otros cultivos de importancia son el ajonjolí, limón, maíz, sorgo, frijol, trigo y otros (INEGI, 1985).

4.5.2 Vegetación transformada

Este tipo de vegetación incluye pastizales inducidos, pastizales cultivados y bosque cultivado. Según INEGI (2006), la cantidad de superficie empleada para esta vegetación representa el 6.28 % del territorio estatal. Las zonas más importantes están en la altiplanicie templada y en la cuenca del Río Tepalcatepec (INEGI, 1985). El pastizal inducido surge espontáneamente al ser eliminada la vegetación original, puede ser a consecuencia del abandono de un área agrícola, por sobrepastoreo o un incendio. Por su parte el pastizal y bosque cultivado son introducidos intencionalmente con fines diversos, como producir forraje, disponer de árboles frutales, etc.

4.5.3 Vegetación natural

Gran parte de la superficie del estado esta cubierta por extensas masas boscosas constituidas principalmente por pino, encino y oyamel, mismas que sustentan actividades forestales de importancia a nivel estatal y nacional (INEGI, 1985). Entre los tipos de vegetación natural destacan los siguientes:

- *Bosques*, que representan el 26.65 % de la superficie estatal (INEGI, 2006). Esta vegetación esta conformada por bosques de *oyamel*, *pino* y *encino*, los cuales se desarrollan en zonas con altitudes variadas, tal es así que los de oyamel crecen principalmente en áreas que van de 2,500 a 3,500 msnm; los de pino crecen entre los 1,500 y 3,500 msnm; y los de encino entre los 500 y 2,800 msnm.
- *Selvas*, que representan el 33.08 % de la superficie estatal (INEGI, 2006). Esta formada por *selva baja caducifolia* y *selva mediana subcaducifolia*. La primera se desarrolla en zonas con climas cálidos secos y semisecos, que van desde el nivel del mar hasta los 1,500 msnm y la segunda se desarrolla en zonas cálidas subhúmedas que van desde el nivel del mar hasta los 1,000 msnm, aproximadamente.
- *Matorrales*, que representan el 3.47 % de la superficie estatal (INEGI, 2006). Ocupan el altiplano michoacano, desarrollándose en suelos superficiales y pedregosos, en las laderas de los cerros y por abajo del nivel de los bosques de pino y encino. El aspecto del matorral es regularmente denso dominado por arbustos entre los 3 y 5 m de alto.

4.6 Necesidades de regionalización del estado de Michoacán

Como se mencionó en la Sección 4.4.2, el estado de Michoacán cuenta en su territorio con 22 acuíferos, de los cuales 9 se encuentran en condición de sobreexplotados. Adicional a los problemas de sobreexplotación se tienen los originados por la contaminación. Ambas situaciones han puesto en riesgo la disponibilidad de agua subterránea, la cual en el 2005 fue calificada por Juárez (2005) como una de las fuentes sujetas a un gran desperdicio. Varias acciones han sido tomadas para evitar los problemas de sobreexplotación y contaminación, entre las cuales destacan la veda de los acuíferos para su extracción a fin de que se recuperen los mantos acuíferos y, en términos de calidad del agua, se contempla el saneamiento de las aguas residuales para evitar que el agua contaminada alcance los acuíferos (Rueda, 2007). Si bien las políticas públicas pueden ser las adecuadas, la población no se ve obligada al cuidado de los recursos naturales, situación que pone en un mayor riesgo al recurso al demandarse y desperdiciarse una gran cantidad de agua.

Dada la sobreexplotación y contaminación de los mantos acuíferos, surgió el interés por parte del Gobierno de Michoacán (2005) en desarrollar modelos de manejo que permitieran la utilización, protección y el cuidado de los principales acuíferos del estado. En particular, se señaló como prioritario el estudio de los 9 acuíferos con problemas de sobreexplotación, los cuales se encuentran al oeste del estado colindando con el estado de Jalisco y un acuífero (Hidalgo-Tuxpan) al este colindando con el Estado de México, como se muestra en la *Figura 4.6*.

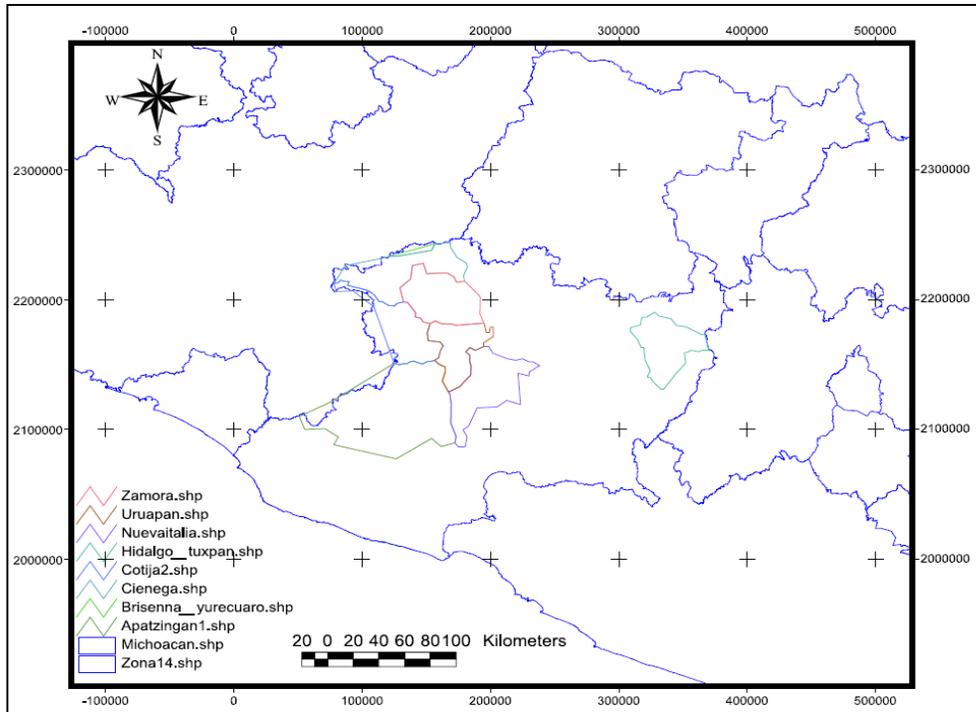


Figura 4.6. Ubicación de los acuíferos sobreexplotados en Michoacán y sus estados vecinos.

La *Tabla 4.1* presenta los acuíferos con problemas de sobreexplotación reportados por la Comisión Nacional del Agua (CONAGUA, 2005).

Tabla 4.1. Acuíferos con problemas de sobreexplotación.

Clave	Acuífero	Recarga anual	Descarga Natural	Volumen de extracción	Disponibilidad Natural	Disponibilidad media anual	Déficit
1605	Pastor Ortiz - La Piedad	28.7	0.11	127.43	24.7	0	-98.84
1607	Ciénega de Chapala	14.4	0	72.01	18.7	0	-57.61
1608	Zamora	308.5	180.18	77.59	61.4	50.73	0
1609	Briseñas - Yurecuaro	121	2	132.97	79.8	0	-13.96
1610	Ciudad Hidalgo - Tuxpan	38	13.66	66.13	24.4	0	-41.79
1614	Uruapan	97.3	29.5	17.35	12.8	50.45	0
1616	Nueva Italia	99.2	0.28	5.03	44.2	93.89	0
1620	Apatzingán	494.4	94.61	221.84	229.8	177.95	0
1622	Cotija	134.8	92.73	28.09	27	13.98	0

Cifras en millones de metros cúbicos anuales

Fuente: CONAGUA, 2005.

En la *Tabla 4.1* se observa que los acuíferos donde la extracción del agua es mucho mayor (>100 %) a su disponibilidad natural son:

- *Pastor Ortiz-La Piedad*, con un déficit de 98.84 millones de metros cúbicos anuales ($\text{Mm}^3/\text{año}$), esto es, el volumen de extracción es del orden de 516 % con respecto a la recarga natural.
- *Ciénega de Chapala*, presenta una recarga natural de $14.4 \text{ Mm}^3/\text{año}$ mientras que el volumen de extracción es de $72.01 \text{ Mm}^3/\text{año}$, ocasionado un déficit de 57.61 Mm^3 .
- *Ciudad Hidalgo-Tuxpan*, tiene un balance anual con un déficit de $41.79 \text{ Mm}^3/\text{año}$, esto debido a que su disponibilidad es de 24.4 Mm^3 mientras que su extracción es de 66.13 Mm^3 .
- *Briseñas–Yurecuaro*, las extracciones casi duplican (167 %) la demanda natural del recurso, con un déficit de $13.96 \text{ Mm}^3/\text{año}$

De los cinco acuíferos restantes, tres de ellos presentan diversos grados de sobreexplotación: Zamora ($16.19 \text{ Mm}^3/\text{año}$), Uruapan ($4.55 \text{ Mm}^3/\text{año}$) y Cotija ($1.09 \text{ Mm}^3/\text{año}$). Aunque es importante mencionar que en todos ellos, el volumen extraído no sobrepasa por mucho a la disponibilidad, al ser estas regiones con un aumento poblacional y económico alto los coloca en acuíferos muy vulnerables. Finalmente, los acuíferos de Nueva Italia y Apatzingán, lejos de presentar una sobreexplotación muestran una disponibilidad mayor que el volumen extraído (principalmente el acuífero de Nueva Italia); sin embargo, al igual que los acuíferos de Zamora, Uruapan y Cotija, estas regiones presentan un aumento poblacional e industrial considerable donde el agua representa la clave para su desarrollo, colocando a estos acuíferos como altamente vulnerables.

Debido a que los acuíferos con problemas de sobreexplotación están localizados en una región extensa al oeste del estado y también al este del mismo, resulta difícil contar con datos meteorológicos que permitan obtener un balance de agua real para cada acuífero. Como se mostró en el Capítulo III, las estaciones climatológicas convencionales (CLICOM) se encuentran dispersas en el Norte del Estado dejando

la parte centro y sur con poca cobertura. De igual forma las estaciones como Observatorios y Estaciones Meteorológicas Automáticas (EMAs) se encuentran principalmente en el Norte. Por tal motivo, el disponer de regiones homogéneas permitirá contar con datos hidrológicos necesarios para establecer un balance de agua confiable que ayude en la toma de decisiones para proteger y aprovechar de manera racional el recurso hídrico.

Capítulo V.

Metodología

En el capítulo IV se dieron a conocer las principales características de la zona de estudio, la cual sobresale por su accidentado relieve formado por dos grandes provincias fisiográficas, Sierra Madre del Sur y Eje Neovolcánico, así como por su variedad de suelos y climas.

En la Sección 2.5 se describió la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*), la cual se siguió mediante la aplicación del software Clementine para el análisis de la calidad de los datos. En este capítulo de nuevo se recurrirá a la metodología CRISP-DM para lograr la regionalización de zonas climatológicamente similares en el estado de Michoacán. Las fases del CRISP-DM comprenden: el *análisis del problema y de los datos* (ambos resueltos en los capítulos III y IV); la *preparación de datos* en donde se integran, transforman y normalizan los datos disponibles con la finalidad de preparar la vista minable a ser utilizada por las técnicas de minería de datos; la *construcción del modelo* en la cual mediante las técnicas de agrupamiento se clasifican las estaciones significativas en regiones climatológicamente homogéneas; y la *validación de resultados* donde mediante indicadores de validación como Ancho de la Silueta, Dunn y Davies Bouldin se validan las regiones climáticas establecidas para el estado de Michoacán.

5.1 Análisis del problema

Como se mencionó en la Sección 4.6, en Michoacán, 9 de los 22 acuíferos identificados se encuentran en condición de sobreexplotados. Aunado a los problemas de sobreexplotación, también existen los originados por la contaminación. Ante estas situaciones, surgió en el Gobierno del estado de Michoacán el interés en desarrollar modelos de manejo que permitan la utilización, protección y el cuidado de los principales mantos acuíferos del estado; en particular, se dio mayor importancia a los 9 acuíferos con problemas de sobreexplotación.

El manejo y control de los recursos hídricos disponibles en una región extensa como es el caso del estado de Michoacán (se consideró todo el estado por la distribución de los acuíferos prioritarios) resulta complicado por la dispersión de estaciones con series completas y continuas; la mayoría de las estaciones muestran series incompletas y no continuas. Debido a la importancia de las variables meteorológicas para definir la disponibilidad de agua de una región, una alternativa es la identificación de zonas o regiones climatológicamente homogéneas. Disponer de regiones homogéneas permitirá contar con datos necesarios para establecer un balance de agua confiable que ayude en la toma de decisiones para proteger y aprovechar de manera racional el recurso hídrico.

5.2 Análisis de los datos

El proceso de *análisis de datos* se llevo a cabo en el Capítulo III, esto con el propósito de determinar la calidad de la serie de datos a emplearse en el proceso de regionalización climática en Michoacán. La principal consideración fue determinar cuántas y cuáles son las estaciones climatológicas apropiadas para el estudio. Los datos analizados fueron de las estaciones tradicionales (CLICOM) y observatorios de Zamora y Morelia; series adquiridas a través del Servicio Meteorológico Nacional (SMN).

De las estaciones tradicionales analizadas (239), 52 fueron las consideradas significativas para el proceso de regionalización (ver *Tabla 3.5*). En el caso de los observatorios, sólo el observatorio de Morelia se consideró apropiado para el análisis, dado que el de Zamora no alcanzó el mínimo porcentaje establecido (85%) para registros válidos (Sección 3.3.4). En general, el conjunto de estaciones consideradas significativas disponen de datos continuos para un periodo de 20 años, esto es, series completas con una escasa o nula cantidad de registros faltantes.

Adicionalmente al número de estaciones apropiadas para el estudio, se analizó la frecuencia de las variables medidas y su comportamiento en el periodo de análisis. De esta manera se determinó para las estaciones CLICOM (Sección 3.3.2) que las variables de influencia serían: precipitación (P), temperatura observada (Tob), temperatura máxima (Tmx) y temperatura mínima (Tmn), en parte debido a que las variables de evaporación, cobertura de nubosidad, y velocidad y dirección del viento mostraron un alto porcentaje de registros nulos o incompletos.

En cuanto al observatorio de Morelia, se analizaron las temperaturas de bulbo seco, húmedo, humedad relativa y precipitación para establecer la calidad de los datos, pero los datos serán utilizados una vez que se obtenga la validación de la regionalización, debido a que la serie de datos cuenta con registros horarios de una mayor cantidad de variables como: insolación, presión en la estación y velocidad y dirección del viento, que serán ocupadas para procesos posteriores como el cálculo de la evapotranspiración (ET).

5.3 Preparación de datos

En esta fase se integraron, transformaron y normalizaron las series de datos de las 52 estaciones consideradas significativas para la regionalización. Esto con la finalidad de preparar la vista minable (archivo de datos) a ser utilizada por las técnicas de minería de datos.

5.3.1 Integración de datos

Para trabajar con los datos de las 52 estaciones obtenidas del CLICOM se deben considerar primeramente como un todo, para lo cual se integró la información de las mismas en un archivo único conformado por los campos que se presentan en la *Tabla 5.1*.

Tabla 5.1. Campos disponibles para la preparación de datos.

Campo	Descripción	Tipo
Identificador	Identifica a la estación climatológica	Carácter
Año	Registra el año en que se midió la variable climatológica	Entero
Mes	Registra el mes en que se midió la variable climatológica	Entero
Día	Registra el día en que se midió la variable climatológica	Entero
Precipitación	Registra la cantidad de precipitación suscitada en el día	Real
Temperatura Observada	Registra la temperatura ambiente medido en el día	Real
Temperatura Máxima	Registra la cantidad máxima de calor medido en el día	Real
Temperatura Mínima	Registra la cantidad mínima de calor medido en el día	Real

Como se observa, tres fueron los tipos de datos utilizados en este trabajo: carácter, enteros y reales, los cuales para la lectura de datos en el *Clementine* fueron instanciados (declarados) como tipos de datos de entrada.

Es importante señalar que cuando se tiene un gran número de variables en un proyecto de minería de datos, lo recomendable es hacer una evaluación de las variables de estudio para determinar aquellas más relevantes que pasarán a formar parte de la vista minable. Esta evaluación se hace mediante un análisis correlacional de datos, el cual consiste en crear una matriz de correlaciones que permite identificar la relación entre pares de variables (Hernández *et al.*, 2004). Este no fue el caso para la vista minable creada en este trabajo, puesto que el número de variables era de cuatro (temperatura -observada, máxima y mínima- y precipitación).

5.3.2 Transformación de datos

En la Sección 2.7 se determinó que una de las tareas de minería de datos que permite construir modelos de regionalización aplicados a hidrología es el *agrupamiento (clustering)*. De esta manera se puede dividir una población heterogénea en un número de subgrupos homogéneos de acuerdo a la similitud de sus registros.

Debido a que en la actualidad existen diversas técnicas de *agrupamiento*, se seleccionaron aquellas cuyas características permitieran trabajar con series de datos no completas (<100%), así los tipos de agrupamiento empleados fueron el Bietápico (agrupamiento jerárquico), K-medias (agrupamiento particional) y las redes de Kohonen. El empleo de tres técnicas permitirá comparar datos y establecer cual de ellas es la que mejor agrupa las estaciones en zonas homogéneas.

Debido a que las técnicas de *agrupamiento* requieren de instancias únicas en el conjunto de datos a ser minados, la vista minable obtenida en la integración de los datos fue transformada. La principal razón para esta transformación se debe a que se tenían en promedio más de 7,300 registros con el mismo identificador por estación, esto es, el total aproximado de registros de una estación en un periodo de 20 años. La transformación de los datos consistió en establecer un único identificador por estación mediante promedios mensuales de las tres variables de temperatura –observada, máxima, y mínima- y la suma mensual de precipitación. Estos promedios y sumas fueron integrados en campos mensuales de enero a diciembre, obteniendo un total de 48 campos producto del desglose mensual de cada variable, es decir 12 campos por variable.

5.3.3 Normalización de datos

A partir de la vista minable transformada se generaron dos configuraciones o nuevas vistas minables para lograr la regionalización:

- (a) Vista establecida por los 48 campos producto de los promedios mensuales de las tres variables de temperatura -observada, máxima y mínima- y la suma mensual de precipitación, y
- (b) Vista formada por los 48 campos antes mencionados y la variable *Altitud*. Se adicionó esta última variable con la finalidad de mejorar la calidad de la nueva vista minable.

Debido a que los datos de las variables en ambas vistas minables presentan valores de tamaños diferentes, esto es, cantidades que van de 1 a 4 dígitos enteros, fue necesario normalizar los datos de cada una de las configuraciones a una escala numérica entre 0 y 1 con base en la siguiente fórmula:

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

donde: x'_i es el valor normalizado, x_i es el valor original, x_{\min} es el valor mínimo de la variable y x_{\max} es el valor máximo de la variable.

5.4 Construcción del modelo de aplicación

Como se mencionó en la Sección 5.3.2, las técnicas de *agrupamiento* seleccionadas para la regionalización fueron: Bietápico, K-medias y redes de Kohonen (Mapa Autoorganizativo).

5.4.1 *Bietápico (TwoStep)*

El algoritmo Bietápico es un tipo de agrupamiento jerárquico desarrollado por Zhang, Ramakrishnan y Livny en 1996 (Clementine, 2006a). Este algoritmo se basa en una metodología denominada BIRCH (*Balanced Iterative Reducing and Clustering using Hierarchies*), cuyas características principales son las siguientes (Fonfría, 2005):

- Permite el análisis de valores atípicos (outliers).
- Permite tratar variables continuas y categóricas, de manera que la distancia para las variables continuas se calculan mediante la media y la varianza, mientras que para las categóricas se calculan las frecuencias.

El algoritmo de agrupamiento *Bietápico* implica dos etapas (Clementine, 2006a): el pre-agrupamiento y el agrupamiento:

- Pre-agrupamiento de los registros, siendo definido cada uno de ellos en pequeños subgrupos. Este procedimiento es implementado en forma de un árbol, también conocido como dendograma.
- Los *subgrupos* resultantes de la primera etapa (pre-agrupamiento) son agrupados comparando sus distancias con un umbral específico, de manera que si la distancia es mayor que el umbral, los dos grupos se fusionan. Completada esta etapa, se obtiene el número de grupos.

La medida de distancia entre dos grupos i y j se define de la siguiente manera (Clementine, 2006a):

$$d(i, j) = \xi_i + \xi_j - \xi_{(i,j)} \quad (2)$$

donde:

$$\xi_v = -N_v \left(\sum_{k=1}^{K^A} \frac{1}{2} \right) \log(\sigma_k^2 + \sigma_{vk}^2) + \sum_{k=1}^{K^B} \hat{E}_{vk} \quad (3)$$

y

$$\hat{E}_{vk} = - \sum_{l=1}^{L_k} \frac{N_{vkl}}{N_v} \log \frac{N_{vkl}}{N_v} \quad (4)$$

donde:

K^A	número de variables de tipo continuo
K^B	número de variables de tipo categórico o simbólico
L_k	número de categorías para la k -ésima variable categórica
N_v	número de registros en el <i>grupo</i> v
N_{vkl}	número de registros en el <i>grupo</i> v , cuya k -ésima variable categórica forma parte de la l -ésima categoría
σ_k^2	varianza estimada de la k -ésima variable continua
σ_{vk}^2	varianza estimada de la k -ésima variable continua para todos los registros en el <i>grupo</i> v
$\langle i, j \rangle$	índice que representa la unión de los <i>grupos</i> i y j .

En el método Bietápico se tiene flexibilidad para especificar el número óptimo de grupos o bien se deja que sea el propio algoritmo quien determine la cantidad de grupos. Para determinar el número óptimo de grupos, el algoritmo utiliza el Criterio de Información Bayesiana (BIC), el cual se calcula como (Clementine, 2006a):

$$BIC(J) = -2 \sum_{j=1}^J \xi_j + m_j \log(N) \quad (5)$$

donde:

$$m_j = J \left(2K^A + \sum_{K=1}^{K^B} (L_K - 1) \right) \quad (6)$$

Por lo general, este tipo de agrupamiento se utiliza para encontrar el número óptimo de grupos, es decir, no existe la necesidad de proporcionarle el número de

grupos que se desea obtener, sino que con base en los registros de la vista minable es el propio algoritmo quien determina la cantidad de grupos.

5.4.2 *K-Medias*

El K-medias es una de las técnicas de agrupamiento particional más usadas en la minería de datos propuesta por J. B. MacQueen en 1967 (Berry y Linoff, 2004). Una característica de este tipo de agrupamiento es el establecer *a priori* el número de grupos de entrada (K), por lo que en la práctica es necesario repetir la prueba considerando diferentes números de grupos, hasta obtener la solución que mejor se adapte al objetivo del problema (Sección 2.6).

El procedimiento que sigue la técnica de *K-medias* es el siguiente (Jain *et al.*, 1999; Larose, 2005):

1. Selección aleatoria de K puntos o elementos, haciendo que éstos representen los “centros” de los grupos.
2. Asignación de cada uno de los elementos restantes al centro más cercano. Esto es, la distancia mínima entre el elemento y el centro. Por lo general, la medida de distancia utilizada es la euclidiana.
3. Una vez que todos los elementos han sido asignados, se recalculan los *K* centros.
4. Se repiten los pasos 2 y 3 hasta que los centros ya no se modifiquen.

Para asignar los registros a los grupos cuyo centro es el más cercano, se utiliza la distancia euclidiana cuadrática definida como (Clementine, 2006a):

$$d_{ij} = \| X_i - C_j \|^2 = \sum_{q=1}^Q (x_{qi} - c_{qj})^2 \quad (7)$$

donde:

- X_i vector de los campos de entrada para el registro i
- C_j centro del grupo para la región j
- Q número de campos de entrada
- x_{qi} valor del q -ésimo campo de entrada para el i -ésimo registro
- c_{qj} valor del q -ésimo campo de entrada para el j -ésimo registro

Para actualizar el valor de los centros en los grupos, éstos se calculan como el vector medio de los registros establecidos en dicho grupo, $C_j = \bar{X}_j$, donde los campos del vector medio \bar{X}_j son calculados de acuerdo con la siguiente ecuación:

$$\bar{x}_{qj} = \frac{\sum_{i=1}^{n_j} x_{qi}(j)}{n_j} \quad (8)$$

Donde n_j es el número de registros en el grupo j , y $x_{qi}(j)$ es el q -ésimo valor para el registro i que es asignado al grupo j .

