

**CENTRO DE INVESTIGACIÓN EN GEOGRAFÍA Y  
GEOMÁTICA “ING. JORGE L. TAMAYO”, A.C.  
CentroGeo**

**Centro Público de Investigación CONACYT**

**APLICACIÓN DEL PROCESO DE DESCUBRIMIENTO DE PATRONES  
GEOESPACIALES AL ESTUDIO SOBRE COMPONENTES DEL  
BALANCE HÍDRICO.**

**TESIS**

Que para obtener el grado de Maestra en Geomática

Presenta  
**CLAUDIA CORONEL ENRÍQUEZ**

Supervisor Principal  
***Dr. Franz Eduardo Mora Flores***

Examinador externo:  
***Dr. Jaime Garatuza Payán***

Comité Supervisor:  
***Dra. María del Carmen Reyes Guerrero***

México, D.F., enero, 2008

© CentroGeo. Derechos reservados. El autor otorga a CentroGeo el permiso de reproducir y distribuir copias de esta tesis en su totalidad o en partes.

## Resumen

El Balance hídrico es un proceso biofísico de vital importancia en el mantenimiento de los ecosistemas naturales y humanos. Por ello, monitorear su comportamiento ofrece elementos de información para el manejo y planeación de recursos hídricos. Desafortunadamente, en México hay serias limitaciones en la adquisición de observaciones de los diferentes componentes del balance hídrico, a excepción de la precipitación y la temperatura del aire, el resto de las variables como la evaporación, la evapotranspiración, el almacenamiento en suelos, etc., no presentan un seguimiento adecuado de su comportamiento o hay restricciones respecto al acceso público.

Estas deficiencias explican en parte la falta de información sobre la capacidad de recarga de los acuíferos y contribuyen a la sobreexplotación del agua en varias regiones hidrológicas del país. Ejemplo de ello es la región del Bajío, en el Lerma Medio, una de las zonas agrícolas más desarrolladas del país donde se privilegia la producción agrícola para la agroindustria (principalmente de capital extranjero).

No obstante que el uso de suelo es un proceso más bien local, la producción extensiva y exhaustiva en el Bajío, ha dado lugar a la disminución del nivel medio de las aguas subterráneas. A su vez, la deforestación acrecienta el riesgo de azolve y de infiltración. Por ello, es importante evaluar el comportamiento del balance hídrico en esta región para ofrecer elementos de información a los tomadores de decisiones.

Desde el contexto de la generación y de la transferencia de información, es importante considerar que los componentes del balance hídrico son variables regionalizables, es decir que se dan en el espacio geográfico y que presentan una estructura espacial intrínseca. Por ello, presentan propiedades como dependencia espacial y temporal, heterogeneidad y un dominio de escala, entre otras propiedades, las cuales deben ser reconocidas y representadas mediante modelos intuitivos y comprensibles para los usuarios de la información.

Estos modelos de representación reconocidos en la literatura como patrones geoespaciales, son herramientas útiles para describir y extraer propiedades intrínsecas de variables geográficas y facilitan un proceso de transferencia de información sobre dichas variables.

Se realizó este trabajo con el objetivo general de extraer patrones geoespaciales en algunos componentes del balance hídrico para iniciar un proceso de transferencia de información sobre este fenómeno a usuarios especializados en primera instancia. Se delimitó este objetivo por dos razones importantes, en primer lugar, las limitaciones de datos sobre variables del balance hídrico y después las deficiencias en la modelación hidrológica actual sobre este proceso, en cuanto al uso de métodos estadísticos tradicionales para el manejo de datos hidrometeorológicos y para la calibración de modelos.

A través de la incorporación del proceso de descubrimiento de patrones fue posible identificar el dominio de escala espacial y temporal de la precipitación y la evaporación y demostrar que la caracterización de estas variables a nivel de unidades de captación pluvial en dos épocas del año diferentes, presenta patrones de aglomeración o de conformación de regiones debidos a la estructura espacial propia de la precipitación y la evaporación. Las regiones obtenidas por clasificación responden también al efecto de diferenciación funcional o heterogeneidad ambiental, a partir del cual se establece que la relación de causalidad o asociación entre variables depende de la porción del territorio que se esté analizando. De esta forma, se determinó un umbral de escala (mayor a 1: 800000) a partir del cual la calibración de modelos hidrológicos en la región del Bajío para el verano y el invierno, debe realizarse considerando al menos la estructura espacial de la variable dependiente o en su defecto, realizar un análisis de regresión a nivel de subregiones y no partiendo del supuesto de que el área es homogénea ambientalmente.

Se realizó paralelamente una investigación bibliográfica sobre los modelos hidrológicos para el balance hídrico hasta la actualidad, encontrándose que los modelos conceptuales semiempíricos semidistribuidos son más factibles de implementar en México, dadas las carencias de datos de entrada y de variables biofísicas explicatorias. En este sentido, se implementó el modelo NOPEX con datos de una zona piloto ubicada en Suecia con el propósito de avanzar en la implementación de este modelo empleando datos del país.

Se considera que se cumplió con el objetivo general del trabajo, ya que se ofrece una metodología alternativa para lidiar con las propiedades intrínsecas de variables geográficas, como lo son los componentes del balance hídrico y variables biofísicas asociadas. Al mismo tiempo, se extrajeron patrones espacio-temporales de la precipitación y la evaporación para dos periodos del

año 2002, los cuales evidencian la estructuración espacio-temporal de estas variables, ofreciendo elementos de información a especialistas interesados, para la adecuación y mejora de modelos hidrológicos.

## **Dedicatoria**

A la vida, por tener el privilegio de compartir, aprender y amar

A Gil, por compartir conmigo esta oportunidad de construir una familia

A Itzmaná y Samael por aceptarme, aún sin comprenderme

A Leila por su sonrisa y su inocencia

A mis profesores, asesores y compañeros por su colaboración en mi preparación

## **Reconocimientos**

Un reconocimiento especial a mis padres por su ejemplo de esfuerzo día a día,

Agradezco al Dr. Franz Mora por su paciencia y guía durante la elaboración de este trabajo,

Un reconocimiento a la Dra. Carmen Reyes por su valiosa contribución en la estructuración de este trabajo,

Un agradecimiento especial al Dr. Jaime Garatuza porque su visión del problema complementó el enfoque general del trabajo,

Agradezco la asesoría y tiempo del Dr. Fernando Ávila en la implementación de métodos de Estadística espacial incorporados en esta tesis,

Un reconocimiento al CentroGeo y al CONACYT por su apoyo económico y esfuerzo en la formación de capital humano

## Tabla de contenido

<b>1. Introducción.....</b>	<b>1</b>
1.1. Contexto general .....	2
1.2. Justificación y relevancia .....	4
<b>2. Planteamiento del problema .....</b>	<b>6</b>
2.1. Características generales del balance hídrico.....	6
2.2. Estrategia general para abordar la búsqueda de patrones espacio-temporales del Balance Hídrico.....	8
2.3. Marco espacio-temporal.....	9
<b>3. Marco teórico .....</b>	<b>11</b>
3.1. El enfoque de análisis espacial para abordar fenómenos geoespaciales .....	11
3.2. Modelos geográficos .....	13
<b>4. Enfoques metodológicos para la extracción de patrones geoespaciales .....</b>	<b>17</b>
4.1. El proceso general de descubrimiento de patrones.....	17
4.2. Herramientas de Geovisualización .....	18
4.3. El descubrimiento de patrones geoespaciales o DPG .....	20
4.4. Aplicación del proceso de DPG en soluciones computacionales.....	21
<b>5. Antecedentes.....</b>	<b>22</b>
5.1. Enfoques metodológicos y conceptuales para la modelación del balance hídrico.....	22
5.2. Ejemplos de implementación de modelos.....	22
5.3. Desarrollo de herramientas de geovisualización y de exploración de datos hidrológicos	24
5.4. Métodos generales para obtener registros de componentes del Balance Hídrico y variables causales asociadas .....	25
5.5. Modelación del balance hídrico .....	26
<b>6. Aportes.....</b>	<b>30</b>
6.1. Alcances.....	30
6.2. Objetivos.....	30
<b>7. Incorporación del proceso de descubrimiento de patrones geoespaciales al estudio del balance hídrico.....</b>	<b>32</b>
7.1. Consideraciones generales para la incorporación del proceso de DPG.....	32
7.2. Incorporación del proceso de DPG.....	33
7.2.1. Organización, manejo y visualización de datos.....	33
7.2.2. Análisis exploratorio o estructural.....	50
7.2.3. Construcción de hipótesis.....	75

7.2.4. Análisis confirmatorio .....	83
<b>8. Consideraciones finales y Conclusiones.....</b>	<b>110</b>
8.1. <i>Detalles de implementación del modelo hidrológico NOPEX</i> .....	112
8.2. <i>Limitaciones en la adecuación de modelos conceptuales al estudio del balance hídrico</i>	113
8.3. <i>Limitaciones en la incorporación del proceso de descubrimiento de patrones geoespaciales.....</i>	114
8.4. <i>Líneas de investigación a futuro.....</i>	116
<b>9. Bibliografía.....</b>	<b>117</b>
<b>10. Apéndice .....</b>	<b>122</b>
10.1. <i>Implementación del diagrama de procedimientos para estimar balance hídrico mediante el modelo NOPEX.....</i>	122



## Lista de Tablas

Tabla 7.1 Resumen de atributos de productos de MODIS. ....	36
Tabla 7.2 Referencias a data warehouses. ....	37
Tabla 7.3 Conjunto de operaciones a realizar con datos preprocesados. ....	38
Tabla 7.4 Matriz de autocorrelación lineal. ....	41
Tabla 7.5 Métodos para la estimación de variables en formato raster. ....	50
Tabla 7.6 Índices de heterogeneidad, Moran I y Geary c. ....	52
Tabla 7.7 Coeficientes de correlación lineal de Pearson. ....	58
Tabla 7.8 Parámetros de variogramas teóricos. ....	63
Tabla 7.9 Parámetros de variogramas teóricos ajustados a variables independientes. ....	66
Tabla 7.10 Fracciones de variación explicada. ....	68
Tabla 7.11 Parámetros de variogramas teóricos ajustados a datos de evaporación promedio diaria .....	71
Tabla 8.1 Resumen de los métodos, operaciones y herramientas, empleadas en la incorporación del proceso de DGP en el estudio del balance hídrico. ....	111

## Lista de Figuras

Fig. 1.1 Ciclo hidrológico.....	3
Fig. 2.1 Áreas de captación .....	6
Fig. 2.2 Mapa del Distrito 011 de la región hidrológica del Lerma, Zona del Bajío. ....	10
Fig. 3.1 Representación del paisaje como modelo jerárquico .....	15
Fig. 4.1 Proceso de descubrimiento de conocimiento en datos (KDD).....	18
Fig. 4.2 Matriz de integración de los métodos de KDD y geovisualización. ....	19
Fig. 4.3 Proceso de descubrimiento de patrones geoespaciales. ....	21
Fig. 5.1 Diagrama del balance hídrico a nivel de unidades hidrológicas. ....	27
Fig. 5.2 Representación gráfica del balance hídrico.....	29
Fig. 7.1 Representación lógica de datos puntuales.....	34
Fig. 7.2 Representación lógica de datos como superficies.....	35
Fig. 7.3 Representación lógica de datos como superficies.....	35
Fig. 7.4 Contrastación de perfiles climáticos. ....	40
Fig. 7.5 Perfiles climáticos. ....	41
Fig. 7.6 Diagramas de Probabilidad Normal para variables climáticas.....	43
Fig. 7.7 Diagramas de dispersión precipitación vs. coordenadas.....	44
Fig. 7.8 Diagramas de dispersión temperatura vs. coordenadas.....	45
Fig.7.9 Imagen 2D o nadir representando el modelo del terreno en la región del Bajío.....	46
Fig. 7.10 Imagen de 3 dimensiones mostrando rasgos del relieve. ....	47
Fig. 7.11 Simulación de vuelo sobre la zona del Bajío. ....	48
Fig. 7.12 Animación de 2 dimensiones. ....	49
Fig. 7.13 Correlogramas direccionales de variables explicatorias. ....	53
Fig. 7.14 Variogramas experimental y teórico. ....	55

Fig. 7.15 Variogramas experimentales para precipitación y temperatura del aire. ....	56
Fig. 7.16 Variogramas experimentales de altitud y temperatura de superficie. ....	57
Fig. 7.17 Diagramas de coordenadas paralelas.....	59
Fig. 7.18 Mapa de asociación entre precipitación y elevación, verano 2002. ....	59
Fig. 7.19 Diagramas de coordenadas paralelas, temperatura del aire.....	60
Fig. 7.20 Mapas de asociación, temperatura y variables físicas, invierno de 2002.....	61
Fig. 7.21 Variografía para datos de precipitación acumulada y temperatura del aire. ....	62
Fig. 7.22 Mapas de precipitación acumulada. ....	64
Fig. 7.23 Mapas de temperatura del aire promedio diaria. ....	65
Fig. 7.24 Mapas de evapotranspiración potencial. ....	70
Fig. 7.25 Esquema de unidades de drenaje tipo. ....	73
Fig. 7.26 Mapas de áreas de captación pluvial. ....	74
Fig. 7.27 Mapas de promedios areales para unidades de captación pluvial. ....	76
Fig. 7.28 Mapas de caja de promedios areales para unidades de captación mayores a 300 km <sup>2</sup> . .	79
Fig. 7.29 Plantilla de mapas de caja para unidades de captación mayores a 1000 km <sup>2</sup> . ....	82
Fig. 7.30 Diagrama de dispersión de Moran. ....	84
Fig. 7.31 Mapa de autocorrelación espacial. ....	85
Fig. 7.32 Distribución del índice de Moran respecto de una distribución aleatoria. ....	85
Fig. 7.33 Mapa de autocorrelación espacial con regiones de precipitación acumulada, unidades mayores a 300 km <sup>2</sup> , verano 2002. ....	87
Fig. 7.34 Mapa de autocorrelación espacial con regiones de precipitación acumulada, unidades mayores a 300 km <sup>2</sup> , invierno de 2002. ....	88
Fig. 7.35 Mapa de autocorrelación espacial con regiones de evapotranspiración potencial promedio diario, unidades mayores a 300 km <sup>2</sup> , verano de 2002.....	89
Fig. 7.36 Mapa de autocorrelación espacial con regiones de evapotranspiración potencial promedio diario, unidades mayores a 300 km <sup>2</sup> , invierno de 2002. ....	90

Fig. 7.37 Diagramas de dispersión del índice Moran para precipitación y para evapotranspiración vs. promedios ponderados por distancia.....	91
Fig. 7.38 Mapa de clases de precipitación acumulada, verano de 2002.....	94
Fig. 7.39 Mapa de clases de precipitación acumulada promedio areal, invierno de 2002. ....	95
Fig. 7.40 Mapa de clases de evaporación promedio areal, verano de 2002. ....	96
Fig. 7.41 Mapa de clases de evaporación promedio areal, invierno de 2002.....	97
Fig. 7.42 Mapa de clases de precipitación acumulada promedio areal, verano de 2002, unidades mayores de 1000 km <sup>2</sup> . ....	98
Fig. 7.43 Mapa de clases de precipitación acumulada promedio areal, invierno de 2002, unidades mayores de 1000 km <sup>2</sup> . ....	99
Fig. 7.44 Mapa de clases de evaporación promedio areal, verano de 2002, unidades mayores de 1000 km <sup>2</sup> .....	99
Fig. 7.45 Mapa de clases de evaporación promedio areal, invierno de 2002, para unidades mayores de 1000 km <sup>2</sup> . ....	100
Fig. 7.46 Mapa de proporción de varianza explicada por variación ambiental para regiones de precipitación acumulada en verano. ....	101
Fig. 7.47 Dominio de escala de variables empleadas en el análisis de regresión de datos de precipitación a nivel de regiones. ....	102
Fig. 7.48 Mapa de proporción de varianza explicada por variación ambiental para regiones de precipitación acumulada en invierno.....	103
Fig. 7.49 Mapa de proporción de varianza explicada por variación ambiental para regiones de evaporación promedio en verano.....	104
Fig. 7.50 Dominio de escala de variables empleadas en el análisis de regresión de datos de evaporación a nivel de regiones.....	105
Fig. 7.51 Mapa de proporción de varianza explicada por variación ambiental para regiones de evaporación promedio en invierno. ....	106
Fig. 7.52 Mapa de proporción de varianza explicada por variación ambiental para regiones de precipitación acumulada en verano, unidades mayores a 1000 km <sup>2</sup> . ....	107
Fig. 7.53 Mapa de proporción de varianza explicada por variación ambiental para regiones de precipitación acumulada en invierno, unidades mayores a 1000 km <sup>2</sup> .....	107
Fig. 7.54 Mapa de proporción de varianza explicada por variación ambiental para regiones de evaporación promedio en verano, unidades mayores a 1000 km <sup>2</sup> .....	108

Fig. 7.55 Mapa de proporción de varianza explicada por variación ambiental para regiones de evaporación promedio en invierno, unidades mayores a 1000 km <sup>2</sup> . .....	109
Fig. 8.1 Diagrama de procedimientos para la estimación del balance hídrico. ....	113

## 1. Introducción

Existen en el país, problemáticas tales como la sobreexplotación de recursos naturales, el crecimiento desordenado de los centros urbanos, el desequilibrio en el desarrollo, etc., las cuales son resultado de la interacción de diferentes procesos y normalmente presentan una evolución en el tiempo y una expresión en el espacio; además, hay un requerimiento urgente por parte de la sociedad de dar solución a estos problemas sin que hasta ahora se cuente con elementos bien sustentados para resolverlos. Esto implica profundizar en la caracterización del proceso de interés, determinar cuál es su dominio de escala espacial y temporal, así como establecer los modelos de representación más adecuados a la planeación y manejo del problema en cuestión. En resumen, partimos del supuesto general de que la mayoría de los procesos que se efectúan en el espacio geográfico pueden ser estudiados desde el enfoque del análisis espacial. En este enfoque, la representación de los procesos geográficos en forma de patrones, rasgos o entidades geoespaciales es la premisa (Haggett, 1965:20). Por ello, se propone en este trabajo enfocarnos en la búsqueda de patrones discernibles y válidos tanto en el proceso de interés como en los factores causales identificados.

En la medida en que profundizamos sobre el comportamiento y las propiedades de los fenómenos espacio-temporales que estudiamos, contribuimos a la generación de información geoespacial sobre tales procesos. Dichos fenómenos se efectúan en el espacio geográfico y presentan características generales que nos permiten identificarlos como objeto de estudio del análisis espacial, presentan un dominio de escala espacio-temporal, pueden ser representados en forma de patrones o rasgos y es posible abordarlos desde una perspectiva holista (Thomas y Edwards, 2001:476). La estrategia metodológica que se plantea para la identificación y extracción de patrones es mediante un proceso de descubrimiento de patrones geoespaciales o DPG (denominado por Gahegan *et al.* (ICA:3), como descubrimiento de conocimiento geográfico o DCG), el cual consiste en descubrir y reconocer patrones geoespaciales válidos y significativos en datos u observaciones de variables geográficas. El enfoque metodológico del descubrimiento de patrones contribuye a identificar y formalizar propiedades sobre las variables estudiadas a través de métodos de geovisualización (Miller, in press:40). De esta forma se puede ofrecer a los usuarios de los datos analizados, elementos para la toma de decisiones y la resolución de problemas.

El objetivo general de este trabajo es ilustrar cómo, a través del enfoque teórico del análisis espacial y del recurso metodológico del DPG, se puede avanzar en la generación de información geoespacial sobre un fenómeno geográfico. Más específicamente, se pretende que mediante la aplicación de herramientas de exploración de datos geoespaciales y operaciones de geovisualización se profundice sobre el comportamiento espacio temporal de los componentes del balance hídrico, ya que el ciclo hidrológico es uno de los fenómenos naturales más relevantes para la administración y manejo de recursos por parte de la sociedad. Se abordará el proceso del balance hídrico a través de un enfoque sistémico, con énfasis en el descubrimiento de patrones y su relación con procesos físicos, bióticos y antropogénicos. La estrategia general es avanzar en la caracterización de componentes del balance hídrico a escala del paisaje ya que es la extensión territorial más idónea para la planeación de recursos (García y Muñoz, 2002:10).

### **1.1. Contexto general**

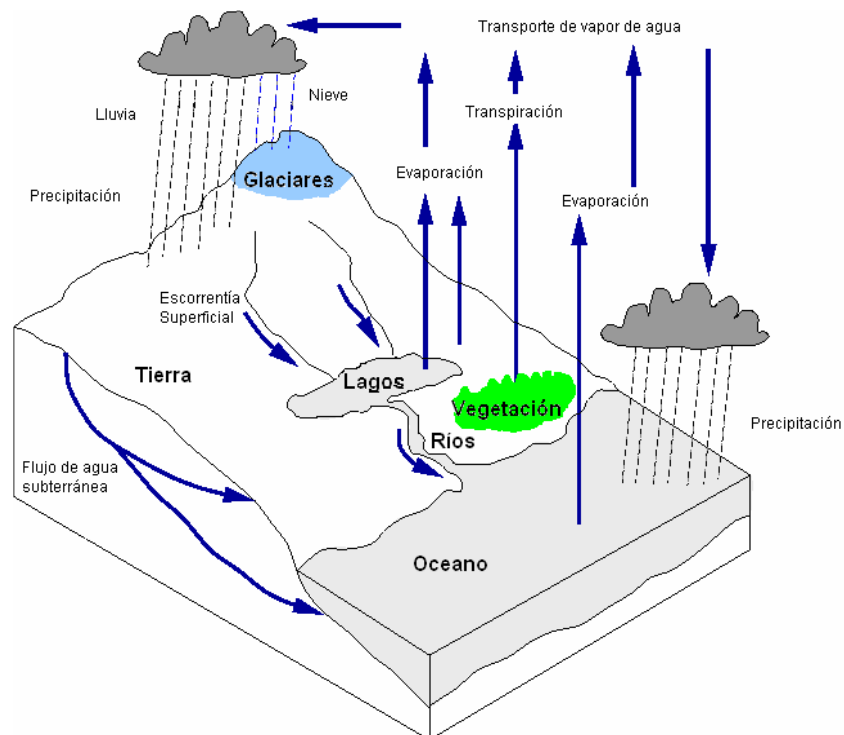
En este estudio, se pretende contribuir a la generación de información geográfica acerca del Balance Hídrico empleando la metodología general del proceso de DPG. Se trata de ofrecer elementos conceptuales y metodológicos para contribuir al descubrimiento de patrones geoespaciales sobre dicho fenómeno, desde la perspectiva del análisis espacial y la exploración de datos geográficos y así profundizar en la caracterización del balance hídrico. A continuación se exponen los planteamientos generales de porqué el balance hídrico debe ser abordado desde dicho enfoque.

En la figura 1.1 se ilustran los diferentes componentes del Balance hídrico, como vemos, la precipitación es la principal entrada al ciclo hidrológico y se compone de dos procesos físicos, la lluvia y la nieve, los cuáles presentan un comportamiento espacio temporal específico, el cual se ha identificado parcialmente mediante herramientas de interpolación de datos de estas variables, considerando la autocorrelación espacial como uno de los factores causales de los patrones registrados (Govaerts, 1999:3).

La otra entrada es el almacenamiento disponible en el suelo, el cual, sumado con la precipitación, se convierte en escurrimientos cuando se rebasa la capacidad de retención de humedad en el suelo y resulta en un componente que reduce la recarga hídrica neta. El almacenamiento disponible depende directamente de las características del suelo y del relieve, su

comportamiento es una medida indirecta de recarga hídrica ya que del balance de entradas y salidas la resultante es igual a la suma del almacenamiento disponible y el volumen de infiltración que potencialmente contribuye a la recarga. El relieve también afecta el volumen y arreglo de los escurrimientos y es a su vez transformado paulatinamente por la acción erosiva del agua superficial o en su caso por la disolución de minerales en la roca por parte de las aguas subterráneas que se infiltran al subsuelo (ver figura 1.1).

**Fig. 1.1 Ciclo hidrológico**



Tomado de [www.waterbalance.ca](http://www.waterbalance.ca)

Otros componentes de sustracción del balance, que tienen que ver con la humedad del aire y la radiación solar neta, son la evaporación directa y la evapotranspiración; para estudiar el comportamiento de la evaporación normalmente se requiere conocer su relación con la temperatura y como esta relación afecta los patrones espacio temporales de la evaporación, por último la evapotranspiración es un proceso derivado de la actividad productiva primaria de la vegetación y si bien reduce el volumen de recarga contribuye a la acumulación de humedad en la atmósfera. La evaporación y la evapotranspiración son afectadas diferenciadamente por la dinámica de la cobertura terrestre, proceso este el más complejo de todos ya que resulta de la



interacción de factores físicos, bióticos y antropogénicos que le imprimen cierta estructura espacio temporal.

El balance hídrico es un problema que ha sido abordado mediante relaciones lineales y no lineales entre sus componentes y diversas variables causales (modelos hidrológicos), para más detalles véase Xu (2000:369-378). Poco se sabe acerca de si estas relaciones son igualmente válidas en cualquier época del año, si son generalizables ya sea que se trate de mediciones puntuales o de estimaciones a nivel de unidades de captación o subcuencas y si los propios componentes exhiben una autocorrelación espacial o temporal variable. El enfoque metodológico del DPG ofrece una alternativa a la resolución de estos cuestionamientos ya que a través de la exploración de datos permite identificar, comparar y extraer patrones espaciales y temporales, los cuales pueden ser más comprensibles y evidentes mediante la visualización de dichos rasgos por diferentes métodos gráficos y de multimedia (Cartrwright, 1999:335-340 y Mennis y Peuquet, 2003:371-380).

Para completar los objetivos generales del trabajo, se propone el análisis de un área piloto dentro de la cuenca del Lerma, por ser una de las regiones hidrológicas con mayor explotación de acuíferos en el país (Peñaloza y López, 2002:35). Paralelamente es una de las cuencas mejor monitoreadas en México pues históricamente ha habido conflictos sociales por la apropiación y derecho de uso del recurso, lo que obliga a las autoridades competentes a llevar un monitoreo continuo de los volúmenes de escurrimientos y de otros componentes del ciclo hidrológico como la precipitación y la evaporación.

## **1.2. Justificación y relevancia**

En México no se realizan estudios veraces sobre el balance hídrico ya que los datos hidrológicos y meteorológicos son inconsistentes, imprecisos o tienen un nivel de procesamiento prácticamente nulo (Mendoza *et al.*, 2002:37). Este es, precisamente, uno de los problemas que se deben resolver: el acceso, manejo y procesamiento de los datos como hasta ahora se registran para explicitar patrones espacio temporales de los componentes del ciclo hidrológico.

El organismo público encargado de administrar los datos hidrometeorológicos, la Comisión Nacional del Agua (CNA), pone a disposición del público sólo datos puntuales de

estaciones meteorológicas o hidrológicas y proporciona datos muy generales, a nivel de subregiones hidrológicas, de carácter administrativo, acerca del grado de explotación de acuíferos, volúmenes de recarga, y estatus de explotación. No hay una liga directa de esta información a mapas o posibles factores causales y no es presentada en forma de gráficos por lo que es difícil inferir algún comportamiento particular en los datos a primera vista. Esto nos indica que debemos acceder a otras fuentes de datos que no necesariamente son más precisas pero que si nos pueden ofrecer información adicional o de contexto a los datos de estaciones meteorológicas, lo cual es el caso de las observaciones de percepción remota para monitorear temperatura de superficie o precipitación.

En vista de las limitantes en cuanto a las fuentes de datos y a la escasez de estudios en la zona tendientes a profundizar sobre el comportamiento del balance hídrico y sus causas, se considera que la presente investigación puede ofrecer importantes elementos para caracterizar este fenómeno en el Bajío y contribuir a la conformación de una metodología para el monitoreo de esta función ecológica en cualquier región del país, aún con las restricciones que presentan las observaciones actualmente.

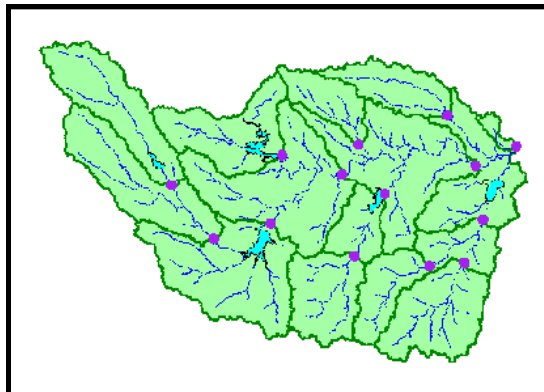
## 2. Planteamiento del problema

Con objeto de contextualizar con más especificidad el problema que nos ocupa plantearemos las implicaciones de tratar con un problema como el balance hídrico, definiremos la estrategia general para abordarlo y por último realizaremos una breve descripción del área piloto.

### 2.1. Características generales del balance hídrico

El balance hídrico no puede estimarse directamente sólo a partir de alguno de sus componentes, esto es, se requiere conocer paralelamente más de dos variables hidrológicas (precipitación, temperatura, evapotranspiración, escurrimiento), que pueden estar o no correlacionadas (Xu, 1999:353-360). Desde el enfoque de la Ecología del paisaje, el balance hídrico es una función ecológica lenta que implica, por un lado, la reincorporación de volúmenes de agua a los depósitos subterráneos y superficiales los cuales retroalimentan el ciclo hidrológico por procesos de evaporación, sublimación y evapotranspiración. El balance también depende de la diferenciación espacio-temporal de la precipitación y de la cubierta terrestre, que son fenómenos más dinámicos. La forma en que se integran procesos tan dispares en la naturaleza es a través de la interacción entre ecosistemas o unidades ecológicas, mediante las conexiones o flujos entre tales unidades, un área de captación de agua puede mantener cierto estado dinámico y realizar las funciones que le son imprescindibles (Coppin *et. al.*, 2004:1570). Para la delimitación de las áreas de captación se requiere establecer el punto de acumulación principal o de salida en el cual se estima el volumen de escurrimientos y por tanto, es la unidad adecuada para estimar el balance hídrico (Fig. 2.1).

**Fig. 2.1 Áreas de captación**



Las áreas de captación pueden ser definidas a partir de puntos de salida localizados en las redes de flujo hidrológico. Tomado de Maidment, (2000:15)

Ya que las áreas de captación son las unidades mínimas de funcionamiento hidrológico no podemos pensar en otro modelo geográfico más adecuado para expresar patrones espacio temporales sobre el balance hídrico. La delimitación de estas unidades bajo criterios estrictamente de funcionamiento es un proceso bien definido pero no hay criterios preestablecidos acerca del tamaño mínimo que debe manejarse, este puede ser, de acuerdo a la zona de que se trate, los objetivos de la regionalización o en su caso los alcances del programa de manejo a implementar. Sean cual sean los criterios para delimitar unidades, sin duda afectarán los patrones del balance hídrico ya que sus diferentes componentes se comportan diferenciadamente en los contextos espacial y temporal (compárese la extensión y variación de eventos de precipitación con los procesos de evapotranspiración que tienen que ver con la productividad primaria y se verifican desde un nivel individual hasta enormes extensiones de vegetación continua).

Las unidades de captación hidrológica, a su vez, conforman un sistema más complejo, conocido como paisaje, el cual se caracteriza precisamente por la heterogeneidad de su composición y por los procesos de interacción entre unidades. Nuevamente, la extensión del paisaje o área de estudio depende de los objetivos del estudio, o del proceso de interés que pretende comprenderse. Finalmente, los patrones del balance hídrico se expresan en la extensión del paisaje y gracias a ello es posible implementar planes de manejo del recurso agua sin impactar de forma relevante la estabilidad de este proceso hidrológico (Ahern, 1999:180).

Para poder explicitar los patrones espacio-temporales del balance hídrico a la escala del paisaje, se requieren observaciones de sus componentes como precipitación, temperatura, escurrimientos, entre otros, variables que son fundamentalmente regionalizables o distribuidas (Cooper e Istok, 1988:270). Desafortunadamente, como se mencionó anteriormente, la red hidrometereológica en la región del Bajío es puntual, inconsistente e insuficiente en cuanto al registro de datos de los componentes del ciclo hidrológico por lo que debemos recurrir a métodos indirectos y estimaciones de las variables del ciclo a partir de Percepción Remota y análisis espacial.

## **2.2. Estrategia general para abordar la búsqueda de patrones espacio-temporales del Balance Hídrico.**

En vista de las limitaciones en los registros de observaciones sobre los componentes del balance hídrico, las tareas de análisis para la representación cartográfica y para la estructuración de variables del ciclo hidrológico se limitarán a algunos componentes del balance hídrico, principalmente porque los datos de escurrimientos están disponibles al público parcialmente y existe una política restrictiva respecto del acceso a estos datos, no se registran observaciones directas de evapotranspiración y los datos de evaporación son escasos. Sin duda el principal aporte en este trabajo en relación al proceso estudiado será explicitar los patrones espaciales de algunos componentes del balance como la precipitación y la evapotranspiración potencial ya que se cuenta con registros de precipitación y de temperatura del aire, la cual está relacionada con la evaporación. Asimismo, se mostrarán ciertas asociaciones de estos componentes con respecto de la dinámica de cobertura terrestre, específicamente de la vegetación, pues como se refiere en la literatura, las observaciones de percepción remota son fuentes de datos adecuadas para el monitoreo regional de la cobertura vegetal (Chuvieco, 1995:156).

En la medida en que pueda avanzarse sobre el establecimiento de hipótesis explicatorias de los patrones observados y sobre su posterior verificación o modificación podrán realizarse nuevos hallazgos sobre el fenómeno en cuestión. A su vez, la caracterización del balance hídrico a escala regional es un reto que difícilmente se habrá de resolver en este trabajo pero se avanzará en la búsqueda de una metodología que facilite la realización de dicha tarea y permita la modelación de balances hídricos a nivel de unidades hidrológicas, aún en cuencas poco monitoreadas.

La estrategia a seguir, es partir del marco o dominio de conocimiento que aborda el balance hídrico y definir a priori los metamodelos o patrones que conceptualmente han sido empleados para representar los componentes del balance hídrico. A partir de este ejercicio de conceptualización de estructuras (como lo denominaron Maceachren *et al.* en 1999:313), estableceremos cómo abordar las diferentes etapas del DPG, desde la definición de la representación lógica de datos más adecuada para representar el comportamiento de los procesos estudiados, hasta la conformación de recursos simplificados de información. Se buscará incorporar en la medida de lo posible modelos de evaluación para explicitar la calidad de la

información generada en cada fase. La intención es que se descubran nuevos hallazgos en cada fase y que esta información pueda emplearse en última instancia, para apoyar la toma de decisiones y conducir a una estrategia adaptativa de planeación del recurso agua.

### **2.3. Marco espacio-temporal**

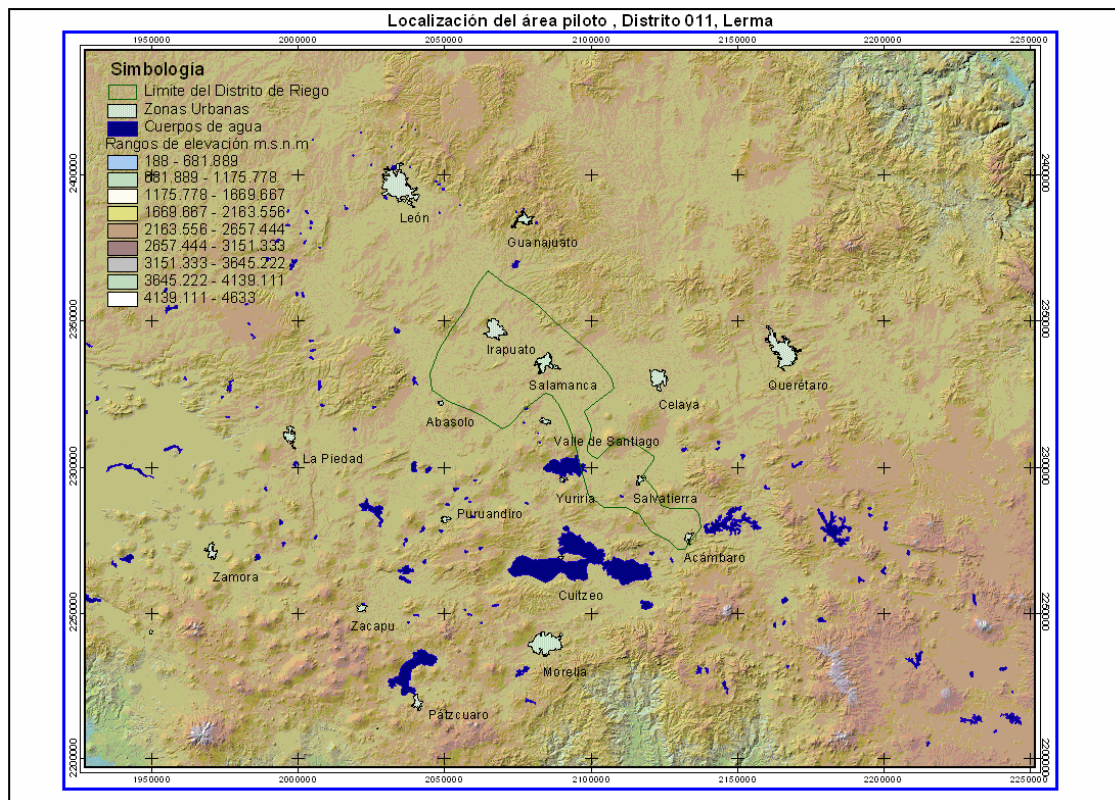
El área reconocida como el Bajío se ubica dentro de la cuenca del Lerma en el Altiplano central mexicano y comprende porciones de los estados de Guanajuato, y Michoacán (43 municipios en total). Casi el 53% de la superficie presenta un alto grado de antropización por actividades agropecuarias, el crecimiento urbano y la actividad industrial. En cuanto al aprovechamiento del agua, el 25% de los pozos de extracción a nivel nacional se encuentran en la cuenca y del abastecimiento total de aguas superficiales y subterráneas para la zona, el 90% es destinado a los usos agrícola y urbano. Esta evidencias nos hablan de que se deben orientar los esfuerzos hacia un manejo adecuado del recurso agua lo cual solo puede realizarse a partir de un conocimiento integrado sobre el funcionamiento hidrológico de la cuenca.

Hasta mediados de la década de los 30 del siglo pasado, la región era considerada el principal productor de granos básicos en México, sin embargo, la conformación del distrito de riego 011 a mediados de la década de los 50 y la posterior instalación de numerosas industrias alimenticias y procesadoras de forrajes, reestructuró la función socioeconómica del Bajío como productor de básicos al de un dinámico polo agroindustrial con una importante participación de inversión extranjera (Soto y Soto, 1990:5), (ver mapa en la Figura 2.2).

Actualmente, el crecimiento económico en el Bajío se considera un motor para el desarrollo regional y del país y la diferenciación espacial del desarrollo da cuenta de ello, baste mencionar que en el estado de Guanajuato se encuentra el 34% de la superficie de riego para toda la cuenca del Lerma y el estado cuenta con más del 50% del abastecimiento. Se distingue un eje carretero que parte del Valle de Toluca hasta la Cd. De Guadalajara atravesando el Bajío al sur de la laguna de Cuitzeo y que conecta a Morelia con los dos extremos del eje. Al norte también se construyó otro eje carretero que parte de la Cd. de México y la conecta con las ciudades industriales más importantes en la región como Querétaro, Celaya, Salamanca, Irapuato, Guanajuato y León. A lo largo de estos ejes se han presentado los principales desarrollos urbanos en el país en los últimos 30 años. Sin embargo, hay suficientes evidencias del deterioro ambiental

que los ecosistemas están manifestando, contaminación y degradación de suelos, sobreexplotación de acuíferos, disminución de la calidad del agua por salinización y endurecimiento, desecación y asolvamiento de cuerpos de agua y respecto de los recursos sociales empobrecimiento, marginación y descapitalización humana de los sistemas de producción (INE-SEMARNAT, 2003:5; Severnay y Prat, 2003:371).

**Fig. 2.2 Mapa del Distrito 011 de la región hidrológica del Lerma, Zona del Bajío.**



Como vemos, la dinámica productiva en la región del Bajío merma de forma importante la capacidad de recarga de los sistemas hidrológicos, sin embargo, no nos es posible evaluar la recarga neta, ya que difícilmente podrían obtenerse estimaciones de los aportes y/o extracciones por riego y otros usos pues la administración de estos datos no obedece a criterios naturales exclusivamente y no hay estudios confiables sobre las propiedades del subsuelo ni de la dinámica de aguas subterráneas. No obstante, es posible obtener indicadores de la recarga a través del balance hídrico, específicamente a partir del almacenamiento disponible. Por ello, se considera, que uno de los objetivos principales de este trabajo es avanzar hacia la implementación de un modelo para la estimación del balance hídrico en cuencas poco monitoreadas.

### **3. Marco teórico**

Dada la necesidad de generar un conocimiento integral y científico sobre el balance hídrico, la estrategia para profundizar sobre el comportamiento de este fenómeno en el presente trabajo es emplear el DPG como recurso metodológico para descubrir patrones geoespaciales y el enfoque teórico del análisis espacial, lo cual permitirá explicitar los patrones del balance hídrico y de sus componentes así como formalizar las relaciones del fenómeno con procesos causales básicos como son algunas variables físicas y la dinámica de la cobertura terrestre. En primer término, se argumenta sobre la conveniencia de estudiar el balance hídrico desde la perspectiva del análisis espacial y la ecología del paisaje y cuáles serían los aportes generales del trabajo desde este enfoque.

#### **3.1. El enfoque de análisis espacial para abordar fenómenos geoespaciales**

En primer término, es necesario establecer claramente si el proceso de interés comparte las propiedades de los objetos que son estudiados desde el enfoque de análisis espacial. En efecto, el balance hídrico se expresa en el espacio geográfico y los elementos básicos de análisis se representan como regiones, por lo que estas unidades adquieren relaciones topológicas de vecindad y de conectividad, entre otras. Además, estudios recientes nos dan evidencias de que las relaciones funcionales (verticales o corológicas) que explican la heterogeneidad de los componentes del balance hídrico también están diferenciadas en los contextos espacial y temporal (Maciek et al. 2005:75-90). A lo largo de este estudio se confirmará si efectivamente los componentes del balance hídrico presentan tales propiedades.

Para establecer el dominio de conocimiento en el cual se puede abordar el balance hídrico, partimos de 3 supuestos generales sobre su comportamiento:

a) para la administración de recursos hídricos se requiere suficiente heterogeneidad espacial del área sujeta a manejo para permitir estrategias adaptativas, esto es, que exista variabilidad espacial en las características generales del área.

b) El propio fenómeno se configura en la escala temporal y espacial de forma diferenciada (heterogeneidad espacio-temporal), por lo que se debe establecer el dominio de escala de los diversos componentes involucrados para abordar adecuadamente el balance hídrico.

c) El funcionamiento hidrológico opera fundamentalmente a nivel de unidades territoriales que presentan una dinámica y estructura propias, de aquí se desprende que la



caracterización del balance a escala del paisaje resuelve los requerimientos de la modelación sistémica del proceso.

De esta forma, al estudiar el balance hídrico a nivel de unidades hidrológicas el paisaje o espacio geográfico queda delimitado por el conjunto de áreas de captación. Sin embargo, como ya se ha puntualizado anteriormente, para caracterizar las diferentes variables a nivel de unidades territoriales se requieren superficies de cada una de éstas, de ahí el uso corriente de métodos de interpolación, de regresión o bien el empleo de imágenes satelitales. Por otro lado, la modelación de áreas de captación implica el análisis de los patrones de escurrimientos basado en la variación espacial del relieve. Además, para formalizar las propiedades de las unidades hidrológicas requerimos herramientas específicas de visualización que resalten las similitudes y diferencias entre unidades. Estas consideraciones nos remiten a un dominio del conocimiento que incluye, además del enfoque de la ecología del paisaje, métodos y conceptualizaciones del análisis espacial, la geocomputación, la geovisualización y la percepción remota.

Parte del ejercicio de implementar el DPG es analizar si los modelos de representación lógica de los datos de estaciones o de imágenes satelitales retoman adecuadamente las propiedades intrínsecas de los componentes del balance hídrico conforme a su conceptualización desde el punto de vista sistémico y de análisis espacial. Esto significa que debemos revisar si los datos tal cual son registrados, ya sea a nivel de estaciones, en formato vector o raster capturan propiedades como la autocorrelación espacial, la heterogeneidad o la colinearidad, elementos que suponemos a priori caracterizan el comportamiento de las variables estudiadas. Además, se deben hacer explícitos, en la medida de lo posible, las fuentes de error y el grado de precisión de las estimaciones de variables.

En lo que se refiere a aspectos computacionales, debe quedar bien definida la representación lógica de los datos de tal forma que se distinga entre las observaciones empíricas y aquéllos datos que ya presentan un nivel de análisis más avanzado, esto es, que las representaciones exhiban elementos interesantes y/o nuevos respecto de las observaciones originales (Coronel y Gutiérrez, 2004:40). Otros elementos a definir son las estrategias de implementación, que pueden contemplar, desde los métodos para la obtención de datos y el bosquejo de técnicas generales para estimar variables, hasta modelos cartográficos o diagramas de procedimiento y programación de algoritmos.

En resumen, los aportes de este trabajo, desde el enfoque de análisis espacial y geocomputación, consisten en:

- a) integrar los enfoques de análisis espacial y la geocomputación, al dominio de conocimiento en el que tradicionalmente se aborda el balance hídrico, el cual comprende la ingeniería hidrológica, la climatología, la agronomía y la modelación por métodos de optimización. En cuanto al enfoque sistémico, ya se mencionó que se manejan modelos hidrológicos desde esta perspectiva, pero no se ha profundizado sobre la estructuración de los componentes del balance hídrico a diferentes escalas (reconocimiento de patrones), ni se ha revisado cómo las relaciones causa-efecto también presentan diferenciación espacio-temporal y cómo esto contribuye a la configuración espacial del balance hídrico.
- b) explicitar algunos patrones espacio-temporales de los componentes del balance hídrico como precipitación y evaporación.
- c) evaluar el desempeño de los modelos o generalizaciones propuestos, abordar aspectos computacionales como la implementación de ciertos algoritmos y modelos cartográficos,
- d) proponer formas de representación de los datos que desplieguen mejor las propiedades de los componentes del balance hídrico.

Una vez planteadas las posibles contribuciones de este trabajo al estudio del balance hídrico desde la perspectiva de análisis espacial y la geocomputación, argumentaremos sobre las representaciones geográficas adecuadas para modelar este proceso.

### **3.2. Modelos geográficos**

En este apartado se argumenta sobre la naturaleza del proceso de Balance hídrico tal como es conceptualizado desde un enfoque sistémico, esto, con el objetivo de reducir el espectro de operaciones de exploración y geovisualización a aquellas que resalten mejor las características particulares del proceso a estudiar. La estrategia que se plantea aquí es definir, a priori, los modelos conceptuales que representan mejor el comportamiento espacio temporal del balance hídrico. Este ejercicio no contradice la filosofía de dejar que los datos hablen, lo que se busca es conducir mejor los análisis de exploración y así facilitar la incorporación del proceso de DPG al estudio del balance hídrico.

En general, al explorar sobre las interacciones entre los componentes del paisaje, al explicitar la dinámica de algunas de sus funciones elementales y estudiar los efectos conjuntos de una serie de procesos sobre la resiliencia de las unidades hidrológicas, estamos abordando el balance hídrico con un enfoque sistémico. La perspectiva sistémica es, sin duda, uno de los pilares más sólidos para la modelación de sistemas como el paisaje y es uno de los enfoques básicos para la resolución integral de problemas utilizada por la Geomática aplicada (Naveh, 2001:269-280).

El paisaje esta compuesto por un conjunto de sistemas complejos (ecosistemas, sistemas productivos o sociales, geosistemas, entre otras acepciones) los cuales presentan una serie de propiedades intrínsecas que les confieren una estructura específica, realizan funciones vitales para su propio mantenimiento y la interacción entre ellos conduce a cierta estabilidad a la matriz heterogénea, dicha estabilidad permite realizar algunos procesos en el contexto y la extensión total del paisaje, a su vez estos procesos son los que determinan los límites del paisaje. En cuanto al caso que nos ocupa, el paisaje está definido principalmente por aquellas áreas de captación que se encuentran dentro de la región del Lerma y que contienen al distrito de riego 011 en los estados de Guanajuato y Michoacán.

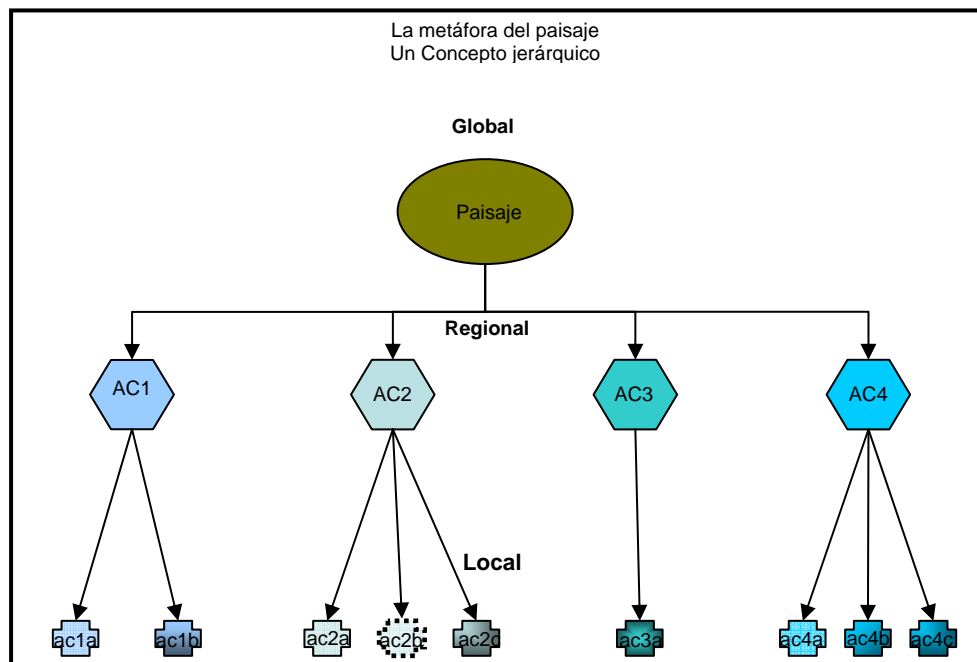
El paisaje es una unidad importante de estudio, entendida también como una construcción social, resultado y expresión de procesos interactuantes e interdependientes, un megasistema abierto, alejado del equilibrio. Se considera que los factores biofísicos y sociales son los principales motores de la dinámica en los sistemas componentes del paisaje, sin embargo, la importancia de los diversos factores varía sobre el tiempo y el espacio. Conforme ha avanzado la construcción teórica desde la perspectiva sistémica, ha sido posible emplear algunos conceptos y modelos para analizar el paisaje (jerarquías, teorías de interacción, resiliencia). Lo que se intenta aquí, es retomar conceptos de la teoría de la complejidad y de la teoría de los sistemas dinámicos como herramientas para el entendimiento de la evolución del paisaje y para apoyar su planificación integral (Miguel, 2002:17).

El paisaje se construye a partir de un **modelo jerárquico** en el que cada nivel corresponde a una escala o dominio, donde ciertos procesos operan y le confieren al conjunto una configuración espacial y temporal observable y susceptible de formalización mediante clases o

regiones (ver Figura 3.1). De esta manera, una de las operaciones de análisis que deberemos abordar, será explicitar **clases** correspondientes a diferentes escalas.