Mediante el método de K-medias se espera obtener resultados que revelen patrones del conjunto de datos de entrada, es decir, que mediante la vista minable se formen regiones homogéneas con características climáticas similares, donde las estaciones que se agrupen en una misma región tiendan a ser similares entre sí.

5.4.3 Redes de Kohonen (*Mapa autoorganizativo*)

Las redes de Kohonen fueron desarrolladas en 1982 por Teuvo Kohonen (Larose, 2005). Este tipo de redes se utiliza para agrupar un conjunto de datos en distintas regiones, especialmente cuando no se tiene información de los mismos.

Esta técnica también es conocido como *mapa autoorganizativo* (SOM, *Self Organizing Map*) (Hernández *et al.*, 2004), donde sus unidades básicas son las neuronas o nodos que se organizan en dos capas (Figura 5.1): una *capa de entrada*

en la que se introducen los registros, y una *capa de salida*, también denominada mapa de resultados, en la que cada nodo representa a un grupo con registros similares (Clementine, 2006b).

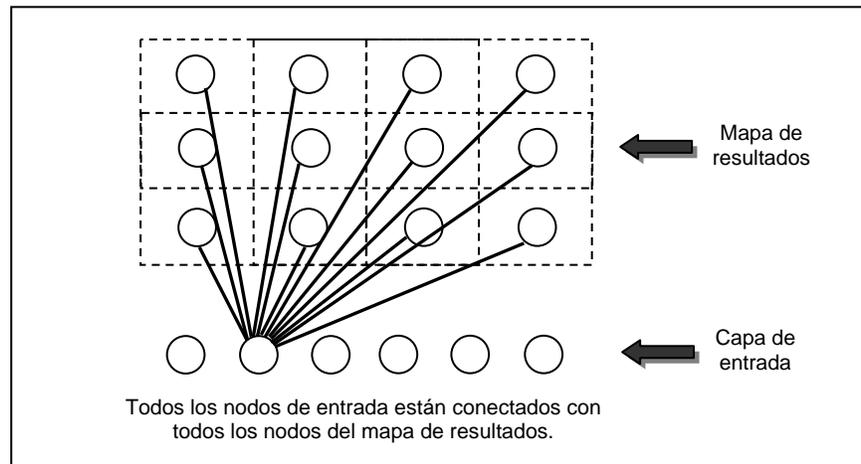


Figura 5.1. Estructura de una red de Kohonen (Mapa Autoorganizativo).
Fuente: Clementine, 2006b.

Los nodos de entrada (Figura 5.1) son conectados a todos los nodos de salida (mapa de resultados) y estas conexiones tienen ponderaciones asociadas a ellos. Durante el entrenamiento, cada nodo compite con los demás para *ganar* cada uno de los registros. A medida que avanza el entrenamiento las ponderaciones en los nodos se ajustan para formar un mapa bidimensional de grupos, de ahí el nombre mapa autoorganizativo (Clementine, 2006b).

El algoritmo de aprendizaje de las redes de Kohonen se resume de la siguiente manera (Hernández *et al.*, 2004):

ALGORITMO Kohonen (E : datos, d : función, m : entero)
 Situar los m grupos en el espacio.
 Inicializar η (es la tasa de aprendizaje)
REPETIR
 Extraer una instancia e de E (se extraen todos los vectores cíclicamente hasta que converja).
 Actualizar η
 Calcular la distancia de e con los grupos usando la función

$$d(x_i, w_j)$$

Seleccionar el grupo w_k con menor distancia.

Aproximar el grupo w_k siguiendo las ecuaciones del aprendizaje.

MIENTRAS $\eta > \text{umbral}$

FIN ALGORITMO

Las distancias en la red de Kohonen se calculan como la *distancia euclidiana* entre el vector de entrada (registro) y el centro del grupo del nodo de salida (Clementine, 2006a); la ecuación matemática es la siguiente:

$$d_{ij} = \sqrt{\sum_k (x_{ik} - w_{jk})^2} \quad (9)$$

donde: x_{ik} es el valor del k -ésimo campo de entrada para el i -ésimo registro, y w_{jk} es el peso para el k -ésimo campo de entrada en el j -ésimo nodo de salida. Así, para activar un grupo (nodo de salida) se emplea la *distancia euclidiana* entre el peso del nodo de salida y el vector de entrada. Finalmente, el grupo con mínima activación es la unidad ganadora. A su vez, para la actualización de los pesos, la magnitud de cambio es determinada por la tasa de aprendizaje η (eta), la cual se define como (Clementine, 2006a):

$$\Delta W = \eta \cdot (W - I) \quad (10)$$

donde: W es el peso vectorial para el nodo de salida, I es el vector de entrada y η es la tasa de aprendizaje.

Por lo general, una red de Kohonen termina con unos pocos nodos que resumen muchos registros, conocidos como *nodos fuertes*. Asimismo, los nodos con poca o nula cantidad de registros son conocidos como *nodos débiles*. Debido a esto, para establecer el número adecuado de grupos es necesario revisar los registros

capturados en cada nodo del modelo generado, identificándose los nodos fuertes (Clementine, 2006b).

5.4.4 Integración del modelo

Para realizar la regionalización de zonas climatológicamente homogéneas, las técnicas Bietápico, K-medias y redes de Kohonen se integraron en un modelo construido en el software Clementine. La estructura final del modelo se presenta en la *Figura 5.2*, donde cada técnica fue configurada en el espacio de trabajo principal, denominado zona de rutas (Sección 2.5). De esta manera, la secuencia de las técnicas en el modelo (Figura 5.2) es la siguiente: (1) corresponde al Bietápico, (2) es el K-medias, y (3) es la red de Kohonen. Para el funcionamiento de éstas técnicas, cada una de ellas leen los archivos de datos (vistas minables) originados tras la *normalización de datos*.

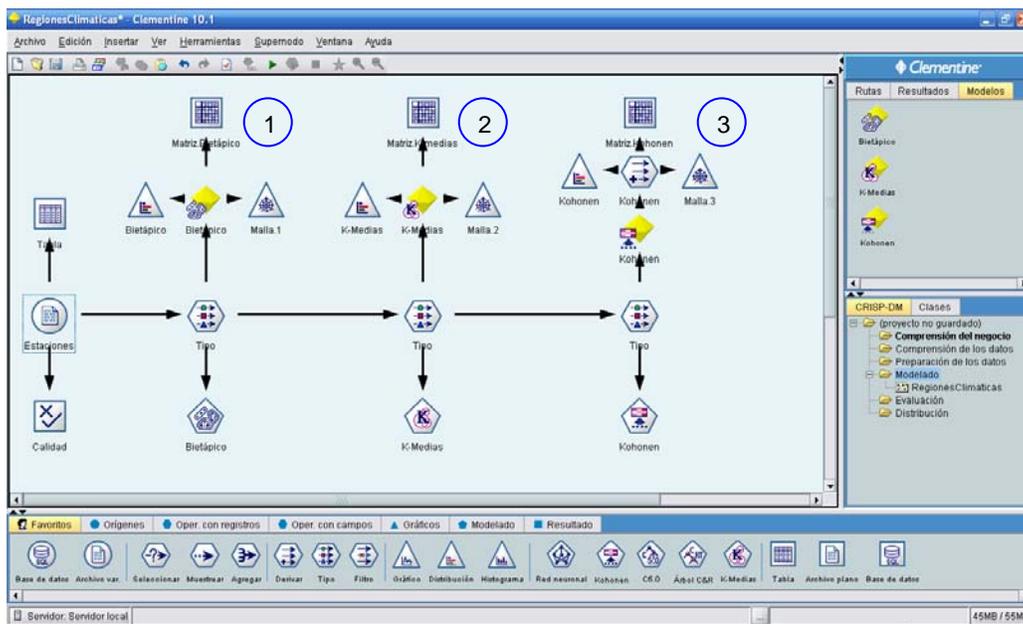


Figura 5.2. Modelo elaborado para obtener los grupos de zonas climatológicamente homogéneas en el estado de Michoacán.

Puesto que una de las características del *K-medias* es establecer *a priori* el número de grupos de entrada, la secuencia establecida en el modelo tuvo como finalidad determinar el número de grupos de entrada mediante un algoritmo jerárquico, que

en este caso es el Bietápico, para posteriormente el resultado de éste usarlo como entrada en las técnicas de K-medias y las redes de Kohonen.

Lo anterior se fundamenta en el método sugerido por Milligan (1980, 1985) y Hair *et al.* (1995), quienes consideran que para determinar el número óptimo de grupos de entrada (k), *primero* se debe utilizar algún algoritmo jerárquico, mediante el cual se obtiene directamente el número deseado de grupos, para luego el resultado de éste ser usado en algún algoritmo particional, como el K-medias (ver Sección 2.3.31).

5.4.5 Ejecución de las técnicas de minería de datos

El número óptimo de grupos obtenido al aplicar el algoritmo Bietápico a las dos vistas minables (*a*) y (*b*), mencionadas en la Sección 5.3.3, fue **4** (número de conglomerados) como se muestra en la *Figura 5.3*.

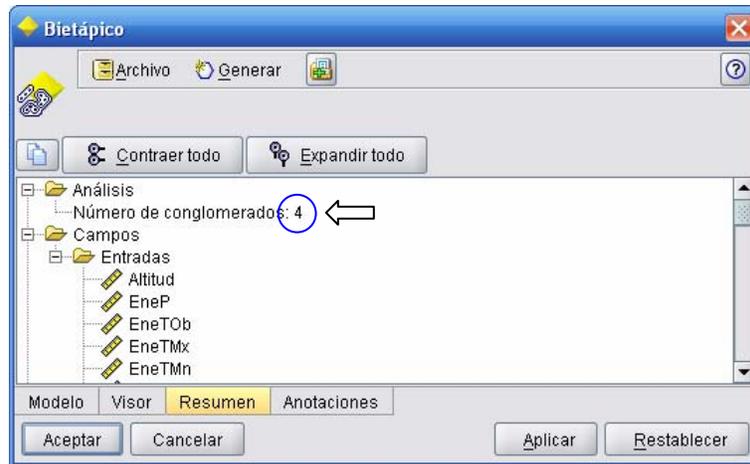


Figura 5.3. Determinación del número deseado de grupos mediante el algoritmo Bietápico.

Una vez establecido el número óptimo de grupos se procedió a ejecutar las tres técnicas de agrupamiento. Puesto que el Bietápico se empleó para establecer el número óptimo de grupos, se consideró como resultado de esta técnica el grupo de 4. En el caso del K-medias se utilizó como valor inicial de agrupación 4, lo mismo

que para las redes de Kohonen se definieron 4 neuronas en la capa de salida, donde cada neurona representa a un grupo.

Los resultados obtenidos por dichas técnicas son presentados en la *Tabla 5.2*, donde los números 1, 2, 3, y 4 representan la pertenencia de la estación CLICOM al grupo correspondiente, y (a) y (b) son las vistas minables utilizadas para la clasificación climática.

Tabla 5.2. Resultados obtenidos por las técnicas de agrupamiento Bietápico, K-medias y redes de Kohonen considerando 4 grupos como valor óptimo para K-medias y redes de Kohonen.

N°	Estación	Bietápico		K-medias		Redes de Kohonen	
		(a)	(b)	(a)	(b)	(a)	(b)
1	St16006	1	1	1	1	1	1
2	St16007	1	1	1	1	1	1
3	St16014	2	2	4	4	3	3
4	St16016	3	3	2	2	4	4
5	St16017	2	2	4	4	3	3
6	St16020	3	3	2	2	4	4
7	St16023	3	3	2	2	4	4
8	St16024	2	2	4	4	3	3
9	St16027	2	2	4	4	3	3
10	St16028	3	3	2	2	4	4
11	St16031	2	2	4	4	3	3
12	St16033	3	3	2	2	4	4
13	St16043	2	2	4	4	3	3
14	St16045	4	4	3	3	3	3
15	St16048	2	2	4	4	2	2
16	St16050	2	2	4	4	3	3
17	St16052	2	2	4	4	3	3
18	St16057	3	3	2	2	4	4
19	St16059	1	1	1	1	1	1
20	St16074	1	1	1	1	1	1
21	St16075	1	1	1	1	1	1
22	St16081	2	2	4	4	3	3
23	St16090	1	1	1	1	1	1
24	St16091	2	2	4	4	4	4
25	St16093	2	2	4	4	2	2
26	St16094	2	2	4	4	3	3
27	St16095	2	2	4	4	3	3
28	St16107	1	1	1	1	1	1
29	St16109	1	1	1	1	1	1
30	St16118	3	3	2	2	4	4
31	St16120	3	3	2	2	4	4
32	St16123	4	4	3	3	2	2
33	St16124	3	3	2	2	4	4
34	St16133	1	1	1	1	1	1
35	St16136	4	4	3	3	2	2
36	St16141	2	2	4	4	2	2
37	St16142	3	3	2	2	4	4
38	St16151	1	1	1	1	1	1
39	St16159	2	2	4	4	3	3

Tabla 5.2. Resultados obtenidos por las técnicas de agrupamiento Bietápico, K-medias y redes de Kohonen considerando 4 grupos como valor óptimo para K-medias y redes de Kohonen.

40	St16162	2	2	4	4	3	3
41	St16165	4	4	3	3	2	2
42	St16188	3	3	2	2	4	4
43	St16192	2	2	4	4	3	3
44	St16194	4	4	3	3	3	3
45	St16208	1	1	1	1	1	1
46	St16219	1	1	1	1	1	1
47	St16225	3	3	2	2	4	4
48	St16228	1	1	1	1	1	1
49	St16248	3	3	2	2	4	4
50	St16254	3	3	2	2	4	4
51	St16255	3	3	2	2	4	4
52	St16515	4	4	3	3	4	4

En la *Tabla 5.2* se observa que para cada una de las técnicas, las vistas minables (*a*) y (*b*) obtuvieron resultados similares, esto es, la misma estación fue asignada al mismo número de grupo en (*a*) y (*b*). Lo cual refleja que la variable *Altitud* no ha influido de manera importante en la formación de los grupos; siendo sólo significativas para la obtención de zonas homogéneas las variables de temperatura –observada, máxima y mínima- y precipitación.

Las estaciones clasificadas por el método Bietápico en el *grupo 1* son las mismas que fueron asignadas en el *grupo 1* de K-medias y de las redes de Kohonen. Para los grupos 2, 3 y 4, se observa que las estaciones clasificadas por K-medias en cada uno de estos grupos formados son similares a las agrupadas por el Bietápico; lo único que varía es el número identificador de los grupos ya que en K-medias es 2 y en Bietápico es 4.

De los grupos formados por las redes de Kohonen, sólo el *grupo 1* coincide con los resultados obtenidos por el Bietápico y el K-medias ya que los otros tres grupos fueron integrados de manera distinta a las anteriores. El *grupo 2* sólo tres estaciones (St16123, St16136 y St16165) coinciden con las agrupadas por el Bietápico y K-medias, las otras tres (St16045, St16194 y St16515) fueron clasificadas en los otros grupos. De manera similar sucede con las estaciones clasificadas en el *grupo 3*, donde de las 16 estaciones agrupadas 14 de ellas, excepto

por St16045 y St16194, coinciden con las 18 estaciones agrupadas tanto el Bietápico como en K-medias. Lo mismo sucede con las 17 estaciones clasificadas en el *grupo* 4, donde 15 de ellas, excepto por St16091 y St16515, coinciden con las agrupadas por las otras dos técnicas mencionadas.

Dado que los resultados obtenidos por las redes de Kohonen (aplicado al mismo conjunto de datos) son diferentes a los proporcionados por el K-medias y Bietápico, un aspecto técnico a destacar es que estas redes requieren de una mayor combinación de variables de entrada para su proceso de aprendizaje, esto es, requieren iniciar con un número amplio de variables y una alta tasa de aprendizaje, para luego ir reduciendo estos valores a medida que avanza el aprendizaje, haciendo que únicamente cambien los pesos de la neurona ganadora.

Lo anterior se refuerza con lo establecido por Balakrishnan *et al.* (1994), quienes compararon sobre las capacidades de agrupamiento de las redes de Kohonen y el K-medias y encontraron que una limitada cantidad de variables de entrada genera una mala clasificación mediante las redes de Kohonen a diferencia de K-medias. Por lo que para tener un buen desempeño empleando las redes de Kohonen se sugiere establecer una mayor cantidad de variables de entrada para el proceso de su aprendizaje.

5.4.6 *Análisis de resultados*

Con base en los resultados presentados en la *Tabla 5.2*, se graficaron los grupos de estaciones CLICOM formados por las técnicas empleadas, de esta manera se facilitó el análisis climatológico de las regiones obtenidas y se comparó con el mapa de climas del INEGI (2003) (Figura 5.4).

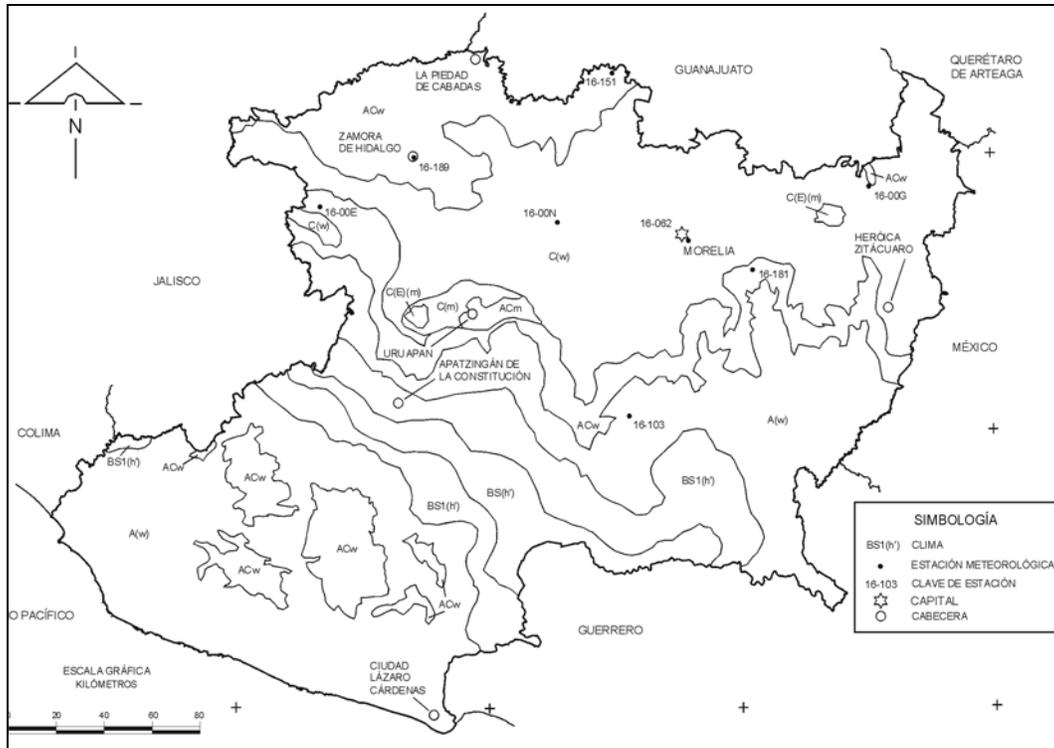


Figura 5.4. Carta de Climas del estado de Michoacán, 1:1'000,000.
Fuente: INEGI, 2003.

Es importante mencionar que se hizo la comparación, pese a que no necesariamente la distribución climática del INEGI debe ser similar a la obtenida en este trabajo. Esto debido a que el mapa del INEGI esta en función del indicador de Thornthwaite (1948), el cual se basa en valores anuales de precipitación y evaporación (ecuación 11), lo cual genera incertidumbre principalmente porque las variables a ese nivel de agregación no dan suficiente información relacionada de como ocurre la variación simultanea de la lluvia-temperatura (factores esenciales en la determinación de la evapotranspiración) en diversos puntos del estado.

$$I = \frac{h - E_p}{E_p} \quad (11)$$

donde I es el indicador, h es promedio anual de lluvia y E_p es el promedio anual de la evapotranspiración potencial, calculada con la ecuación de Turc (1954), la cual se define como:

$$\text{si } \left(\frac{P}{L}\right)^2 > 0.1; \quad ET_{real} = \frac{P}{\sqrt{0.9 + \left(\frac{P^2}{L^2}\right)}} \quad (12)$$

$$\text{si no;} \quad ET_{real} = P$$

$$L = 300 + 25 * Ta_{prom} + 0.05 * Ta_{prom}^3 \quad (13)$$

donde P es la precipitación, Ta es la temperatura ambiente y 0.9 es un factor que representa las condiciones de la región al igual que los valores de L . La ecuación de Turc se emplea porque la variable evaporación no siempre es medida de manera constante durante el tiempo en las estaciones climatológicas.

5.4.6.1 Resultados de la comparación

Los grupos de estaciones CLICOM formados por el algoritmo Bietápico, K-medias y redes de Kohonen se ilustran en las *Figuras 5.5, 5.6 y 5.7*, respectivamente.

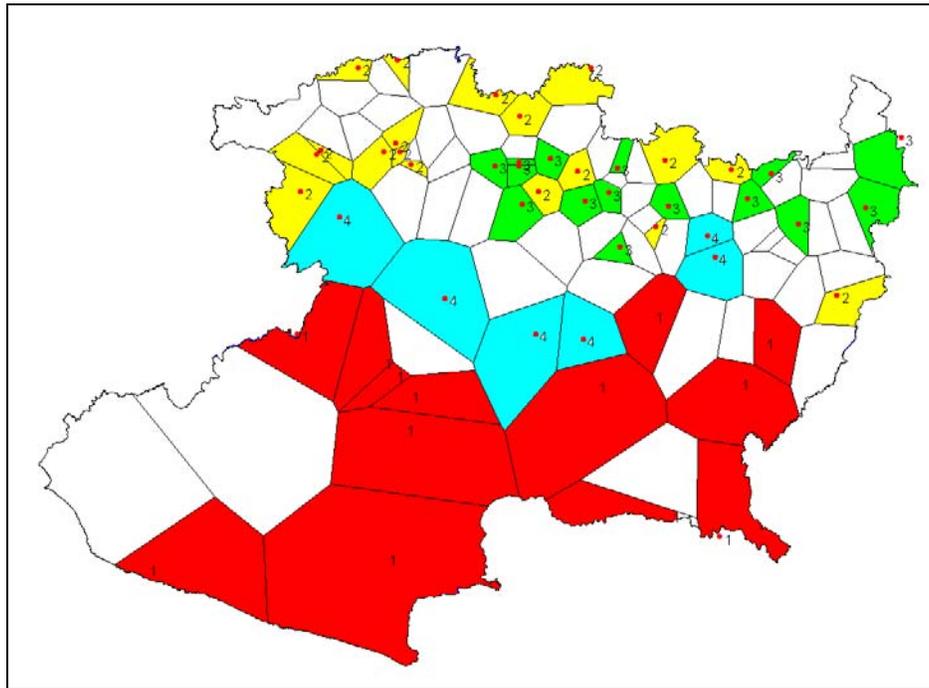


Figura 5.5. Regiones climáticas formadas por el algoritmo Bietápico.

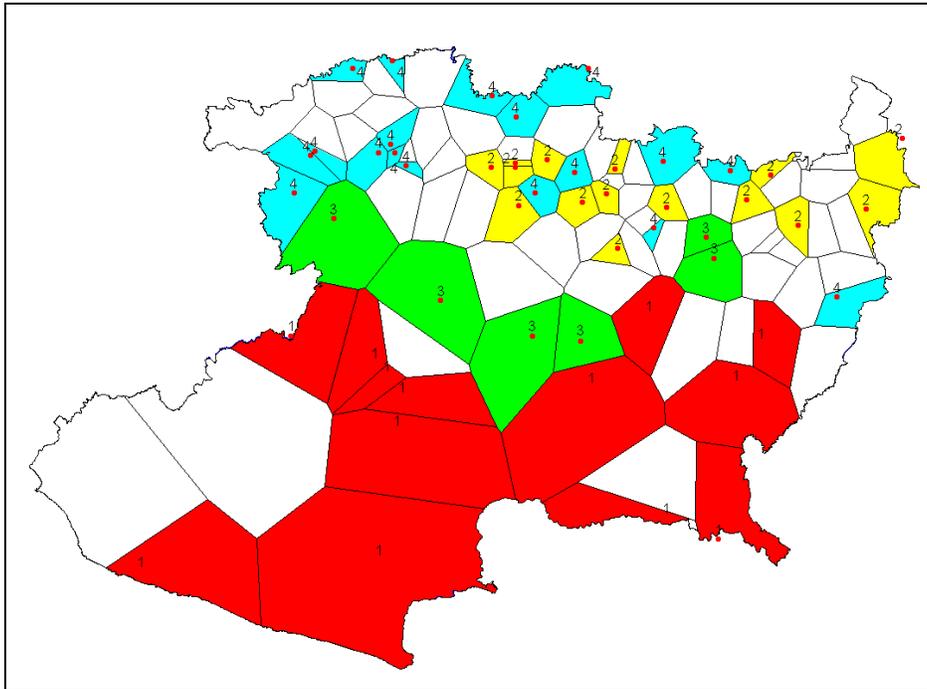


Figura 5.6. Regiones climáticas formadas por el K-medias.

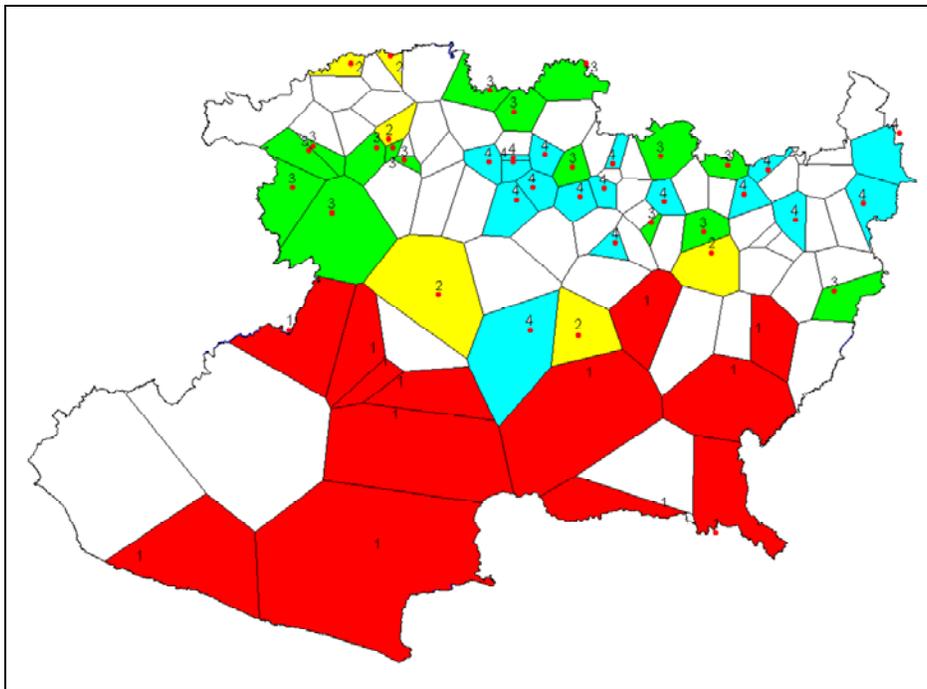


Figura 5.7. Regiones climáticas formadas por las redes de Kohonen.

Dado que los resultados obtenidos por el algoritmo Bietápico y el K-medias son similares, las *Figuras 5.5 y 5.6* son similares excepto por el número (color) asignado a cada grupo. Con respecto al mapa climático del INEGI se observó lo siguiente:

- a. La *región 1* (color rojo en la Figura 5.5 y 5.6) quedó integrada por 13 estaciones climáticas, las cuales se ubican en la zona Sur del estado (zona costera, depresión del Balsas y del Tepalcatepec) cubriendo por lo general la provincia fisiográfica Sierra Madre del Sur. Esta región es la más grande por la gran dispersión de las estaciones debido principalmente a la dificultad de colocar o mantener una estación en zonas de densos bosques y selvas. El tipo de clima que le corresponde a esta región es el *Seco y semiseco muy cálido y cálido y Cálido subhúmedo con lluvias en verano*, con un promedio de precipitación total anual de 500 a 989 mm y una temperatura media anual que oscila entre los 23 y 31 °C. De acuerdo al análisis de variabilidad y distribución de las variables significativas (Sección 3.4), se observó que la temperatura media anual de las estaciones que forman esta región varía entre los 17.6 y 23.4 °C y la precipitación media anual alcanza valores entre 630.3 y 1,296 mm; estos valores revelan una cierta aproximación a los obtenidos a través del mapa climático del INEGI.
- b. Las seis estaciones que integran la *región 2* (color azul en la Figura 5.5 y verde en la Figura 5.6) se encuentran ubicadas en la zona central de estado (este, centro y centro oriente), cubriendo sectores en las subprovincias fisiográficas: Neovolcánica Tarasca, Escarpa Limitrofe del Sur y Mil Cumbres, pertenecientes al Eje Neovolcánico; y Cordillera Costera del Sur y Depresión del Balsas, de la provincia Sierra Madre del Sur. El tipo de clima que presenta esta región es el *Semicálido subhúmedo con lluvias en verano y Templado subhúmedo y húmedo con abundantes lluvias en verano*, donde la temperatura media anual varía entre los 12.5 a 20.9 °C y la precipitación total anual alcanza los 1500 mm. Se observó que temperatura media anual de este grupo de estaciones (Sección 3.4) oscila entre los 11.4 y 17.4 °C, lo cual refleja

una considerable aproximación a la obtenida a través del mapa climático del INEGI (12.5 a 20.9 °C); caso similar ocurre con la precipitación media anual (1,004.2 y 1,293.7 mm) que son valores muy cercanos a lo establecido por el INEGI (<1,500 mm).

- c. Las 18 estaciones agrupadas en la *región 3* (color amarillo en la Figura 5.5 y azul en la Figura 5.6), se ubican en el noroeste, norte y este (una estación, St16192) del estado. Estas estaciones se localizan en parte de las subprovincias fisiográficas del Eje Neovolcánico: Chapala, Bajío Guanajuatense, Sierras y Bajíos Michoacanos y Mil Cumbres, y un pequeño sector de la subprovincia Cordillera Costera del Sur (Sierra Madre del Sur) al noroeste del estado. El tipo de clima de esta región es el *Semicálido subhúmedo y Templado subhúmedo con lluvias en verano*, donde la temperatura media anual alcanza valores entre los 12.5 y 21 °C y la precipitación total anual fluctúa entre 741 y 959 mm. Si bien la estación St16192 es la más alejada de la región, localizándose al este del estado, ésta posee valores de temperatura (14 °C) y precipitación (860.3 mm) muy similares al resto de estaciones que forman esta región. Por lo general, los valores de temperatura (12.4 y 16 °C) y precipitación media anual (652.2 y 888.2 mm) de este grupo de estaciones (Sección 3.4) son muy cercanos a los obtenidos a través del mapa climático del INEGI.
- d. Por último, la *región 4* (color verde en la Figura 5.5 y amarillo en la Figura 5.6) fue establecida por 15 estaciones que se ubican en la zona norte (centro superior) y noreste del estado, cubriendo por lo general las subprovincias fisiográficas del Eje Neovolcánico: Sierras y Bajíos Michoacanos, Neovolcánica Tarasca y Mil Cumbres. El clima en esta región es el *Templado subhúmedo con lluvias en verano y Semifrío húmedo con abundantes lluvias en verano*, donde la temperatura media anual oscila entre los 10.6 y 17.4 °C y la precipitación total anual alcanza valores entre los 741 y 1,534 mm. En esta región, de acuerdo a los valores establecidos en la Sección 3.4, las estaciones presentan una temperatura media anual que oscila entre los 8.9 y 12.5 °C y la

precipitación media anual alcanza valores entre 642.3 y 901.7 mm. Como se puede observar estos valores revelan una aproximación a los obtenidos a través del mapa climático del INEGI.

De lo anterior, desde el punto de vista de la formación de los grupos, se considera que la regionalización obtenida por ambas técnicas (Bietápico y K-medias), permitió dividir el estado de Michoacán en dos regiones claramente diferenciadas (grupos 1 y 2). Las otras dos regiones (grupos 3 y 4) muestran en la parte norte y este del estado estaciones como St16050, St16091, St16181 y St16192, las cuales comparten zonas climáticas en ambas regiones. Por lo que para mejorar estos resultados se sugiere como trabajo futuro emplear otro tipo de combinaciones de las variables atmosféricas disponibles para la zona de estudio. Por ejemplo, se podría organizar una nueva vista minable con:

1. Suma mensual de precipitación, media mensual de temperatura observada y media mensual de los valores máximos y mínimos de la temperatura máxima y mínima, respectivamente.
2. Promedios quincenales de las variables: precipitación, temperatura - observada, máxima y mínima-, esto con la finalidad de disponer de una mayor cantidad de campos a minar, es decir, tener 24 campos por variable, haciendo un total de 96 campos a minar.

En el caso de las redes de Kohonen, dado que sólo el *grupo 1* es similar al obtenido por el Bietápico y K-medias (Figura 5.7), se analizó la formación de los grupos 2, 3 y 4 en relación con los resultados obtenidos por el Bietápico y K-medias.

Como se mencionó, en el grupo 2 (color amarillo en la Figura 5.7) sólo el 50 % de las estaciones (tres) coinciden con las agrupadas por el Bietápico y K-medias, el otro 50 % fueron agrupadas en regiones distintas. Junto a las tres estaciones que coinciden, otras tres (St16048, St16093 y St16141) fueron agrupadas en el centro y noreste del estado, las cuales comparten zonas climáticas con la región 3 (color

verde en la Figura 5.7); esto refleja una agrupación no muy razonable por la falta de claridad en la división de las regiones.

Caso similar al anterior ocurre con las estaciones clasificadas en el *grupo 3* y *4* (color verde y azul en la Figura 5.7). En el *grupo 3* el 77 % de las estaciones coinciden con las agrupadas por el Bietápico y K-medias, mientras que en el *grupo 4* las estaciones coinciden en su totalidad pero existen otras dos (St16091 y St16515) que fueron clasificadas en este grupo; por lo que también no existe una clara división de las regiones formadas por esta técnica.

Por lo anterior, la falta de claridad en la formación y división de los grupos ha llevado a determinar que las regiones formadas por las redes de Kohonen no fueron las más adecuadas para la regionalización climática del estado de Michoacán. Por lo que se sugiere como trabajo posterior ampliar el número de combinaciones de las variables atmosféricas disponibles, esto con fin de mejorar el proceso de aprendizaje de las redes de Kohonen para obtener un mejor desempeño en la formación de los grupos. Una alternativa para ampliar el número de variables de entrada es organizar una nueva vista minable con promedios quincenales de las variables de temperatura -observada, máxima y mínima- y precipitación.

Finalmente, puesto que las técnicas de agrupamiento que obtuvieron una mejor clasificación de zonas o regiones climatológicamente similares en el estado de Michoacán fueron el Bietápico y K-medias (con iguales resultados). Se consideró para este trabajo como regiones con climas similares los establecidos por dichas técnicas, quedando así definido como mapa de regionalización climatológica el estado la *Figura 5.6*. Esto debido a que los índices de validación sólo se aplicaran al K-medias, puesto que en la literatura no se encontró indicadores de validación diseñados para el método Bietápico, ni para las redes de Kohonen.

5.5 Validación

Para validar los resultados obtenidos se usaron *índices de validación* que ayudan a evaluar si la formación de los grupos fue la idónea. Estos índices fueron el Ancho de la Silueta, Dunn y Davies-Bouldin, los cuales se basan en estimaciones que identifican cuan compactos o separados se encuentran los grupos formados especialmente por técnicas particionales, como la de K-medias (Chou *et al.*, 2003; Brock *et al.*, 2007). A continuación se abordan brevemente los índices de validación utilizados en este trabajo:

1. **Ancho de la Silueta** (Rousseeuw, 1987). Este índice se usa para estimar el número deseado de grupos, así como para evaluar la asignación de los registros en los grupos establecidos (Brock *et al.*, 2007). Para estimar el número deseado de grupos se toma la partición (K) con el promedio más alto; mientras que para evaluar la asignación de los registros se calcula $s(i)$ para el i -ésimo registro definido como (Bolshakova y Azuje, 2003):

$$s(i) = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (14)$$

Donde a_i es la distancia promedio entre el i -ésimo registro y todos los otros que están en el mismo grupo, y b_i es la distancia promedio mínima entre el i -ésimo registro y los registros que están en otros grupos.

El valor $s(i)$ se ubica en el intervalo -1 y 1: si $s(i)$ es próximo a 1 se puede inferir que el i -ésimo registro fue asignado a un grupo apropiado, si $s(i)$ se aproxima a cero indica que el i -ésimo registro podría ser asignado a otro grupo más cercano, y si $s(i)$ es próximo a -1 se puede inferir que el i -ésimo registro fue mal agrupado.

Para este trabajo el índice Ancho de la Silueta se utilizó para validar el número deseado de grupos en la regionalización de zonas con climas similares en el estado de Michoacán.

2. **Índice de Dunn** (Dunn, 1974). Este índice indica si los grupos formados son bien compactos y separados. Para estimar el número deseado de grupo, este indicador maximiza la distancia intergrupala y minimiza la distancia intragrupal de los mismos (Saitta *et al.*, 2007). Dada una partición de grupos, donde C_i representa el i -ésimo grupo de la partición, el índice de Dunn se define como (Kovács *et al.*, 2005):

$$D = \min_{1 \leq i \leq n} \left\{ \min_{1 \leq j \leq n, j \neq i} \left(\frac{d(C_i, C_j)}{\max_{1 \leq k \leq n} \text{diam}(C_k)} \right) \right\} \quad (15)$$

donde $d(C_i, C_j)$ es la distancia entre los grupos C_i y C_j (distancia intergrupala), y $\text{diam}(C_k)$ representa la distancia o diámetro *intragrupala* del grupo C_k .

Finalmente, el número óptimo de grupos es aquel que maximiza D .

3. **Índice de Davies–Bouldin** (Davies y Bouldin, 1979). Mediante el índice de Davies-Bouldin se estima el número deseado de grupos a través de una medida de dispersión y desemejanza de los grupos establecidos (Halkidi *et al.*, 2005). Al igual que el índice de Dunn, este índice revela si los grupos formados son compactos y están bien separados (Bolshakova y Azuje, 2003). El índice de Davies-Bouldin (DB) se define como (Boutin y Hascoët, 2004):

$$DB = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} \left\{ \frac{\text{diam}(C_i) + \text{diam}(C_j)}{d(C_i, C_j)} \right\} \quad (16)$$

donde: $diam(C_i)$ y $diam(C_j)$ representan la distancia *intragrupal* de los grupos C_i y C_j , respectivamente, mientras que $d(C_i, C_j)$ es la distancia entre los grupos C_i y C_j (*distancia intergrupala*).

La configuración que minimiza DB se toma como el número deseado de grupos. De acuerdo a Saitta *et al.* (2007), en la actualidad este índice es considerado como uno de los mejores índices de validación por su mejor aproximación del número deseado de grupos.

En este trabajo, los índices de validación descritos sólo pudieron ser aplicados al K-medias. Esto debido a que las distancias en el método Bietápico se calculan mediante la media y la varianza (Sección 5.4.2), por lo que no fue posible disponer de una matriz de distancias euclidianas como la que genera el K-medias. Otro aspecto ha señalar es que en la literatura no se encontraron indicadores de validación para este tipo de técnica en particular.

En el caso de las redes de Kohonen, tampoco en la literatura se contó con *índices de validación* diseñados para este método, por lo que una alternativa para establecer el número adecuado de grupos consiste en revisar los registros capturados en cada nodo del modelo generado, identificando los *nodos fuertes* que resumen la mayor cantidad de registros agrupados. Esta alternativa permitió confirmar que los 4 nodos establecidos por esta técnica son *nodos fuertes* por tener agrupados un determinado número de estaciones.

Al aplicar los indicadores de validación al K-medias, se consideró adecuado analizar el comportamiento de la técnica modificando el número de grupos de entrada, para lo cual con base en las vistas minables utilizadas (a) y (b) para la regionalización se tuvieron siete agrupamientos con particiones diferentes, esto es, $K = 2, 3, \dots, \text{y } 8$. En el *Anexo F* se muestran las gráficas con los grupos obtenidos.

Los resultados obtenidos al aplicar los índices de validación se muestran en la *Tabla 5.3*.

Tabla 5.3. Determinación del número deseado de grupos para K-medias mediante los índices de validación Silueta, Dunn y Davies-Bouldin.

Índice de Validación	Número de Grupos						
	K=2	K=3	K=4	K=5	K=6	K=7	K=8
Silueta	0.57	0.32	0.45	0.32	0.29	0.24	0.25
Dunn	0.44	0.22	0.22	0.27	0.27	0.31	0.21
Davies-Bouldin	0.63	1.12	1.01	1.07	1.08	1.09	1.33

En la *Tabla 5.3* se observa que el número óptimo de grupos sugerido por los índices de validación, *Silueta* (valor más alto = 0.57), *Dunn* (valor más alto = 0.44) y *Davies-Bouldin* (valor más pequeño = 0.63), es de 2. Pero debido al accidentado relieve del Estado, la presencia de las provincias fisiográficas Sierra Madre del Sur y Eje Neovolcánico, y la variedad de sus climas; no es razonable trabajar con sólo dos grupos. El siguiente agrupamiento que cumple con al menos dos índices de validación -*Silueta* (0.45) y *Davies-Bouldin* (1.01)- es el de 4 grupos. Esta selección se basa principalmente en el índice de *Davies-Bouldin*, el cual es uno de los índices más recomendables por su mejor aproximación de acuerdo con Saitta *et al.* (2007).

Así, el número deseado de grupos óptimo es K=4, validándose de esta manera las cuatro regiones climáticas establecidas para el estado de Michoacán.

Capítulo VI.

Conclusiones, recomendaciones y trabajo futuro

En este capítulo se presentan las conclusiones del trabajo realizado y se mencionan recomendaciones, así como las futuras líneas de investigación que se pueden establecer con base en los resultados obtenidos en este trabajo de tesis.

6.1 Conclusiones

En este trabajo para obtener las regiones climatológicamente similares en el estado de Michoacán se utilizaron las bases de datos del CLICOM y de los Observatorios. El CLICOM tiene un total de 239 estaciones de las cuales casi tres cuartas partes contienen datos nulos por falta de registro. Los observatorios analizados fueron los de Morelia y Zamora, aunque estos se emplearán una vez que esté establecida la regionalización, se revisó la cantidad de registros válidos y se analizó el comportamiento de las variables. El resultado mostró que 52 estaciones del CLICOM disponen de datos continuos en un periodo mínimo de 20 años a partir de 1981. Las 52 estaciones utilizadas en el proceso de regionalización representan el 21.7 % del total de 239 estaciones con las que cuenta el estado de Michoacán. Esta falta de datos se puede asociar a que algunas estaciones dejaron de funcionar (St16035, St16040, St16072, St16099, St16140, St16212) no registrando datos desde 1985. Algunas otras pese a tener registros disponen de una limitada cantidad de éstos.

De las 52 estaciones seleccionadas cuatro fueron las variables significativas para la regionalización: temperatura -observada, máxima y mínima- y precipitación. Otras variables como evaporación, cobertura de nubosidad, y velocidad y dirección del

viento presentaron por lo general un alto porcentaje de registros nulos o incompletos, por lo que no fue viable integrar estas variables en las bases de datos a ser empleadas en el proceso de regionalización o vistas minables.

Otro aspecto de gran interés se relaciona a la ubicación geográfica de las 52 estaciones, ya que la mayoría se encuentran situadas en la zona norte del estado, mientras que un grupo menor se ubica en el centro y aún menor al sur. Esta dispersión probablemente se debe a la dificultad de colocar o mantener las estaciones en zonas boscosas y selvas, particularmente en la zona sur del estado. En el norte y centro, predomina un relieve montañoso y de valles, el cual pese a ser una limitación ha permitido la colocación y mantenimiento de diversas estaciones al paso del tiempo.

A partir de los registros de las 52 estaciones se generaron las vistas minables que fueron empleadas por las técnicas de minería de datos para la regionalización. El *agrupamiento (clustering)* de datos fue seleccionado debido a que permite formar regiones homogéneas con características similares, esto es, divide una población heterogénea en un número de subgrupos homogéneos de acuerdo a la similitud de sus registros; el cual se ha manejado exitosamente en hidrología. Tres técnicas fueron seleccionadas de los tipos de agrupamiento: Bietápico (jerárquico), K-medias (particional) y redes de Kohonen (mapa autoorganizativo).

Se mostró la utilidad del algoritmo Bietápico en la obtención del número óptimo de grupos, en particular, para el empleo de la técnica de K-medias y redes de Kohonen. En este caso se obtuvo como número óptimo de grupos **4** el cual fue usado como valor de agrupación en el K-medias y redes de Kohonen. Es importante mencionar que aún aumentando la variable *altitud* a una de las vistas minables, los resultados para cada una de las técnicas de agrupamiento (Bietápico y K-medias) fueron las mismas, es decir, la misma estación fue asignada al mismo número de grupo en ambas vistas; lo cual reflejó que la variable *altitud* no fue preponderante en la formación de las cuatro regiones climáticas.

Las técnicas que ofrecieron mejores resultados en el proceso de regionalización fueron el Bietápico y K-medias, ambos con iguales resultados. Estos se compararon con el mapa de climas del INEGI (2003). La comparación confirmó que las redes de Kohonen no proporcionaron una buena regionalización debido a la falta de claridad en la división de los grupos. Finalmente, se realizó la validación de los resultados. En el caso del Bietápico no se aplicó la validación ya que las distancias de los registros en esta técnica se calculan mediante la media y la varianza por lo que no fue posible disponer de una matriz de distancias euclidianas como el que se obtiene en el K-medias. En el caso de las redes de Kohonen, en la literatura no se encontraron *índices de validación* diseñados para este tipo de técnica aunque se probó un método alterno que dio como valor óptimo 4. De esta manera, la validación fue realizada sólo al K-medias, mediante los índices de validación: Ancho de la Silueta, Dunn y Davies-Bouldin (uno de los índices más recomendables por su mejor aproximación). Los índices Ancho de la Silueta y Davies-Bouldin permitieron validar las cuatro regiones climáticas establecidas para el estado de Michoacán. La validación comprobó que los grupos formados por el K-medias proporcionan una regionalización adecuada. Esto debido a que el agrupamiento cumple con al menos dos de los índices de validación -Silueta (0.45) y Davies-Bouldin (1.01)-.