Desde la visión de la complejidad, las unidades hidrológicas son estructuras disipativas que emplean gran cantidad de energía, recursos e información para conservar su estabilidad. A pesar de entrar en fases de inestabilidad, el paisaje es capaz de alcanzar un nuevo estado estable, en cuyo caso, hablamos de autoorganización (Tyrtania, 1999:65). Esta estrategia adaptativa o mecanismo de evolución no depende de factores externos sólo es resultado de las propias interacciones internas en el paisaje y dentro de sus componentes. La forma como pueden explicitarse las interacciones entre unidades hidrológicas es a través de la modelación de flujos y escurrimientos de volúmenes de agua. Normalmente los flujos se representan como **redes** y el patrón de escurrimientos es el que define finalmente que áreas están alimentando una unidad de captación y como se relaciona ésta con otras unidades. Para el caso particular de este trabajo no se busca modelar la red superficial de escurrimientos sino emplear la variación del relieve para delimitar áreas de captación.

**Fig. 3.1 Representación del paisaje como modelo jerárquico**



En cuanto a los factores externos, el principio de Lotka establece que aquellos sistemas que hagan un uso eficiente y extensivo de los recursos disponibles son los mejor adaptados hasta que

se agote el mínimo disponible de energía (Adams, 2001:157). La principal entrada al sistema hidrológico es la lluvia y la proporción de precipitación que esté disponible al almacenamiento puede ser un indicador de la evolución de la recarga hídrica. Aunque en este trabajo no se realizará un análisis de la variabilidad interanual de las variables del ciclo hidrológico, se avanzará en el reconocimiento de posibles **estructuras temporales** en los componentes del balance considerados y en las variables causales asociadas.

Recordemos también que partimos de datos puntuales para evidenciar posibles patrones espaciales, por ello, en una primera fase, debemos identificar la existencia de **vecindarios** y luego proceder a la delimitación del dominio de escala del proceso. Paralelamente, requerimos identificar las relaciones causa efecto que nos permitan explicar los patrones identificados, por lo tanto, otra tarea a realizar será la búsqueda de **asociaciones**.

En resumen, hemos definido aquellos modelos espaciales que pueden representar más claramente las propiedades de los procesos involucrados en el balance hídrico y que también ofrecen un recurso metodológico estratégico para el manejo y administración del agua.

## **4. Enfoques metodológicos para la extracción de patrones geoespaciales**

Una vez definido el marco teórico desde el cual se conceptualiza el problema del balance hídrico, es necesario plantear una línea metodológica fundamentada científicamente para contribuir al descubrimiento de patrones geoespaciales.

### **4.1. El proceso general de descubrimiento de patrones**

Plantear los métodos más adecuados para la obtención y manejo de datos hidrológicos de tal forma que podamos reconocer patrones espacio-temporales de la precipitación, la temperatura y la evaporación, es uno de los principales aspectos a resolver en la problemática planteada en la presente investigación. Para cubrir esta necesidad requerimos fundamentalmente de un proceso que nos permita identificar patrones nuevos, válidos, potencialmente útiles y en última instancia comprensibles en los datos que pueden obtenerse de diversas fuentes. Este proceso general denominado por Fayyad *et al.* (1996:38) como descubrimiento de conocimiento en bases de datos es el antecedente metodológico para el descubrimiento de patrones geoespaciales. El proceso de descubrimiento de conocimiento en bases de datos o KDD (por sus siglas en inglés) contempla un conjunto de etapas que pueden realizarse o no consecutivamente, según las necesidades del analista, sin embargo no considera las características específicas de datos espaciales (ver la figura 4.1).

Como vemos en la figura 4.1, este proceso no es otro que la aplicación del método científico aplicado a bases de datos cuando el problema es el manejo de un gran volumen de datos a partir del cual se requiere generar información útil y válida a partir del reconocimiento de patrones mediante técnicas de minería de datos. Si bien esta metodología es útil y consistente, parte de un razonamiento abductivo, en cambio, la estrategia que se ha seguido en este trabajo, es formular de forma precisa la cuestión a resolver, más que optimizar los detalles algorítmicos de un método particular de minería de datos o dejar solamente que los datos hablen.

Además de establecer claramente cuáles herramientas de exploración nos pueden conducir a reconocer cierto tipo de patrones, debemos apoyarnos en algunos métodos de despliegue y visualización de datos que nos faciliten dicha tarea. Desde una perspectiva histórica, la convergencia de la metodología de KDD con los requerimientos de herramientas específicas para la visualización de datos geográficos, por parte de geógrafos y cartógrafos, planteó la necesidad de conformar una nueva línea metodológica que considerara los requerimientos específicos de

análisis de datos espaciales, la formalización de los patrones espaciales y temporales identificados y la búsqueda de explicaciones plausibles a los comportamientos observados; la definición del proceso de DPG, buscó suplir tal necesidad. Por esta razón, se revisa en primer lugar la importancia de la geovisualización y su aporte al reconocimiento de patrones espaciales y en segundo término se detallan las características del proceso de DPG.

**Fig. 4.1 Proceso de descubrimiento de conocimiento en datos (KDD).**

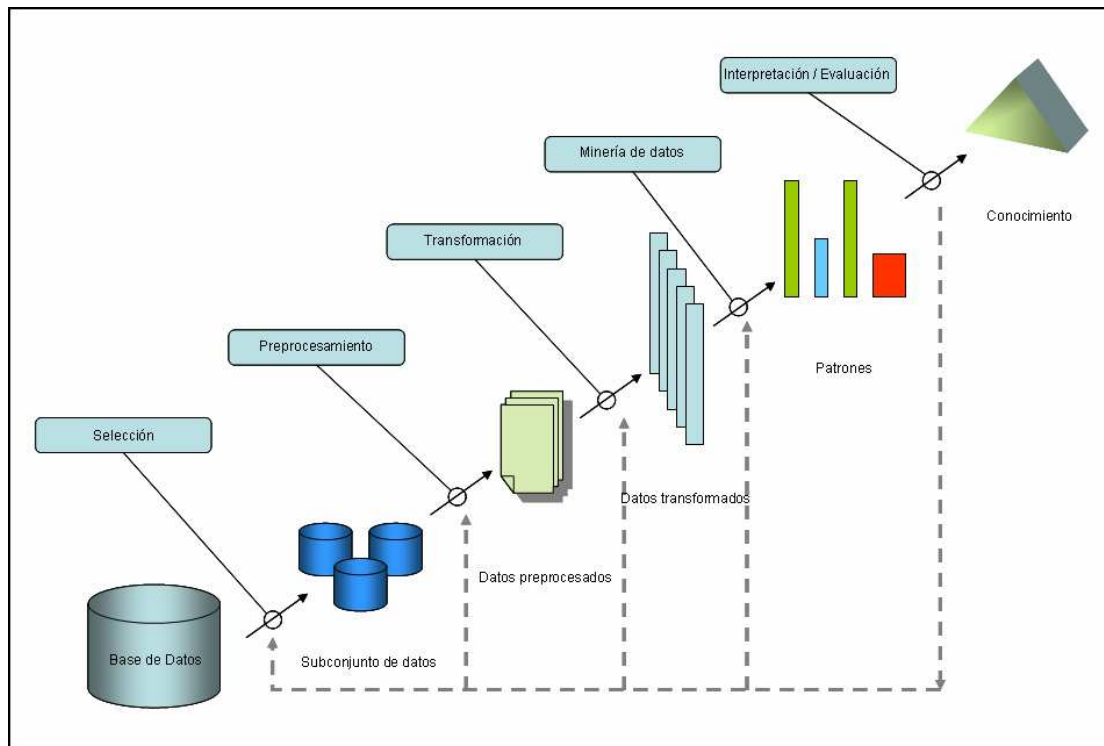


Diagrama que resume las fases que componen el proceso de KDD. Modificado de Fayyad *et al.* (op cit.:40).

## 4.2. Herramientas de Geovisualización

En general, se ha visto que la utilidad de las representaciones de datos u observaciones para realizar un proceso científico o bien de la información generada a través de éste, puede aumentar significativamente al desplegar el contexto o dominio de conocimiento al cual corresponden, en el caso particular de datos geoespaciales nos referimos al espacio y al tiempo. Por ello, no es casualidad que la noción de análisis exploratorio de datos o KDD, desarrollado, como sabemos, en estadística, matemáticas y computación, comenzó a ser considerada indispensable para la cartografía en la década de los 90 (DiBiase, 1990:14). Más recientemente, los cartógrafos han reconocido la necesidad de nuevo software que permita a los especialistas generar mapas y usarlos como herramientas para facilitar el pensamiento visual acerca de datos geoespaciales

(Kraak, 1998:12-14). A su vez, se han efectuado grandes esfuerzos para integrar el poder de los métodos cartográficos de representación al servicio de la exploración de datos espaciales (Andrienko *et al.*, 2001:152; Maceachren y Kraak, 1997:335-338).

Maceachren y otros (1999:312), acuñan el concepto de geovisualización como el proceso que involucra la obtención de información o aprendizaje al interactuar con los datos a través de despliegues visuales que son representaciones de estos datos y de las operaciones que pueden aplicárseles. Estos mismos autores distinguen 3 meta operaciones de visualización que se aplican en un proceso común de visualización cartográfica: identificación, comparación e interpretación de rasgos, objetos o patrones. Al integrar estas meta operaciones con las diversas tareas que se realizan en el proceso de KDD, desarrollan un esquema a partir del cual, se deducen algunos métodos de minería de datos que pueden ser más adecuados para cada una de las meta operaciones de geovisualización.

**Fig. 4.2 Matriz de integración de los métodos de KDD y geovisualización.**

		Metaoperaciones del proceso de KDD		
		Conceptualización de jerarquías y estructuras	Extracción de categorías y clasificación	Extracción de fenómenos
Metaoperaciones de Geovisualización	Identificación de patrones	se pueden distinguir los niveles a los cuales se configuran patrones y apoyar en la conceptualización de jerarquías	contribuye a establecer límites entre categorías, identificar datos que ejemplifican una categoría y aquéllos que son anómalos	ayuda a encontrar instancias de fenómenos reales que cumplen con los criterios derivados de la caracterización de categorías
	Comparación de patrones	contribuye a la creación de una estructura multidimensional que relaciona y compara niveles en jerarquías	facilita la comparación de categoría para determinar aquéllos atributos por los cuales las categorías se relacionan o difieren	puede ser usado para comparar fenómenos y derivar reglas entre ellos, que describan relaciones
	Interpretación de patrones	facilita la liga entre las jerarquías conceptuales o la estructura conceptual que las define a un dominio de conocimiento particular	permite analizar las causales de las categorías extraídas con relación a parámetros del mundo real	puede emplearse para desarrollar una caracterización de los fenómenos y modelar sus comportamientos

Modificado de Maceachren *et al.* (*op cit.*: 320).

Al observar la figura 4.2 podemos darnos cuenta del enorme potencial que deriva de aplicar la geovisualización a las operaciones de KDD y en particular podemos avanzar hacia la caracterización o generalización de las entidades u objetos espaciales que estudiamos. Se



considera que las operaciones de geovisualización pueden acompañar cada una de las etapas en el descubrimiento de patrones. En este trabajo se hará énfasis en cuáles son las operaciones de identificación, comparación y de extracción de patrones que son más adecuadas para la caracterización de componentes del balance hídrico mas no se implementarán técnicas avanzadas de gran interactividad pues ello implica realizar análisis cualitativos de requerimientos de usuarios (Gahegan *et al.*, *op cit.*:18).

### **4.3. El descubrimiento de patrones geospaciales o DPG**

Modificado del método científico tradicional y del proceso de KDD, la figura 4.3 resume las diferentes fases a seguir para el descubrimiento de patrones geospaciales. Es importante resaltar que la importación del modelo metodológico de KDD al proceso de DPG aplica propiamente a las 2 primeras etapas, es decir, sólo aborda las fases de manejo de datos con énfasis en el reconocimiento de patrones. La segunda parte del proceso hace énfasis en la modelación de procesos y sus causas y en la caracterización de los mismos.

En relación con la importación de herramientas de minería de datos, Miller (*op cit.*:37-41), realiza un análisis preliminar sobre las implicaciones de aplicar los métodos de exploración de datos tradicionales a datos espaciales, distinguiendo al menos 5 grupos de herramientas de minería de datos aplicables a observaciones de fenómenos geográficos. Cada grupo corresponde a un tipo genérico de patrón, rasgo u objeto espacial a saber, clases, asociaciones, reglas, aglomeraciones y anomalías. Estos patrones tipo tienen una expresión espacial ya que pueden estar condicionados por su morfología o por las relaciones espaciales entre objetos.

En este enfoque metodológico los avances son incipientes y de acuerdo a un documento propuesto al comité de investigación del UCGIS algunas áreas de investigación de frontera son: desarrollar repositorios de datos (data warehouses), diseñar mejores representaciones espacio-temporales, implementación de interfases para el proceso de DPG, integración con la tecnología de los SIG y análisis espacial (sitio de referencia : <http://kartoweb.itc.nl/icavis/index.html>). Con relación a estos aspectos de innovación, se explorarán diferentes opciones, adecuadas a la naturaleza del balance hídrico.

**Fig. 4.3 Proceso de descubrimiento de patrones geoespaciales.**

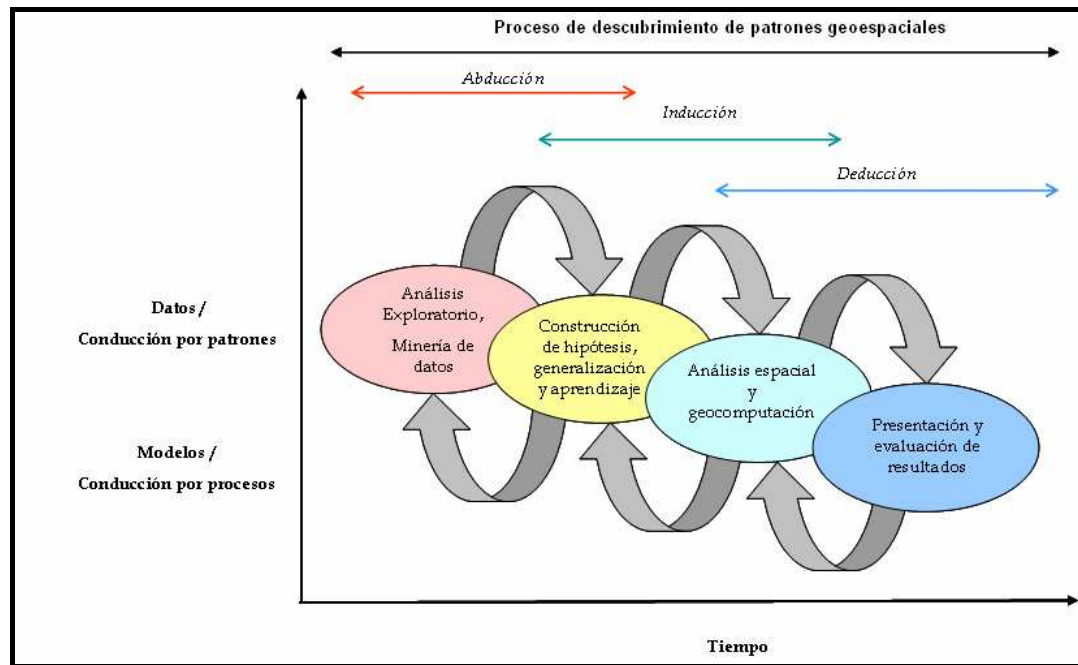


Diagrama del proceso DPG y los modos de inferencia que se aplican en cada fase. Modificado de Gahegan *et al.* (op cit.: 2).

#### 4.4. Aplicación del proceso de DPG en soluciones computacionales

Se han desarrollado diversas soluciones para el análisis exploratorio de datos espaciales, las cuales van desde la concepción y diseño hasta la implementación de algoritmos de minería de datos espaciales, algunas de ellas se caracterizan por ser altamente interactivas y por presentar despliegues dinámicos de datos. (Ester *et al.*, 1999:1-14; Anselin *et al.*, 2004:1-18). La mayoría de estas aplicaciones facilitan la construcción de mapas y el manejo de datos, con objeto de promover en el usuario la exploración de variables, la búsqueda de relaciones causales o de patrones (Cook *et al.*, 1997:372-380). Todos estos trabajos, sin excepción, emplean métodos de inferencia abductivos e inductivos (énfasis en el reconocimiento de patrones), mientras que la evaluación de hipótesis o generalizaciones queda fuera de los alcances de estas aplicaciones.

A continuación se incorpora una revisión general de modelos hidrológicos, para cubrir el objetivo de avanzar hacia la implementación de un modelo de balance hídrico aplicable a cuencas poco monitoreadas. Se revisan también, antecedentes sobre métodos de exploración y geovisualización de datos hidrológicos.

## **5. Antecedentes**

### **5.1. Enfoques metodológicos y conceptuales para la modelación del balance hídrico**

Se han realizado diversos esfuerzos a nivel local y regional desde diversas escuelas y enfoques para abordar la modelación hidrológica. Los diversos componentes del ciclo hidrológico así como otras funciones asociadas como la tasa de descarga máxima de esorrentías y el flujo de escurrimientos han sido estudiados desde dos puntos de vista: la modelación de la recarga y el balance hidrológico o la simulación de volúmenes y flujos de escurrimientos.

La primera aproximación busca la integración de los diferentes componentes hidrológicos para modelar funciones ecológicas complejas como la recarga de acuíferos, el balance hidrológico o el cambio en la humedad del suelo, dicha modelación sirve para apoyar el manejo integrado de cuencas y la programación de actividades que conduzca a un uso racional del agua, ejemplos de este enfoque multidisciplinario se pueden revisar a detalle en Pickett *et al.* (2001:145-155) y Pickett *et al.* (2004: 380-384) quienes integran en una estructura o modelo conceptual diversos procesos ecológicos y sociales que convergen en el espacio y el tiempo, los cuales finalmente configuran de forma específica al paisaje y su funcionamiento hidrológico. Sin embargo, en ninguno de estos trabajos se pone especial énfasis en los métodos de obtención de datos para alimentar los modelos hidrológicos, y mucho menos se hace explícita alguna metodología para estructurar el paisaje y la dinámica hidrológica.

El segundo enfoque tiene una visión más disciplinaria dirigida a la identificación de zonas con alto riesgo de inundaciones, desbordamiento de ríos y otros almacenamientos para apoyar la planeación urbana y la construcción de obras de infraestructura hidráulica y la generación de energía eléctrica. En estos trabajos se da mayor énfasis a la formalización de relaciones entre variables físicas y de estructura del terreno con el comportamiento de los diversos componentes del balance, se hacen esfuerzos importantes para calibrar los modelos empíricos y así minimizar los errores en el diseño de las obras. Lógicamente, esta modelación no puede aplicarse a diversas condiciones ambientales (Singh, 1995:16-20).

### **5.2. Ejemplos de implementación de modelos**

Respecto del desarrollo de aplicaciones computacionales para la modelación hidrológica existen antecedentes desde los dos enfoques metodológicos mencionados arriba. Desde la perspectiva de

ingeniería, los trabajos se enfocan en los problemas de inundaciones, disminución de la calidad del agua y recarga hídrica a nivel de unidades administrativas o hidrológicas, de tal forma que se facilite la ejecución de actividades para la mitigación de impactos negativos a nivel local y regional (Sanders, 1986:361-363, Singh, *op cit.*:265-280). Por su parte, desde el enfoque sistémico Costanza *et al.* (1990:91-107) y Bockstael *et al.* (1995:143-159), diseñan y evalúan modelos ecológicos sociales para evaluar los impactos de políticas de uso de suelo sobre diferentes funciones ecológicas, entre ellas la hidrodinámica, la hidrología y la productividad primaria en ecosistemas acuáticos.

En general los modelos referidos están dirigidos fundamentalmente a usuarios expertos los cuales cuentan con una vasta experiencia acerca de los fenómenos que se pretenden modelar, precisan de un fuerte esfuerzo computacional, una serie de requerimientos de software, hardware y datos que un usuario no experto difícilmente puede cubrir. Los productos finales o resultantes de los procesos de simulación o modelación no son genéricos ni simplificados por lo que haría falta una etapa adicional de generalización para hacerlos accesibles al público interesado.

Estas particularidades en las aplicaciones tienen que ver con cubrir las necesidades de usuarios que tienen un acceso fácil a los datos para la calibración de los modelos y pretenden comprender a mayor profundidad los procesos ecológicos de interés. En este sentido, cabe destacar algunos esfuerzos por simplificar modelos con enfoque sistémico acerca del balance hídrico, sin perder de vista la formalización, lo que facilita con mucho la implementación. Estos modelos hacen un uso eficiente de los datos disponibles y se adecuan a las especificidades del área de estudio o paisaje de interés, por lo que son los más adecuados para cuencas poco monitoreadas donde se da preferencia a la modelación semidistribuida del balance hídrico a nivel de unidades hidrológicas, lo cual es el caso del área piloto propuesta. Las diferencias en estos modelos estriban en cuál es la variable empleada para calibración, los trabajos de Xu, Vandewiele (1995:39), Xu, *et al.* (1996:211) emplean los volúmenes de escurrimientos y más recientemente Mazi *et al.* (2004:243), Fazal *et al.* (2005:56) y Nolan *et al.* (2006), utilizan estimaciones de recarga o registros de los niveles de agua en subsuelo. Normalmente, los modelos calibrados con datos de la dinámica del subsuelo, requieren una buena caracterización de suelos y de las propiedades de infiltración de los materiales consolidados (rocas).

### 5.3. Desarrollo de herramientas de geovisualización y de exploración de datos hidrológicos

En cuanto a enfoques de análisis visual para el manejo de datos hidrológicos se han hecho algunos avances, Kreuseler, (2000:101-107), ofrece algunas herramientas para la exploración de observaciones de variables biofísicas a lo largo de la columna de aguas oceánicas. Se propone el uso de diagramas de profundidad (los cuales funcionan de manera similar a los diagramas de coordenadas paralelas) y perfiles de tiempo, tales herramientas facilitan la visualización de relaciones o asociaciones entre variables; para observar patrones espaciales se propone el uso de visualizaciones de escenas 3D y planos codificados en color (como mapas de cloropletas). Fuhrman (2000:12-18), desarrolla métodos de visualización de datos hidrológicos y la documentación de modelos hidrológicos, conformando una herramienta de visualización cartográfica y de consulta de bases de datos.

Mitas y Mitsova (1997:438-442), demuestran que el uso de animaciones apoya la exploración de patrones en el espacio y el tiempo en los fenómenos a escala del paisaje. En el trabajo de Portoghese *et al.* (2005:15-27) se maneja paralelamente el uso de mapas y perfiles de tiempo asociados a unidades hidrológicas, no obstante, se da más peso a las bondades de los modelos propuestos sin abundar en la caracterización de las unidades ni en la formalización de relaciones causales, aún cuando se manejan suficientes características físicas de las unidades del paisaje. Al contrario, Mazi (*op cit.*:247-253) y Nolan (*op cit.*) ya incorporan el empleo de algoritmos para la caracterización de relaciones causales entre la recarga y diferentes variables ambientales, entre ellas la precipitación, aunque se privilegia la variabilidad temporal de estas asociaciones. Es a partir de las investigaciones de Fazal, (*op cit.*:60-77) y Maciek *et al.* (*op cit.*:80-93), que se revisa someramente el comportamiento diferenciado de las asociaciones entre la recarga y variables ambientales, a escala del paisaje. Al mismo tiempo, estos trabajos plantean el uso de perfiles de tiempo para la variable observada y su estimación, comparándolos mediante el uso de plantillas. En estas referencias se ha encontrado que la recarga tiene una resolución temporal mínimamente anual a diferencia de los patrones de precipitación que son más dinámicos, sin embargo, el grado de asociación entre estos dos componentes del ciclo hidrológico a esta escala temporal, es fuertemente lineal a nivel global. Cuando se maneja una escala temporal mensual y se caracterizan las asociaciones entre la precipitación y la recarga a nivel del paisaje, las unidades hidrológicas presentan una diferenciación espacial y temporal que tiene que ver, por un lado, con la dinámica de la precipitación a lo largo del año y por otro lado, está

relacionada con las características específicas del subsuelo que pueden llegar a presentar una variación muy local. No hay referencias sobre el efecto de la cobertura del terreno en la recarga en áreas como el Bajío, con una fuerte intervención del hombre sobre los patrones de escurrimientos. La revisión bibliográfica confirma que no se han diseñado soluciones para la incorporación de procesos de análisis confirmatorio sobre los patrones observados en las variables hidrometeorológicas o hacia la generación de recursos simplificados de información.

#### **5.4. Métodos generales para obtener registros de componentes del Balance Hídrico y variables causales asociadas**

Como se mencionó en el capítulo anterior es a través de Percepción remota y observaciones directas de estaciones hidrometeorológicas como se obtienen las variables para la calibración de modelos de la dinámica y la función hidrológica. Existen gran cantidad de trabajos que se han enfocado en la obtención de variables del ciclo hidrológico a partir de la aplicación de la percepción remota, estos incluyen la estimación de la precipitación, la humedad del suelo, el cálculo de la evapotranspiración real, escurrimientos y corrientes subterráneas, entre otros (Hall *et al.*, 1995:138-156; Saraf y Choudhury, 1998:1825-1828; Lettenmaier, 2004). En todos estos estudios se resalta la posibilidad de realizar monitoreos consistentes, con un cubrimiento territorial a nivel regional y facilitar la comparación entre patrones a lo largo del tiempo.

En cuanto a la estructuración espacio temporal de los componentes del ciclo como precipitación, temperatura y evapotranspiración, medidos directamente por instrumentos, los métodos de interpolación para la estimación de estas variables regionalizables ha sido explorada intensamente por Rossi *et al.* (1992:277-300) y Myers (1994:283-315). Legendre y Fortín (1989:107-138), hacen una revisión de diversos métodos para explicitar la auto correlación espacial, entre otros, los variogramas, los correlogramas y algunos índices de covariación espacial como el índice de Moran. Por su parte, Mora e Iverson (2002:160-165), sugieren una forma de deducir la escala de los procesos estudiados a través de la variografía.

En los últimos años, la configuración del paisaje y los patrones de cambio de la estructura de la cubierta vegetal se han estudiado ampliamente a través de procesos de fragmentación de bosques (Ritters *et al.*, 2000:3-20) y estudios multitemporales (Hurd *et al.*, 2003). Estos fenómenos tienen una incidencia directa sobre el volumen y patrones de escurrimientos, sobre las

tasas de evapotranspiración y afectan la capacidad de retención de humedad del suelo. Una cuestión importante a resolver que queda fuera del alcance de este trabajo, es conocer los efectos de la fragmentación de bosques sobre los componentes del balance, ya que es un proceso que recién comienza a estudiarse en paralelo al funcionamiento hidrológico.

Más recientemente, mediante la tipificación del crecimiento urbano y el monitoreo de superficies impermeables (Hoffine *et al.*, 2002) se ha comenzado a reconocer la importancia de relacionar estos fenómenos antropogénicos con la caracterización de funciones hidrológicas, dada la tendencia de urbanizar áreas boscosas. Sin duda, este es otro de los nichos de investigación que requiere atención en el largo plazo, dada su afectación a los servicios ambientales en zonas rurales. Por último, se realiza una revisión general de los modelos del balance hídrico desde la perspectiva de la hidrología.

### **5.5. Modelación del balance hídrico**

De acuerdo con Xu y Vandewiele (*op cit.*:40), las principales líneas de investigación en la construcción de modelos de balance hídrico son:

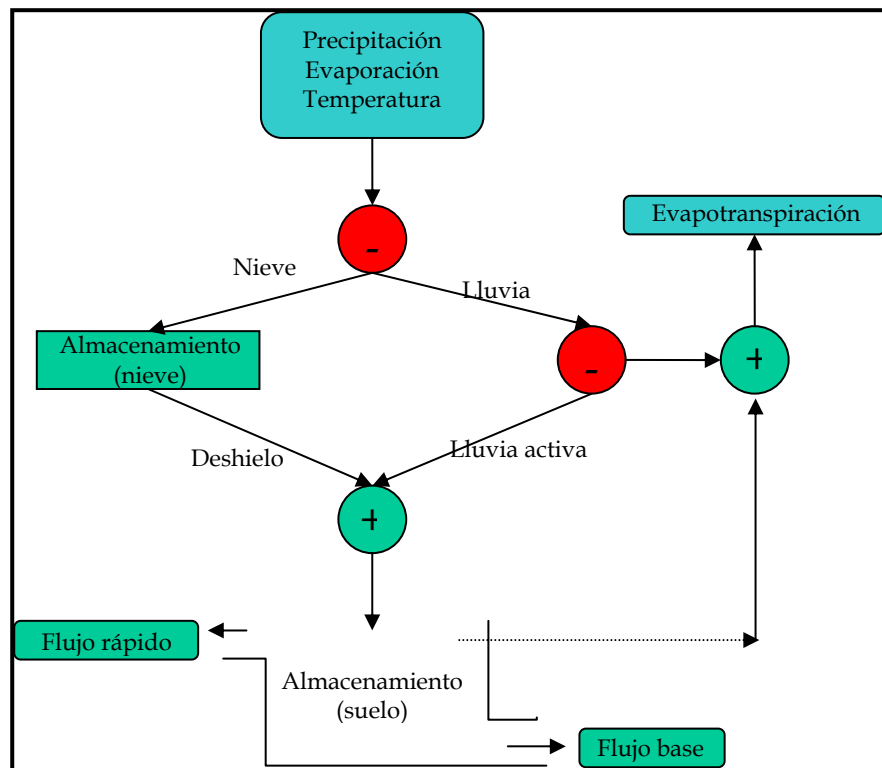
- Reconstrucción de la hidrología de subcuencas.
- Evaluación de cambios por impacto climático.
- Evaluación de los patrones estacionales y geográficos de la distribución de agua y la demanda de irrigación.

Entre el conjunto de modelos de balance hídrico es conveniente emplear modelos conceptuales ya que las funciones que consideran los procesos físicos actuantes sobre las entradas y salidas tienen forma simple. Entre estos, los modelos periódicos serían los más útiles ya que capturan la temporalidad del proceso y tienen menor requerimiento de datos que modelos más complejos.

Los modelos conceptuales de balance hídrico conceptualizan una unidad hidrológica como un agregado de almacenamientos conectados, a través de los cuales entra el agua como lluvia y sale como flujos a la salida de la unidad, exploran además las relaciones que controlan el proceso del balance hídrico (Fig. 5.1). Por estas características se consideran adecuados para estudios a escala del paisaje desde una perspectiva sistémica.

Estos modelos manejan los procesos hidrológicos como espacialmente promediados o agregados. Algunos de sus parámetros se estiman empíricamente a partir de precipitación y flujos. Su principal objetivo es simular flujos. Tienen una estructura simple y un pequeño número de parámetros comparados con modelos con resolución temporal fina. La base es la ecuación de balance hídrico.

**Fig. 5.1 Diagrama del balance hídrico a nivel de unidades hidrológicas.**



Tomado de Xu, *et al.* (1996:237).

Si no hay información espacial sobre la dinámica temporal de los diferentes componentes del balance hídrico, los modelos conceptuales semiempíricos facilitan el cálculo de algunas variables como evapotranspiración, almacenamientos y flujos a partir de algunos parámetros climáticos y características del terreno como la cobertura, el relieve y la permeabilidad. La ventaja de los modelos conceptuales semiempíricos es que utilizan pocos parámetros y es posible incrementar la cantidad de datos que se requieren para la calibración y obtener predicciones más confiables.



Xu y Singh (1998:35-48), realizan una revisión exhaustiva de los modelos mensuales y reportan algunos métodos para calcular o estimar algunos componentes del balance hídrico:

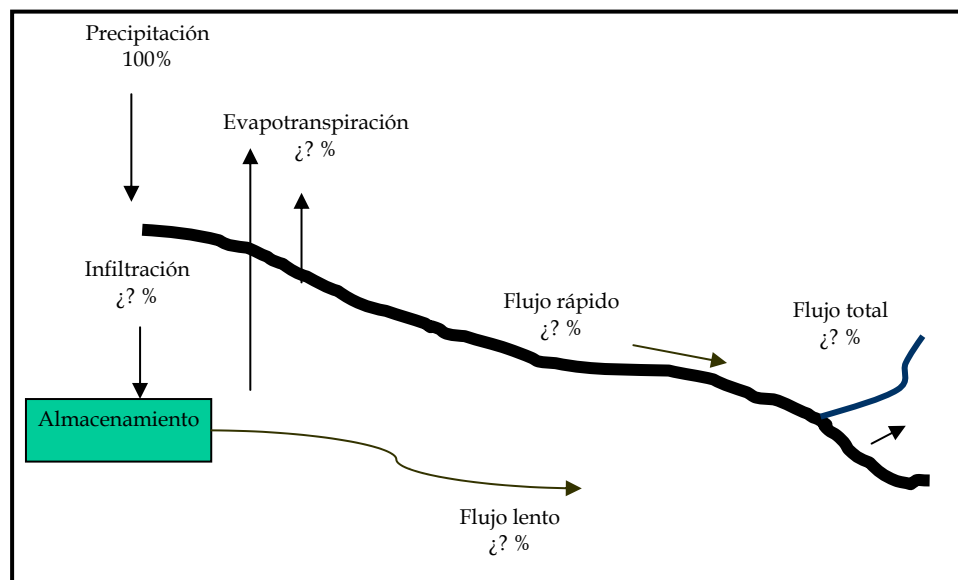
- La precipitación es la variable con mayor heterogeneidad espacial y temporal. Un cálculo confiable del balance hídrico no puede ser posible con un conocimiento insuficiente de los patrones espaciales de la lluvia. Los métodos de interpolación y la estimación empleando datos satelitales o de radar son las opciones para obtener la superficie que modele la variación espacial de la intensidad de la precipitación.
- La evapotranspiración [ET] es la segunda variable más importante en el balance hídrico. Se requieren predicciones precisas de la ET para los modelos de balance. El cálculo de esta variable se realiza a través de tratamientos simples o complejos, entre estos, podemos mencionar los modelos de Thornthwaite y Penman.
- La evapotranspiración real [ETa] es una función de la potencial (evaporación) y de la humedad del suelo. En el procedimiento para estimar ETa se debe considerar la evapotranspiración local del sistema suelo-planta y generalizar los datos para las unidades hidrológicas, determinar la evapotranspiración areal utilizando el concepto de relación complementaria entre ambos tipos de evapotranspiración, puntual y areal. Al igual que la precipitación, los métodos de interpolación y técnicas de PR son empleadas para estimar la evapotranspiración aunque en el último caso se utilizan datos de temperatura del aire y de vegetación.
- La habilidad de los modelos de balance hídrico para predecir descargas es uno de sus principales objetivos. Es importante conocer los diferentes componentes de los escurrimientos y sus regímenes. Se reconocen 4 componentes: flujo superficial, interflujo rápido, (componentes rápidos), interflujo lento y flujo base (componentes lentos).

El objetivo general de los modelos conceptuales mensuales semiempíricos es modelar el balance hídrico a nivel de unidades hidrológicas y obtener las proporciones correspondientes a cada componente en el sistema (Ver Fig. 5.2).

A partir de lo revisado en esta sección, se considera que un modelo conceptual semiempírico de balance hídrico, aplicable a subcuencas poco monitoreadas, que permita estimar parámetros a través de técnicas de exploración de datos, sería el más adecuado para estudiar el comportamiento del balance hídrico y para explicitar en la medida de lo posible las relaciones

empíricas de sus componentes respecto de variables físicas causales. En este caso se sugiere el modelo NOPEX diseñado por Xu, Seibert y Halldin (1996:211), pues permite estimar evapotranspiración real y almacenamiento disponible en suelo, empleando como entrada la precipitación y la temperatura del aire y para la calibración, volúmenes de escurrimientos. Más adelante se presentarán los avances en la implementación de este modelo aún con las limitaciones de datos ya referidas.

**Fig. 5.2 Representación gráfica del balance hídrico.**



## **6. Aportes**

### **6.1. Alcances**

En el presente trabajo se busca profundizar en las relaciones entre los componentes del balance hídrico y otras variables, se proponen operaciones de geovisualización y análisis espacial confirmatorio que permitan extraer patrones espacio-temporales de los componentes del balance hídrico. Para cubrir esta meta se busca cubrir los siguientes aspectos:

Caracterización espacio temporal de componentes del balance hídrico a nivel de unidades del paisaje en el distrito de Riego 011, considerando aquellas áreas de captación que lo contengan, por medio de herramientas de geovisualización y exploración.

Evaluar la importancia de los métodos de análisis espacial en el cálculo de superficies para la estimación de variables regionalizadas.

Diseñar y probar métodos de procesamiento digital de imágenes de PR para la estimación de elementos del ciclo hidrológico.

Proponer los métodos más adecuados para el análisis exploratorio de datos hidrológicos y su posible aplicación en fases avanzadas del DPG.

### **6.2. Objetivos**

Considerando los alcances de la presente investigación, el objetivo general a realizar es incorporar el proceso de descubrimiento de patrones geospaciales en el estudio del balance hídrico a escala del paisaje, para lo cual es necesario cumplir con los siguientes objetivos:

1. Evaluar las limitaciones de los modelos de representación lógica de los datos hidrometeorológicos de precipitación y temperatura del aire.
2. Definir algunos métodos de obtención y preprocesamiento de los datos hidrometeorológicos y su transformación a componentes del balance hídrico como precipitación, temperatura del aire, evaporación y variables explicatorias
3. Identificar las herramientas de análisis exploratorio y de geovisualización útiles para la identificación y comparación de clases, asociaciones, estructuras temporales y vecindarios en la precipitación, temperatura del aire, evaporación y en variables explicatorias.
4. Definir los atributos y propiedades de los datos generados a partir de la aplicación de las diferentes fases del proceso de descubrimiento de patrones espaciales.

5. Proponer modelos para la representación de hipótesis o generalizaciones plausibles para explicar los patrones identificados en los componentes del balance hídrico.
6. Identificar las herramientas de análisis confirmatorio y espacial para la extracción de clases, asociaciones y estructuras temporales en los componentes del balance hídrico.
7. Avanzar en la implementación de un modelo de balance hídrico semidistribuido y semiempírico para cuencas poco o no monitoreadas.

## **7. Incorporación del proceso de descubrimiento de patrones geospaciales al estudio del balance hídrico**

### **7.1. Consideraciones generales para la incorporación del proceso de DPG**

El dominio de conocimiento que contextualiza el proceso del balance hídrico involucra, como ya vimos, diversas disciplinas, tales como, el análisis espacial, la geocomputación, la geovisualización, la modelación hidrológica y el enfoque sistémico. No obstante, las propias características del balance hidrológico y la naturaleza de las observaciones hidrometeorológicas, nos conducen a enfatizar sobre las propiedades y comportamiento de ciertos objetos geográficos: regiones, vecindarios, jerarquías, estructuras temporales.

Delimitarnos a cierto dominio de conocimiento implica contar con experiencia en cuanto al manejo de datos hidrometeorológicos, su comportamiento y el manejo de modelos o conceptos a priori sobre el balance hidrológico para poder apoyar o refutar posibles hipótesis que se construyan a partir de la exploración de datos. Esto implica que el analista sea capaz de tomar decisiones respecto de la validez de categorías o patrones resultantes de los análisis y que defina de acuerdo a sus observaciones cuáles métodos de exploración son los más adecuados para la identificación de los objetos geográficos propuestos.

Como se resaltó en el planteamiento del problema, la disponibilidad de observaciones de los componentes del balance hídrico es muy limitada y poco consistente, por lo que debemos plantear e implementar tareas de preprocesamiento para transformar los registros de observaciones tal como son entregadas por la CNA u otras fuentes de datos hidrometeorológicos a los formatos de representación lógica que modelen mejor las propiedades de los componentes del balance y de las variables relacionadas.

Para definir las tareas de análisis más adecuadas, con respecto del universo de herramientas de geovisualización, de métodos de análisis exploratorio y de modelación disponibles, se debe disponer de una muestra de datos del área de interés y comenzar con la selección de técnicas para resaltar las propiedades y comportamiento de los objetos geospaciales definidos.

## **7.2. Incorporación del proceso de DPG**

### **7.2.1. Organización, manejo y visualización de datos**

Se parte de los datos tal como son obtenidos de observaciones o registros en fuentes oficiales. En esta etapa se define la estructura básica de los datos y las operaciones básicas a realizar (consultas, selección, despliegue).

#### **7.2.1a. Definición de datos de entrada y de formatos de representación lógica**

Los componentes del balance hídrico son variables regionalizables, lo cual significa que presentan variación en una o varias direcciones sobre una extensión de terreno. Normalmente la media de las variables no es constante en las diferentes porciones del área de interés, lo cual conduce al principio de heterogeneidad espacial que se expresa en forma de rangos o clases de precipitación, temperatura, escurrimientos, etc. (Cooper e Istok *op cit.*:270-299 y Fortín 1999:253). Este supuesto general nos lleva a inferir que la mejor forma de representar el comportamiento de estas variables es a través de superficies y para realzar los cambios en toda la extensión del área de interés, la representación en 3D es la más adecuada por ser el espacio en el cual nos manejamos cotidianamente.

La CNA entrega datos de estaciones hidrometereológicas en formato alfanumérico. Si se pretende visualizar la heterogeneidad espacial de los datos de estaciones se requiere ubicarlos en un plano o superficie de acuerdo a sus coordenadas de georeferencia. El formato vector de representación lógica, permite representar los datos de estaciones metereológicas como puntos y con referencia a un sistema de proyección de coordenadas planas, esto permite la medición de distancias y direcciones entre puntos. Para identificar posibles estructuras temporales en los datos, los valores de las variables deben ser promediados o sumariados a periodos más amplios como quincenas o meses ya que los estudios regionales del balance hídrico se manejan en estos periodos de tiempo.

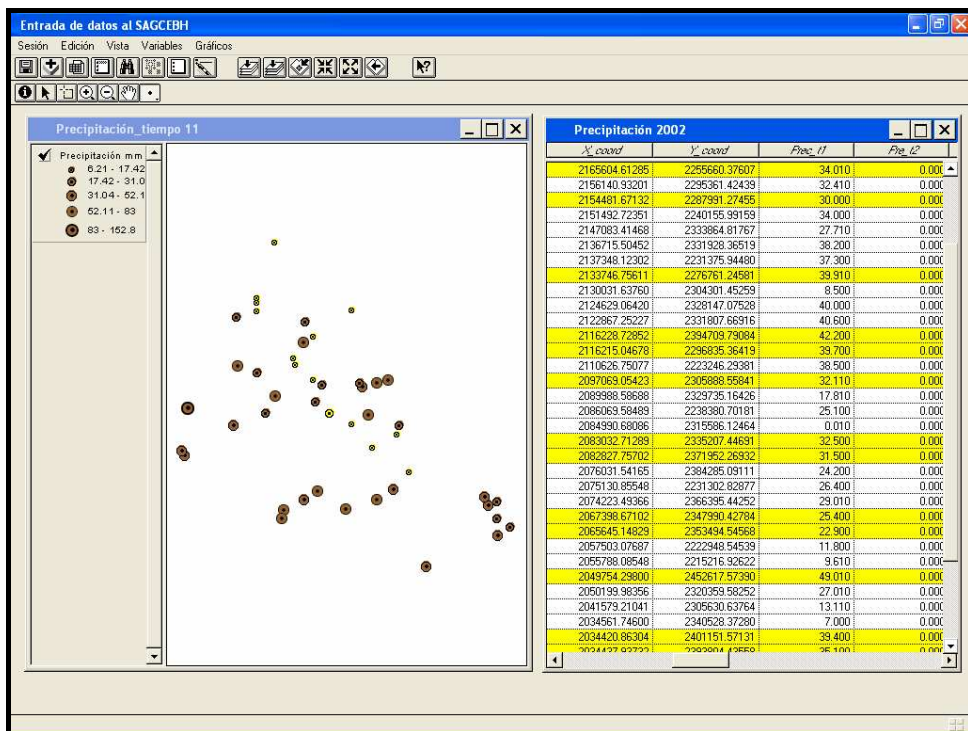
Un ejemplo de los datos en el formato vector puede visualizarse en la figura 7.1. Al añadir elementos simples de representación como simbología de mapas podemos distinguir posibles patrones de comportamiento de las variables. La figura 7.1 muestra la representación gráfica de los datos dentro de un contexto geográfico y los atributos asociados a los puntos de observación que en este caso se refieren a la precipitación

acumulada de un periodo 'x' del año. Pueden observarse gradientes desde diversas direcciones lo que indica que la variable no se comporta de la misma forma en toda la extensión del área.

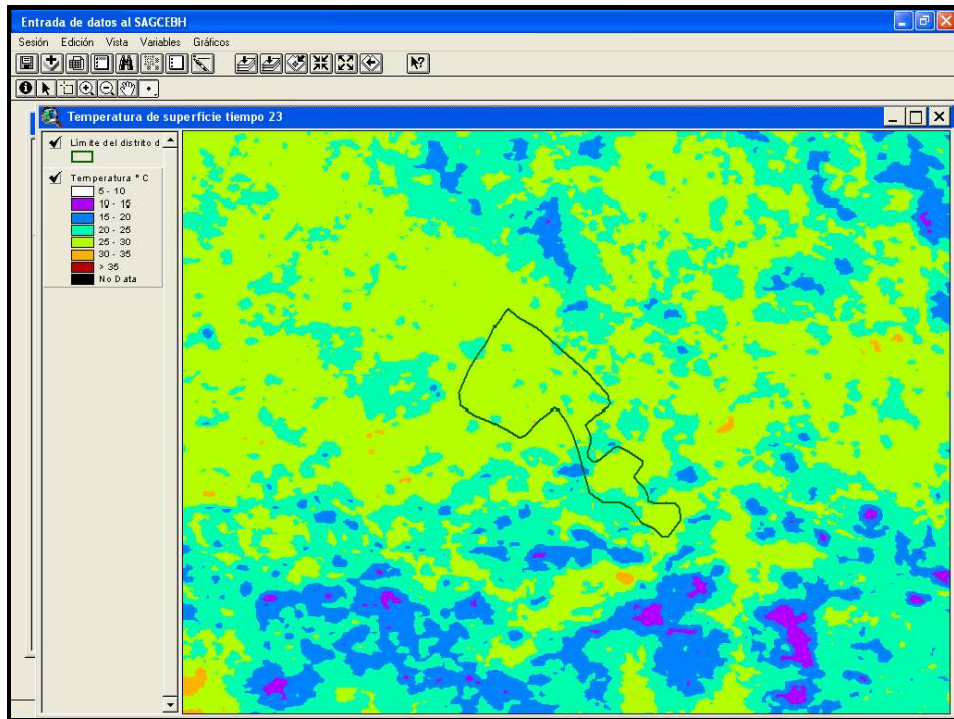
El formato raster permite visualizar mejor variables regionalizables o que presentan estructura espacial. La simple visualización de las superficies nos muestra la heterogeneidad espacial de las variables en un punto en el tiempo, lo cual no es tan de explícito con los datos puntuales a menos que se empleen técnicas de graficación y de acceso a los datos.

En las figuras 7.2 y 7.3 se representan modelos raster de temperatura de superficie y de reflectancia en el rojo respectivamente para el tiempo 23, este periodo corresponde a los últimos días del año 2002 y los primeros del 2003, puede apreciarse que las zonas con temperaturas bajas también presentan una reflectancia en el rojo menor a 0.11 (esto es, 11 %), que es el rango más bajo (Para detalles sobre las características de las fuentes de datos de estas superficies, consultar la tabla 7.1 (productos MOD13A2 y MOD13Q1 de MODIS).

**Fig. 7.1 Representación lógica de datos puntuales.**

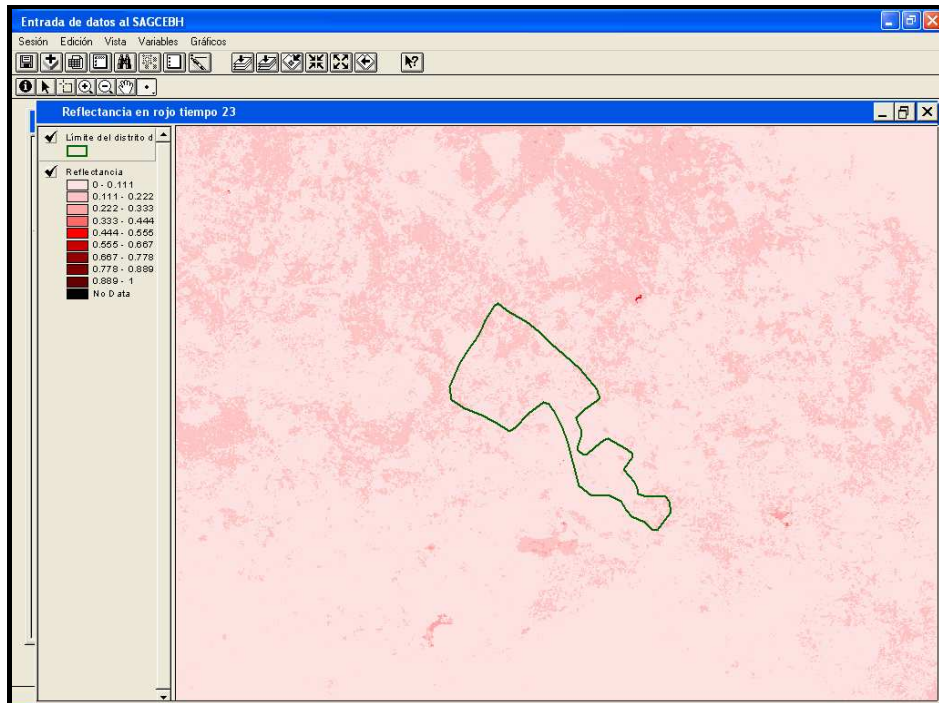


**Fig. 7.2 Representación lógica de datos como superficies.**



Temperatura de superficie en ° C, la superficie representa un periodo de 16 días en invierno del año 2002.

**Fig. 7.3 Representación lógica de datos como superficies.**



Reflectancia en el rojo en porcentaje, la superficie representa un periodo de 16 días en invierno del año 2002.



### 7.2.1b. Otras fuentes de datos sobre componentes del balance hídrico y variables relacionadas

Hay datos de imágenes de percepción remota, disponibles gratuitamente, que registran variables relacionadas con componentes del balance hídrico, por ejemplo, imágenes de temperatura de superficie que deben ser preprocesadas para convertirlas a °C o imágenes de reflectancia en el rojo y el infrarrojo cercano, útiles para caracterizar la cobertura de vegetación. Al igual que los datos de estaciones se requieren datos de temperatura de superficie promedio y compuestos de reflectancia para periodos quincenales o mensuales.

Las características de las imágenes MODIS se consideran apropiadas para un estudio a escala del paisaje, ya que en escenarios rurales como el que estudiamos es posible capturar la variabilidad espacial y temporal de la cobertura vegetal en el terreno y la de la temperatura de superficie. Además tienen un cubrimiento espacial nacional y la resolución temporal con que son obtenidas permite realizar un seguimiento continuo de estas variables. En la tabla 7.1 se detallan las características generales de las imágenes referidas.