El resultado final de la regionalización divide el mapa del estado en cuatro regiones climáticas (Figura 6.1):

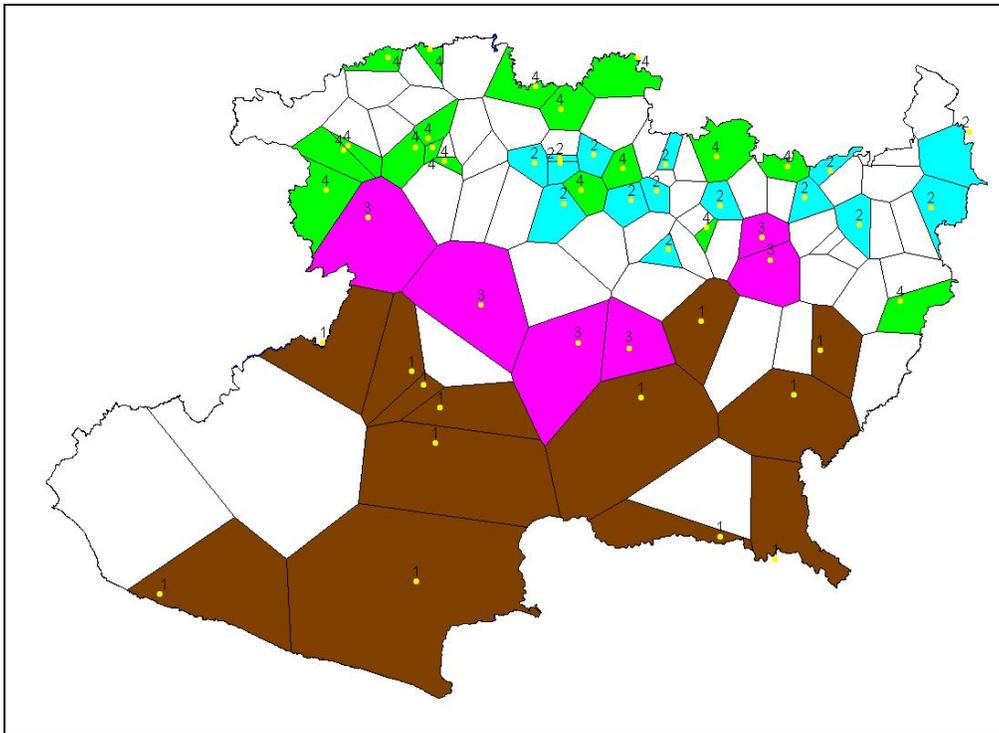


Figura 6.1. Mapa de regionalización climática del estado de Michoacán.

- Dos regiones fueron claramente separadas (color marrón y rosado en la Figura 6.1) que coinciden con los tipos de clima del INEGI: Seco y semiseco muy cálido y cálido y Cálido subhúmedo con lluvias en verano (color marrón) y Semicálido subhúmedo con lluvias en verano y Templado subhúmedo y húmedo con abundantes lluvias en verano (color rosado).

- Dos regiones (color verde y azul en la Figura 6.1) que comparten algunas zonas climáticas en la zona norte y este del estado, específicamente sectores en las que se ubican las estaciones St16050, St16091, St16181 y St16192. Estas regiones coinciden con los climas (INEGI, 2003): Semicálido subhúmedo y Templado subhúmedo con lluvias en verano (color verde) y Templado subhúmedo con lluvias en verano y Semifrío húmedo con abundantes lluvias en verano (color azul).

Pese a la separación que se obtuvo, para mejorar la división de las dos últimas regiones (color verde oscuro y verde claro) se sugiere ampliar la vista minable con promedios quincenales de: precipitación, temperatura observada, y valores máximos y mínimos de temperatura máxima y mínima, respectivamente.

6.2 Conclusiones particulares

- El uso del software Clementine permitió generar modelos basados en la metodología CRISP-DM en forma estructurada, lo que ayuda a explorar y analizar grandes conjuntos de datos de forma organizada, genera reportes y dispone de herramientas visuales para su fácil interpretación.
- El número de estaciones a ser usadas para el análisis de regionalización fue una limitante importante, ya que en su mayoría se contó con el 79% de series incompletas y no continuas.
- Para el análisis de regionalización se analizaron y prepararon los datos mediante la construcción de vistas minables, lo que permite acceder a las técnicas de minería de datos de manera concisa para generar conocimiento útil.
- El proceso de análisis y preparación de datos fue el más complejo debido a la información disponible, así se pasó de análisis horarios, a diarios para finalmente aceptar análisis mensuales.
- El análisis de variabilidad y distribución de las variables atmosféricas registradas en el CLICOM permitió establecer aquellas variables significativas para la regionalización: temperatura -observada, máxima y mínima-, y precipitación. Otro tipo de variable no se usó debido a la falta de registros presentes en las bases de datos.

- El *agrupamiento* proporciona ventajas sobre otras tareas de minería de datos para la identificación de zonas o regiones climatológicamente similares. En este caso tanto el método jerárquico (Bietápico) y no jerárquico (K-medias) logran resultados similares.
- Las redes de Kohonen obtuvieron resultados menos favorables que las de agrupamiento jerárquicos y no jerárquicos, ya que el número de variables de entrada para el proceso de aprendizaje fue una limitante.
- La comparación con el mapa climatológico del INEGI (2003) fue satisfactoria por su proximidad en valores de temperatura y precipitación, sin embargo se debe considerar que estos mapas están generados con bases de datos anuales.
- Para la validación se consideraron el Ancho de la Silueta y Davies-Bouldin como aquellos indicadores con la validez necesaria para proporcionar el número óptimo de grupos formados. Así se estableció un número de grupos igual a 4, mismo que se obtuvo aplicando el algoritmo Bietápico.
- La validación mediante *índices de validación* sólo fue aplicada a K-medias, lo cual no permitió establecer cual sería la confiabilidad de los métodos Bietápico y redes de Kohonen con respecto a los grupos formados.
- La validación mostró que la técnica de agrupamiento K-medias permite la regionalización de zonas climatológicamente homogéneas con alto grado de confiabilidad.
- Finalmente, el trabajo realizado implicó retos importantes como: el análisis de series temporales no continuas e incompletas, así como el manejo de técnicas de agrupamiento para regionalizar zonas homogéneas, y el empleo de métodos de validación; los cuales ayudaron a tener una visión más amplia

de minería de datos y su aplicación a problemas de diverso índole, en este caso aplicado a la hidrología.

6.3 Recomendaciones

Con base en los resultados obtenidos y a las conclusiones llegadas, se presentan las siguientes recomendaciones por medio de las cuales se considera que los resultados logrados pueden tener un mayor grado de confiabilidad:

- Debido a que el número de estaciones con registros completos y continuos disponible para el análisis de regionalización fue de 52, se recomienda incrementar dicho número. Para ello se sugiere revisar aquellas estaciones con registros incompletos y encontrar la causa, de manera que se vuelva a contar con registros completos. Asimismo se sugiere la instalación de estaciones en el centro y sur del estado donde la dispersión es muy alta.
- Los resultados obtenidos fueron comparados con el mapa de climas del INEGI (2003), lo que permitió definir el grado de confiabilidad de la regionalización obtenida. Sin embargo, la distribución climática del INEGI está basada en series anuales que proporcionan variación al compararla con series mensuales y más si estas son diarias. Es por ello, que se debe contar con un método alternativo para comparar los valores obtenidos con mayor precisión.
- Para mejorar el desempeño de las redes de Kohonen, y asimismo mejorar la división de las dos regiones que comparten zonas climáticas (color verde y azul en la Figura 6.1), se sugiere realizar nuevos agrupamientos empleando otro tipo de combinaciones de las variables climatológicas disponibles para el estado de Michoacán, por ejemplo se podrían generar nuevas vistas minables con:

- Promedios quincenales de precipitación, temperatura observada, y valores máximos y mínimos de temperatura máxima y mínima, respectivamente.
 - Promedios quincenales de temperatura -observada, máxima y mínima- y precipitación.
- Es recomendable emplear otro tipo de índices de validación que permitan comparar los resultados obtenidos al aplicar diversos tipos de agrupaciones: jerárquicas y no-jerárquicas.

6.4 Líneas de trabajo futuro

En esta tesis se han empleado técnicas de minería de datos para regionalizar zonas climatológicamente similares en el estado de Michoacán. Si bien las regiones obtenidas fueron validadas, la división de las dos regiones ubicadas al norte del estado (color verde y azul en la Figura 6.1) ha dejado abiertas futuras líneas de investigación:

- Probar agrupamientos con métodos distintos (con el fin de comparar resultados), como las *máquinas de soporte vectorial* (SVM) que son técnicas que tratan de maximizar el margen entre los grupos y las clases formadas, y *algoritmos genéticos* (AG) que permiten la homogenización de una población, es decir, encuentran una representación adecuada de los individuos o elementos con características similares.
- Emplear indicadores de heterogeneidad basado en L-momentos, mediante los cuales se estima el grado de similitud en un grupo de estaciones y se evalúa si éstas pueden ser tratadas como una región homogénea. Hosking y Wallis (1997) han considerado que el método L-momentos es adecuado para identificar regiones homogéneas a partir de muestras de datos, permitiendo describir la región a la que pertenecen; esto hace que este método sea

confiable, especialmente frente a datos anómalos. Para validar el desempeño del método L-momentos, los autores señalan que ésta se debe realizar mediante la estimación de cuantiles y el análisis de la medida de error.

Referencias Bibliográficas

- Balakrishnan P., Cooper M., Jacob V. y Lewis P. (1994). A study of the classification capabilities of neural networks using unsupervised learning: A comparison with K-means clustering. *Psychometrika*, 59, 4, 509-525.
- Berkhin P. (2006). A Survey of Clustering Data Mining Techniques. *Grouping Multidimensional Data*, editado por Springer-Verlag, pp. 25-71, ISBN: 978-3-540-28349-2, Heidelberg, Alemania.
- Berry M. y Linoff G. (2004). Data Mining Techniques: for marketing, sales, and customer relationship management. 2da edición por Wiley Publishing, Inc., pp. 643, ISBN: 0-471-47064-3, Indiana, Estados Unidos.
- Bolshakova N. y Azuaje F. (2003). Improving expression data mining through cluster validation. 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, pp. 19-22, ISBN: 0-7803-7667-6, Irlanda.
- Boutin F. y Hascoët M. (2004). Cluster Validity Indices for Graph Partitioning. Proceedings of the Information Visualisation, Eighth International Conference on (IV'04), pp. 376-381, ISBN: 1093-9547, Washington, Estados Unidos.
- Brock G., Pihur V., Susmita D. y Somnath D. (2007). clValid, an R package for cluster validation. Department of Bioinformatics and Biostatistics, University of Louisville, pp. 28, Estados Unidos.
- Burn D. (1990). An appraisal of the “region of influence” approach to flood frequency analysis. *Hydrological Sciences Journal*, 35, 2, 149-165.
- Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T., Shearer C. y Wirth R. (2000). CRISP-DM 1.0 Step-by-step Data Mining Guide. <<http://www.crisp-dm.org/CRISPWP-0800.pdf>>. Última consulta el 28.04.2007.
- Chiu B., Keogh E. y Lonardi S. (2003). Probabilistic Discovery of Time Series Motifs. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 493-498, ISBN: 1-58113-737-0, Washington, Estados Unidos.
- Chou C., Su M. y Lai E. (2003). A New Cluster Validity Measure for Clusters with Different Densities. Department of Electrical Engineering, Tamkang University, pp. 6, Taiwan.

- CIIFEN - OMM (2003). Informe Final del Seminario - Taller. Rescate y Gerencia, Monitoreo, Aplicación y Predicción de Datos Climáticos. Centro Internacional para la Investigación del Fenómeno de El Niño, Organización Meteorológica Mundial, pp. 7, Guayaquil, Ecuador.
- Clementine (2006a). Clementine 10.1 Algorithms Guide, editado por Integral Solutions Limited. SPSS Inc., pp. 152, Estados Unidos.
- Clementine (2006b). Referencia de Nodos de Clementine 10.1, editado por Integral Solutions Limited. SPSS Inc., pp. 687, Estados Unidos.
- Clementine (2006c). Ayuda de Clementine, editado por Integral Solutions Limited, SPSS Inc. 2006. <c:\Archivos de programa\SPSS Clementine\10.1>, Estados Unidos.
- CONAGUA (2005). Agua Subterránea. Comisión Nacional del Agua. <http://www.cna.gob.mx/eCNA/Espaniol/Programas/Subdirecciones/Html-GAS/displ_gas/disponibilidad.htm>. Última consulta el 07.07.2007.
- Davies D. y Bouldin D. (1979). A cluster separation measure. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 1, 224-227.
- Dunn J. (1974). Well separated clusters and optimal fuzzy partitions. *Journal on Cybernetics*, 4, 95-104.
- Elías F. y Castellví F. (2001). Agrometeorología. 2da ed. Departamento de Medio Ambiente y Ciencias del Suelo, editores Mundi-Prensa, pp. 517, ISBN: 84-7114-973-7, Madrid, España.
- Fonfría A. (2005). Objetivos tecnológicos y de internacionalización de las políticas de apoyo a las pyme en Europa. Instituto Complutense de Estudios Internacionales y Departamento de Economía Aplicada II, Universidad Complutense de Madrid, pp. 25, ISSN: 1578-0244, España.
- Gaál L., Kyselý J. y Štěpánek P. (2006). Region-of-influence approach to the frequency analysis of heavy k-day precipitation amounts. In: Bioclimatology and Water in the Land. Proc. Int. Bioclimatological Conference, pp. 13, Eslovaquia.
- Garza M. y Asociados (2003). Elaboración del Catalogo de Información Disponible en las diversas dependencias estatales y federales relacionadas con el Agua. Consejo del sistema Veracruzano del Agua. Sistema de Información Hidráulica, pp. 22, Veracruz, México.
- Geng L. y Hamilton H. (2007). Choosing the Right Lens: Finding what is Interesting in Data Mining. *Studies in Computational Intelligence (SCI)*, 43, 3-24. Editado por Guillet F. y Hamilton H. Quality Measures in Data

- Mining, editado por Springer-Verlag, pp. 313, ISBN: 3-540-44911-6, Heidelberg, Alemania.
- Gondar J. (2004). Metodologías para la Realización de Proyectos de Data Mining. Data Mining Institute, S.L. Artículos estadísticos. <<http://www.estadistico.com/arts.html?20040426>>. Última consulta el 10.06.2007.
- Hair J., Anderson R., Tatham R. y Black W. (1995). Multivariate data analysis. Englewood Cliffs, 4ta edition por Prentice Hall, pp. 500, ISBN: 0023490209, Estados Unidos.
- Halkidi M., Batistakis Y. y Vazirgiannis M. (2002). Clustering Validity Checking Methods: Part II. *ACM SIGMOD Record*, 31, 3, 19-27, ISSN: 0163-5808, New York, Estados Unidos.
- Hall M. y Minns A. (1999). The classification of hydrologically homogeneous regions. *Hydrological Sciences Journal*, 44, 5, 693-704.
- Hand D., Mannila H. y Smyth P. (2001). Principles of Data Mining, editado por The Massachusetts Institute of Technology Press., pp. 546, ISBN: 0-262-08290-x, Massachusetts, Estados Unidos.
- Hernández J., Ramírez M. J. y Ferri C. (2004). Introducción a la Minería de Datos. Pearson Educación. Editorial Pearson Prentice Hall, pp. 680, ISBN: 84-205-4091-9, Madrid, España.
- Hosking J. y Wallis J. (1997). Regional Frequency Analysis: An Approach Based on L-moments. Cambridge University Press, pp 240, ISBN: 0521019400, New York, Estados Unidos.
- INEGI (1985). Síntesis Geográfica del Estado de Michoacán. Instituto Nacional de Estadística, Geografía e Informática, pp. 319, México.
- INEGI (2003). Anuario Estadístico del Estado de Michoacán, 2003. Instituto Nacional de Estadística, Geografía e Informática, Aguascalientes, México.
- INEGI (2006). Anuario Estadístico del Estado de Michoacán, 2006. Instituto Nacional de Estadística, Geografía e Informática, Aguascalientes, México.
- Jain A., Murty M. y Flynn P. (1999). Data Clustering: A Review. *ACM Computing Surveys*, 31, 3, 264-323, ISBN: 0360-0300, New York, Estados Unidos.
- Juárez A. (2005). Sobreexplotación de los mantos acuíferos en Michoacán. Agencia de Información y Análisis Quadratín, 13 de junio de 2005, México. <<http://www.mimorelia.com/vernota.php?id=11698>>. Última consulta el 06.07.2007.

- Juárez A. (2006). Contaminado, 95% del agua que llega a los hogares michoacanos. Cambio de Michoacán, 14 de marzo de 2006, México. <<http://www.cambiodemichoacan.com.mx/vernota.php?id=40531>>. Última consulta el 09.07.2007.
- Keogh E., Lin J. y Truppel W. (2003). Clustering of Time Series Subsequences is Meaningless: Implications for Previous and Future Research. Proceedings of the Third IEEE International Conference on Data Mining, pp. 115, ISBN: 0-7695-1978-4, Florida, Estados Unidos.
- Kessler M. (2003). Apuntes de métodos estadísticos de la Ingeniería y Apuntes de estadística industrial. Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena, pp. 73, España.
- Klösgen W. y Zytlow J. M. (2002). Handbook of Data Mining and Knowledge Discovery. Oxford University Press, Inc., pp. 980, ISBN: 0-19-511831-6, New York, Estados Unidos.
- Kohonen T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59-69.
- Kohonen T. (1986). Learning Vector Quantization for Pattern Recognition. Helsinki University of Technology, Department of Technical Physics. Technical Report TKK-F-A601.
- Kovács F., Legány C. y Babos A. (2005). Cluster Validity Measurement Techniques. Department of Automation and Applied Informatics, Budapest University of Technology and Economic, pp. 11, Budapest, Hungría.
- Kyselý J., Pícek J. y Huth R. (2007). Formation of homogeneous regions for regional frequency analysis of extreme precipitation events in the Czech Republic. *Studia Geophysica et Geodaetica*, 51, editado por Springer Verlag, pp. 327-344, República Checa.
- Larose D. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons, Inc., pp. 222, ISBN: 0-471-66657-2, New Jersey, Estados Unidos.
- MacQueen J. (1967). Some methods for classification and analysis of multivariate observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1, 281-297.
- Mapas de México (2006). Mapas de México. Code Network Media Group Ltd. <<http://www.map-of-mexico.co.uk/espanola/mapa-de-michoacan.htm>>. Última consulta el 02.07.2007.

- Milligan G. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, 45, 3, 325-342.
- Milligan G. y Cooper M. (1985). An Examination of Procedures for Determining the Number of Clusters in a Data Set. *Psychometrika*, 50, 159-179.
- MIT (2001). The Technology Review Ten, MIT Technology Review, January/February. <<http://www.techreview.com>>. Última consulta el 08.02.2007.
- Muñoz D. y Rodrigo F. (2002). Aplicación del análisis cluster para el estudio de la relación Nao-Precipitaciones de invierno en el sur de la Península Ibérica. Departamento de Física Aplicada, Universidad de Almería, pp. 283-292, España.
- Pelczer I., Ramos J., Domínguez R. y González F. (2007). Establishment of regional homogeneous zones in a watershed using clustering algorithms en proceedings of 32nd IAHR Congress “Harmonizing the Demands of Art and Nature in Hydraulics”, artículo 1118, Venecia, Italia.
- Puerto J. y Paz M. (2001). Análisis descriptivo de series temporales aplicadas al precio medio de la vivienda en España. Management Mathematics for European Schools, pp. 41, España.
- Rao A. y Srinivas V. (2006). Regionalization of watersheds by hybrid-cluster analysis. *Journal of Hydrology*, 318, 37-56.
- Rousseeuw P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.
- Rueda J. (2007). Más del 50% de cuerpos acuíferos en Michoacán, contaminados. Agencia de Información y Análisis Cuadratín, 28 de febrero de 2007, México. <<http://www.mimorelia.com/vernota.php?id=25501>>. Última consulta el 06.07.2007.
- Saitta S., Raphael B. y Smith I. (2007). A Bounded Index for Cluster Validity. Machine Learning and Data Mining in Pattern Recognition, editado por Springer-Verlag, pp. 174-187, ISBN: 978-3-540-73498-7, Alemania.
- SMN (2006). Resumen Sinóptico Decenal. Servicio Meteorológico Nacional. <<http://smn.cna.gob.mx/productos/observatorios/rdecenal.html>> Última consulta el 22.06.2007.
- SMN (2007a). Guía de Instrumentos y Métodos de Observación Meteorológicos. Servicio Meteorológico Nacional, Subdirección General Técnica, pp. 189, México.

- SMN (2007b). Términos Meteorológicos. Servicio Meteorológico Nacional. <<http://smn.cna.gob.mx/glosario/glosario.html>>. Última consulta el 06.07.2007.
- SMN (2007c). Información Sinóptica de los Observatorios. Servicio Meteorológico Nacional. <<http://smn.cna.gob.mx/productos/observatorios/informacion.html>>. Última consulta el 11.04.2007.
- Sumathi S. y Sivanandam S. (2006). Introduction to Data Mining and its Applications. *Studies in Computational Intelligence*, 29, editado por Springer-Verlag, pp. 828, ISBN: 3-540-34350-4, Heidelberg, Alemania.
- Thorntwaite C.W. (1948). An approach toward a rational classification of climate. *Geographical Review*, 38, 1, 55-94.
- Turc L. (1954). Le bilan d'eau des sols: relations entre les précipitation, l'évaporation et l'écoulement. *Annales Agronomiques*, 5, 491-569.
- Vazirgiannis M., Halkini M. y Gunopulos D. (2003). Uncertainty Handling and Quality Assessment in Data Mining. *Advanced Information and Knowledge Processing*, editado por Springer-Verlag, pp. 226, ISBN: 1-85233-655-2, Heidelberg, Alemania.
- Witten I. y Frank E. (2005). Data Mining: Practical machine learning tools and techniques. 2da ed. por Morgan Kaufmann Series in Data Management Systems, pp. 525, ISBN: 0-12-088407-0, Estados Unidos.
- Wong M. y Leung K. (2002). Data Mining Using Grammar Based Genetic Programming And Applications. *Genetic Programming Series*, editado por Kluwer Academic Publishers, pp. 213, ISBN: 0-792-37746-X, Estados Unidos.
- Xu L. y Yao L. (2001). Applications of data mining in hydrology. ICDM, Proceedings of the 2001 IEEE International Conference on Data Mining, pp 617-620, ISBN: 0-7695-1119-8, Washington, Estados Unidos.
- Xu R. y Wunsch D. (2005). Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*, 16, 3. pp. 645-678.
- ZDNet (2001). Data mining: Digging user info for gold, ZDNet News, February 8. <<http://zdnet.com.com/2100-11-528032.html?legacy=zdn>>. Última consulta el 07.02.2007.
- Zhang T., Ramakrishnan R. y Livny M. (1996). BIRCH: An efficient data clustering method for very large databases. ACM SIGMOD Conference 1996, pp. 103-114, Montreal, Canadá.

Anexo A.

En este apartado se presentan las 239 estaciones del Sistema CLICOM establecidas en el estado de Michoacán. Se indica el identificador, nombre oficial, ubicación geográfica, año de inicio y año final de los registros disponibles por estación (Tabla A.1).

Tabla A.1. Lista de las 239 estaciones del Sistema CLICOM establecidas en el estado de Michoacán.