**Tabla 7.1 Resumen de atributos de productos de MODIS.**

Variable	Producto MODIS	No. de Banda Sensor	Resolución Espacial (m)	Resolución Espectral	Resolución Temporal (días)	Resolución Radiométrica bits
Temperatura de superficie	MOD12A2	31	1000	10.780 - 11.280 $\mu\text{m}$	8	16
Reflectancia rojo	MOD13Q1	1	250	620-670 nm	16	16
Reflectancia infrarrojo	MOD13Q1	2	250	841-876 nm	16	16

En cuanto a otros datos hidrológicos como evapotranspiración potencial, real, humedad del suelo, escurrimientos y flujos, existen graves carencias a nivel nacional; aún cuando la cuenca del Lerma es una de las más monitoreadas no hay un acceso público eficiente a bases de datos en las instancias correspondientes como la CNA y el Servicio Meteorológico Nacional (SMN). Tan es así que no se ha podido acceder a estos datos con la suficiente rapidez para poder incorporarlos al presente trabajo. Una alternativa viable es el acceso a data warehouses activos donde es posible obtener imágenes o datos en formato raster de acumulados diarios de precipitación, modelos del terreno para la

modelación de áreas de captación y diferentes variables climáticas estimadas a partir de estaciones. Debe atenderse a las características de los datos con el propósito de prever si son adecuados para representar los objetos espaciales que pretendemos modelar.

En la tabla 7.2 se listan las propiedades de los datos obtenidos por diversas fuentes vía Internet, se especifican los patrones que pueden identificarse de acuerdo con las propiedades de las variables registradas. Por ejemplo, cuando hay un cubrimiento temporal a lo largo del año es posible identificar estructuras espaciales, además de los patrones espaciales como clases y vecindarios.

**Tabla 7.2 Referencias a data warehouses.**

Variable	Referencia del data warehouse	Dirección electrónica	Resolución Espacial (m)	Resolución Temporal	Resolución Radiométrica bits	Rasgos identificables
Climáticas	Worldclim	<a href="http://www.worldclim.org">www.worldclim.org</a>	~1000	NA	8	Vecindarios, regiones
Modelos del terreno	SRTM C-Band	<a href="http://www2.jpl.nasa.gov/srtm/cbanddataproducts.html">http://www2.jpl.nasa.gov/srtm/cbanddataproducts.html</a>	30	NA	16	Vecindarios, regiones y flujos
Precipitación	TRMM	<a href="http://disc.sci.gsfc.nasa.gov/data/datapool/TRMM/">http://disc.sci.gsfc.nasa.gov/data/datapool/TRMM/</a>	Variable	Variable	Variable	Vecindarios, regiones, estructuras temporales

Características de datos disponibles y patrones espaciales que pueden identificarse con dichos datos.

### **7.2.1c. Aportes de la fase de definición de datos de entrada y de formatos de representación**

El contenido de información se refiere a la medida en que los datos evidencian o representan las propiedades o atributos de los objetos geográficos y comienza con la identificación de patrones. Para diferenciar entre los datos resultantes de cada etapa del DPG se propone un orden jerárquico. Se inicia en un nivel I que corresponde a la conformación de bases de datos de observaciones sobre el proceso estudiado, en este nivel se define la forma de representación lógica de los datos. El nivel I se compone de dos subniveles: el nivel IA se refiere a la remoción de ruido o inconsistencia de los datos y en segundo término el nivel IB comprende además la calibración de observaciones y la definición de proyección cartográfica. Estos subniveles tienen que ver con determinar la capacidad de los formatos de representación de datos geoespaciales para capturar la realidad del proceso.

Los datos de salida de esta fase presentan un contenido de información nivel IB donde se ha realizado un proceso de validación por calibración y se ha establecido el formato de representación lógica más adecuado para las variables a analizar el cual es, como vimos, el formato raster.

### 7.2.1d. Análisis exploratorio de datos preprocesados Nivel I B

Para presentar con coherencia el conjunto de operaciones que debemos realizar con los datos una vez que presentan los formatos adecuados, diferenciamos en un primer nivel las tareas básicas de acceso y manejo como selecciones gráficas y queries, entre otras, de las tareas para la extracción de estructuras y la construcción de jerarquías a nivel conceptual, que son operaciones básicas de exploración. En la tabla 7.3 se enumeran las operaciones de visualización aplicables a una base datos multidimensional en diferentes niveles (espacial, temporal o de agregación).

**Tabla 7.3 Conjunto de operaciones a realizar con datos preprocesados.**

Tipo de dato	Representación	Acceso, operaciones básicas	Herramientas de exploración	Metaoperación de geovisualización
Imagen y modelo digital del terreno	Raster	Sel. gráfica, Querys, (Brushing, Enhancing) Acceso a atributos	Descripción, mapas de percentiles, histograma, 3D, vuelo sobre un perfil y perfil vertical	Identificación de rasgos
Imagen y modelo digital del terreno	Raster	Sel. gráfica, Querys, (Brushing, Enhancing) Acceso a atributos	Diagramas de dispersión, animación 2D,	Comparación de rasgos
Est. Meteorológicas Puntos de muestreo	Vector	Sel. gráfica, Querys, (Brushing, Enhancing), Acceso a atributos en BD ext.	Mapas de percentiles, descripción, diagrama de probabilidad	Identificación de rasgos
Est. Meteorológicas Puntos de muestreo	Vector	Sel. gráfica, Querys, (Brushing, Enhancing), Acceso a atributos en BD ext.	Diagramas de caja, diagramas de dispersión, diagramas de coordenadas paralelas, plantillas de firmas o perfiles de tiempo	Comparación de rasgos

Las operaciones básicas enumeradas en el cuadro corresponden al proceso de selección de submuestras de los datos para su posterior exploración. En general, estas operaciones nos conducen a la identificación de rasgos o patrones, su comparación a través de una jerarquía conceptual y finalmente de forma implícita nos apoyan en la generación de posibles interpretaciones cuando asociamos estos patrones con el conocimiento a priori

que manejamos sobre el fenómeno. Lo que sigue, es identificar patrones en los datos de estaciones, caracterizando la variación espacial y temporal de una muestra de datos.

Dada la dificultad de estudiar en el continuo espacio-tiempo recurrimos a un proceso que transforma una variable continua en el tiempo a una representación discreta en forma de rangos o periodos de tiempo. El universo de datos es segmentado y caracterizado en forma de sumalizaciones mensuales, quincenales o compuestos promediados, modales o extremos (donde se toma el valor máximo o mínimo del periodo). Aún con esta simplificación, corremos el riesgo de perdernos en la identificación de una gran cantidad de posibles patrones, (23 periodos, ya que se dividen los 365 días en compuestos de 16 días cada uno para poder comparar con los datos de temperatura de superficie y los compuestos de reflectancia de MODIS, multiplicado por el número de componentes del balance hídrico).

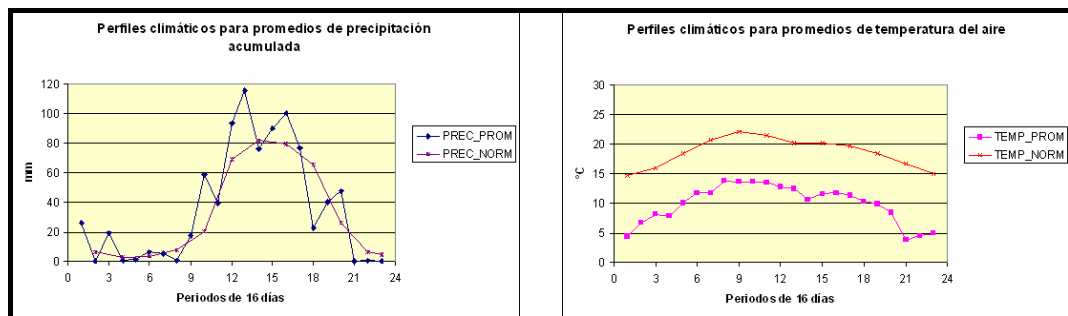
Nuestro propósito es generalizar a partir de una muestra de datos que represente la variación anual de la precipitación y la temperatura como mínimo para caracterizar categorías climáticas temporales. El área de estudio se encuentra en una zona semiárida con una precipitación acumulada anual de 750 mm. y una temperatura promedio anual de 18.6 °C. El año consta de 6 meses con tasas de precipitación menores a 20 mm. mensuales, los cuales se reconocen como la temporada de estiaje, el resto de los meses comenzando en Mayo y terminando en Octubre conforman el temporal o periodo de verano (Datos del registro histórico del SMN, 1950-1980). Puesto que no se cuenta con datos históricos de percepción remota o con datos actuales de referencia sobre normales climáticas sería muy difícil caracterizar categorías climáticas típicas, por tanto se define el año 2002 como muestra por ser un año más húmedo y frío que la normal según indica un estudio reciente sobre el área (Coronel y Mora, 2006:80-85).

Se parte de datos de precipitación para el año 2002 en forma de acumulados de 16 días y de promedios de temperatura del aire para esos mismos periodos. Para comprobar que el año 2002 difiere de la normal, se toman datos de estaciones en la zona del Bajío, se realizan promedios y se comparan con los promedios para normales climáticas históricas como referencia. Este ejercicio nos permite inferir si el subconjunto de observaciones

empíricas es una instancia de las categorías temporales conceptuales, para ello, empleamos la comparación de perfiles climáticos o temporales (Ver figura 7.4).

Mediante una simple comparación de plantillas de firmas o perfiles temporales podemos ver que el año 2002 es un año húmedo ya que presenta mayor volumen de lluvias y también es más frío respecto de las normales, sin embargo se observan dos estructuras temporales correspondientes al temporal y a la época de estiaje.

**Fig. 7.4 Contrastación de perfiles climáticos.**



Se comparan promedios de temperatura del aire y de precipitación acumulada de normales climáticas con las mismas variables para periodos de 16 días en el año 2002, datos de estaciones metereológicas en la región del Bajío.

Estas estructuras temporales obedecen a la presencia de autocorrelación temporal entre observaciones de periodos consecutivos y pueden identificarse revisando la correlación en los datos, de existir una buena correlación temporal se pueden definir componentes temporales mediante la transformación de los datos por métodos de ordenación. Se observa en la tabla 7.4, que la correlación lineal nos da indicios de estructuras temporales, se pueden distinguir dos estructuras a simple vista, la primera que comprende prácticamente los primeros 9 meses del año y la segunda que toma sólo los últimos 4 periodos del año. Existen 2 periodos totalmente independientes de estas dos estructuras, el 18 y el 19, que pueden considerarse una estructura de transición entre el temporal y la época de estiaje. Estas son claras evidencias de un año atípico.

El análisis de correlación lineal para los acumulados de precipitación mostró que no hay autocorrelación temporal significativa entre los periodos y que por tanto las observaciones son independientes para este componente del balance hídrico. Debido a que se emplean acumulados no se requiere identificar el efecto de la autocorrelación temporal

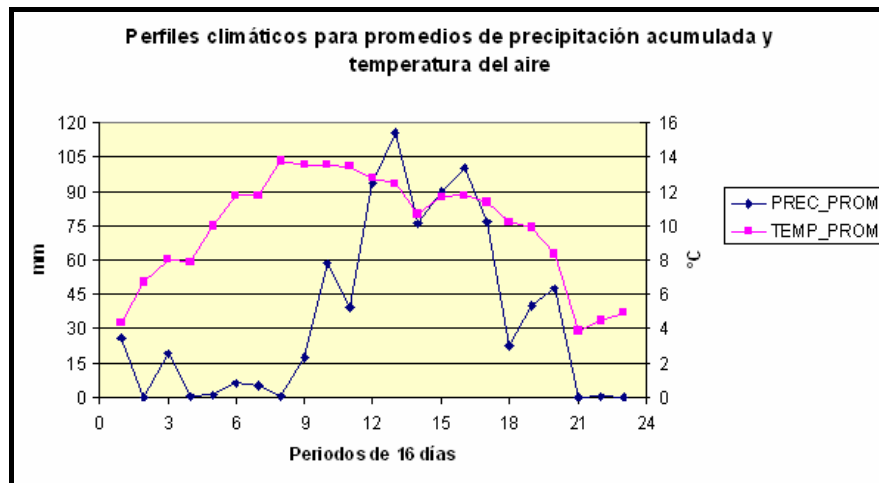
de los datos, a diferencia de variables que son promediadas, donde habría sesgo en los parámetros característicos (media y varianza). Estas diferencias entre las variables climáticas indican que la precipitación es más dinámica que la temperatura del aire y por lo tanto sería inadecuado proceder a la reducción de la dimensionalidad de la muestra en periodos más gruesos. Considerando la presencia de dependencia temporal de los datos en lo que respecta a la temperatura del aire, bastará con designar al menos un periodo correspondiente a las dos estructuras temporales identificadas, el periodo de secas y el temporal.

**Tabla 7.4 Matriz de autocorrelación lineal.**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
1	1																							
2	0.94	1																						
3	0.90	0.97	1																					
4	0.89	0.96	0.95	1																				
5	0.90	0.95	0.95	0.96	1																			
6	0.86	0.90	0.90	0.92	0.96	1																		
7	0.89	0.94	0.94	0.91	0.95	0.96	1																	
8	0.80	0.91	0.93	0.90	0.93	0.92	0.97	1																
9	0.82	0.89	0.91	0.89	0.92	0.94	0.97	0.96	1															
10	0.81	0.89	0.89	0.90	0.92	0.93	0.95	0.95	0.98	1														
11	0.78	0.85	0.89	0.85	0.89	0.92	0.94	0.92	0.97	0.96	1													
12	0.80	0.83	0.85	0.82	0.87	0.91	0.94	0.91	0.96	0.97	0.96	1												
13	0.71	0.74	0.77	0.72	0.77	0.83	0.86	0.83	0.91	0.89	0.93	0.95	1											
14	0.77	0.81	0.82	0.79	0.83	0.86	0.89	0.86	0.92	0.93	0.94	0.97	0.97	1										
15	0.77	0.80	0.81	0.76	0.79	0.84	0.86	0.83	0.90	0.90	0.92	0.95	0.97	0.99	1									
16	0.74	0.76	0.79	0.72	0.76	0.79	0.80	0.78	0.85	0.85	0.87	0.90	0.94	0.95	0.97	1								
17	0.75	0.70	0.72	0.64	0.72	0.77	0.77	0.70	0.78	0.77	0.81	0.87	0.91	0.90	0.93	0.96	1							
18	0.64	0.61	0.62	0.61	0.63	0.63	0.66	0.63	0.67	0.67	0.68	0.67	0.66	0.62	0.60	0.63	0.56	1						
19	0.67	0.61	0.63	0.60	0.64	0.65	0.68	0.64	0.68	0.68	0.68	0.70	0.67	0.64	0.62	0.65	0.60	0.99	1					
20	0.02	0.02	0.02	0.05	0.07	0.07	0.11	0.07	0.15	0.18	0.14	0.17	0.16	0.13	0.12	0.13	0.07	0.05	0.06	1				
21	0.60	0.60	0.57	0.60	0.61	0.59	0.66	0.61	0.67	0.69	0.62	0.66	0.59	0.61	0.59	0.54	0.47	0.50	0.51	0.76	1			
22	0.52	0.51	0.46	0.53	0.53	0.50	0.58	0.51	0.57	0.61	0.51	0.57	0.49	0.51	0.49	0.44	0.38	0.37	0.39	0.80	0.96	1		
23	0.48	0.46	0.40	0.47	0.48	0.47	0.54	0.46	0.54	0.58	0.47	0.54	0.47	0.49	0.47	0.41	0.36	0.34	0.36	0.82	0.95	0.99	1	

Coefficiente de autocorrelación entre promedios de temperatura del aire para periodos de 16 días, año 2002.

**Fig. 7.5 Perfiles climáticos.**



Se comparan promedios de precipitación y temperatura del aire para el año 2002, en la región del Bajío.

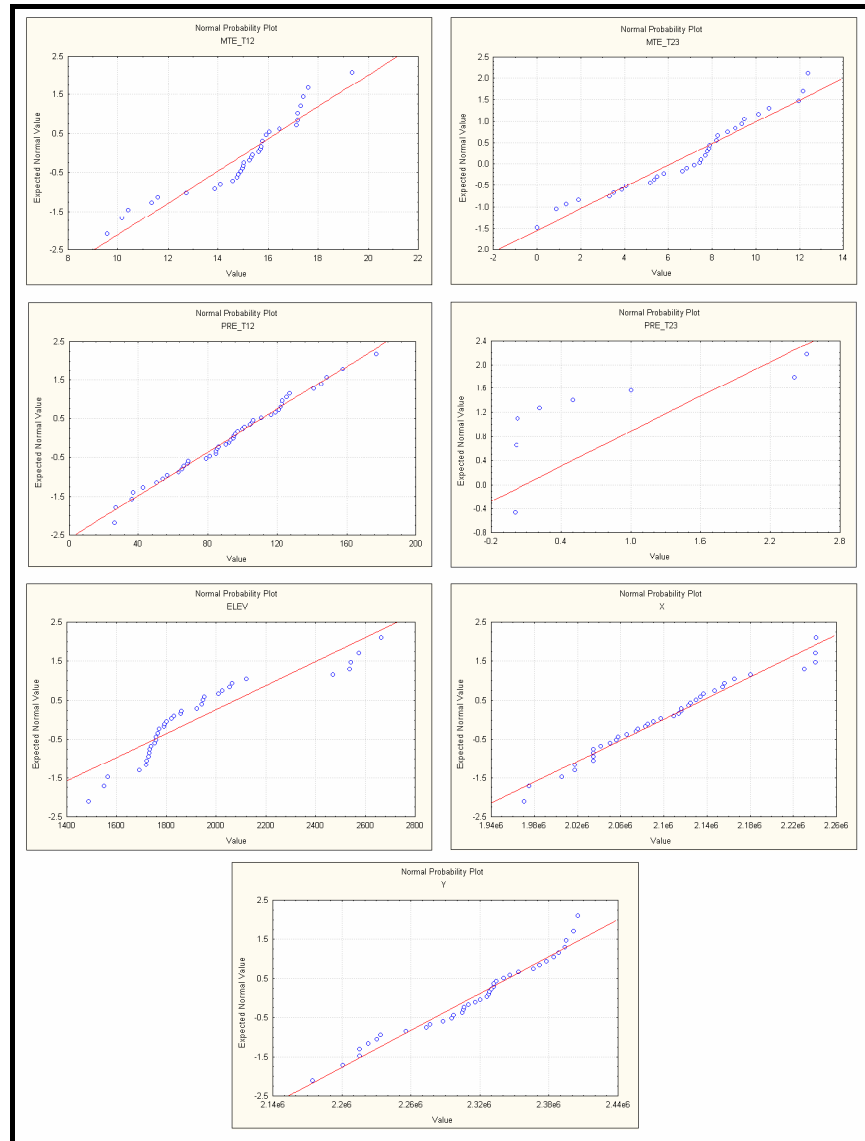
El gráfico de la figura 7.5 presenta los promedios de temperatura del aire y los acumulados de precipitación para los periodos de 16 días. Podemos distinguir que los periodos posteriores al temporal son más representativos de las condiciones de secas, ya que prácticamente no se registra precipitación. Por ello, se propone el periodo 23 que corresponde al invierno del año 2002 como instancia de la temporada de secas. En cuanto a la época de temporal, no se presentan coincidencias entre los máximos de temperatura (periodos 8 al 11) y los máximos de precipitación (12, 13 y 16). Aparentemente, la humedad afecta la temperatura del aire disminuyendo los valores por debajo de la normal. Para definir un periodo de verano más distinguible del periodo de secas, es necesario disminuir el efecto combinado de la temperatura y la precipitación y escoger aquel periodo que presente los máximos de estas variables. Para esta muestra de datos el periodo 12 presenta un valor alto de temperatura aún cuando ya se manifestaron tasas de precipitación representativas del verano (más de 90 mm promedio en el periodo). No se pretende que los dos periodos representen de forma efectiva los patrones climáticos a lo largo del año, sin embargo, se considera que no hay suficiente información sobre la asociación de las variables climáticas analizadas con variables causales por lo que se da prioridad a la búsqueda de relaciones causa efecto.

Para comenzar a caracterizar los dos periodos identificados se presenta una plantilla de diagramas de probabilidad para variables como la precipitación, la temperatura del aire, las coordenadas geográficas y la elevación (Fig. 7.6), esto con el objetivo de establecer si pueden ser analizadas con métodos estadísticos que asumen una distribución normal en los datos.

Otra de las propiedades intrínsecas de variables regionalizables o fenómenos geográficos es la autocorrelación o dependencia espacial que asume que las observaciones en un punto están correlacionadas con las observaciones en puntos vecinos. Esta propiedad viola el supuesto de independencia entre observaciones fundamental en la estadística tradicional por lo que para evitar sesgos en análisis estadísticos subsecuentes se requiere considerar el efecto de la dependencia espacial en la variación total de las variables analizadas. Para visualizar si hay autocorrelación espacial en las variables y esta es capturada por la muestra de estaciones se utilizan diagramas de dispersión. Este análisis

se sugiere para explorar cómo se distribuyen las observaciones en los gradientes de coordenadas geográficas y al mismo tiempo, para identificar la presencia de autocorrelación espacial de las variables.

**Fig. 7.6 Diagramas de Probabilidad Normal para variables climáticas.**



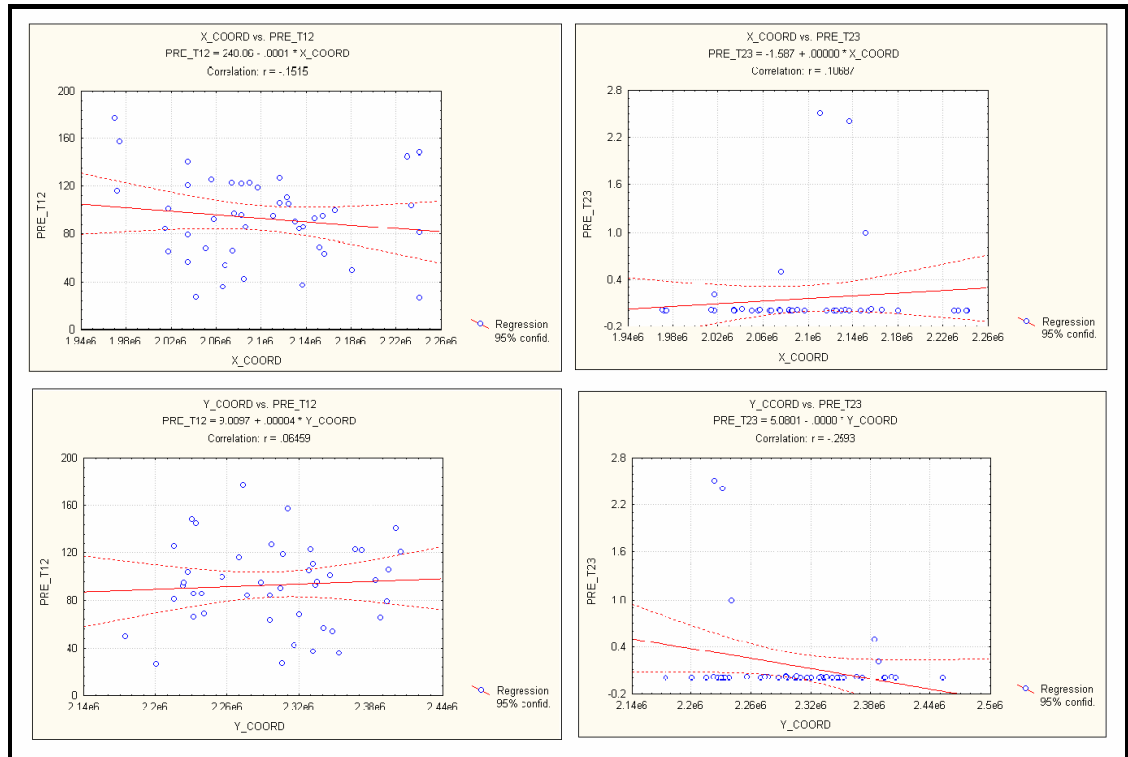
Temperatura promedio del aire verano, temperatura promedio del aire invierno, precipitación acumulada verano, precipitación acumulada invierno, los periodos representados son de 16 días en cada temporada, continúa altitud, coordenadas geográficas en el eje X y por último coordenadas en el eje Y. Se observa que sólo la precipitación en invierno presenta una distribución de frecuencias diferente a la normal, lo que sugiere que se compare con otros modelos de distribución de frecuencias. En otras palabras, encontraremos en estos datos algunas propiedades específicas en los análisis subsecuentes.

De acuerdo a la figura 7.7, la dispersión de los datos de precipitación en verano es mucho mayor que la de los datos en invierno, la precipitación en invierno es prácticamente



invariante. El grado de correlación lineal con coordenadas geográficas es muy pequeño para ambas épocas. Por último, se observa que las estaciones están sesgadas más bien al centro y noroeste de la región.

**Fig. 7.7 Diagramas de dispersión precipitación vs. coordenadas.**



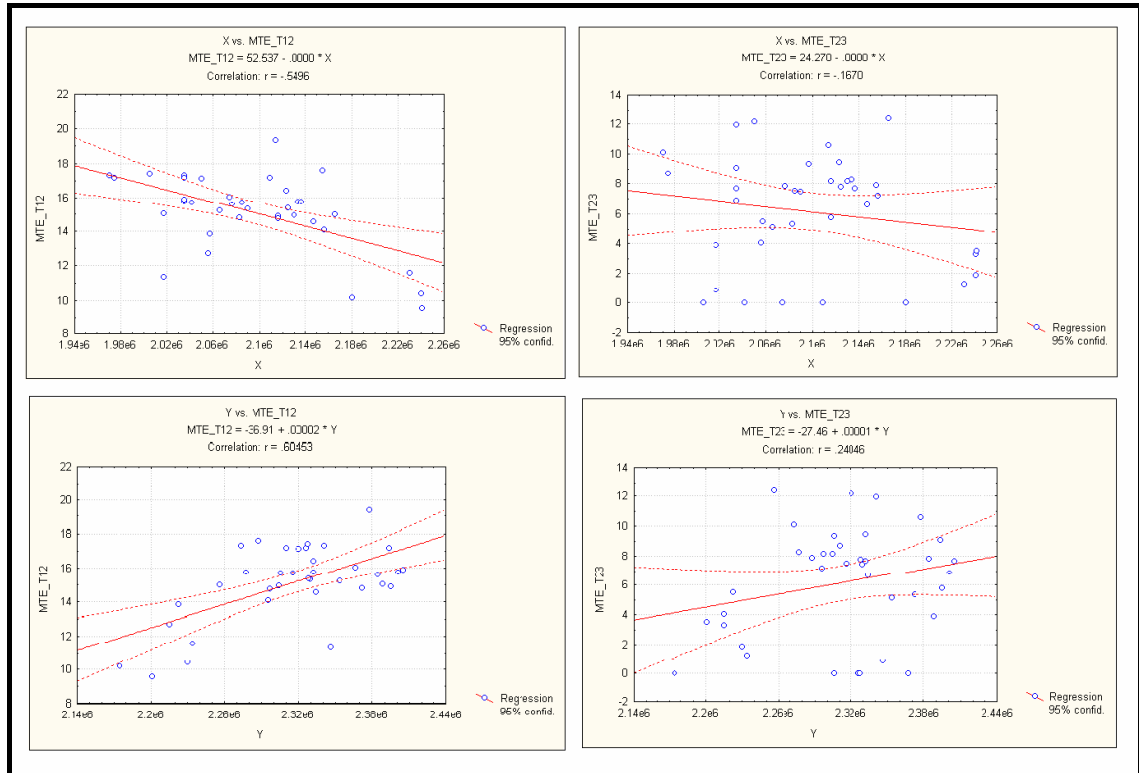
Datos de precipitación (mm) en estaciones climatológicas respecto de coordenadas geográficas X y Y, se compara la asociación para periodos en verano e invierno. Las gráficas de la izquierda representan la dispersión de los datos para el verano y en la columna derecha los datos para el invierno.

En términos generales, se observa en la figura 7.8, que la temperatura del aire está más correlacionada linealmente con las coordenadas geográficas que la precipitación y aunque la forma de la correlación es similar para verano que para invierno, sin duda la magnitud de la correlación es el elemento de diferenciación entre ambos periodos para esta variable, lo cual seguramente se expresará en forma de patrones espaciales específicos para cada temporada.

Este primer análisis sobre la autocorrelación espacial nos da una indicación de que encontraremos diferencias en cuanto a la porción de variación explicada por la estructura espacial entre la precipitación acumulada y la temperatura del aire y entre periodos para

una misma variable. Aunque la magnitud de la correlación lineal es pequeña para la precipitación no podemos descartar la presencia de una estructura de variación espacial en esta variable sin realizar análisis más contundentes.

**Fig. 7.8 Diagramas de dispersión temperatura vs. coordenadas.**

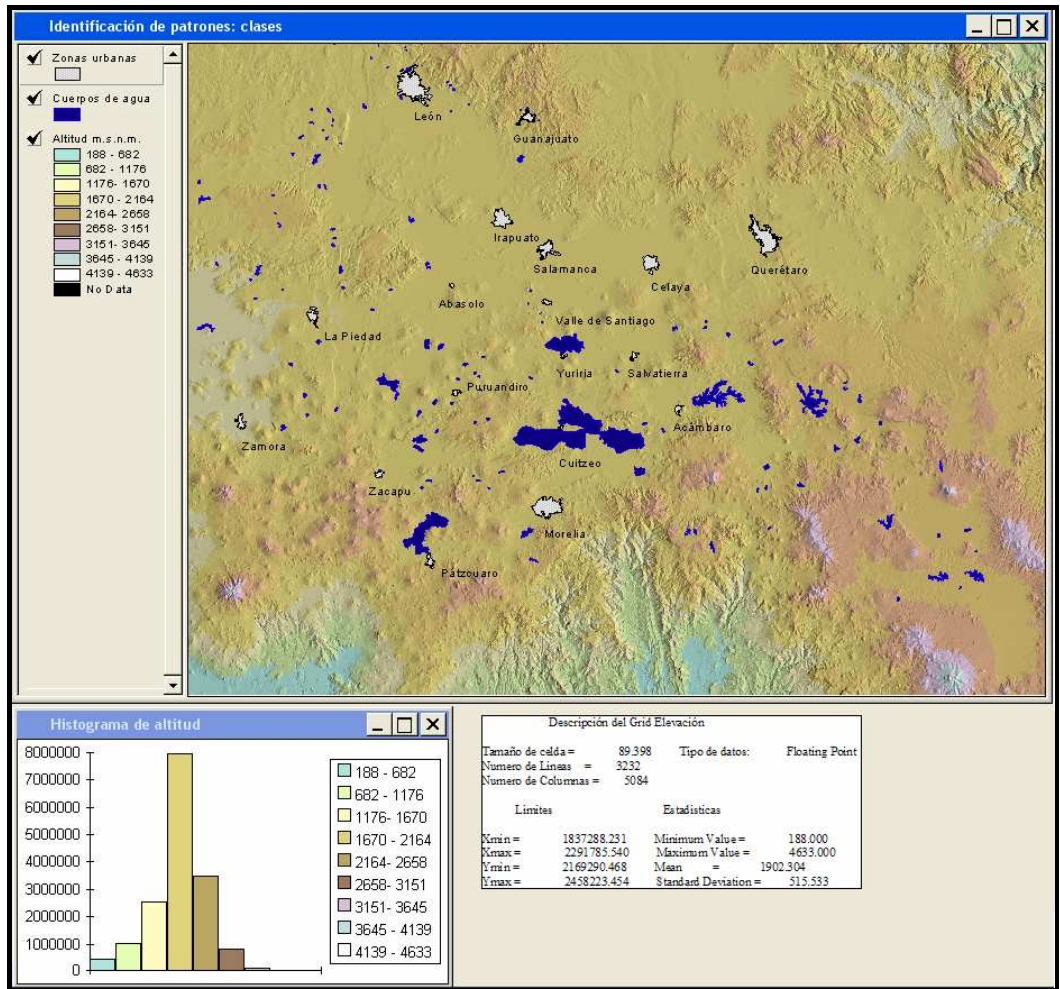


Datos de temperatura promedio del aire (°C) en estaciones climatológicas respecto de coordenadas geográficas X y Y, se compara la asociación entre estas variables para periodos en verano e invierno. Las gráficas de la izquierda representan la dispersión de los datos para el verano y en la columna derecha los datos para el invierno.

A través de la observación de superficies de variables representadas en formato raster es posible visualizar la heterogeneidad de toda la extensión del área y comparar entre periodos cuando se cuenta con observaciones. Nos familiarizaremos con la apariencia general de la región usando como referencia el modelo del terreno, iluminado desde un ángulo particular para observar los cambios del relieve, la visualización es directamente desde el nadir o la vertical del plano. Además, al emplear otras herramientas como el histograma y estadísticas generales asociadas a la imagen podemos generalizar el comportamiento de la variable. La altitud es una variable muy importante para la delimitación de áreas de captación, ya que a partir de la rugosidad y la pendiente del

terreno se definen los patrones de escurrimientos y sus conexiones, los cuales son los insumos para definir áreas que funcionan como una unidad de captación pluvial.

**Fig.7.9 Imagen 2D o nadir representando el modelo del terreno en la región del Bajío.**



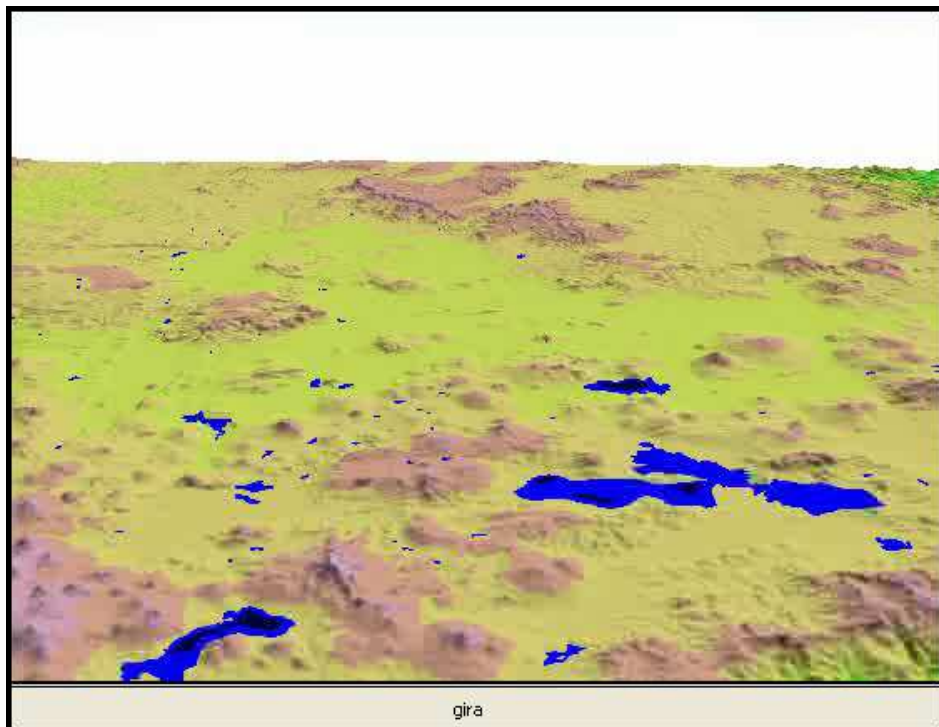
Para entender mejor el comportamiento de la variable se presenta el histograma y la descripción general del conjunto de datos.

En la figura 7.9 se presenta una imagen 2D o nadir del área de estudio, en ella podemos identificar claramente el rango más frecuente de altitud que corresponde a la zona centro donde se localizan la mayor parte de los cuerpos de agua pero ésta zona se encuentra dividida por diversas formaciones con altitudes en el rango inmediato superior al modal, al sureste se ubican los rangos de altitud superiores, mientras que las vertientes al sur, representan los rangos de altitud más bajos. El valor de la media queda fuera del rango más frecuente lo que indica que existen valores raros pero suficientemente grandes para

desviar este descriptor del valor modal y por tanto deberían ser considerados separadamente del análisis si lo que se persigue es caracterizar patrones. Los rangos menores de la altitud se ubican en áreas que funcionalmente no contribuyen a la cuenca del Lerma por lo que de antemano estas zonas no deben considerarse para la estimación de los componentes del balance hídrico.

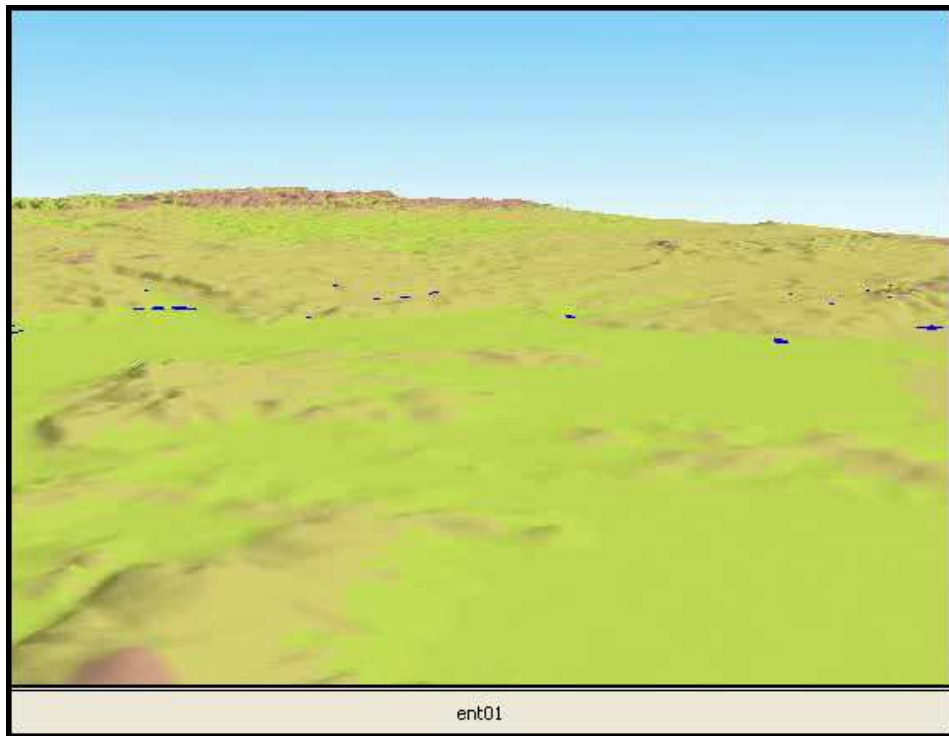
Las herramientas de visualización nos permiten observar además desde todos los ángulos posibles los rasgos previamente identificados, aquí se proponen dos herramientas, la primera conocida como imagen en 3 dimensiones que permite la manipulación de una representación en 3D de la variable a observar y nos facilita la identificación de rasgos (Fig. 7.10). En esta figura la escena es girada pero se mantiene horizontal. En comparación se simula un vuelo a partir de la entrada noroeste a la zona del Bajío, destacándose el relieve que rodea la planicie y conforma diversas áreas de captación que drenan esta región agrícola (Fig. 7.11).

**Fig. 7.10 Imagen de 3 dimensiones mostrando rasgos del relieve.**



Se aprecia al centro la zona del Bajío al norte de los lagos de Cuitzeo y Pátzcuaro.

**Fig. 7.11 Simulación de vuelo sobre la zona del Bajío.**

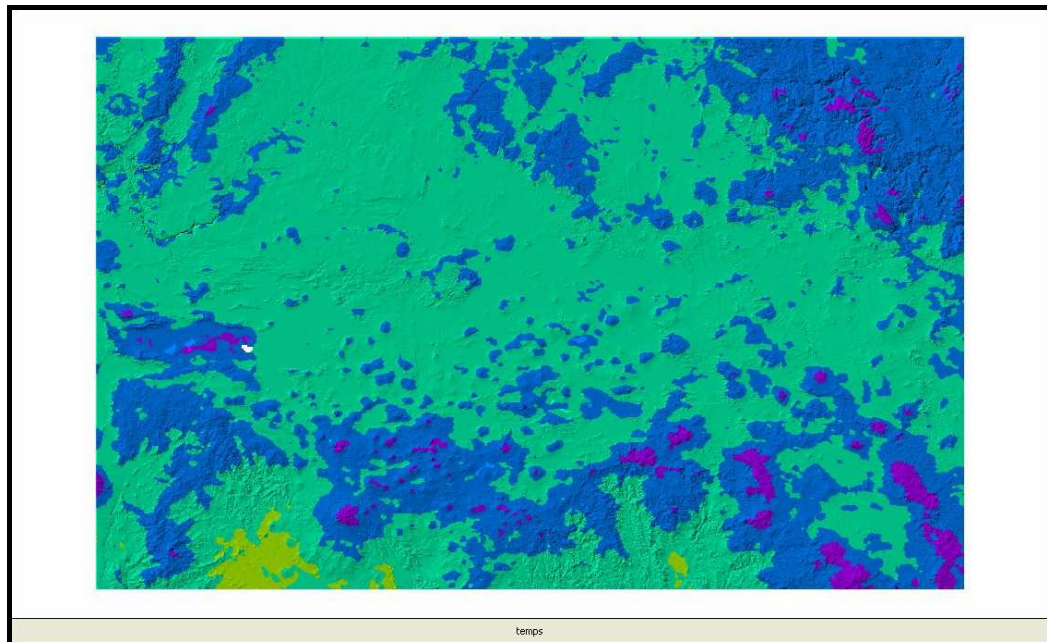


Se muestran las diferencias entre esta extensa planicie y los diferentes sistemas de lomeríos que la rodean y la separan de otros valles situados en altitudes mayores, como el área de captación de la laguna de Yuriria o del Lago de Cuitzeo.

La segunda herramienta se denomina animación de dos dimensiones y permite la comparación entre rasgos, consiste en la presentación en secuencia de escenas desde la perspectiva nadir que corresponden a diferentes fechas, lo que permite ver la evolución del fenómeno en 3 dimensiones, los dos componentes espaciales (X y Y) y el tiempo. En esta herramienta se da énfasis a la dinámica y se pueden observar patrones espacio-temporales. A través de la observación de superficies de temperatura a lo largo del año podemos visualizar ciertos patrones que pueden ser más o menos estables e identificar patrones emergentes, usualmente efímeros, puede haber direccionalidad (ver figura 7.12).



**Fig. 7.12 Animación de 2 dimensiones.**



Se muestran los cambios en la temperatura de superficie a lo largo de 23 periodos de 16 días del año 2002 en la región del Bajío.

### **7.2.1e. Aportes de la fase de organización y manejo de los datos**

En esta etapa de acceso y manejo a los datos aplicamos principalmente métodos de inferencia abductivos, buscamos que los datos nos hablen y nos den evidencia sobre la presencia de patrones espaciales y temporales. Se ha conformado una muestra de datos con la que se busca probar la presencia de tales patrones, hay evidencias de que parte de la variación en la precipitación acumulada y la temperatura del aire puede ser debida al espacio. La selección de la muestra de datos y su análisis mediante herramientas de geovisualización nos refieren a un nivel de información II donde continúa el proceso de identificación de algunos de los objetos espaciales que se establecieron como metamodelos a nivel conceptual, hablamos de las asociaciones y estructuras temporales. En este nivel los formatos lógicos se transforman en instancias de dichos modelos y adquieren por tanto las propiedades intrínsecas de los objetos geográficos que representan. En el caso de las estructuras temporales nos referimos a la autocorrelación temporal y a la escala temporal que los datos capturan sobre el proceso estudiado. En cuanto a las asociaciones nos referimos a la inclusión del tiempo y el espacio como variables explicatorias de fundamental importancia para entender el comportamiento de los componentes del balance hídrico.

## 7.2.2. Análisis exploratorio o estructural

En esta fase se manejan dos niveles de abstracción, uno que parte del reconocimiento de patrones a nivel de registros (puntuales y raster) y otro que intenta generalizar el comportamiento de objetos geográficos. Estos dos niveles tienen como objetivo extraer patrones, rasgos o categorías e incorporan un proceso general de clasificación.

### 7.2.2a. Datos requeridos

Para caracterizar los diferentes rasgos espaciales sugeridos (clases, asociaciones, redes, vecindarios y jerarquías) se requieren superficies de temperatura, precipitación, evapotranspiración potencial, fracción de cobertura, elevación. Como se ha mencionado anteriormente, la estimación de superficies a partir de puntos se realiza tradicionalmente a partir de métodos de interpolación, este es el caso de los datos de precipitación y temperatura del aire (Hijmans *et al.*, 2005:1965-1970). El resto de las variables se estima a partir de datos de imágenes de percepción remota tales como los modelos digitales del terreno o la temperatura de superficie, también pueden derivar de operaciones entre superficies. En la tabla 7.5 se resumen diferentes métodos para realizar las estimaciones de estas superficies.

**Tabla 7.5 Métodos para la estimación de variables en formato raster.**

Variable	Método
Elevación	<a href="ftp://e0srp01u.ecs.nasa.gov/srtm/version2/SRTM30/srtm30_documentation.pdf">ftp://e0srp01u.ecs.nasa.gov/srtm/version2/SRTM30/srtm30_documentation.pdf</a>
Fracción vegetal	Coronel y Mora (2006)
Evapotranspiración potencial	Thornthwaite, citado por Serafín, Y. V. (1992)

### 7.2.2b. Reconocimiento e identificación de patrones a nivel de registros

En esta etapa, los datos con los que partimos para el reconocimiento de patrones deberían haber sido explorados y por tanto deberemos tener una idea general de su comportamiento en el espacio y en el tiempo, el siguiente paso es realizar un proceso de generalización de los registros para la extracción de rasgos.

#### I. Estructuración de variables: autocorrelación espacial y dominio de escala

Se parte del supuesto de que los componentes del balance hídrico se estructuran a diferentes niveles, esto significa que exhiben patrones espaciales y temporales específicos

dependiendo de la escala, suponemos también que esto se debe a que diferentes procesos causales están operando en cada nivel o bien que tienen un diferente dominio de escala y su efecto es menor o mayor sobre los componentes del ciclo hidrológico. Lo anterior es el fundamento sobre el cual se considera apropiado aplicar un modelo jerárquico para estudiar el comportamiento del balance. Para probar tales argumentos debemos determinar el dominio de escala de las variables causales y de los componentes del ciclo hidrológico mediante funciones de estructura, estas funciones estiman la covarianza de la variable en función de rangos de distancia.

I.a. Análisis exploratorio para probar la autocorrelación espacial y describir estructuras espaciales o temporales

Se intenta probar la existencia de autocorrelación espacial a través de la exploración de **vecindarios** u observaciones situadas a ciertos rangos de distancia y básicamente la diferencia entre los métodos de exploración de autocorrelación espacial estriba en la manera como se consideran los vecindarios y el método de estimación de la función de estructura.

La autocorrelación espacial global puede ser medida por el índice de Moran ( $I$ ) o el de Geary ( $c$ ). La fórmula de Moran se comporta como el coeficiente de correlación de Pearson ya que su numerador consiste de la suma de productos cruzados de los valores centrales (lo cual es un término de covarianza), comparando los valores encontrados en todos los pares de puntos a un rango de distancia dado. El valor de  $I$  es significativamente diferente de 0 y el de  $c$  diferente de 1. El coeficiente de Geary es una función de distancia ya que el numerador suma todas las diferencias cuadradas entre los valores encontrados en los pares de puntos comparados.

En la tabla 7.6 se presentan los valores de los índices  $I$  y  $c$  globales para la fracción de cobertura y la temperatura de superficie en los periodos de verano e invierno y la altitud. El valor de estos índices sugiere que existe autocorrelación espacial y es un indicador de heterogeneidad o “patchness”. El índice fue calculado con el módulo de GRID de ArcInfo v. 8.1 el cual evalúa la autocorrelación espacial entre celdas adyacentes en una superficie raster y resume la dependencia espacial en un solo índice.



**Tabla 7.6 Índices de heterogeneidad, Moran I y Geary c.**

	<i>I</i>	<i>c</i>
<b>Altitud</b>	0.9993	0.0004
<b>F_veg_ver</b>	0.8179	0.1791
<b>F_veg_inv</b>	0.7684	0.2278
<b>Tems_ver</b>	0.9939	0.0026
<b>Tems_inv</b>	0.9968	0.0022

Índices para las variables altitud, de fracción de vegetación y de temperatura de superficie en verano e invierno.  $p = 0.05$

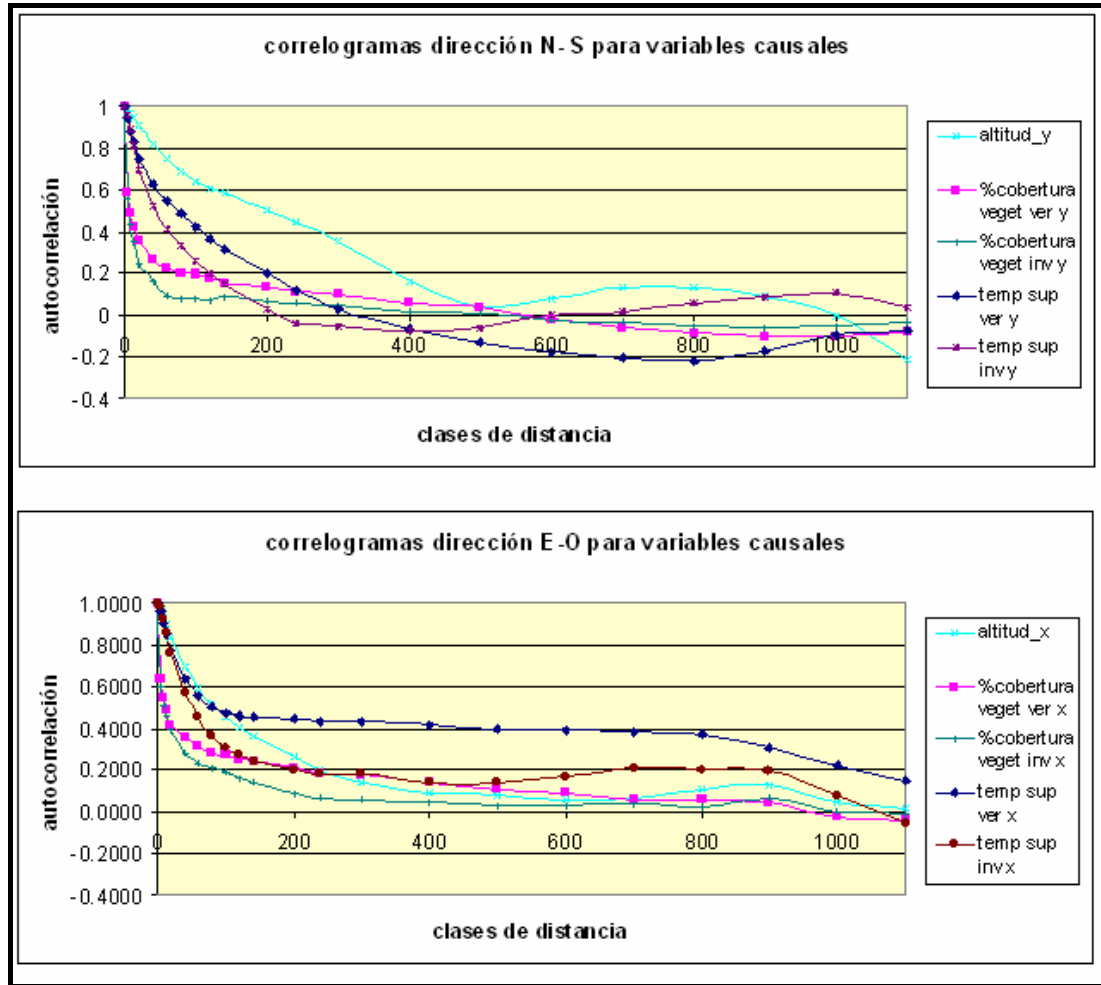
De acuerdo con los valores de los índices *I* y *c*, los datos de temperatura de superficie en invierno presentan la mayor autocorrelación espacial global del conjunto de variables. Podríamos inferir que esta variable presenta mayor heterogeneidad y por tanto su dominio de escala es más local que el resto de las variables dentro del área de estudio. Sin embargo, debido a la forma en que se calculan, los índices *I* y *c* dependen de la resolución espacial del raster. Esto significa que debe explorarse más a profundidad sobre la autocorrelación espacial de las variables.