Nº	Identificador	Nombre	Latitud		Longitud		Altitud Msnm	Fecha	
			g	min	g	min		Inicio	Fin
1	16001	ACUITZIO DEL CANJE, (SMN)	19	29	101	20	2,070	1961-01	2004-07
2	16002	AGOSTITLAN, CD. HIDALGO	19	34	100	41		1950-07	2004-02
3	16003	AGUILILLA, AGUILILLA	18	44	102	15	970	1943-03	1982-08
4	16004	ALVARO OBREGON, (SMN)	19	50	101	2		1964-10	1986-01
5	16005	ANGAMACUTIRO, (SMN)	20	7	101	42		1923-09	1980-02
6	16006	ANTUNEZ, PARACUARO	19	0	102	12	360	1962-01	2002-12
7	16007	APATZINGAN, APATZINGAN	19	5	102	16	682	1922-09	2003-02
8	16008	AQUILA, AQUILA	18	25	103	23	100	1961-05	1999-04
9	16009	ARTEAGA, ARTEAGA (SMN)	18	21	102	17	860	1936-02	1988-08
10	16010	BARTOLINAS, TACAMBARO	19	12	101	26		1961-01	1976-12
11	16011	BUENAVISTA, BUENAVISTA	19	6	102	33		1969-02	1980-07
12	16012	CAJONES, GABRIEL ZAMORA	19	18	101	55	480	1970-08	1999-12
13	16013	CALETA DE CAMPOS,	18	5	102	45	12	1961-01	1984-04
14	16014	CAMECUARO, TANGANCICUARO	19	55	102	12	1,708	1951-06	2003-04
15	16015	CARAPAN, CHILCHOTA	19	51	102	2	1,980	1950-01	2004-09
16	16016	CARRILLO PUERTO, A. O.	19	55	101	20	1,840	1969-01	2004-10
17	16017	CASA BLANCA, PURUANDIRO	20	19	101	27	1,694	1949-03	2000-08
18	16018	CASA FUERTE, SAHUAYO	20	4	102	41	1,523	1965-08	1988-01
19	16019	CERRITO PELON, JIQUILPAN	20	2	102	42	1,540	1969-01	1972-08
20	16020	CIUDAD HIDALGO (DGE)	19	42	100	34	2,020	1961-01	2004-10
21	16021	COALCOMAN, COALCOMAN	18	48	103	15		1963-01	1975-04
22	16022	COINTZIO, MORELIA	19	37	101	16	2,096	1940-04	2005-01
23	16023	COPANDARO DE GALEANA,	19	57	101	37	1,500	1969-12	2001-04
24	16024	CORRALES (A. LOS CORRALES)	20	12	101	51	1,693	1934-10	2003-07
25	16025	COTIJA, COTIJA	19	43	102	40	1,640	1924-07	1985-07
26	16026	CUATRO CAMINOS, NVA. I.	18	55	102	50	350	1962-09	1985-09
27	16027	CUITZEO, CUITZEO	19	57	101	8	1,831	1923-03	2003-01
28	16028	CUITZILLO GRANDE,	19	46	101	7	1,840	1969-07	2005-01
29	16029	CUMUATILLO, V. CARRANZA	20	8	102	38	1,522	1969-01	1991-09
30	16030	CUMUATO, BRIENAS D. MATAM	20	16	102	35	1,524	1951-07	1999-08
31	16031	CHAPARACO, ZAMORA	19	58	102	15	1,632	1948-02	2002-05
32	16032	CHARAPAN, CHARAPAN	19	42	102	12	1,560	1952-08	1973-10
33	16033	CHINCUA, SENGUO	19	46	100	17	2,415	1966-11	2004-07
34	16034	EL JACAL, CHUCANDIRO	19	53	101	20		1969-01	2002-12
35	16035	CHURUMUCO, CHURUMUCO	18	39	101	38	251	1933-02	1994-09
36	16036	EL BOSQUE, ZITACUARO	19	23	100	23	1,750	1952-02	1989-08
37	16037	EL CAJON, TEPACALTEPEC	18	58	102	43	290	1962-08	1978-04
38	16038	EL CARRIZO, TEPALCATEPEC	19	3	102	59		1969-07	1985-09
39	16039	P. BLANCAS, BUENAVISTA	19	2	102	35	344	1961-01	1985-09
40	16040	EL GIGANTE, MARAVATIO	19	58	100	27	1,961	1970-08	2003-09
41	16041	EL PEJO, HUETAMO	18	41	100	57	380	1964-07	1985-09
42	16042	EL PLATANAL, JACONA	19	57	102	16		1969-03	1970-10
43	16043	EL PUERTO, COTIJA	19	48	102	40	1,581	1943-08	2000-04
44	16044	EL SALTO, LA PIEDAD	20	22	102	7	1,650	1948-06	1997-09

Tabla A.1. Lista de las 239 estaciones del Sistema CLICOM establecidas en el estado de Michoacán.

45	16045	EL TEMASCAL, CHARO	19	39	100	57	2,240	1965-01	2004-11
46	16046	EL VASO INFIERNILLO	18	16	101	53	180	1965-04	1994-09
47	16047	EL ZAPOTE, LA HUACANA	18	55	102	46	920	1951-03	2001-02
48	16048	ZAMORA, ZAMORA (DGE)	20	0	102	16	1,567	1947-01	2003-04
49	16049	ETUCUARO, MADERO (SMN)	19	53	102	13	1,612	1944-06	1988-09
50	16050	HUANIQUEO, HUANIQUEO	19	54	101	30	1,950	1948-07	2004-10
51	16051	HUETAMO, HUETAMO (SMN)	18	35	100	53	300	1922-05	2004-12
52	16052	HUINGO, ZINAPECUARO	19	55	100	51	1,830	1941-01	2004-12
53	16053	INGENIO SAN SEBASTIAN,	19	36	102	36		1961-01	1975-12
54	16054	IXTLAN DE LOS HERVORES,	20	10	102	20		1961-01	1988-02
55	16055	JESUS DEL MONTE, MORELIA	19	40	101	7	2,100	1935-11	2004-01
56	16056	JICALAN, URUAPAN	19	23	102	4	1,610	1963-05	1998-06
57	16057	JIMENEZ, VILLA JIMENEZ	19	56	101	45	1,980	1970-07	1999-03
58	16058	JUNGAPEO, JUNGAPEO	19	26	100	30	1,300	1940-09	1988-05
59	16059	LA CAIMANERA, HUETAMO	18	27	100	53	305	1961-01	1999-12
60	16060	LA ESTANZUELA, IXTLAN	20	7	102	21	1,566	1951-01	1999-12
61	16061	LAGUNA DEL FRESNO,	19	28	100	25	2,070	1947-01	2004-12
62	16062	LA HUACANA, LA HUACANA	19	8	101	58	530	1938-11	1982-09
63	16063	LA PALMA, V. CARRANZA	20	9	102	45	1,522	1935-02	2002-08
64	16064	LA PASTORIA, NVA. ITALIA	18	55	102	3	269	1965-09	1991-07
65	16065	LA PIEDAD CABADAS (DGE)	20	20	102	10	1,684	1925-01	2004-12
66	16066	LA PIEDAD CABADAS (SMN)	20	8	102	2		1961-01	1968-05
67	16067	LA RAYA, SAHUAYO	20	50	102	43	1,577	1948-07	2003-03
68	16068	LAS FUENTES, SAHUAYO	20	0	102	51	2,101	1948-02	1998-12
69	16069	LA ESTANCIA, HUETAMO	18	38	100	55		1964-07	1972-09
70	16070	LA VILLITA, MELCHOR O.	18	2	102	10	61	1951-03	1991-12
71	16071	LOS AZUFRES, CD. HIDALGO	19	46	100	39	2,860	1964-10	1997-01
72	16072	LOS CHORROS DEL VARAL,	19	30	102	34	882	1965-07	1994-09
73	16073	LOS LIMONES, LOS REYES	19	1	102	31	1,225	1954-10	1998-08
74	16074	LOS PANCHES, LA HUACANA	18	52	102	13	207	1964-11	1999-12
75	16075	LOS PINZANES, HUETAMO	18	32	101	6	372	1964-10	1999-12
76	16076	LOS POZOS, ARTEAGA	18	26	102	23	1,000	1972-01	1985-09
77	16077	LOS REYES, LOS REYES	19	35	102	29	1,280	1923-06	1987-09
78	16078	MARAVATIO, MARAVATIO	19	54	100	25		1923-05	1984-12
79	16079	MELCHOR OCAMPO, M. OCAMPO	17	56	102	12		1961-05	1988-08
80	16080	MORELIA, MORELIA	19	42	101	11	1,915	1986-05	2002-07
81	16081	MORELIA, MORELIA (DGE)	19	41	101	10	1,903	1947-05	2004-02
82	16082	NUEVA ITALIA, ZARAGOZA	19	0	102	5	460	1941-04	1989-05
83	16083	PAJACUARAN, PAJACUARAN	20	7	102	34	1,538	1969-01	1974-08
84	16084	PANINDICUARO,	19	58	101	45	1,638	1946-08	2002-12
85	16085	PARACUARO, PARACUARO	19	9	102	13	498	1969-06	2001-08
86	16086	PASTOR ORTIZ, PURUANDIRO	20	18	101	35	1,692	1967-09	2004-06
87	16087	PATZCUARO, PATZCUARO	19	38	101	36	2,035	1969-01	2004-12
88	16088	PERIBAN, PERIBAN	19	31	102	26	1,630	1969-02	1998-06
89	16089	PLANTA H. EL COBANO, G. Z.	19	9	101	59		1953-02	1983-10
90	16090	P. BLANCAS, BUENAVISTA	19	14	102	40	304	1963-02	2002-12
91	16091	PLANTA BOMBEO A. ZINZIMEO	19	49	101	40	1,384	1966-10	2005-01
92	16092	PSA. AGOSTITLAN, CD. H.	19	34	100	36	2,380	1961-01	1984-08
93	16093	PSA. GONZALO, VISTAHERMOSA	20	18	102	26	1,533	1969-01	2003-01
94	16094	PRESA GUARACHA, VILLAMAR	19	58	102	35	1,570	1944-01	1999-08
95	16095	PRESA JARIPO, VILLAMAR	19	57	102	36	1,624	1948-06	2000-08
96	16096	PSA. MALPAIS, QUERENDARO	19	49	100	52	1,859	1940-12	2004-12
97	16097	PSA. PUCUATO, CD. HIDALGO	19	37	100	41	2,480	1950-07	2004-10
98	16098	PSA. SABANETA, CD. HIDALGO	19	36	100	40	2,500	1952-12	2004-10
99	16099	PRESA TUXPAN, TUXPAN	19	33	100	20		1952-11	1995-06
100	16100	PUENTE SAN ISIDRO,	19	51	101	31	2,001	1947-01	1993-01
101	16101	PUNTA DE AGUA, BUENAVISTA	19	12	102	41	279	1969-11	1985-09
102	16102	PUNTA SAN TELMO, AGUILA	18	19	103	29	70	1961-04	1988-08
103	16103	PUREPERO, PUREPERO	19	54	102	1	1,950	1948-06	1986-09
104	16104	PURUANDIRO, PURUANDIRO	20	51	101	30	1,848	1942-10	1997-07
105	16105	QUIRIO, INDAPARAPEO	19	49	101	0	1,858	1963-06	2005-01
106	16106	SAN ANGEL, TINGUINDIN	19	53	102	32	1,870	1954-05	1991-06
107	16107	SAN CARLOS, TIQUICHEO	19	14	100	43	685	1969-01	2001-07
108	16108	SAN CRISTOBAL, IXTLAN	20	12	102	26	1,523	1969-01	2000-08
109	16109	SAN DIEGO CURUCUPATZEO,	19	20	101	11	1,020	1922-01	2004-12
110	16110	SAN JERONIMO, HUETAMO						1964-09	1979-06

Tabla A.1. Lista de las 239 estaciones del Sistema CLICOM establecidas en el estado de Michoacán.

111	16111	SAN JOSE, MARAVATIO	19	54	100	19	2,043	1955-06	2004-03
112	16112	SAN JUAN DE LIMA,	18	35	103	39	150	1969-01	1985-10
113	16113	SAN LUCAS, LUCAS	18	34	100	47	265	1964-06	1985-09
114	16114	S. MIGUEL DEL MONTE, MOR.	19	16	101	0	2,127	1963-12	2004-11
115	16115	SAN NICOLASITO, MARAVATIO	19	55	100	24		1955-04	1971-06
116	16116	SAN SEBASTIAN, QUERENDARO	19	51	100	56	2,070	1969-07	1991-12
117	16117	SANTA FE DEL RIO,	20	12	101	49	1,681	1946-09	1997-09
118	16118	SANTA FE, QUIROGA	19	49	101	22	2,056	1963-06	2003-02
119	16119	STA. RITA, COPANDARO DE G.	19	54	101	16	1,880	1964-10	2004-08
120	16120	SANTIAGO UNDAMEO, MORELIA	19	36	101	19	2,048	1953-12	2004-06
121	16121	SENGUIO, SENGUIO	19	44	100	22	2,222	1969-01	2001-08
122	16122	SUSUPUATO DE GUERRERO,	19	11	100	25	1,560	1945-08	2004-12
123	16123	TACAMBARO, TACAMBARO	19	14	101	28	1,755	1922-01	2003-06
124	16124	TEMASCALES, CONTEPEC	20	3	100	8	2,409	1943-06	2004-03
125	16125	TANGANCICUARO A. ETUCUARO	19	53	102	13	1,690	1947-07	2005-01
126	16126	TANHUATO, TANHUATO (SMN)	20	16	102	21	1,540	1943-06	1978-06
127	16127	TARETAN, TARETAN	19	20	101	55	1,170	1961-01	1998-06
128	16128	TEPALCATEPEC,	19	11	102	50	974	1943-01	1983-12
129	16129	TEPUXTEPEC, CONTEPEC	19	59	100	13	2,380	1947-02	2001-08
130	16130	E.N.M.G.F. URUAPAN,	19	23	102	4		1973-02	1975-04
131	16131	TICUITACO, LA PIEDAD	20	18	102	1	1,750	1965-02	1971-06
132	16132	TIQUICHEO, TIQUICHEO (CFE)	18	51	100	40		1955-05	1971-07
133	16133	TURICATO, TURICATO	19	3	101	25	900	1969-01	2004-06
134	16134	TUXPAN, TUXPAN	19	34	100	27	1,800	1941-04	1985-04
135	16135	TUZANTLA, TUZANTLA (SMN)	19	13	100	35	666	1940-08	1986-06
136	16136	TZITZIO, TZITZIO	19	34	100	55	1,540	1969-01	2004-11
137	16137	UREPETIRO, TANGANCICUARO	19	58	102	8		1949-01	2001-12
138	16138	URUAPAN, URUAPAN (DGE)	19	23	102	3	1,611	1951-03	1985-10
139	16139	VILLA MADERO, MADERO	19	24	101	16	2,227	1943-05	1984-09
140	16140	VILLA MADERO, MADERO	19	23	101	16	2,134	1961-01	1994-09
141	16141	YURECUARO, YURECUARO	20	20	102	16	1,524	1923-03	2003-11
142	16142	ZACAPU, ZACAPU (DGE)	19	48	100	47	1,986	1968-08	2001-12
143	16143	ZAMORA, ZAMORA (SMN)	19	59	102	17	1,561	1922-05	1960-12
144	16144	ZAMORA, ZAMORA (SMN)	20	0	102	7	1,567	1969-01	1989-03
145	16145	ZINAPECUARO, ZINAPECUARO	19	53	100	40	1,920	1923-08	2004-12
146	16146	ZIRAHUEN, VILLA ESCALANTE	19	28	101	45	2,228	1947-10	2003-03
147	16147	ZIRITZICUARO, LA HUACANA	19	40	102	0	525	1952-08	1985-09
148	16148	ZITACUARO, ZITACUARO	19	26	100	22	1,981	1921-10	1983-01
149	16149	APATZINGAN, APATZINGAN	19	5	102	21	320	1969-01	1985-09
150	16150	APATZINGAN, APATZINGAN	19	5	102	21	3,200	1969-01	1975-12
151	16151	ARTEAGA, ARTEAGA (CFE)	18	21	102	17	940	1957-11	2000-02
152	16152	CD. HIDALGO, CD. HIDALGO	19	42	100	34	2,060	1922-06	1988-09
153	16153	COALCOMAN, COALCOMAN DGE	18	46	103	9	1,100	1961-01	1973-11
154	16154	COALCOMAN, COALCOMAN SMN	18	46	103	9	1,100	1926-12	1989-10
155	16155	COPANDARO, JIMENEZ	19	56	101	0		1969-11	2001-12
156	16156	CHARAPENDO, G. ZAMORA	19	16	102	4	1,000	1969-07	1985-09
157	16157	CHILA, AGUILILLA	18	56	102	37	300	1969-03	1985-09
158	16158	EL CAJON, TEPALCATEPEC	18	58	102	45	296	1952-07	1997-12
159	16159	EL ROSARIO, ANGAMACUTIRO	20	7	101	45	1,720	1972-07	2003-01
160	16161	LA VILLITA, MELCHOR O.	18	1	101	0		1961-04	1985-05
161	16162	ORANDINO, JACONA	19	58	102	19	1,633	1961-01	2000-12
162	16163	TUZANTLA, TUZANTLA (DGE)	19	12	100	34		1969-01	1985-06
163	16164	URUAPAN, URUAPAN (SMN)	19	25	102	4	1,634	1921-11	1988-09
164	16165	URUAPAN, URUAPAN (CFE)	19	23	102	3	1,611	1962-10	1999-12
165	16168	ARIO DE ROSALES, (SMN)	19	12	101	44	1,840	1974-01	1988-09
166	16169	LA GARITA A. INFIERNILLO	18	15	101	50		1955-06	1969-04
167	16171	ZACAPU, ZACAPU (SMN)	19	49	101	47	1,980	1923-01	1969-07
168	16174	SAHUAYO, SAHUAYO	20	4	102	43		1972-11	1996-10
169	16175	TEPUXTEPEC, CONTEPEC	20	0	100	14		1969-02	2002-10
170	16178	ESC. NAL. GUARDIAS FOREST.	19	28	100	5		1969-01	1991-04
171	16183	MELCHOR O., LAZARO C.	18	1	102	12		1973-01	1980-12
172	16184	AGRO-URUAPAN, URUAPAN	19	25	102	3	1,611	1973-01	1974-04
173	16185	COAHUAYANA, COAHUAYANA	18	45	103	39	30	1987-01	2001-04
174	16186	EMILIANO ZAPATA, VILLAMAR	20	0	102	34	1,550	1973-01	1999-12
175	16187	ESC. TEC. AGRO. 139,	19	0	101	7	505	1973-04	1981-07
176	16188	TIRINDARO, ZACAPU	19	46	101	44	2,500	1973-12	2003-02

Tabla A.1. Lista de las 239 estaciones del Sistema CLICOM establecidas en el estado de Michoacán.

177	16189	C.R.E.D.F.F.A., ZITACUARO	19	32	100	12	2,150	1973-11	1983-06
178	16190	ARCEO, TUMBISCATIO	18	43	102	16	320	1974-07	1985-09
179	16191	CUATRO ESQUINAS, BRISEÑAS	20	19	102	23	1,522	1974-07	1988-01
180	16192	ENCARNACION, ZITACUARO	19	25	100	24	1,720	1974-09	2004-06
181	16193	E.T.A. 36 LA HUACANA,	18	55	101	50		1974-03	1985-10
182	16194	E.T.A. 39 TOCUMBO, (SMN)	19	42	102	30	1,600	1975-01	1999-12
183	16195	E.T.A. 137 COAHUYANA,	18	45	103	40		1974-01	1978-07
184	16196	SAN GREGORIO, PAJACUARAN	20	11	102	32	1,523	1974-09	1985-10
185	16197	JIQUILPAN, JIQUILPAN	20	0	102	42	1,536	1975-03	1988-02
186	16198	LA ARENA, V. CARRANZA	20	11	102	41	1,527	1975-03	1983-09
187	16199	SAN MIGUEL CURAHUANGO,	19	53	100	27	2,020	1975-03	2002-09
188	16201	C.N.I.E.A. I. LOPEZ RAYON	19	25	102	4		1976-11	1979-12
189	16203	TINGAMBATO ESC. SEC. FED.	19	25	102	4		1975-01	1982-05
190	16204	E.T.A. 132, VILLAMAR	20	0	102	34	1,500	1975-01	1978-12
191	16206	SAN JOSE PURUA, JUNGAPEO	19	29	100	29	1,470	1977-05	1989-05
192	16208	EL CACHAN, AQUILA	18	17	103	17	45	1977-05	2000-12
193	16212	BARRANCA DE CUPATITZIO	19	28	102	5	1,700	1977-08	1995-09
194	16213	IRIMBO, CIUDAD HIDALGO	19	42	100	29	2,235	1978-01	2004-12
195	16216	BUENAVISTA, BUENAVISTA	19	12	102	36	450	1976-01	1999-07
196	16217	EL CANTIL, L. CARDENAS	17	52	102	20		1980-01	1985-09
197	16219	SAN PEDRO, TIQUICHEO	19	4	100	49	530	1979-08	2003-10
198	16220	SAN SIMON, IXTLAN	20	3	102	24		1979-03	2001-01
199	16221	UCAREO UNION DE FRUT.	19	55	100	50		1980-05	1988-07
200	16222	RINCONADA, RINCONADA	20	2	102	17		1979-02	1989-03
201	16223	LA ESPERANZA, CHAVINDA	20	3	102	24		1979-02	1986-04
202	16224	TANHUATO, TANHUATO (DGE)	20	17	102	20	1,540	1979-04	1999-08
203	16225	VILLA JIMENEZ, V. JIMENEZ	19	55	101	45	1,980	1980-01	2003-02
204	16226	QUETZERIA, HUETAMO	18	35	100	53		1980-08	1981-11
205	16227	LAZARO CARDENAS, L. C.	18	2	102	11	50	1981-01	1999-07
206	16228	ACAHUATO, APATZINGAN	19	8	102	19	1,000	1981-01	2002-09
207	16229	ADJUNTAS, TANGANDICUARO	19	57	102	12	1,630	1981-08	2000-12
208	16230	LAS CRUCES, TURICATO	18	35	101	5	800	1981-05	2003-05
209	16231	LAS CRUCES, PURUANDIRO	19	56	101	23	2,070	1982-07	2003-05
210	16232	EL DEVANADOR, TZITZIO	19	23	100	49		1982-01	1988-08
211	16233	EPITACIO HUERTA, E. H.	20	8	100	17	2,035	1982-07	2004-12
212	16234	GALEANA, PURUANDIRO	20	2	101	34	1,550	1981-05	2002-09
213	16235	HUAJUMBARO, CD. HIDALGO	19	40	100	44	2,285	1982-07	2005-01
214	16236	EL JOVERO, AGUILILLA	18	35	102	40		1981-06	1985-08
215	16237	EL LIMON, TIQUICHEO	18	52	100	38		1982-01	1988-07
216	16238	MELCHOR OCAMPO, TUZANTLA	19	5	100	39		1982-01	1983-12
217	16239	MEXCALHUACAN, LAZARO C.	18	4	102	39	5	1981-03	1997-05
218	16240	EL PUEBLITO, ZACAPU	19	49	101	55	1,100	1981-06	2003-04
219	16241	TEFETAN, TZITZIO	19	24	100	54	2,235	1978-09	1988-08
220	16242	TUMBISCATIO, TUMBISCATIO	18	32	102	22	900	1981-01	1985-09
221	16243	ZIQUITARO, PENJAMILLO	20	3	101	53	1,920	1981-06	2004-12
222	16244	TIQUICHEO, TIQUICHEO (DGE)	18	54	100	44		1978-06	1988-08
223	16245	SAN CARLOS, TIQUICHEO DGE	19	14	100	47		1978-08	1983-11
224	16247	CAPULA, MORELIA	19	40	101	23	1,708	1981-07	2005-01
225	16248	CAURIO DE GUADALUPE	19	55	101	51	2,020	1981-07	2004-12
226	16249	EL CHICO, AQUILA	18	35	103	31		1982-01	1983-12
227	16250	HUANDACAREO, CUITZEO	19	49	101	16	2,285	1982-07	2004-11
228	16251	PATAMBAN, TANGANCICUARO	19	53	102	13		1982-01	2001-12
229	16252	SAN LORENZO, HIDALGO	19	52	101	30	2,059	1982-01	1990-12
230	16253	TANACO, CHERAN	19	55	102	4	1,850	1982-01	1999-12
231	16254	TEREMENDO, MORELIA	19	47	101	28	2,350	1982-01	2004-12
232	16255	UCAREO, MORELIA	19	54	100	41	1,800	1981-01	2003-12
233	16257	STA. ISABEL DE AJUNO	19	30	101	35	2,250	1982-07	1988-08
234	16258	PRESA TUXPAN, TUXPAN	19	32	100	28		1982-07	1990-12
235	16259	TZINDIO, SAN JUAN NUEVO	19	24	102	6	1,850	1982-10	1985-06
236	16512	EL COLEGIO, MORELIA	19	42	101	11	1,900	1986-01	2004-08
237	16513	EL JACAL						1991-01	2004-08
238	16514	JARIPEO, LA PUNTA	19	27	100	30	1,300	1991-10	2004-06
239	16515	LA SANDIA, PATZCUARO	19	15	101	40	1,910	1980-04	2004-12

Fuente: Servicio Meteorológico Nacional (SMN), 2006.