Una herramienta para medir la autocorrelación como función de la distancia consiste en estimar la autocorrelación empleando diferentes ventanas de análisis y generar un gráfico similar a un correlograma (en éste se grafica el índice *I* respecto de clases de distancia), donde se represente el índice de Pearson a diferentes rangos de distancia. Al presentar con más detalle la forma como varía la autocorrelación lineal, se puede identificar de forma aproximada el dominio de escala de las variables. Debido a que el espacio está organizado en 2 ejes de coordenadas, es decir latitud y longitud o bien X y Y para el caso de coordenadas métricas, se distinguen los correlogramas en dirección norte-sur y este-oeste.

En los correlogramas de la figura 7.13 se pueden apreciar las diferencias en los valores de la autocorrelación entre las dos direcciones, lo cual da indicios de anisotropía en las variables consideradas. La anisotropía indica que el grado de autocorrelación está diferenciado en la distancia y dirección entre observaciones. Los correlogramas en conjunto evidencian que las variables consideradas presentan una forma de variación conocida como paso marcado según la clasificación de Sokal (Legendre y Fortín, 1989:127). En esta estructura se observa una pendiente pronunciada seguida por un

cambio a una pendiente mucho más suave. Este cambio tiene mucho que ver con el rango de distancia al cual la autocorrelación es significativa.

**Fig. 7.13 Correlogramas direccionales de variables explicatorias.**



Agregados o promedios para dos periodos de 16 días del año 2002, en verano y en invierno. (Cada clase o rango de distancia corresponde a 1 Km.).

Es claro que existen diferencias en cuanto a la forma y magnitud de la autocorrelación espacial, por ejemplo, la altitud presenta el cambio en la pendiente a muy diferentes rangos de distancia, en la dirección N-S se presenta después de los 150 km. y en la dirección E-O no parece haber cambios importantes en todo el eje (más allá de los 280 Km.).

En la dirección N-S la fracción vegetal de la cobertura presenta cambios importantes después de los 30 Km. Esto corrobora que la estructura espacial de la variable

no es capturada adecuadamente por la resolución espacial de las imágenes ya que el índice  $I$  es más bajo que el de la altitud la cual es una variable menos local. En la dirección E-O, la fracción vegetal de la cobertura se comporta de manera semejante.

La temperatura de superficie en verano presenta una variación más local que el invierno en los primeros 100 Km., especialmente en dirección E-O. La resolución espacial de 1 Km. de las imágenes de temperatura parece capturar bien la estructura espacial de esta variable ya que el índice  $I$  es muy cercano al de la altitud aún cuando parece tener un dominio de escala más local. Los índices de Moran y Geary, junto con el correlograma, nos dan evidencias de estructura espacial en las variables, no obstante, para identificar mejor su dominio de escala, se requieren más análisis.

Para describir estructuras espaciales, se sugiere realizar variografía, especialmente si los registros se encuentran muy espaciados entre sí o no están distribuidos sobre toda la extensión de la ventana de análisis, lo cual es el caso de las estaciones meteorológicas. No obstante, antes de utilizar este método, es necesario conocer si los datos son estacionarios lo cual significa que la media y la varianza son constantes en toda el área de estudio. Este supuesto, difícilmente se cumple en el caso de procesos geográficos, por lo que debemos asumir que, al menos, se cumple la hipótesis intrínseca, lo cual significa que los incrementos (o diferencias en la variable) entre todos los pares de puntos localizados a una distancia  $d$  entre sí tienen una media de cero y una varianza finita y constante.

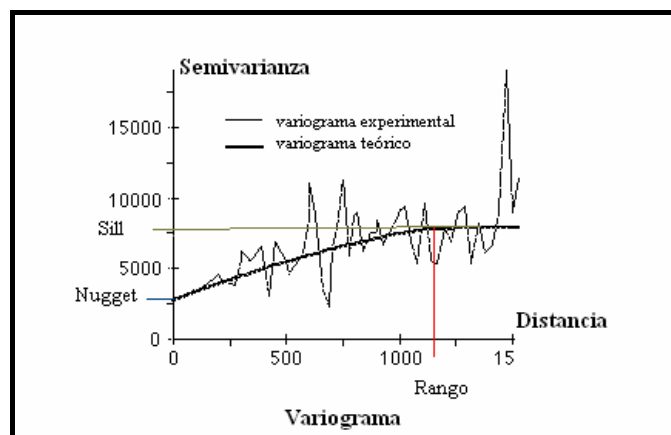
El valor de la varianza, para una clase de distancia  $d$ , es dos veces el valor de la función de semivarianza  $\gamma(d)$  o variograma. Existen dos tipos de variogramas, el experimental y el teórico, el primero modela el comportamiento de  $\gamma(d)$  en función de clases de distancia ( $d$ ):

$$\gamma(d) = \frac{1}{2} (2n_d) \sum [y_{(i+d)} - y_{(i)}]^2$$

donde  $n_d$  es el número de pares de puntos localizado a una distancia  $d$  de otros. Este variograma puede ser usado como una descripción de la función de estructura de la variable.

El variograma teórico se calcula a partir del experimental y puede tener varias formas: lineal, esférico, circular, gaussiano, exponencial, entre otros Figura. A través del ajuste del variograma teórico al experimental (ver figura 7.14), se obtienen los parámetros de la función de estructura, el *rango* de influencia de la estructura espacial, la cual es la distancia donde la semivarianza se estabiliza, el *sill* (meseta o silla), donde la semivarianza ya no es más una función de la distancia o la dirección y corresponde a la varianza de los datos y el efecto *nugget* donde se despliega una cantidad de variación aún a una distancia de cero, lo cual puede ser causado por una variabilidad intrínseca de los datos (varianza no explicada) o es una indicación de que el muestreo no se ha desarrollado a la escala espacial correcta (anomalía local). Para mayores detalles sobre el uso de variografía consultar a Legendre y Fortín (*op cit.*:134-136).

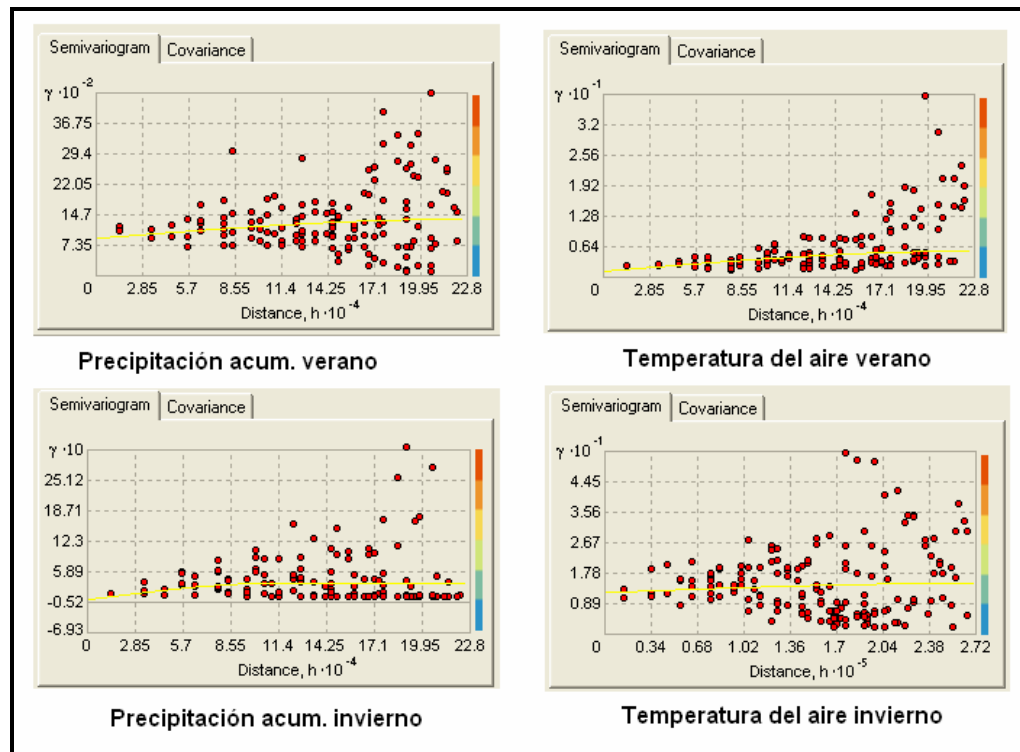
**Fig. 7.14 Variogramas experimental y teórico.**



Se distinguen los parámetros que definen la escala de de la variable según la capturan los datos. La presencia del efecto pepita (nugget) puede ser indicativo de anomalías locales.

En la figura 7.15, se observan variogramas experimentales para los datos de temperatura y de precipitación para la época de invierno y la de verano. Para probar el efecto de la dirección sobre la autocorrelación, se recomienda explorar variogramas experimentales en diferentes direcciones para determinar si la estructura espacial es constante en todas direcciones (isotrópica) o no. Un indicativo de tendencia global es un variograma experimental con forma de función cuadrática, lo que nos conduce inmediatamente a la remoción de esta deriva externa mediante métodos de regresión. Se representaron variogramas en diferentes direcciones pero se observó en todos la presencia de deriva en la media.

**Fig. 7.15 Variogramas experimentales para precipitación y temperatura del aire.**

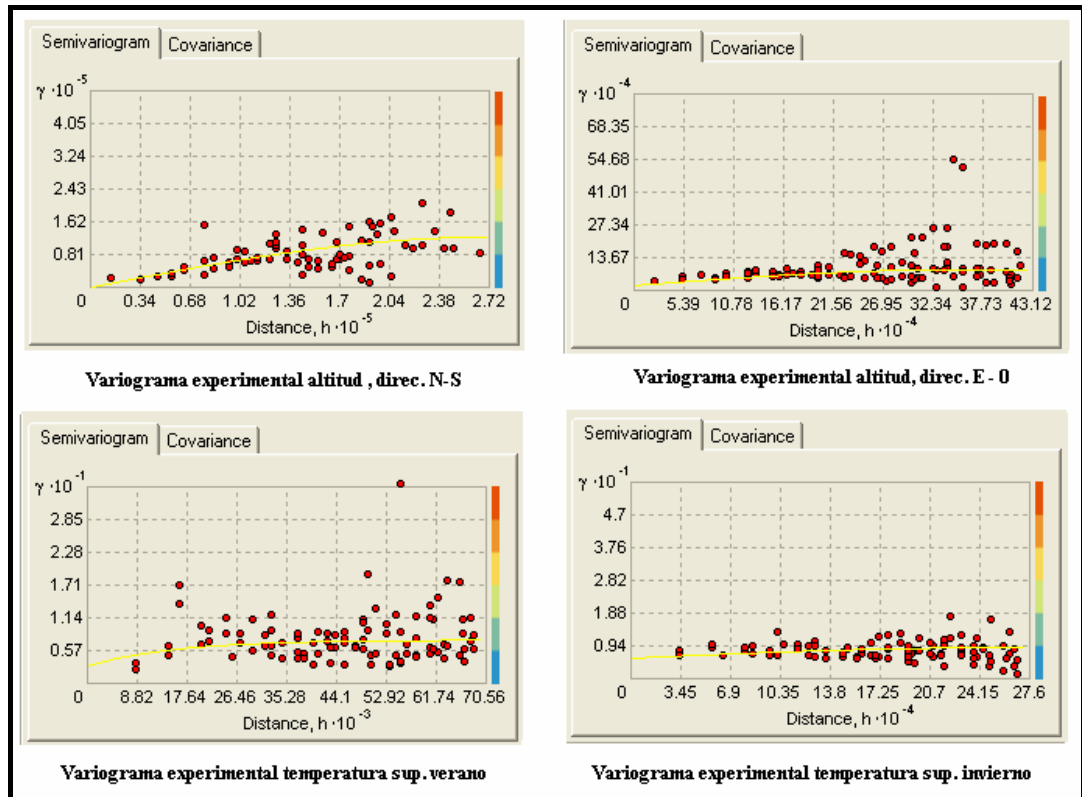


Se presentan dos periodos de 16 días en verano e invierno, en el 2002, para datos de estaciones metereológicas en la región del Bajío.

Para identificar si los datos de altitud y temperatura de superficie presentan una estructura de variación similar o no a la de los datos de precipitación y de temperatura del aire, podemos explorar el variograma experimental que se obtiene con una muestra de estas variables explicatorias registrada en las locaciones de estaciones metereológicas. Si es así, parte de la correlación entre variables puede ser debida a la estructura espacial de estas variables explicatorias. Los variogramas de la figura 7.16 dan indicios de estructura espacial en la altitud y la temperatura de superficie, capturada por la muestra de datos en la ubicación de las estaciones.

Con este análisis, esperamos que parte de la asociación que se observe entre las variables dependientes y las variables explicatorias, sea debida a la estructura espacial de las variables explicatorias. Esta estructura, denominada estructura espacial ambiental (b), por Legendre (1993:1661), sólo puede ser conocida identificando la función de estructura que ajusta a la autocorrelación espacial observada.

**Fig. 7.16 Variogramas experimentales de altitud y temperatura de superficie.**



### I.b. Estimación de superficies de temperatura y precipitación

Para generar superficies de estas variables se requiere interpolar el valor de temperatura y precipitación en locaciones no monitoreadas a partir de los registros existentes. Se conocen una gran variedad de métodos de interpolación, algunos como el algoritmo IDW o Spline que consideran implícitamente la distancia entre observaciones y los más sofisticados como Kriging, cokriging o kriging con deriva externa que utilizan los parámetros del variograma para estimar las superficies. En este trabajo se empleará Kriging ordinario una vez que se ha removido la deriva externa, para mayores detalles sobre esta metodología consultar a Govaerts (*op cit.*:1-17) y Hengl *et. al.* (2003:1-17).

El interpolador de Kriging usa un estimador local que considera las observaciones dentro de la vecindad del punto donde una variable va a ser estimada así como la estructura de autocorrelación proveída por el variograma. Esto significa que se estima un punto considerando aquéllos que se encuentran localizados en el elipse del variograma (dado por la dirección y rango) y se ponderan empleando los valores proporcionados por

el variograma teórico a las distancias apropiadas, el computo se hace de tal forma que el resultado no depende de la densidad local de puntos, sino de su distribución en el espacio.

Como vimos, antes de poder realizar un buen ajuste al variograma experimental se requiere remover la tendencia o deriva en la media global de una variable. Para ello, se revisará la correlación de la temperatura del aire y la precipitación con variables que suponemos están asociadas. Respecto de la temperatura se explora la asociación con la altitud y la temperatura de superficie y se hace lo propio con la precipitación respecto de la altitud de acuerdo con Govaerts (*op cit.1*). Es importante resaltar que las variables explicatorias presentan de por sí una estructura de variación espacial como lo identificamos en secciones anteriores.

Ahora procederemos a identificar la asociación entre variables explicatorias con la temperatura del aire y con la precipitación acumulada. En la tabla 7.7 se presenta el coeficiente de correlación lineal de Pearson para cada variable dependiente respecto de la altitud y la temperatura de superficie para los periodos de 16 días durante el verano y el invierno. Como se aprecia en la tabla 7.7, no existe una correlación importante entre las variables de entrada al modelo de balance hídrico y las variables causales sugeridas, especialmente en lo que respecta a la precipitación.

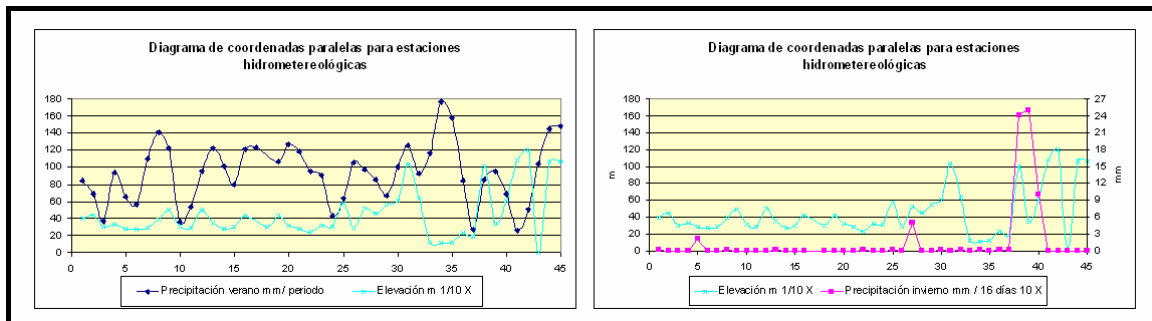
**Tabla 7.7 Coeficientes de correlación lineal de Pearson**

	Precip_ver	Precip_inv	Temp_a_ver	Temp_a_in
Altitud	0.0384	0.2283	-0.5459	-0.1977
Temp_sup	NA	NA	0.5620	0.2108

Altitud y temperatura de superficie respecto de precipitación acumulada y temperatura promedio del aire para periodos de 16 días en verano e invierno del 2002, región del Bajío.

Otra herramienta utilizada para identificar asociaciones son los diagramas de coordenadas paralelas, los cuales nos dan elementos para explicar por que las correlaciones lineales entre variables son bajas. En éstos se representa la variabilidad de los datos conjuntamente, de tal manera que puedan reconocerse patrones coincidentes en cuanto a la forma de variación y asociación entre variables (Ver figura 7.17).

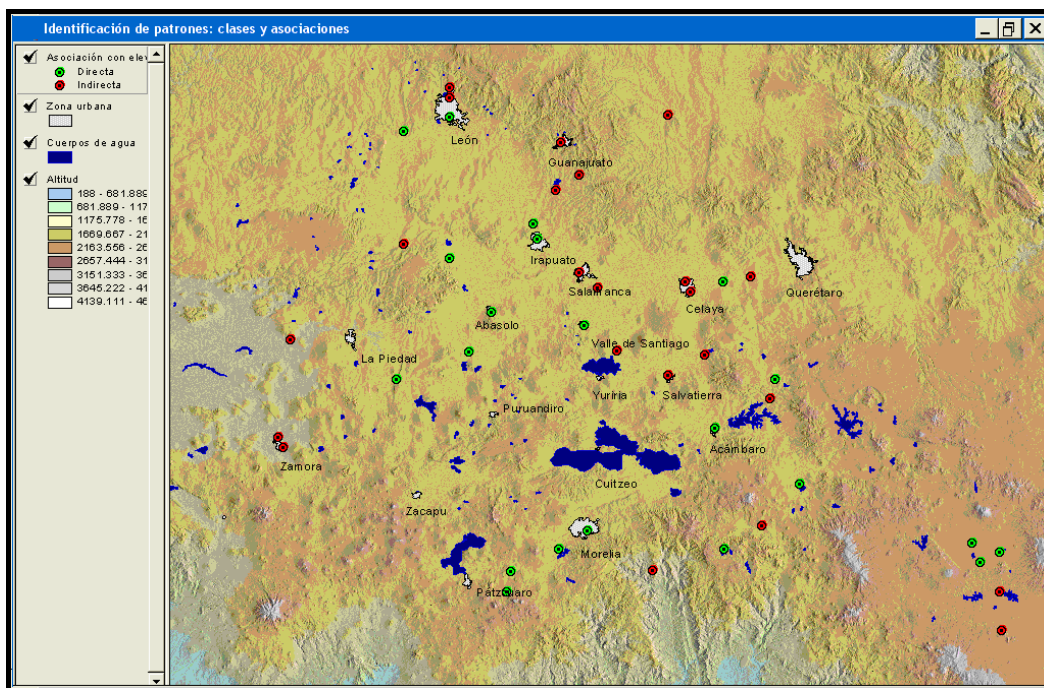
**Fig. 7.17 Diagramas de coordenadas paralelas.**



Se presenta la variación de los datos de estaciones meteorológicas para la precipitación de dos periodos, en azul marino para el verano y en rosa para invierno del año 2002. Se explora la asociación con la elevación para facilitar la comparación, ésta se calculó a partir de las diferencias de altitud entre puntos y se re-escaló a 1/10 X.

Para la precipitación acumulada en verano la asociación con la elevación es muy variable, en algunas estaciones es directamente proporcional y para otras es inversamente proporcional. Según el mapa de la figura 7.18, hay estaciones que presentan correlación inversa (aparecen en rojo) y se ubican en zonas altas pero más secas o muy cerca de zonas urbanas. Esto nos sugiere que la correlación entre la precipitación y la elevación varía según la localización geográfica.

**Fig. 7.18 Mapa de asociación entre precipitación y elevación, verano 2002.**

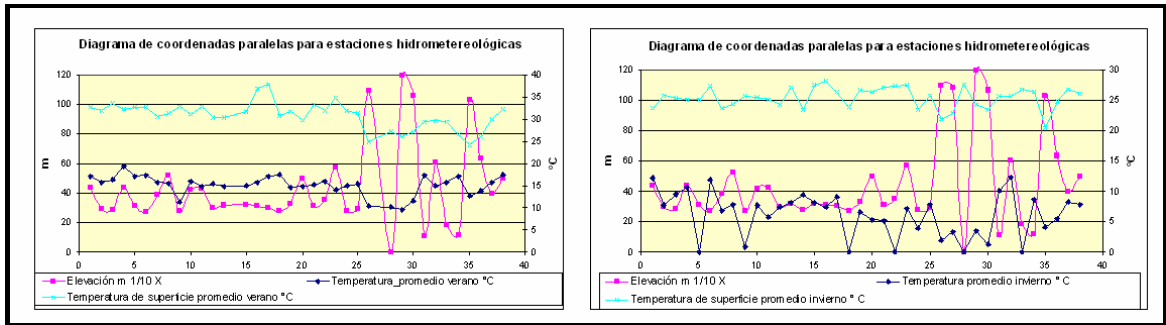


Las estaciones en verde presentan una correlación directa con la elevación y los puntos en rojo representan una asociación inversamente proporcional.



Respecto de la temperatura del aire se observa claramente en la figura 7.19 una correlación lineal con la temperatura de superficie. En cuanto a la asociación con la elevación, la temperatura del aire en verano está más correlacionada que la temperatura en invierno pues presenta en conjunto una menor variación en las asociaciones.

**Fig. 7.19 Diagramas de coordenadas paralelas, temperatura del aire.**

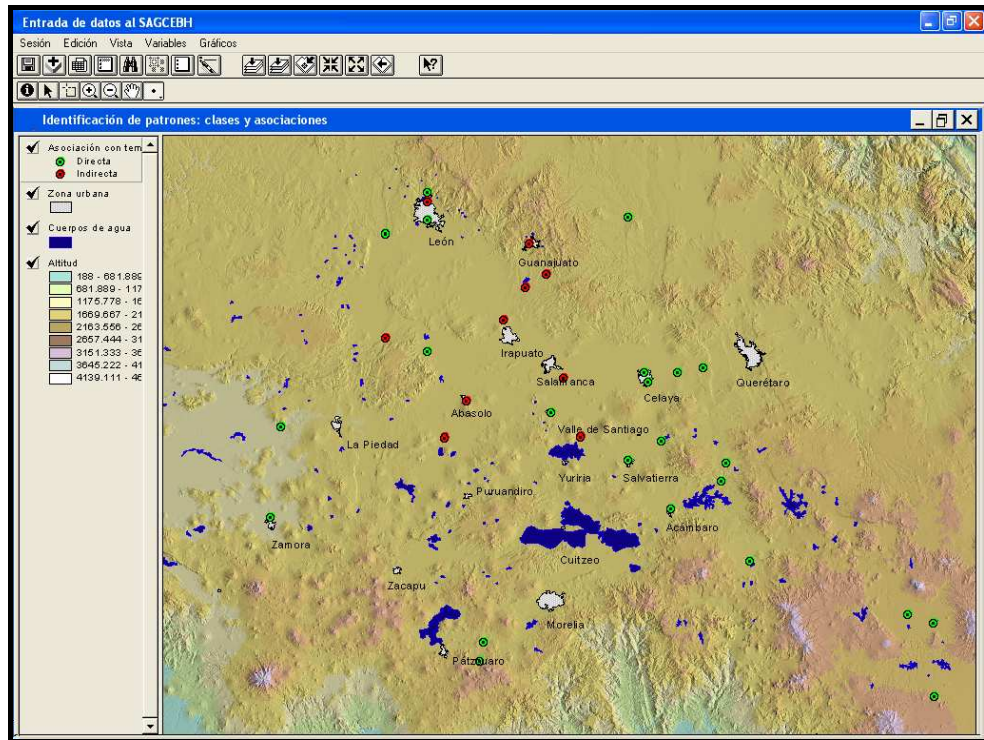


Se presenta la variación de los datos de estaciones meteorológicas para la temperatura promedio del aire en dos periodos, verano (gráfico a la izquierda) e invierno (derecha) del año 2002. En los diagramas se representa la variación de la temperatura del aire promedio en azul oscuro y su asociación con la temperatura de superficie en azul claro y la elevación re-escalada en rosa.