Anexo B.

En este apartado se listan (Tabla B.1) las 111 estaciones seleccionadas de acuerdo al análisis preliminar efectuado a las 239 estaciones del Sistema CLICOM. El análisis consistió en seleccionar aquellas estaciones con registros suficientes, esto es, que tuvieran periodos de datos continuos con al menos 20 años a partir de 1981.

Tabla B.1. Lista de las 111 estaciones del Sistema CLICOM seleccionadas para analizar la calidad de la serie de datos.

Nº	Identificador	Nombre	Latitud		Longitud		Altitud msnm	Fecha	
			g	min	g	min		Inicio	Fin
1	16001	ACUITZIO DEL CANJE, (SMN)	19	29	101	20	2,070	1961-01	2004-07
2	16002	AGOSTITLAN, CD. HIDALGO	19	34	100	41		1950-07	2004-02
3	16006	ANTUNEZ, PARACUARO	19	0	102	12	360	1962-01	2002-12
4	16007	APATZINGAN, APATZINGAN	19	5	102	16	682	1922-09	2003-02
5	16008	AQUILA, AQUILA	18	25	103	23	100	1961-05	1999-04
6	16014	CAMECUARO, TANGANCICUARO	19	55	102	12	1,708	1951-06	2003-04
7	16015	CARAPAN, CHILCHOTA	19	51	102	2	1,980	1950-01	2004-09
8	16016	CARRILLO PUERTO, A. O.	19	55	101	20	1,840	1969-01	2004-10
9	16017	CASA BLANCA, PURUANDIRO	20	19	101	27	1,694	1949-03	2000-08
10	16020	CIUDAD HIDALGO (DGE)	19	42	100	34	2,020	1961-01	2004-10
11	16022	COINTZIO, MORELIA	19	37	101	16	2,096	1940-04	2005-01
12	16023	COPANDARO DE GALEANA,	19	57	101	37	1,500	1969-12	2001-04
13	16024	CORRALES (A. LOS CORRALES)	20	12	101	51	1,693	1934-10	2003-07
14	16027	CUITZEO, CUITZEO	19	57	101	8	1,831	1923-03	2003-01
15	16028	CUITZILLO GRANDE,	19	46	101	7	1,840	1969-07	2005-01
16	16030	CUMUATO, BRIENAS D MATAM	20	16	102	35	1,524	1951-07	1999-08
17	16031	CHAPARACO, ZAMORA	19	58	102	15	1,632	1948-02	2002-05
18	16033	CHINCUA, SENGUIO	19	46	100	17	2,415	1966-11	2004-07
19	16034	EL JACAL, CHUCANDIRO	19	53	101	20		1969-01	2002-12
20	16040	EL GIGANTE, MARAVATIO	19	58	100	27	1,961	1970-08	2003-09
21	16043	EL PUERTO, COTIJA	19	48	102	40	1,581	1943-08	2000-04
22	16045	EL TEMASCAL, CHARO	19	39	100	57	2,240	1965-01	2004-11
23	16047	EL ZAPOTE, LA HUACANA	18	55	102	46	920	1951-03	2001-02
24	16048	ZAMORA, ZAMORA (DGE)	20	0	102	16	1,567	1947-01	2003-04
25	16050	HUANIQUEO, HUANIQUEO	19	54	101	30	1,950	1948-07	2004-10
26	16051	HUETAMO, HUETAMO (SMN)	18	35	100	53	300	1922-05	2004-12
27	16052	HUINGO, ZINAPECUARO	19	55	100	51	1,830	1941-01	2004-12
28	16055	JESUS DEL MONTE, MORELIA	19	40	101	7	2,100	1935-11	2004-01
29	16057	JIMENEZ, VILLA JIMENEZ	19	56	101	45	1,980	1970-07	1999-03
30	16059	LA CAIMANERA, HUETAMO	18	27	100	53	305	1961-01	1999-12
31	16060	LA ESTANZUELA, IXTLAN	20	7	102	21	1,566	1951-01	1999-12
32	16061	LAGUNA DEL FRESNO,	19	28	100	25	2,070	1947-01	2004-12
33	16063	LA PALMA, V. CARRANZA	20	9	102	45	1,522	1935-02	2002-08
34	16065	LA PIEDAD CABADAS (DGE)	20	20	102	10	1,684	1925-01	2004-12
35	16067	LA RAYA, SAHUAYO	20	50	102	43	1,577	1948-07	2003-03
36	16074	LOS PANCHES, LA HUACANA	18	52	102	13	207	1964-11	1999-12
37	16075	LOS PINZANES, HUETAMO	18	32	101	6	372	1964-10	1999-12
38	16080	MORELIA, MORELIA	19	42	101	11	1,915	1986-05	2002-07
39	16081	MORELIA, MORELIA (DGE)	19	41	101	10	1,903	1947-05	2004-02
40	16084	PANINDICUARO,	19	58	101	45	1,638	1946-08	2002-12
41	16085	PARACUARO, PARACUARO	19	9	102	13	498	1969-06	2001-08
42	16086	PASTOR ORTIZ, PURUANDIRO	20	18	101	35	1,692	1967-09	2004-06
43	16087	PATZCUARO, PATZCUARO	19	38	101	36	2,035	1969-01	2004-12
44	16090	P. BLANCAS, BUENAVISTA	19	14	102	40	304	1963-02	2002-12
45	16091	PLANTA BOMBEO A. ZINZIMEO	19	49	101	40	1,384	1966-10	2005-01
46	16093	PSA. GONZALO, VISTAHERMOSA	20	18	102	26	1,533	1969-01	2003-01

Tabla B.1. Lista de las 111 estaciones del Sistema CLICOM seleccionadas para analizar la calidad de la serie de datos.

47	16094	PRESA GUARACHA, VILLAMAR	19	58	102	35	1,570	1944-01	1999-08
48	16095	PRESA JARIPO, VILLAMAR	19	57	102	36	1,624	1948-06	2000-08
49	16096	PSA. MALPAIS, QUERENDARO	19	49	100	52	1,859	1940-12	2004-12
50	16097	PSA. PUCUATO, CD. HIDALGO	19	37	100	41	2,480	1950-07	2004-10
51	16098	PSA. SABANETA, CD. HIDALGO	19	36	100	40	2,500	1952-12	2004-10
52	16105	QUIRIO, INDAPARAPEO	19	49	101	0	1,858	1963-06	2005-01
53	16107	SAN CARLOS, TIQUICHEO	19	14	100	43	685	1969-01	2001-07
54	16108	SAN CRISTOBAL, IXTLAN	20	12	102	26	1,523	1969-01	2000-08
55	16109	SAN DIEGO CURUCUPATZEO,	19	20	101	11	1,020	1922-01	2004-12
56	16114	S. MIGUEL DEL MONTE, MOR.	19	16	101	0	2,127	1963-12	2004-11
57	16118	SANTA FE, QUIROGA	19	49	101	22	2,056	1963-06	2003-02
58	16119	STA. RITA, COPANDARO DE G.	19	54	101	16	1,880	1964-10	2004-08
59	16120	SANTIAGO UNDAMEO, MORELIA	19	36	101	19	2,048	1953-12	2004-06
60	16121	SENGUIO, SENGUIO	19	44	100	22	2,222	1969-01	2001-08
61	16122	SUSUPUATO DE GUERRERO,	19	11	100	25	1,560	1945-08	2004-12
62	16123	TACAMBARO, TACAMBARO	19	14	101	28	1,755	1922-01	2003-06
63	16124	TEMASCALES, CONTEPEC	20	3	100	8	2,409	1943-06	2004-03
64	16125	TANGANCICUARO A. ETUCUARO	19	53	102	13	1,690	1947-07	2005-01
65	16129	TEPUXTEPEC, CONTEPEC	19	59	100	13	2,380	1947-02	2001-08
66	16133	TURICATO, TURICATO	19	3	101	25	900	1969-01	2004-06
67	16136	TZITZIO, TZITZIO	19	34	100	55	1,540	1969-01	2004-11
68	16137	UREPETIRO, TANGANCICUARO	19	58	102	8		1949-01	2001-12
69	16141	YURECUARO, YURECUARO	20	20	102	16	1,524	1923-03	2003-11
70	16142	ZACAPU, ZACAPU (DGE)	19	48	100	47	1,986	1968-08	2001-12
71	16145	ZINAPECUARO, ZINAPECUARO	19	53	100	40	1,920	1923-08	2004-12
72	16146	ZIRAHUEN, VILLA ESCALANTE	19	28	101	45	2,228	1947-10	2003-03
73	16151	ARTEAGA, ARTEAGA (CFE)	18	21	102	17	940	1957-11	2000-02
74	16159	EL ROSARIO, ANGAMACUIRO	20	7	101	45	1,720	1972-07	2003-01
75	16162	ORANDINO, JACONA	19	58	102	19	1,633	1961-01	2000-12
76	16165	URUAPAN, URUAPAN (CFE)	19	23	102	3	1,611	1962-10	1999-12
77	16175	TEPUXTEPEC, CONTEPEC	20	0	100	14		1969-02	2002-10
78	16186	EMILIANO ZAPATA, VILLAMAR	20	0	102	34	1,550	1973-01	1999-12
79	16188	TIRINDARO, ZACAPU	19	46	101	44	2,500	1973-12	2003-02
80	16192	ENCARNACION, ZITACUARO	19	25	100	24	1,720	1974-09	2004-06
81	16194	E.T.A. 39 TOCUMBO, (SMN)	19	42	102	30	1,600	1975-01	1999-12
82	16199	SAN MIGUEL CURAHUANGO,	19	53	100	27	2,020	1975-03	2002-09
83	16208	EL CACHAN, AQUILA	18	17	103	17	45	1977-05	2000-12
84	16213	IRIMBO, CIUDAD HIDALGO	19	42	100	29	2,235	1978-01	2004-12
85	16216	BUENAVISTA, BUENAVISTA	19	12	102	36	450	1976-01	1999-07
86	16219	SAN PEDRO, TIQUICHEO	19	4	100	49	530	1979-08	2003-10
87	16220	SAN SIMON, IXTLAN	20	3	102	24		1979-03	2001-01
88	16224	TANHUATO, TANHUATO (DGE)	20	17	102	20	1,540	1979-04	1999-08
89	16225	VILLA JIMENEZ, V. JIMENEZ	19	55	101	45	1,980	1980-01	2003-02
90	16227	LAZARO CARDENAS, L. C.	18	2	102	11	50	1981-01	1999-07
91	16228	ACAHUATO, APATZINGAN	19	8	102	19	1,000	1981-01	2002-09
92	16229	ADJUNTAS, TANGANCICUARO	19	57	102	12	1,630	1981-08	2000-12
93	16230	LAS CRUCES, TURICATO	18	35	101	5	800	1981-05	2003-05
94	16231	LAS CRUCES, PURUANDIRO	19	56	101	23	2,070	1982-07	2003-05
95	16233	EPITACIO HUERTA, E. H.	20	8	100	17	2,035	1982-07	2004-12
96	16234	GALEANA, PURUANDIRO	20	2	101	34	1,550	1981-05	2002-09
97	16235	HUAJUMBARO, CD. HIDALGO	19	40	100	44	2,285	1982-07	2005-01
98	16240	EL PUEBLITO, ZACAPU	19	49	101	55	1,100	1981-06	2003-04
99	16243	ZIQUITARO, PENJAMILLO	20	3	101	53	1,920	1981-06	2004-12
100	16245	SAN CARLOS, TIQUICHEO DGE	19	14	100	47		1978-08	1999-11
101	16247	CAPULA, MORELIA	19	40	101	23	1,708	1981-07	2005-01
102	16248	CAURIO DE GUADALUPE	19	55	101	51	2,020	1981-07	2004-12
103	16250	HUANDACAREO, CUITZEO	19	49	101	16	2,285	1982-07	2004-11
104	16251	PATAMBAN, TANGANCICUARO	19	53	102	13		1982-01	2001-12
105	16253	TANACO, CHERAN	19	55	102	4	1,850	1982-01	1999-12
106	16254	TEREMENDO, MORELIA	19	47	101	28	2,350	1982-01	2004-12
107	16255	UCAREO, MORELIA	19	54	100	41	1,800	1981-01	2003-12
108	16512	EL COLEGIO, MORELIA	19	42	101	11	1,900	1986-01	2004-08
109	16513	EL JACAL						1991-01	2004-08
110	16514	JARIPEO, LA PUNTA	19	27	100	30	1,300	1991-10	2004-06
111	16515	LA SANDIA, PATZCUARO	19	15	101	40	1,910	1980-04	2004-12

Anexo C.

En este anexo se describen los nodos utilizados en el software *Clementine* para evaluar la calidad de la serie de datos de las estaciones CLICOM y observatorios de Zamora y Morelia. En el proceso de evaluación, cada nodo fue asociado a uno o más nodos formando así la ruta que une el flujo de datos. Estos nodos representan las operaciones realizadas para la exploración y análisis del conjunto de datos. *Para información relacionada con los nodos utilizados en el Clementine se recomienda al lector revisar Clementine (2006b, 2006c).*

1. Nodo Archivo variable

El nodo denominado *Archivo Variable* fue utilizado para leer el conjunto de datos de los archivos disponibles. Archivos cuyas columnas contienen una colección de registros almacenados en el tiempo, de forma diaria para el caso de los archivos del sistema CLICOM, y horaria para los archivos obtenidos de los observatorios de Zamora y Morelia. En la *Figura C.1* se muestra el cuadro de opciones de configuración que dispone el nodo *Archivo Variable*.

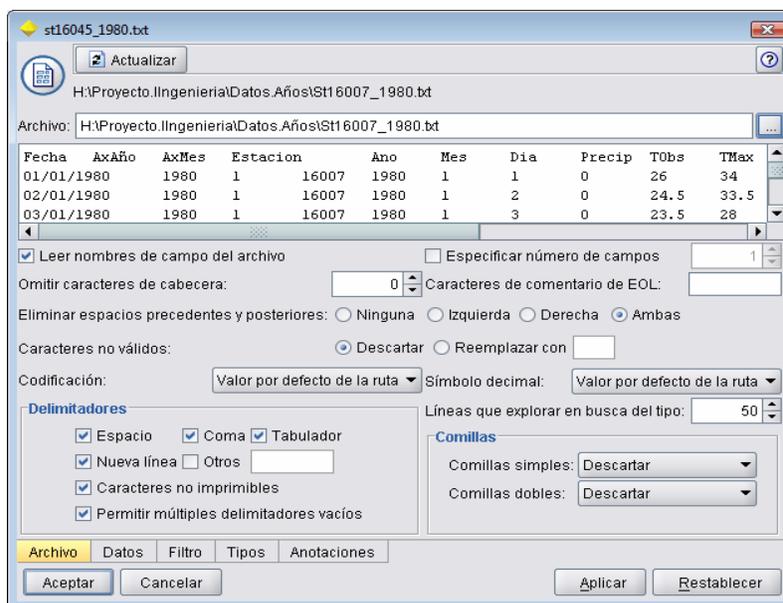


Figura C.1. Cuadro de opciones de configuración del nodo Archivo Variable.

De la *Figura C.1*, las opciones de configuración utilizadas para la lectura del conjunto de datos fueron:

- a. *Archivo*. En esta opción se estableció la ruta y el nombre del archivo de datos a ser utilizado.
- b. *Leer nombres de campo del archivo*. Esta opción permitió especificar que la primera fila del archivo sea considerada como nombres de las columnas. Si no se dispone de nombres en las variables se deshabilita esta opción, siendo el sistema quien asigna nombres como Campo1, Campo2, etc.
- c. *Eliminar espacios precedentes y posteriores*. Esta opción se seleccionó para descartar los espacios anteriores y posteriores en las cadenas de datos, es decir, eliminar los espacios de la derecha, izquierda o ambas del registro.
- d. *Caracteres no válidos*. Se seleccionó *Descartar* para eliminar los caracteres no válidos en el conjunto de datos, ej. nulos.
- e. *Codificación*. Es donde se especificó el método de codificación de textos, esto es, tomar por defecto el valor del sistema o si se trabaja en red del equipo servidor.
- f. *Delimitadores*. En esta opción se seleccionaron los delimitadores para los campos del archivo de datos, tales como: coma, espacio, tabulador, etc.
- g. *Líneas que explorar en busca del tipo*. Es donde se especificó el número de filas que debe explorar el software para verificar los tipos de datos de todas las variables.

En su ejecución, si el archivo no existe o no está disponible para leerlo se produce un mensaje de error y la ejecución finaliza. De no producirse ningún problema, se leen todos los registros existentes en él. El mensaje de error se puede asociar a problemas de ubicación en la ruta de datos o por el cambio de nombre del archivo.

2. Nodo *Tabla*

El nodo *Tabla* permitió crear una matriz a partir del conjunto de datos. Fue útil para examinar los valores de los registros y evaluar la cronología de la serie de datos. En este nodo se estableció dos variables de comparación: *fecha* y *hora*. La variable *fecha* se utilizó para la evaluación cronológica de las estaciones CLICOM, mientras que para los observatorios se emplearon ambas variables. La comparación consistió en verificar la secuencia de datos de las variables *Año*, *Mes*, *Día* y *Hora* según sea el caso, tomando como referencia el calendario del sistema. En la *Figura C.2* se muestra la matriz con el conjunto de datos generada por el nodo *Tabla*.

	Fecha	AxAño	AxMes	Estacion	Año	Mes	Día	Precip	TObs	TMax	TMin	Evap	c
1	01/01/1980	1980	1	16014	1980	1	1	0.000	8.000	24.0...	4.500	3.600	
2	02/01/1980	1980	1	16014	1980	1	2	0.000	6.000	24.5...	5.000	3.500	
3	03/01/1980	1980	1	16014	1980	1	3	0.000	9.000	21.5...	6.500	2.800	
4	04/01/1980	1980	1	16014	1980	1	4	0.000	1.500	22.5...	0.000	2.200	
5	05/01/1980	1980	1	16014	1980	1	5	0.000	6.000	23.0...	1.500	4.200	
6	06/01/1980	1980	1	16014	1980	1	6	0.000	6.000	24.0...	1.500	3.800	
7	07/01/1980	1980	1	16014	1980	1	7	0.000	4.000	25.5...	3.000	3.500	
8	08/01/1980	1980	1	16014	1980	1	8	0.000	6.000	25.5...	4.000	2.200	
9	09/01/1980	1980	1	16014	1980	1	9	0.000	4.500	26.5...	3.000	5.100	
10	10/01/1980	1980	1	16014	1980	1	10	0.000	5.000	26.5...	4.500	4.100	
11	11/01/1980	1980	1	16014	1980	1	11	0.000	4.000	28.5...	3.500	3.500	
12	12/01/1980	1980	1	16014	1980	1	12	0.000	4.500	29.0...	3.500	4.400	
13	13/01/1980	1980	1	16014	1980	1	13	0.000	6.000	26.0...	4.500	3.600	
14	14/01/1980	1980	1	16014	1980	1	14	0.000	4.500	26.5...	4.500	3.300	

Figura C.2. Matriz de datos generada por el nodo *Tabla*.

A partir de la matriz de datos (Figura C.2) se pueden realizar operaciones como: selección y copia de celdas, reordenar columnas y guardar e imprimir la tabla. Un factor relevante en la matriz fue la herramienta de búsqueda, la cual permitió examinar valores máximos, mínimos, y otros en el conjunto de datos.

3. Nodo *Tipo*

Después de realizada la importación de datos y corroborada la secuencia cronológica, se utilizó el nodo *Tipo* para especificar el tipo de datos de cada una de

las variables. Inicialmente el sistema toma por defecto los tipos establecidos en el archivo fuente, pero fue necesario verificar y determinar el tipo adecuado para evitar interrupciones en la ejecución del modelo. Asimismo, se permitió establecer los índices de exploración para la búsqueda de registros válidos, esto es, registros sin valores nulos y/o faltantes (null, -9999). Los índices fueron *Nombre, Año, Mes y Día*, que juntos identificaron de manera única a los registros de la serie, evitando la redundancia de datos. En la *Figura C.3* se muestran las opciones de configuración para este nodo.

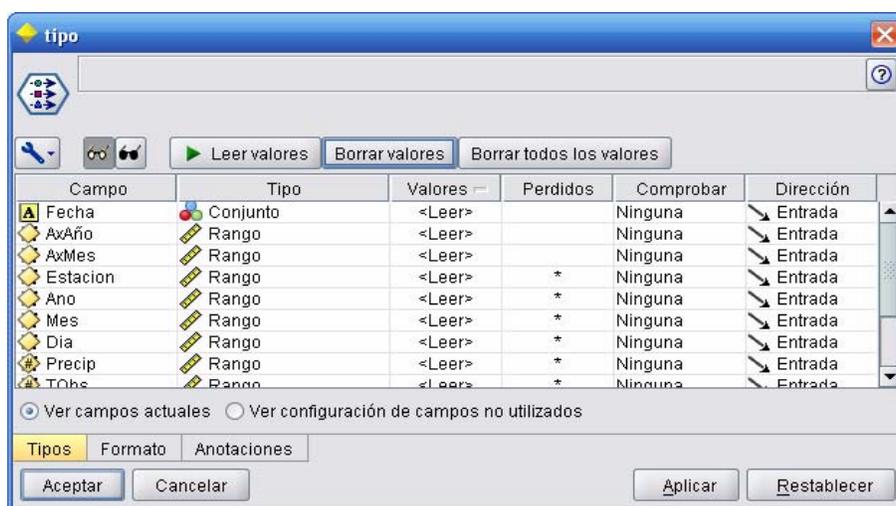


Figura C.3. Cuadro de opciones de configuración del el nodo Tipo.

Las propiedades de configuración utilizadas en el nodo Tipo fueron (*Figura C.3*):

- Tipo*. Se usó esta propiedad para especificar el tipo de dato en el conjunto de variables. Los tipos especificados fueron: cadena, número entero, número real, fecha y hora (marcas de tiempo).
- Dirección*. Se usó para especificar si las variables son de entrada o salida. En este caso para el análisis de datos todos los campos fueron de entrada.
- Valores nulos o perdidos*. Se usó para especificar qué valores dentro del archivo deben ser tomados como vacíos o nulos. Los valores nulos considerados para el análisis fueron: null, -9999 y espacios en blanco.

- d. *Instancias.* En la columna *Valores* se especificó la opción de lectura de datos para todas las variables. Esta lectura de datos permitió identificar los valores válidos, nulos y vacíos.

4. **Nodo *Seleccionar***

Una vez establecidos los tipos de datos del conjunto de variables, mediante el nodo *Seleccionar* se eligieron subconjuntos de datos de la serie temporal, esto con la finalidad de determinar la cantidad de registros válidos por año y variable, es decir, se filtraron los registros para cada año, iniciándose por lo general a partir de 1980 y 1981 en adelante. La *Figura C.4* muestra el cuadro de opciones del nodo *Seleccionar*, donde la condición de selección para elegir el subconjunto de datos fue la variable *Año*.

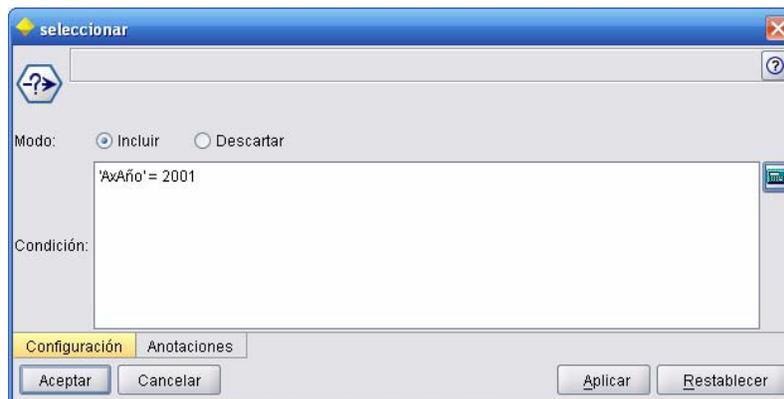


Figura C.4. Cuadro de opciones para filtrar registros mediante el nodo *Seleccionar*.

De la *Figura C.4*, las propiedades de configuración utilizadas en el nodo *Seleccionar* fueron:

- Modo.* Es el lugar donde se eligió la opción *incluir* para filtrar el subconjunto de registros.
- Condición.* Es donde se estableció el parámetro de selección para filtrar los registros por año, ej. $Año=2001$.

5. Nodo *Calidad*

En este nodo se configuraron las condiciones para determinar cuales valores son considerados como no válidos, ej. nulos –null, 9999-, cadenas vacías y espacios en blanco. Esta configuración permitió que todas las columnas sean analizadas registro por registro, con lo que una vez terminado el proceso de ejecución se emitió un reporte con el nombre de la variable, la cantidad porcentual y numérica de registros válidos y nulos, la cantidad de cadenas vacías y el número de espacios en blanco. En la *Figura C.5* se muestra el cuadro de opciones para este nodo.

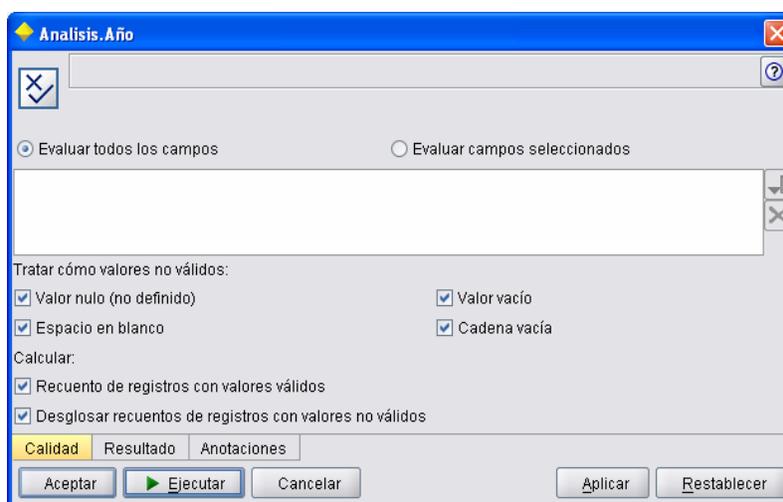


Figura C.5. Cuadro de opciones de configuración del nodo *Calidad*.

De la *Figura C.5*, las opciones utilizadas fueron:

- a. *Valor nulo*. Se eligió esta opción para que el sistema identifique valores nulos (null).
- b. *Cadena vacía*. Se consideró esta opción para que el sistema detecte cadenas vacías y lo considere como valores no válidos.
- c. *Espacio en blanco*. Mediante esta opción se estableció como valores no válidos los registros con espacios en blanco (cadenas sin caracteres visibles).
- d. *Recuento de registros con valores válidos*. Se seleccionó esta opción para mostrar el número de registros con valores válidos de todas las variables.

- e. *Desglosar recuentos de registros con valores no válidos.* Esta opción fue elegida para mostrar el número de registros con valores no válidos de todas las variables.
- f. *Resultados de análisis.* Mediante esta opción se permitió ver la cantidad de registros válidos por variable.

6. **Nodo Estadístico**

Este nodo se utilizó para obtener información estadística del subconjunto de datos. Esta información se obtuvo para cada una de las variables y se estimaron mediciones como: valor medio, mínimo, máximo, varianza, desviación típica, error típico, mediana, moda y recuento. El cuadro de opciones para configurar este nodo se presenta en la *Figura C.6*.



Figura C.6. Cuadro de opciones de configuración del nodo Estadístico.

De la *Figura C.6*, las opciones configuradas fueron:

- a. *Examinar.* En este lugar se seleccionó las variables para calcular las estimaciones estadísticas. Estas estimaciones se resumen de manera individual para cada una de las variables o campos seleccionados.

- b. *Estadísticos*. En esta parte se seleccionaron algunas opciones de cálculo estadístico. Las estimaciones fueron: recuento, media, valor mínimo, valor máximo, rango, varianza, desviación típica, error típico de la media, mediana y moda.

7. **Nodo Gráfico**

Se utilizó el nodo Gráfico para mostrar de manera representativa la relación de las variables numéricas y el tiempo en años. Estas representaciones agrupadas en una colección de puntos y en trazas lineales fueron útiles para analizar la secuencia temporal y las fluctuaciones de las variables durante el año. En la *Figura C.7* se muestra el cuadro de opciones que permiten generar gráficos a partir de los datos.



Figura C.7. Cuadro de opciones de configuración del nodo gráfico.

De la *Figura C.7*, las opciones de configuración fueron:

- Campo X*. Permitted selecting a variable from the data file for its representation on the x-axis.
- Campo Y*. Permitted selecting the variable *Año* for its representation on the y-axis.

Anexo D.

En este apartado se presentan los resultados de las estimaciones mensuales de las variables seleccionadas como significativas para el proceso de regionalización de las 52 estaciones CLICOM. Las estimaciones realizadas fueron: para la precipitación (P) la suma mensual y la media anual, ambas expresadas en milímetros (mm). En el caso de la temperatura observada (Tob), se calcularon las medias mensuales y anuales, expresadas grados centígrados ($^{\circ}\text{C}$). Para la temperatura -máxima (Tmx) y mínima (Tmn), se estimaron las medias mensuales y anuales de los máximos y mínimos, respectivamente, expresados en grados centígrados $^{\circ}\text{C}$. Los resultados fueron organizados en la *Tabla D.1* con el nombre oficial de la estación, altitud sobre el nivel del mar, variable climatológica, unidad de medida, meses del año y media anual.

Tabla D.1. Resultados de las estimaciones mensuales de las variables significativas para el proceso de regionalización de las 52 estaciones CLICOM.

Nº	ESTACIÓN	ELEM.	UM	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC	ANUAL
1	St16006 ANTUNEZ, PARACUARO 360 m.s.n.m.	P	mm	28.2	1.3	1.3	2.3	17.0	124.8	186.7	179.1	190.0	70.0	13.3	7.2	821.2
		Tob	°C	19.6	20.8	23.2	24.7	25.2	24.6	23.8	23.8	23.4	23.0	21.9	20.5	22.9
		Tmx	°C	36.1	38.4	40.2	41.5	42.3	42.2	40.5	38.2	37.5	37.1	37.2	36.6	39.0
		Tmn	°C	7.8	9.2	11.3	12.8	13.2	13.9	13.8	13.1	12.1	10.2	9.3	7.2	11.2
2	St16007 APATZINGAN, APATZINGAN 682 m.s.n.m.	P	mm	21.5	2.6	2.1	0.5	10.8	118.2	169.8	176.5	155.4	76.0	11.6	5.1	749.8
		Tob	°C	17.4	18.6	20.4	22.5	24.8	25.2	23.8	23.6	23.4	22.9	21.0	18.5	21.8
		Tmx	°C	37.3	39.5	41.0	42.7	43.3	43.2	41.3	39.8	38.7	38.1	37.8	37.4	40.0
		Tmn	°C	5.9	7.8	9.8	12.5	14.2	14.8	14.3	13.7	13.4	13.1	11.5	9.6	11.7
3	St16014 CAMECUARO, TANGANCICUARO 1708 m.s.n.m.	P	mm	18.9	5.9	4.3	7.4	39.2	147.8	211.1	211.1	158.7	56.3	15.2	12.2	888.2
		Tob	°C	7.1	8.2	10.0	13.2	16.0	17.6	16.7	16.3	15.7	13.2	10.2	7.9	12.7
		Tmx	°C	30.5	32.5	35.2	37.4	37.5	38.0	35.3	33.5	32.2	31.2	30.5	30.4	33.7
		Tmn	°C	-3.3	-2.5	-1.5	1.5	4.6	5.4	5.4	5.2	4.3	0.5	-0.5	-1.6	1.5
4	St16016 CARRILLO PUERTO, A. O. 1840 m.s.n.m.	P	mm	19.5	6.4	7.1	13.2	35.4	124.1	156.7	152.3	118.0	42.9	15.3	7.3	698.1
		Tob	°C	6.3	7.9	10.0	12.6	15.0	15.5	14.9	14.6	14.3	12.5	10.0	7.3	11.7
		Tmx	°C	30.5	32.5	35.2	37.4	37.5	38.0	35.3	33.5	32.2	31.2	30.5	30.4	33.7
		Tmn	°C	-2.5	-1.0	0.3	2.9	5.0	6.6	5.6	5.5	4.3	2.1	0.7	-2.0	2.3
5	St16017 CASA BLANCA, PURUANDIRO 1694 m.s.n.m.	P	mm	18.0	6.6	6.8	8.1	28.2	117.8	217.3	164.7	127.3	48.6	11.5	9.1	764.0
		Tob	°C	7.0	8.4	10.8	13.9	16.9	18.3	16.6	16.3	15.8	13.3	10.3	7.8	13.0
		Tmx	°C	29.8	32.5	36.2	38.0	39.2	37.5	34.3	33.1	32.0	31.5	31.0	30.8	33.8
		Tmn	°C	-2.7	-2.0	0.7	2.7	5.7	6.5	7.1	5.7	4.6	1.6	-1.1	-2.8	2.2
6	St16020 CIUDAD HIDALGO (DGE) 2020 m.s.n.m.	P	mm	23.9	9.1	7.3	20.3	59.9	138.5	167.0	172.9	128.6	61.4	17.7	8.6	815.3
		Tob	°C	5.2	6.0	7.2	9.7	12.1	13.9	13.2	13.1	13.1	11.1	8.0	6.1	9.9
		Tmx	°C	31.3	33.4	35.0	36.5	37.1	37.9	34.8	33.1	32.8	32.2	31.7	31.2	33.9
		Tmn	°C	-3.2	-2.6	-1.1	1.5	4.6	5.6	5.4	4.7	3.2	0.1	-1.1	-2.5	1.2
7	St16023 COPANDARO DE GALEANA 1500 m.s.n.m.	P	mm	13.8	5.2	5.4	7.4	31.6	145.9	191.9	203.9	170.7	56.2	14.7	15.4	862.1
		Tob	°C	5.2	6.4	8.6	11.9	14.7	15.1	14.5	14.0	13.3	10.7	8.4	5.9	10.7
		Tmx	°C	26.6	28.5	31.1	32.6	33.9	34.1	31.0	29.7	28.6	28.2	27.6	26.9	29.9
		Tmn	°C	-4.1	-2.2	0.6	2.9	4.6	6.3	5.9	5.2	3.4	1.4	-1.2	-3.2	1.6
8	St16024 CORRALES (A. LOS CORRALES) 1693 m.s.n.m.	P	mm	14.5	2.9	3.2	7.7	26.0	130.4	182.2	149.3	107.6	45.8	9.6	5.5	684.7
		Tob	°C	7.2	8.7	11.2	14.1	16.8	17.7	16.5	16.2	15.4	12.9	10.1	8.1	12.9
		Tmx	°C	28.5	30.6	34.2	36.5	38.5	38.9	36.2	35.5	34.2	33.5	32.6	31.9	34.3
		Tmn	°C	0.2	1.3	2.2	4.6	8.0	8.8	8.1	7.2	4.9	2.9	1.3	0.3	4.1
9	St16027 CUITZEO, CUITZEO 1831 m.s.n.m.	P	mm	22.3	4.7	5.1	8.1	29.8	122.9	158.8	179.2	108.1	44.0	8.6	7.3	698.8
		Tob	°C	9.8	10.9	13.2	15.3	17.4	17.8	16.5	16.3	15.9	14.6	12.9	11.2	14.3
		Tmx	°C	29.3	32.5	35.0	37.2	38.3	36.5	34.3	33.5	32.3	30.7	30.1	29.6	33.3
		Tmn	°C	-2.2	-1.0	0.8	4.0	6.4	7.5	7.1	6.6	5.3	3.0	1.1	-0.8	3.2
10	St16028 CUITZILLO GRANDE 1840 m.s.n.m.	P	mm	13.1	6.4	7.5	18.2	33.0	122.4	154.7	142.5	114.4	42.3	14.5	6.0	674.8
		Tob	°C	5.6	6.5	8.8	11.9	14.8	16.0	15.0	14.7	14.0	11.6	7.9	5.8	11.1
		Tmx	°C	29.5	32.0	34.1	35.8	37.1	37.6	34.3	31.7	30.3	29.3	28.7	28.1	32.4
		Tmn	°C	-2.7	-1.2	1.8	3.3	5.9	6.7	6.1	5.9	3.8	1.3	-1.0	-3.1	2.2

Tabla D.1. Resultados de las estimaciones mensuales de las variables significativas para el proceso de regionalización de las 52 estaciones CLICOM.

11	St16031 CHAPARACO, ZAMORA 1632 m.s.n.m.	P	mm	23.0	3.9	3.4	3.9	33.4	139.2	173.1	174.3	126.6	38.6	9.8	7.9	737.1
		Tob	°C	8.4	9.6	11.7	14.5	17.6	18.5	16.8	16.3	15.8	13.7	11.1	9.3	13.6
		Tmx	°C	29.6	32.5	34.6	36.5	38.1	37.5	36.2	32.5	31.6	30.7	30.1	29.7	33.3
		Tmn	°C	0.9	2.3	4.3	6.1	9.5	10.3	10.0	9.4	8.2	6.1	3.8	2.2	6.1
12	St16033 CHINCUA, SENGUIO 2415 m.s.n.m.	P	mm	24.0	6.2	7.0	18.6	47.6	142.3	223.3	177.1	138.6	55.8	16.8	7.9	865.3
		Tob	°C	6.0	6.6	7.7	9.2	10.7	12.0	11.1	11.2	11.5	10.2	8.5	7.0	9.3
		Tmx	°C	26.8	29.4	31.2	32.9	34.2	34.2	32.7	29.6	28.3	28.1	27.4	27.3	30.2
		Tmn	°C	-1.2	-0.3	1.6	3.4	6.7	7.2	6.7	5.6	4.3	2.2	1.3	0.8	3.2
13	St16043 EL PUERTO, COTIJA 1581 m.s.n.m.	P	mm	35.2	3.3	3.2	7.5	35.2	141.8	170.1	149.9	102.8	57.0	20.1	11.1	737.1
		Tob	°C	6.9	7.9	9.7	12.8	16.2	18.2	17.2	16.7	16.2	14.0	10.8	8.2	12.9
		Tmx	°C	30.1	32.2	34.7	35.4	36.8	36.2	34.5	32.3	31.2	30.7	30.3	29.6	32.8
		Tmn	°C	-1.2	0.1	2.2	4.9	7.6	8.4	7.6	7.1	6.1	3.5	1.3	-0.6	3.9
14	St16045 EL TEMASCAL, CHARO 2240 m.s.n.m.	P	mm	30.2	7.5	8.7	11.9	53.6	219.6	290.6	293.4	246.0	99.0	23.9	9.3	1293.7
		Tob	°C	11.5	12.2	13.6	15.5	16.4	15.7	14.7	14.5	14.6	14.1	13.1	12.0	14.0
		Tmx	°C	26.8	29.1	31.1	33.9	34.2	33.7	30.1	28.6	28.1	27.5	26.9	26.3	29.7
		Tmn	°C	1.2	2.5	4.7	6.1	7.7	8.4	7.7	7.0	5.6	4.1	2.4	1.2	4.9
15	St16048 ZAMORA, ZAMORA (DGE) 1567 m.s.n.m.	P	mm	20.5	6.2	3.8	5.4	36.3	154.1	183.3	198.8	135.7	44.3	12.8	8.5	809.6
		Tob	°C	9.4	10.6	12.3	14.8	17.0	18.2	17.4	17.2	16.6	14.8	12.1	10.3	14.2
		Tmx	°C	31.6	33.5	35.3	37.0	38.2	38.5	36.2	34.5	34.1	33.1	33.1	32.3	34.8
		Tmn	°C	0.5	1.1	2.6	4.6	7.8	8.5	8.1	6.9	5.5	3.8	2.1	1.5	4.4
16	St16050 HUANIQUEO, HUANIQUEO 1950 m.s.n.m.	P	mm	20.4	6.5	6.0	15.2	44.1	172.0	205.7	172.8	153.8	60.6	18.6	8.7	884.2
		Tob	°C	8.1	9.3	11.7	14.8	17.0	17.2	16.0	15.8	15.2	13.2	10.7	9.0	13.2
		Tmx	°C	27.5	30.3	32.5	33.6	34.7	35.2	31.7	29.5	29.0	28.6	27.9	27.6	30.7
		Tmn	°C	-0.6	1.3	2.3	4.2	7.0	8.6	8.2	7.5	6.2	3.5	2.0	0.6	4.2
17	St16052 HUINGO, ZINAPECUARO 1830 m.s.n.m.	P	mm	19.2	4.4	6.2	11.2	28.2	129.7	191.1	185.1	136.4	45.4	11.5	6.3	774.5
		Tob	°C	8.1	9.2	11.1	14.7	16.5	17.1	16.8	16.6	16.1	14.6	11.9	9.1	13.5
		Tmx	°C	28.0	30.3	32.5	34.6	35.9	36.2	34.7	32.6	31.3	30.3	28.5	28.1	31.9
		Tmn	°C	-4.1	-3.5	0.8	2.8	4.7	6.9	7.0	5.5	3.0	0.5	-1.4	-3.5	1.6
18	St16057 JIMENEZ, VILLA JIMENEZ 1980 m.s.n.m.	P	mm	13.5	4.6	6.8	8.9	36.6	154.5	218.8	204.5	175.4	55.4	13.8	9.1	901.7
		Tob	°C	4.4	5.5	7.8	10.8	13.9	15.8	14.8	14.4	13.9	11.0	7.8	5.7	10.5
		Tmx	°C	30.2	31.7	33.6	35.0	36.4	36.2	32.8	31.3	31.1	30.2	30.0	29.4	32.4
		Tmn	°C	-4.2	-2.5	-0.5	1.3	3.7	6.6	6.5	5.8	4.2	0.4	-1.7	-3.1	1.4
19	St16059 LA CAIMANERA, HUETAMO 305 m.s.n.m.	P	mm	20.6	1.8	2.8	3.5	18.1	211.6	235.5	214.1	167.9	62.1	14.5	5.3	957.8
		Tob	°C	16.8	18.9	21.7	25.2	27.6	26.2	24.0	23.6	23.2	22.4	20.1	17.5	22.3
		Tmx	°C	38.7	40.2	42.1	43.4	44.6	43.7	42.0	40.2	39.5	39.1	38.4	38.2	40.8
		Tmn	°C	5.6	8.2	10.7	13.8	16.0	16.4	15.9	15.3	13.5	11.7	8.7	6.7	11.9
20	St16074 LOS PANCHES, LA HUACANA 207 m.s.n.m.	P	mm	31.4	1.8	2.0	1.5	19.5	93.8	133.4	145.4	129.7	71.6	11.1	6.5	647.7
		Tob	°C	15.6	16.5	18.0	21.2	24.5	26.4	25.0	25.0	24.3	23.2	20.5	17.5	21.5
		Tmx	°C	38.5	40.3	42.4	43.7	44.7	44.6	42.0	40.2	39.5	39.5	39.1	38.7	41.1
		Tmn	°C	7.9	8.9	10.1	11.5	13.8	15.2	14.7	14.3	13.7	11.9	9.7	7.4	11.6
21	St16075 LOS PINZANES, HUETAMO 372 m.s.n.m.	P	mm	19.5	1.7	2.8	3.3	17.6	173.5	195.9	173.0	164.6	65.9	8.8	5.1	831.8
		Tob	°C	19.8	21.1	22.7	25.8	28.2	26.7	24.3	24.0	23.7	23.0	21.7	19.8	23.4
		Tmx	°C	39.2	41.1	42.8	45.0	45.5	45.2	42.3	40.5	40.0	39.5	39.3	39.4	41.7

Tabla D.1. Resultados de las estimaciones mensuales de las variables significativas para el proceso de regionalización de las 52 estaciones CLICOM.

		Tmn	°C	9.3	10.7	12.9	14.8	17.9	18.1	17.9	17.7	16.3	13.9	12.0	10.4	14.3
22	St16081	P	mm	19.4	6.5	7.1	11.8	45.3	142.9	170.7	167.2	125.9	49.0	12.1	5.0	762.7
	MORELIA, MORELIA (DGE)	Tob	°C	8.3	9.7	11.6	13.8	15.4	15.8	15.2	15.2	14.8	13.3	10.9	9.1	12.8
	1903 m.s.n.m.	Tmx	°C	30.1	32.4	34.2	35.8	37.2	36.2	33.5	32.9	31.8	31.2	30.9	30.3	33.0
		Tmn	°C	-0.2	0.2	1.7	4.1	6.6	7.3	7.1	6.1	5.3	3.2	0.5	-0.5	3.4
23	St16090	P	mm	29.0	5.6	2.0	0.7	14.8	117.7	145.6	135.1	105.0	52.7	16.0	6.2	630.3
	P. BLANCAS, BUENAVISTA	Tob	°C	19.3	20.0	21.3	23.3	25.3	25.7	23.9	23.7	23.4	23.0	21.7	20.2	22.6
	304 m.s.n.m.	Tmx	°C	38.4	40.1	42.2	44.2	45.1	43.1	42.1	40.5	39.7	39.8	39.2	38.8	41.1
		Tmn	°C	10.2	11.1	13.2	14.8	17.5	17.5	17.4	17.0	16.1	13.9	12.1	10.7	14.3
24	St16091	P	mm	17.3	5.8	6.6	14.9	33.3	112.6	153.1	136.0	114.7	39.3	13.0	5.6	652.2
	PLANTA BOMBEO A. ZINZIMEO	Tob	°C	7.2	8.8	10.5	12.8	15.0	16.5	16.2	15.8	15.3	13.2	9.9	8.0	12.4
	1384 m.s.n.m.	Tmx	°C	30.2	31.2	34.8	35.6	37.3	37.1	34.8	33.7	33.0	32.1	31.2	30.4	33.5
		Tmn	°C	-1.1	0.6	1.4	4.0	6.7	7.2	6.6	6.0	4.6	2.1	0.2	-1.2	3.1
25	St16093	P	mm	13.9	3.2	4.7	3.2	26.6	171.9	211.6	184.6	129.5	46.9	10.6	10.7	817.4
	PSA. GONZALO, VISTA HERMOSA	Tob	°C	9.1	10.3	11.9	15.1	18.1	19.4	18.6	18.1	17.7	16.0	12.2	9.7	14.7
	1533 m.s.n.m.	Tmx	°C	32.3	34.6	36.3	37.8	39.9	39.4	37.6	35.0	34.5	34.1	33.3	33.0	35.6
		Tmn	°C	-1.9	-0.3	1.6	3.9	5.5	7.6	7.4	6.5	3.7	1.5	0.2	-1.6	2.8
26	St16094	P	mm	32.1	4.0	2.4	6.6	43.6	187.5	187.9	161.7	130.1	60.5	17.9	10.1	844.5
	PRESA GUARACHA, VILLAMAR	Tob	°C	8.2	9.3	12.1	14.7	17.2	18.5	17.2	16.7	16.4	14.4	11.7	9.3	13.8
	1570 m.s.n.m.	Tmx	°C	34.4	36.6	37.7	39.1	39.6	39.3	37.6	35.4	34.5	33.3	33.1	32.8	36.1
		Tmn	°C	0.6	1.7	2.8	6.5	8.3	9.3	9.1	7.8	6.3	4.0	2.3	0.8	5.0
27	St16095	P	mm	29.3	3.9	2.7	7.0	37.3	165.1	182.5	154.6	111.7	54.4	17.8	8.4	774.8
	PRESA JARIPO, VILLAMAR	Tob	°C	8.1	9.3	12.0	14.5	16.9	18.0	16.7	16.3	15.9	13.8	11.5	9.3	13.5
	1624 m.s.n.m.	Tmx	°C	30.4	32.5	35.1	36.1	38.0	37.4	34.2	33.3	32.6	31.8	31.3	30.5	33.6
		Tmn	°C	-0.3	0.8	2.7	5.5	7.6	8.9	8.8	8.4	6.5	4.3	2.8	0.7	4.7
28	St16107	P	mm	19.7	2.9	4.2	2.6	28.4	208.1	330.3	309.7	267.0	101.3	17.6	4.2	1296.0
	SAN CARLOS, TIQUICHEO	Tob	°C	13.5	14.6	16.3	19.7	22.5	22.8	20.9	20.4	20.3	19.2	16.6	14.0	18.4
	685 m.s.n.m.	Tmx	°C	35.9	38.6	40.5	42.7	44.1	43.9	39.9	37.8	36.5	36.2	35.8	35.7	39.0
		Tmn	°C	4.9	6.0	7.7	10.3	11.9	12.2	12.0	11.8	10.1	8.4	6.1	4.6	8.8
29	St16109	P	mm	15.6	2.8	3.8	1.9	33.7	189.8	260.4	245.3	193.5	86.0	18.6	2.0	1053.3
	SAN DIEGO, CURUCUPATZEO	Tob	°C	15.1	16.1	18.3	20.1	21.4	20.8	19.8	19.9	19.5	18.6	17.5	15.9	18.6
	1020 m.s.n.m.	Tmx	°C	33.8	35.9	38.0	40.0	41.0	40.6	39.3	37.5	36.6	36.3	35.1	34.2	37.4
		Tmn	°C	6.6	8.3	9.4	12.0	13.7	14.2	13.8	13.6	12.5	10.5	9.1	7.4	10.9
30	St16118	P	mm	28.9	4.6	2.9	3.3	27.7	128.1	152.2	139.9	95.7	46.6	7.4	5.1	642.3
	SANTA FE, QUIROGA	Tob	°C	7.6	8.8	10.4	12.8	13.6	14.1	13.4	14.0	13.6	12.3	9.8	8.2	11.6
	2056 m.s.n.m.	Tmx	°C	31.1	32.7	34.7	36.0	38.2	37.1	34.4	32.2	32.1	31.2	30.9	30.6	33.4
		Tmn	°C	-3.1	-0.7	2.4	4.3	6.1	7.1	6.5	6.1	4.0	2.3	0.8	-2.2	2.8
31	St16120	P	mm	15.2	8.3	8.2	11.3	37.6	134.8	183.8	186.7	153.9	52.5	18.9	7.7	684.0
	SANTIAGO UNDAMEO, MORELIA	Tob	°C	4.3	5.6	8.0	11.0	13.7	15.4	14.5	14.3	13.9	11.2	7.5	5.2	10.4
	2048 m.s.n.m.	Tmx	°C	29.5	32.2	34.3	36.8	38.4	37.6	34.6	33.0	31.3	30.4	29.6	29.1	33.1
		Tmn	°C	-3.9	-2.1	1.1	3.5	5.3	6.1	6.0	5.7	3.7	1.0	-1.8	-3.4	1.8
32	St16123	P	mm	34.8	5.0	4.2	6.4	56.5	207.8	239.5	221.9	210.6	83.7	18.2	3.7	1092.2
	TACAMBARO, TACAMBARO	Tob	°C	14.8	15.4	16.8	18.7	19.5	18.6	17.8	17.9	17.8	17.5	17.4	16.6	17.4

Tabla D.1. Resultados de las estimaciones mensuales de las variables significativas para el proceso de regionalización de las 52 estaciones CLICOM.

	1755 m.s.n.m.	Tmx	°C	29.9	32.1	36.0	37.3	38.7	38.4	36.1	34.6	33.9	32.1	31.5	30.8	34.3
		Tmn	°C	-1.9	0.4	2.3	4.4	6.5	6.9	6.6	5.9	5.0	3.1	1.5	-0.8	3.3
33	St16124 TEMASCALES, CONTEPEC 2409 m.s.n.m.	P	mm	16.6	11.9	13.6	30.5	49.3	119.2	183.6	149.2	129.9	66.1	18.5	6.3	794.7
		Tob	°C	2.1	3.7	6.7	10.4	13.1	14.3	13.4	12.9	12.2	9.3	5.9	3.1	8.9
		Tmx	°C	25.1	27.5	29.8	31.7	33.1	32.8	29.9	28.2	26.9	26.2	26.1	26.0	28.6
		Tmn	°C	-4.9	-3.2	-0.4	1.2	4.3	5.1	5.3	4.5	2.4	0.3	-2.2	-4.0	0.7
34	St16133 TURICATO, TURICATO 900 m.s.n.m.	P	mm	21.7	2.8	5.6	1.7	29.0	164.9	217.2	181.6	140.9	56.0	24.0	5.4	850.8
		Tob	°C	12.7	13.2	15.2	17.9	20.4	22.0	21.0	20.6	20.3	18.9	16.5	14.1	17.7
		Tmx	°C	35.5	37.3	39.5	41.7	42.0	40.6	39.3	37.5	36.6	36.3	36.1	35.8	38.2
		Tmn	°C	2.6	3.8	5.2	7.3	10.2	11.5	11.5	11.0	8.4	5.9	4.0	2.2	7.0
35	St16136 TZITZIO, TZITZIO 1540 m.s.n.m.	P	mm	31.8	7.6	4.7	9.5	46.0	186.2	267.5	251.8	211.2	85.4	23.1	9.9	1134.4
		Tob	°C	12.8	13.9	16.1	19.2	20.3	19.4	18.4	18.2	18.1	17.5	16.1	13.8	17.0
		Tmx	°C	28.1	30.5	32.6	34.1	35.5	35.7	32.3	31.3	30.5	29.1	28.8	28.1	31.4
		Tmn	°C	3.8	4.9	7.0	9.5	10.1	11.6	11.1	10.9	9.2	8.2	6.8	4.8	8.2
36	St16141 YURECUARO, YURECUARO 1524 m.s.n.m.	P	mm	16.0	4.6	3.7	6.2	22.3	166.0	203.0	168.8	130.0	45.9	9.2	4.9	780.4
		Tob	°C	10.5	12.0	13.8	16.9	19.8	20.4	18.9	18.7	18.3	16.9	14.1	12.0	16.0
		Tmx	°C	29.6	32.9	36.1	37.4	38.0	38.1	36.5	34.7	33.3	32.2	31.8	31.2	34.3
		Tmn	°C	-2.1	-1.0	1.8	3.7	6.1	6.5	6.2	5.5	4.8	2.4	0.9	-1.1	2.8
37	St16142 ZACAPU, ZACAPU (DGE) 1986 m.s.n.m.	P	mm	22.4	6.4	6.3	13.6	41.6	141.8	217.5	197.7	157.6	68.6	12.5	13.9	899.9
		Tob	°C	5.1	6.3	9.7	13.0	15.6	16.9	15.3	14.6	13.5	11.2	8.8	6.3	11.4
		Tmx	°C	29.1	31.0	33.2	35.4	36.1	35.5	33.5	32.1	31.8	31.3	30.3	29.9	32.4
		Tmn	°C	-3.8	-1.7	2.5	4.6	6.7	7.6	7.5	6.1	4.4	2.5	0.4	-3.1	2.8
38	St16151 ARTEAGA, ARTEAGA (CFE) 940 m.s.n.m.	P	mm	48.6	2.6	1.9	0.6	22.5	123.0	175.1	171.9	180.2	121.5	24.0	8.7	880.8
		Tob	°C	13.7	13.7	14.4	16.1	18.9	21.1	20.9	20.3	19.9	19.3	17.7	15.7	17.6
		Tmx	°C	33.6	35.0	36.8	38.2	39.0	39.3	38.1	36.0	35.3	35.2	34.8	34.2	36.3
		Tmn	°C	3.3	4.3	5.9	7.9	10.8	11.5	11.1	10.4	8.4	6.6	4.8	3.0	7.3
39	St16159 EL ROSARIO, ANGAMACUTIRO 1720 m.s.n.m.	P	mm	10.3	5.2	3.3	6.0	32.3	132.1	200.1	177.9	134.4	46.8	11.5	5.1	765.0
		Tob	°C	10.1	11.5	13.4	15.4	17.4	17.6	16.9	16.4	16.1	14.0	12.6	11.3	14.4
		Tmx	°C	31.3	34.4	36.7	38.6	40.1	40.5	35.7	32.6	31.5	30.6	30.4	29.9	34.4
		Tmn	°C	0.9	1.9	3.4	5.5	6.9	7.8	7.6	6.1	5.3	3.3	2.4	1.1	4.4
40	St16162 ORANDINO, JACONA 1633 m.s.n.m.	P	mm	16.1	5.4	4.4	5.5	48.7	154.0	205.3	190.3	125.6	39.7	12.5	8.8	816.1
		Tob	°C	8.6	9.5	11.1	13.8	16.3	17.3	16.3	16.3	15.7	13.7	10.9	8.9	13.2
		Tmx	°C	30.1	32.2	34.7	36.3	38.8	39.2	34.8	32.2	31.4	31.0	30.1	29.3	33.3
		Tmn	°C	-0.4	0.5	3.1	5.3	7.5	8.4	8.0	7.7	6.3	3.4	1.5	-0.1	4.3
41	St16165 URUAPAN, URUAPAN (CFE) 1611 m.s.n.m.	P	mm	52.3	6.4	4.7	4.7	28.7	271.1	298.0	320.0	264.6	136.9	28.4	12.6	1428.4
		Tob	°C	9.7	10.1	11.7	14.1	15.9	17.3	16.6	16.4	16.3	14.6	12.8	11.0	13.9
		Tmx	°C	26.5	28.4	30.7	32.8	34.8	35.2	32.1	30.1	28.2	27.0	26.4	26.2	29.9
		Tmn	°C	1.6	2.6	3.2	5.7	7.1	8.1	7.9	6.9	5.5	4.2	2.8	1.8	4.8
42	St16188 TIRINDARO, ZACAPU 2500 m.s.n.m.	P	mm	26.2	4.1	5.4	13.0	44.0	117.2	167.9	180.2	150.5	59.5	14.8	9.2	792.0
		Tob	°C	4.1	5.3	8.1	11.6	14.1	15.6	15.2	14.9	14.4	11.1	7.6	5.0	10.6
		Tmx	°C	27.3	29.7	31.5	33.5	35.0	35.2	32.4	30.0	29.3	29.0	28.3	28.0	30.8
		Tmn	°C	-4.5	-3.1	-0.7	2.0	4.3	5.9	5.6	5.4	4.2	2.0	-0.6	-2.6	1.5
43	St16192	P	mm	33.8	6.6	16.9	9.6	31.0	172.6	185.3	170.8	150.1	59.7	15.3	8.6	860.3

Tabla D.1. Resultados de las estimaciones mensuales de las variables significativas para el proceso de regionalización de las 52 estaciones CLICOM.

	ENCARNACION, ZITACUARO	Tob	°C	10.1	10.7	12.1	14.4	16.7	17.4	16.2	15.7	15.6	14.6	13.0	11.1	14.0
	1720 m.s.n.m.	Tmx	°C	29.9	32.3	34.6	36.3	37.3	36.3	34.3	32.2	31.3	31.0	30.3	30.1	33.0
		Tmn	°C	1.8	2.7	4.4	5.9	7.7	8.4	8.2	6.9	5.5	4.1	2.7	1.8	5.0
44	St16194	P	mm	39.3	6.0	2.9	3.8	24.0	175.6	266.6	272.3	204.0	101.0	26.8	10.4	1132.8
	E.T.A. 39 TOCUMBO, (SMN)	Tob	°C	9.4	9.7	11.4	13.6	15.2	15.8	15.1	15.5	15.9	14.7	12.6	10.7	13.3
	1600 m.s.n.m.	Tmx	°C	30.2	33.2	34.3	37.4	37.9	36.1	33.3	32.5	31.1	30.4	29.8	29.1	32.9
		Tmn	°C	-3.3	-1.0	1.5	3.7	5.6	6.4	6.3	5.6	3.8	2.8	1.2	-1.0	2.6
45	St16208	P	mm	42.8	1.5	0.1	0.0	12.2	123.7	142.2	193.3	124.2	88.6	24.6	12.4	865.6
	EL CACHAN, AQUILA	Tob	°C	18.5	18.5	18.9	20.4	23.1	25.4	25.2	24.6	24.2	23.6	22.2	19.9	22.0
	45 m.s.n.m.	Tmx	°C	34.6	37.1	39.1	40.4	41.8	40.4	38.2	36.3	35.2	34.7	34.1	33.7	37.1
		Tmn	°C	8.8	10.1	12.2	14.3	16.1	17.1	16.9	16.4	14.5	12.8	10.7	8.7	13.2
46	St16219	P	mm	16.8	1.5	4.5	1.4	26.6	174.3	219.3	202.2	184.6	78.4	14.9	5.3	929.8
	SAN PEDRO, TIQUICHEO	Tob	°C	14.2	16.4	18.9	22.2	24.8	24.9	22.8	22.6	22.4	21.3	18.0	15.3	20.3
	530 m.s.n.m.	Tmx	°C	36.7	39.3	42.5	43.7	45.3	44.1	41.2	39.3	37.6	37.1	37.0	36.4	40.0
		Tmn	°C	4.2	5.4	7.5	9.8	11.9	12.5	12.1	11.6	9.5	7.5	5.5	3.8	8.4
47	St16225	P	mm	14.5	6.0	6.5	10.8	35.0	156.2	210.0	204.8	160.9	56.5	15.3	8.7	885.4
	VILLA JIMENEZ, V. JIMENEZ	Tob	°C	4.2	5.4	7.9	10.7	13.8	15.6	14.6	14.1	13.8	11.0	7.3	5.3	10.3
	1980 m.s.n.m.	Tmx	°C	31.5	33.0	35.5	37.1	38.4	37.2	34.3	33.8	33.4	32.2	31.8	31.7	34.1
		Tmn	°C	-3.2	-1.0	1.5	3.2	5.7	7.0	6.8	5.5	3.9	1.6	-0.3	-3.5	2.3
48	St16228	P	mm	21.4	1.7	2.4	0.2	16.2	187.8	245.0	237.9	204.6	74.4	13.2	4.8	1009.6
	ACAHUATO, APATZINGAN	Tob	°C	17.2	18.1	19.0	21.3	22.5	22.1	20.9	20.8	20.5	20.4	19.5	18.4	20.1
	1000 m.s.n.m.	Tmx	°C	36.2	38.3	40.9	42.8	43.0	39.9	37.5	37.1	36.6	36.1	35.8	35.5	38.3
		Tmn	°C	8.5	9.2	11.5	13.0	15.4	15.8	15.3	14.9	13.8	12.1	10.0	8.1	12.3
49	St16248	P	mm	9.8	6.0	5.5	7.4	39.5	123.3	173.8	157.1	140.5	56.9	16.4	5.9	742.1
	CAURIO DE GUADALUPE	Tob	°C	8.7	8.5	9.9	12.5	14.9	16.3	15.4	15.3	14.8	13.4	11.2	9.8	12.5
	2020 m.s.n.m.	Tmx	°C	29.1	31.9	34.6	36.9	37.1	37.2	34.4	33.8	32.3	31.3	31.0	30.3	33.3
		Tmn	°C	-4.0	-2.1	0.6	3.8	5.5	6.3	5.8	5.1	4.0	1.5	-1.2	-3.4	1.8
50	St16254	P	mm	13.4	4.0	2.7	8.8	28.4	140.5	174.2	161.3	117.0	61.0	12.3	3.7	727.3
	TEREMENDO, MORELIA	Tob	°C	6.9	8.3	10.0	12.2	13.6	13.8	13.0	12.8	12.5	10.8	9.1	7.5	10.9
	2350 m.s.n.m.	Tmx	°C	29.4	32.2	34.7	36.6	37.9	36.5	35.5	32.5	31.6	30.5	29.1	28.8	32.9
		Tmn	°C	-1.3	0.8	2.8	4.5	6.5	7.2	6.9	6.5	5.4	3.2	1.5	-0.8	3.6
51	St16255	P	mm	15.8	5.8	3.4	7.6	35.8	151.1	210.7	220.2	130.8	52.6	16.8	5.3	855.9
	UCAREO, MORELIA	Tob	°C	8.1	8.6	10.1	12.2	13.5	12.9	12.3	12.4	12.0	11.1	9.7	8.0	10.9
	1800 m.s.n.m.	Tmx	°C	25.5	27.5	30.0	32.3	34.1	33.2	30.6	28.7	27.4	26.8	26.4	25.5	29.0
		Tmn	°C	-1.4	0.8	3.1	5.4	7.9	8.0	7.9	6.6	4.6	2.8	1.8	-0.8	3.9
52	St16515	P	mm	33.0	6.6	5.9	3.1	36.3	184.9	192.9	179.7	179.3	131.1	37.4	13.8	1004.2
	LA SANDIA, PATZCUARO	Tob	°C	7.3	7.8	9.5	12.2	13.8	14.5	13.5	13.4	13.3	12.3	10.4	8.1	11.4
	1910 m.s.n.m.	Tmx	°C	28.2	30.7	33.2	35.3	35.8	34.6	32.6	31.3	30.1	29.4	28.8	28.4	31.5
		Tmn	°C	-1.1	0.9	2.9	4.7	7.0	7.9	7.1	6.3	4.9	2.8	1.3	0.8	3.8

Anexo E.

En este anexo se presentan los resultados de sumas y promedios mensuales de las variables atmosféricas medidas para el observatorio de Morelia. Los elementos atmosféricos analizados fueron: temperatura de bulbo seco (101), temperatura de bulbo húmedo (102), precipitación horaria (104), humedad relativa (105) y presión en la estación (106). Tanto para la temperatura de *bulbo seco* como para la de *bulbo húmedo* se calcularon medias mensuales y anuales, ambas expresadas en grados centígrados °C; para la *precipitación horaria* se estimó la suma mensual y la media anual, expresadas en milímetros (mm); y para la *humedad relativa* y *presión en la estación* se calcularon las medias mensuales y anuales, la precipitación expresada en porcentaje (%) y la presión en milibares (mbar) (Tabla E.1).

Tabla E.1. Resultados de sumas y promedios mensuales de las variables atmosféricas medidas para el observatorio de Morelia.

OBSERVATORIO	ELEM.	UM	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC	ANUAL
76665	101	°C	14.6	16.5	18.6	20.7	21.6	20.3	18.8	18.9	18.6	17.7	16.3	15.1	18.1
Observatorio de Morelia	102	°C	9.8	10.4	11.3	12.5	14.3	16.0	15.7	15.8	15.5	14.0	12.0	10.7	13.1
1,913 m.s.n.m.	104	mm	6.7	5.4	6.0	13.0	38.7	135.4	179.3	162.2	111.6	43.9	13.0	6.0	720.9
	105	%	58.0	50.9	44.1	41.5	48.3	66.4	73.7	74.7	74.5	69.8	65.1	61.8	60.7
	106	mbar	777.5	810.9	810.1	811.3	787.3	766.7	812.7	810.8	810.3	767.6	811.7	810.9	798.6

Anexo F.

En este apartado se muestran (Tabla F.1) las gráficas de los grupos de estaciones CLICOM formados por el K-medias. Estas gráficas corresponden a siete agrupamientos con particiones diferentes, esto es, $K = 2, 3, \dots, \text{y } 8$, donde cada color representa a un determinado grupo.

Tabla F.1. Agrupamientos con particiones diferentes ($K = 2, 3, \dots, \text{y } 8$) formados por la técnica K-medias.

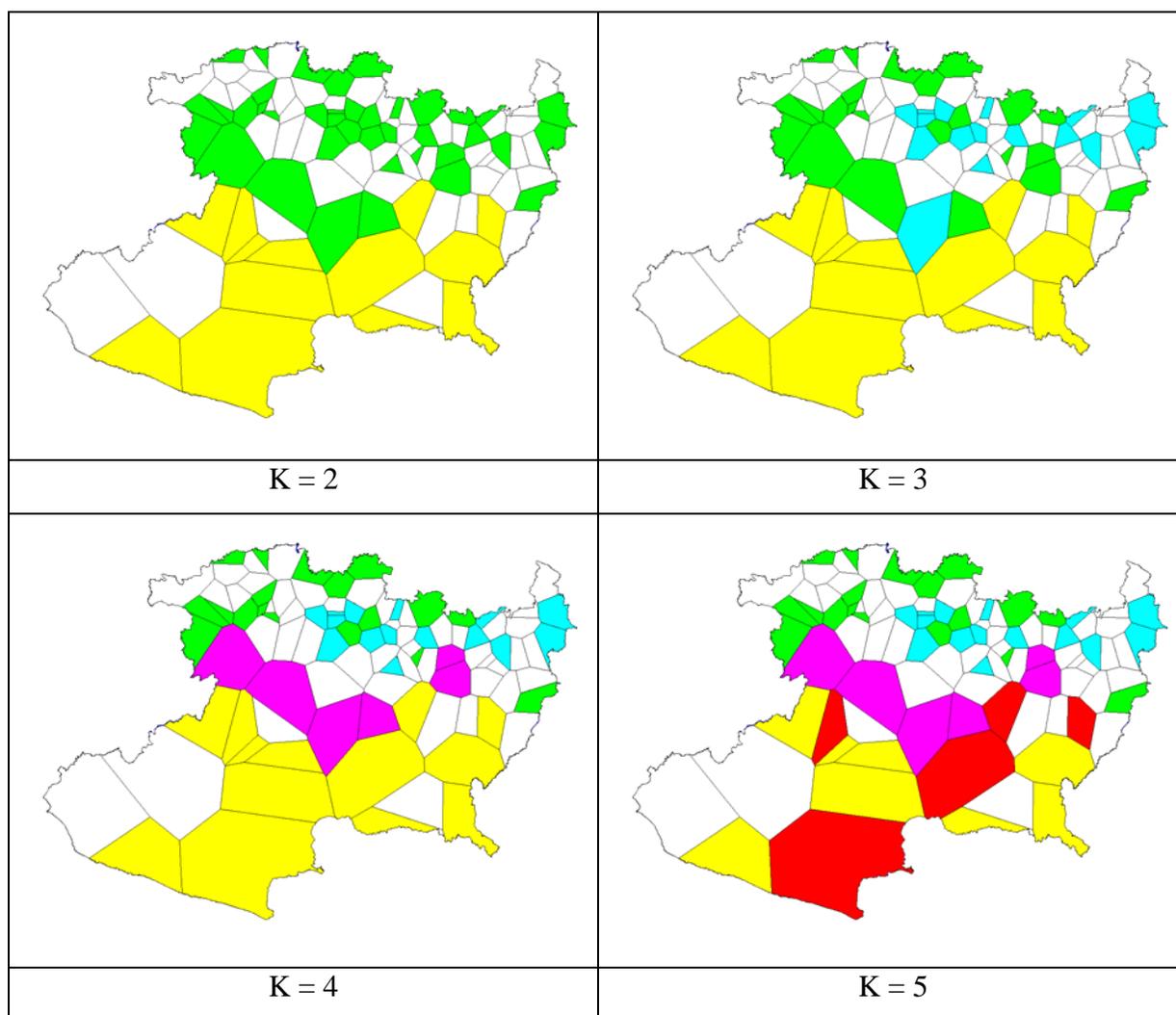


Tabla F.1. Agrupamientos con particiones diferentes ($K = 2, 3, \dots$, y 8) formados por la técnica K-medias.