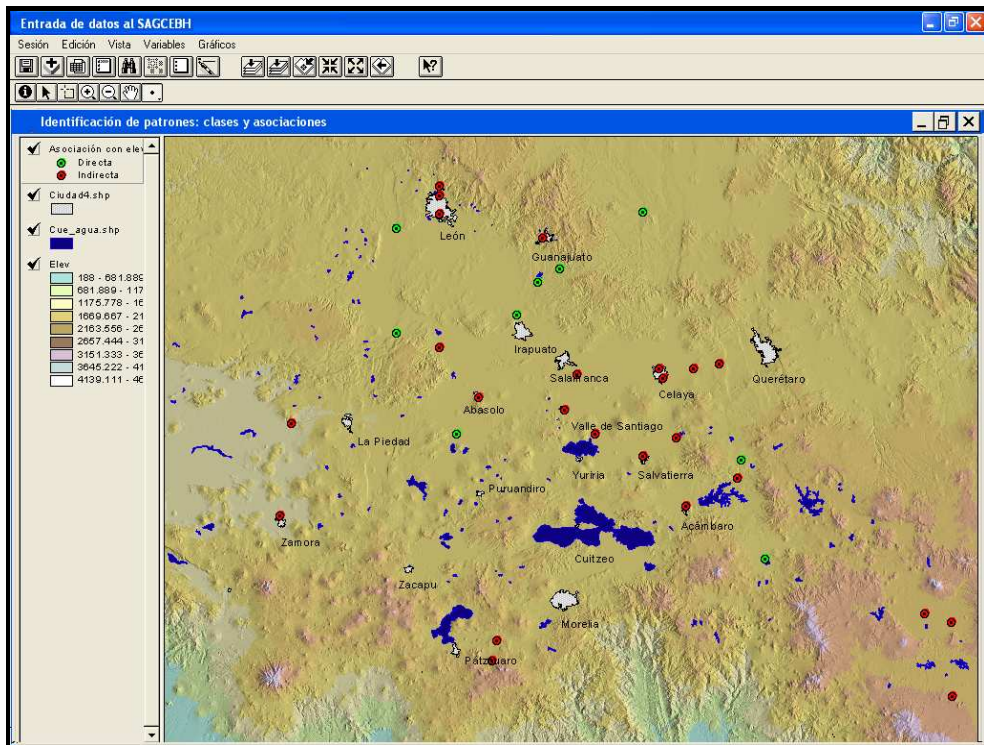
De acuerdo al mapa en la figura 7.20a, las estaciones que presentan una asociación inversamente proporcional entre la temperatura del aire y la de superficie, en invierno, se encuentran cerca de zonas urbanas o en áreas abiertas muy probablemente desprovistas de vegetación por lo que se espera que la temperatura de superficie sea sensiblemente mayor que la temperatura del aire. Recordemos que tanto el suelo como las zonas cubiertas con superficies impermeables como cemento, pavimento o concreto retienen por más tiempo el calor que las áreas cubiertas con vegetación.

Conforme al mapa en la figura 7.20b, encontramos estaciones en verde que se comportan de una manera muy particular ya que registran temperaturas bajas en invierno y se encuentran a una altitud por debajo de la media en la zona, todas se localizan en la porción norte del área y se encuentran expuestas a cualquier fenómeno climático como podrían ser ondas frías. El resto de las estaciones, en rojo, presentan temperaturas muy bajas y se localizan a elevaciones superiores a la media o se localizan en lugares donde la temperatura del aire es superior a la media pero la elevación es baja.

Fig. 7.20 Mapas de asociación, temperatura y variables físicas, invierno de 2002.



a)



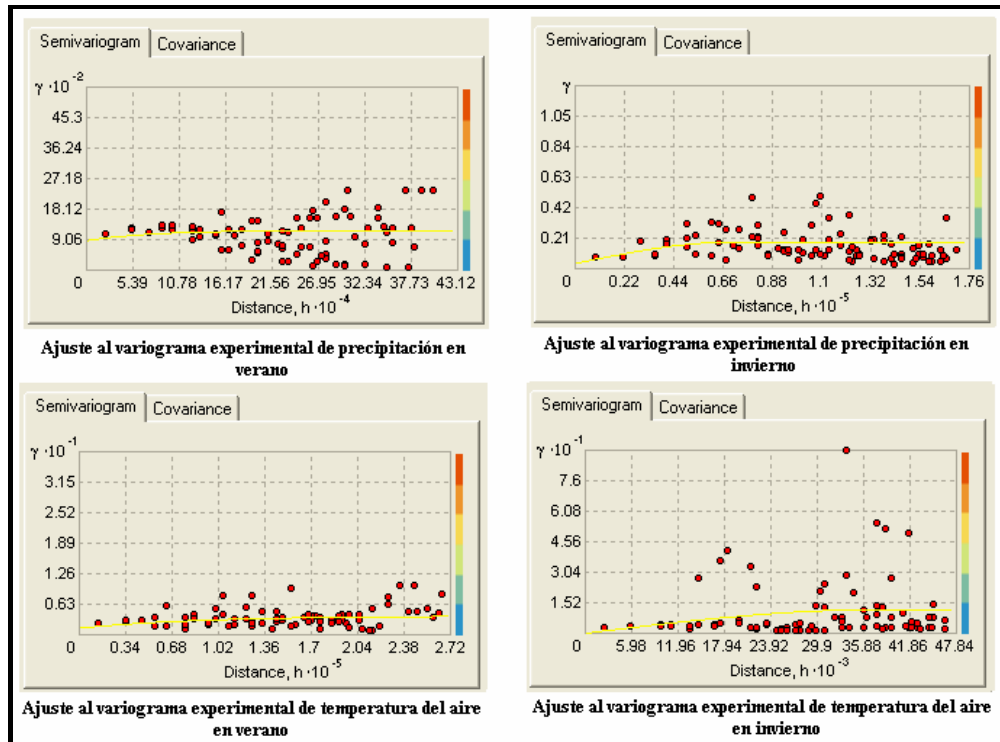
b)

Las estaciones en verde presentan una correlación directa entre variables y los puntos en rojo presentan una asociación inversamente proporcional. a) Correlación entre temperatura de superficie y temperatura del aire, b) correlación entre elevación y temperatura del aire.

La representación de la forma de las asociaciones nos permite iniciar la identificación de posibles patrones del tipo clases o regiones, ya que el área de estudio parece diferenciarse entre el centro de la región coincidiendo con las zonas planas y las porciones al norte y sur que rodean dicho centro. La extracción de estos rasgos o patrones continuará en la siguiente sección.

Para continuar con el proceso de interpolación de datos de estaciones, se obtienen los modelos de regresión lineal para temperatura y precipitación respecto de las variables causales mencionadas. Una vez obtenida la superficie de regresión se toman los residuales y se procede a identificar la autocorrelación espacial mediante variografía. A continuación, se muestra el ajuste de los variogramas experimentales de la variación residual a modelos teóricos (ver figura 7.21).

**Fig. 7.21 Variografía para datos de precipitación acumulada y temperatura del aire.**



Comparativamente con la figura 7.15, se observa un mejor ajuste después de la remoción de la deriva externa en la media.

Se observó que los modelos que más se ajustan a los variogramas experimentales son el exponencial y el esférico (ver tabla 7.8), por lo que se procede a la medición del error

medio de la estimación y la correspondiente desviación por validación cruzada y se obtienen los parámetros de error indicados en el mismo cuadro.

**Tabla 7.8 Parámetros de variogramas teóricos.**

Variable	Modelo	Nugget ( $C_0$ )	Sill Parcial (C)	Sill	Dirección	Rango mayor	Rango menor	Media error	Desv. Error
T. aire verano	Exponencial	1.1660	3.759	4.926	307.800	267330.000	259580.000	0.004	1.500
T. aire invierno	Esférico	0.0000	11.761	11.761	332.800	44706.000	18977.000	0.041	3.688
Prec. ac. veran	Exponencial	874.8600	315.040	1189.900	338.400	403100.000	207170.000	1.945	N/A
Prec. ac. Invier	Exponencial	0.0034	0.16768	0.171	321.300	164370.000	69931.000	-0.009	N/A

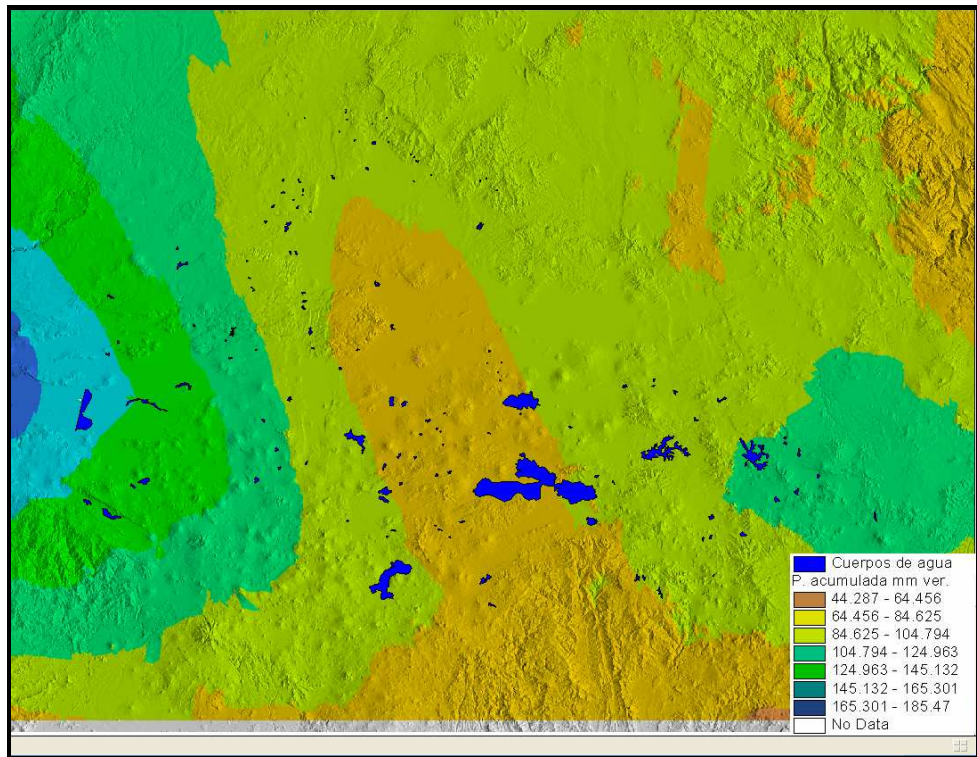
Representan el ajuste a datos de estaciones metereológicas para temperatura del aire y precipitación acumulada en periodos de 16 días en verano e invierno de 2002. Las unidades de los rangos son metros.

Los modelos indican que la extensión de las variables es específica, según la variable y la época del año. En verano, las variables presentan un dominio de escala más regional llegando hasta los 400 Km. en el caso de la precipitación. La autocorrelación espacial más extensa se registra en la dirección NO – SE en todos los casos, por ello, se espera que en esa dirección y en su respectiva perpendicular, se identifiquen patrones respecto de estas variables.

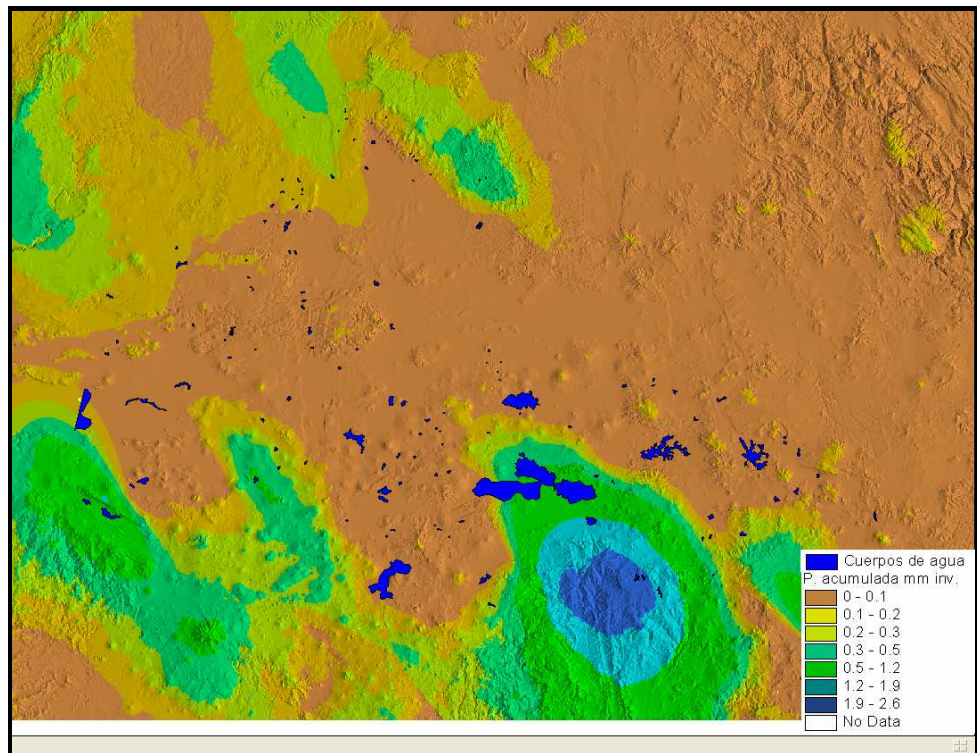
Ya que la validación sugiere que el error de la estimación es mínimo, se obtienen las superficies de residuales y se suman a las superficies de interpolación por regresión para cada variable y periodo. Mediante este procedimiento se obtienen las superficies de precipitación y temperatura del aire que se presentan en los mapas de las figuras 7.22 y 7.23, respectivamente. Los mapas hablan del dominio de escala de las variables, presentándose patrones regionales de precipitación en verano en la dirección noroeste sureste principalmente (ver figura 7.22a), en contraste, la precipitación acumulada en invierno presenta patrones desde aquella dirección y la dirección suroeste – noreste hacia el centro de la región (ver figura 7.22b).



Fig. 7.22 Mapas de precipitación acumulada.



a)

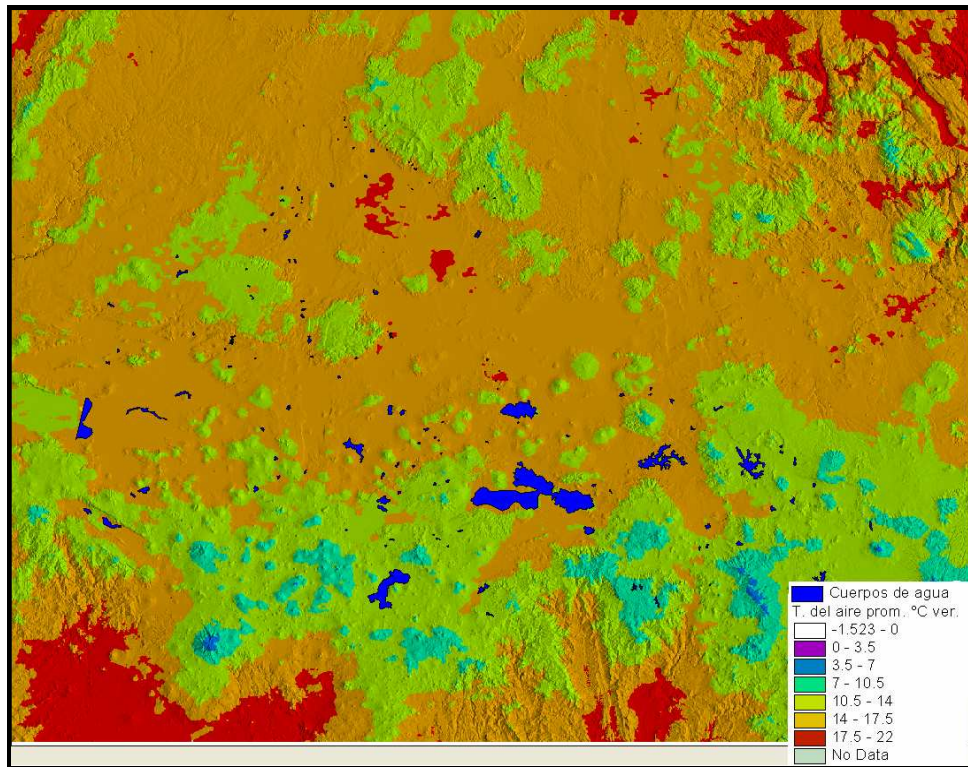


b)

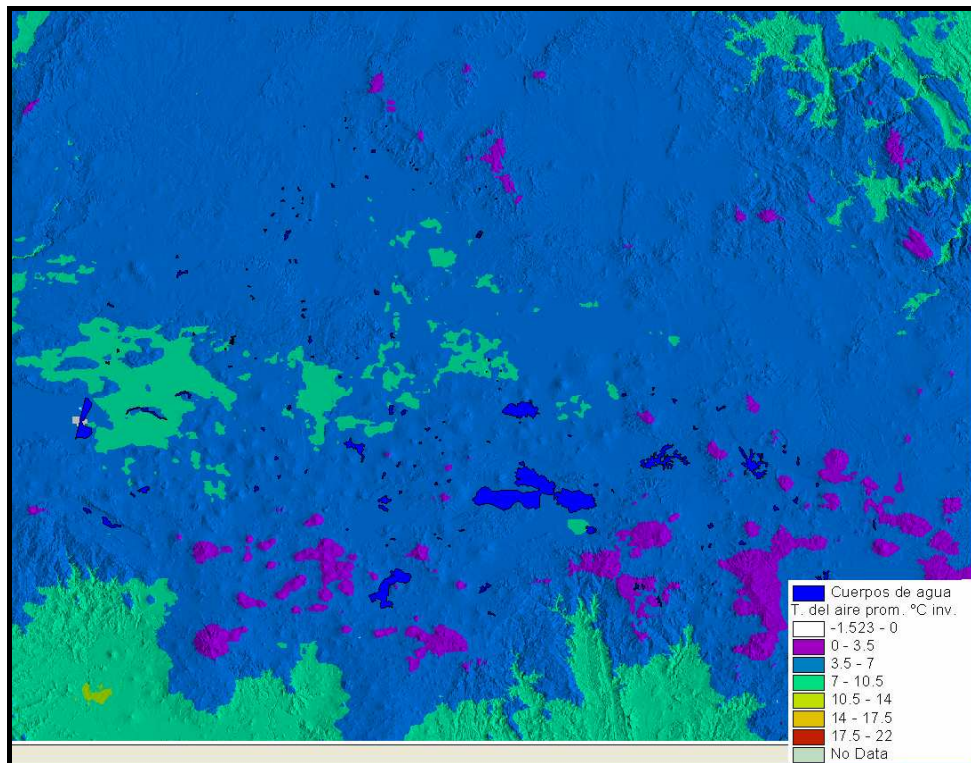
a) Precipitación en verano, b) Precipitación en invierno, mm. / 16 días.



Fig. 7.23 Mapas de temperatura del aire promedio diaria.



a)



b)

a) Temperatura en verano, b) Temperatura en invierno, °C / día para periodos de 16 días.

En cuanto a la temperatura del aire en verano (ver mapa de la figura 7.23a), observamos patrones mucho más locales en direcciones similares a las de precipitación, aunque en sentido opuesto, pues el gradiente va del centro a la periferia. Se distinguen islas de calor en la depresión del Balsas hacia el sur de la región y al noreste en las estribaciones occidentales de la sierra madre oriental, las cuales corresponden a la cuenca del Pánuco. Estos sistemas de calor no contribuyen a la dinámica hidrológica del Lerma por lo que no serán considerados de aquí en adelante. En el mapa de la figura 7.23b se observa que la temperatura del aire en invierno es más fría al sureste de la región y las zonas más calientes coinciden con la depresión del Balsas, las zonas cercanas a ciudades y a los valles donde se asientan algunos cuerpos de agua.

Para identificar el dominio de escala de las variables de temperatura de superficie, altitud y fracción vegetal de la cobertura terrestre, se resumen, en la tabla 7.9, los parámetros de los variogramas teóricos para estas variables.

**Tabla 7.9 Parámetros de variogramas teóricos ajustados a variables independientes**

Variable	Modelo	Nugget (C <sub>0</sub> )	Sill Parcial (C)	Sill	Dirección	Rango mayor	Rango menor	Media error
Altitud	Exponencial	20918.0000	66416.000	87334.000	286.600	403090.000	189360.000	14.290
T. sup. Verano	Exponencial	2.9644	4.575	7.539	7.900	65988.000	45399.000	-0.219
T. sup. Invierno	Esférico	5.8564	2.991	8.847	80.300	258170.000	132670.000	-0.210
Frac. Veg. Ver.	Exponencial	0.0007	0.021	0.021	22.600	34171.000	23257.000	-0.005
Frac. Veg. Inv.	Esférico	0.0004	0.013	0.013	333.100	41613.000	19476.000	-0.002

Datos registrados en verano e invierno de 2002. Las unidades de los rangos son metros.

Los variogramas confirman la presencia de estructura espacial en las variables independientes, por lo que, efectivamente, parte de la variación en la precipitación y la temperatura del aire puede ser explicada por la estructura espacial ambiental. Además, como se mencionó anteriormente, la fracción vegetal es la variable con el dominio de escala más local, seguida de la temperatura de superficie en verano, en invierno y por último, la altitud, con el dominio de escala más regional.

## II. Formalización de asociaciones mediante regresión parcial

Mediante la técnica de regresión parcial formalizamos las relaciones entre variables y explicitamos la porción de variación, en los componentes del balance, que es explicada por las diferentes variables ambientales y la proporción correspondiente a la estructura espacial.

La estrategia general del análisis de regresión parcial sugerido por Legendre (1993:1662-1668), consiste en determinar el potencial explicatorio de una matriz de variables causales respecto de una variable dependiente  $y$ , a esta porción de la variación se le denomina varianza ambiental ( $a + b$ ). La porción de variación explicada por el espacio se denomina estructura espacial. La varianza remanente en la variable dependiente se conoce como varianza no explicada.

Sin embargo, parte del poder explicatorio de las variables ambientales tiene que ver con la estructura espacial que presenta la matriz de las variables independientes y que es capturada por la muestra de datos analizados. Esta fracción corresponde a la estructura espacial ambiental ( $b$ ).

En primer término la fracción de variación ambiental es identificada mediante la regresión de la variable  $y$  respecto de la matriz de variables independientes esta fracción está compuesta por la variación ambiental no espacial ( $a$ ) y por la variación ambiental espacialmente estructurada ( $b$ ).

Para separar la fracción ( $a$ ) de la variación ambiental ( $a + b$ ), se mide el efecto combinado de variación espacialmente estructurada de las variables ambientales ( $b$ ), el cual fue definido por los respectivos variogramas, y se resta del total de la variación ambiental.

Una vez removida la fracción ambiental queda un remanente de variación correspondiente a la estructura espacial de la variable  $y$  ( $c$ ) y a la varianza no explicada ( $d$ ). El residual  $c$  queda definido por el variograma teórico que ajusta a la estructura de la variable  $y$ .

En la tabla 7.10 se describen las funciones para calcular cada una de las fracciones descritas arriba, los resultados corresponden al análisis de regresión parcial que se hizo paralelo al proceso de interpolación de las superficies de precipitación y temperatura del aire para verano e invierno.



**Tabla 7.10 Fracciones de variación explicada.**

	var_nesp_amb (a)	var_esp_amb (b)	var_esp_int (c)	var_nexp (d)
Formulación	$y = [(var_1, \dots, var_n) - (var_1, \dots, var_n) \gamma(d)]$	$y = [(var_1, \dots, var_n) \gamma(d)]$	$y = [ (var_0) \gamma(d)]$	$y = [((var_1, \dots, var_n), (var_1, \dots, var_n) \gamma(d), (var_0) \gamma(d)) \text{ res}]$
Prec.ac. ver	0.0104	0.0009	0.2149	0.7738
Prec.ac. inv	0.0605	0.0130	0.4799	0.4466
T. aire verano	0.1493	0.1291	0.1648	0.5568
T. aire inviern	0.0176	0.0152	0.0106	0.9566

Para acumulados de precipitación y promedio de temperatura del aire para periodos de 16 días en verano e invierno de 2002, región del Bajío. La estructura espacial de variación se obtiene sumando las parciales b + c. Los valores representan porcentajes decimales.

A excepción del promedio de temperatura del aire para el periodo de invierno, la estructura espacial explica una fracción importante de la variación de las variables analizadas. La precipitación en invierno es la que cuenta con mayor porcentaje de variación respecto del espacio (con 49.29 %). Ello resalta la importancia de considerar el espacio como una fuente de variación en los procesos analizados.

Por otra parte, se había previsto que la altitud y la temperatura de superficie contribuirían poco a la variación total en las variables de entrada al balance hídrico, dada la magnitud del coeficiente r de correlación lineal entre los datos. La mayor contribución se estimó para la temperatura del aire en verano con casi un 18 % de su varianza como función las variables ambientales.

### III. Estimación de la evapotranspiración potencial

La evaporación o evapotranspiración potencial es la otra entrada importante del balance hídrico que preferiblemente requiere ser registrada directa o indirectamente a través de ciertos modelos que revisan su relación con la temperatura del aire y otras variables, de aquí la necesidad de contar con superficies precisas de temperatura.

El método sugerido para el cálculo de evaporación es el de Thornthwaite tal como se plantea en Serafini y Mintz (*op cit.*15,16), donde se precisa de la estimación de un índice de calor *I* que es el promedio diario anual de temperatura del aire.

Con base a relaciones empíricas, Thornthwaite obtuvo en 1948 la regresión de la evaporación  $E^*$  respecto de  $T_A$ :

$$E^* \begin{cases} = 0 & \text{para } T_A < 0^\circ\text{C} \\ = 0.444 h (10 T_A / I)^a & \text{para } 0 \leq T_A < 26.5^\circ\text{C} \\ = -13.862 + 1.0747 T_A - 0.01442 T_A^2 & \text{para } T_A \geq 26.5^\circ\text{C} \end{cases}$$

$$I = \sum_{i=1}^{12} i, \quad i = (T_M / 5)^{1.514}, \quad i_{\min} = 0$$

$$a = (6.75 \times 10^{-7} I^3) - (7.71 \times 10^{-5} I^2) + (1.792 \times 10^{-2} I) + 0.49239$$

donde ,

$E^*$  = evapotranspiración potencial por periodo de 16 días, mm. / día

$h$  = longitud del día en horas

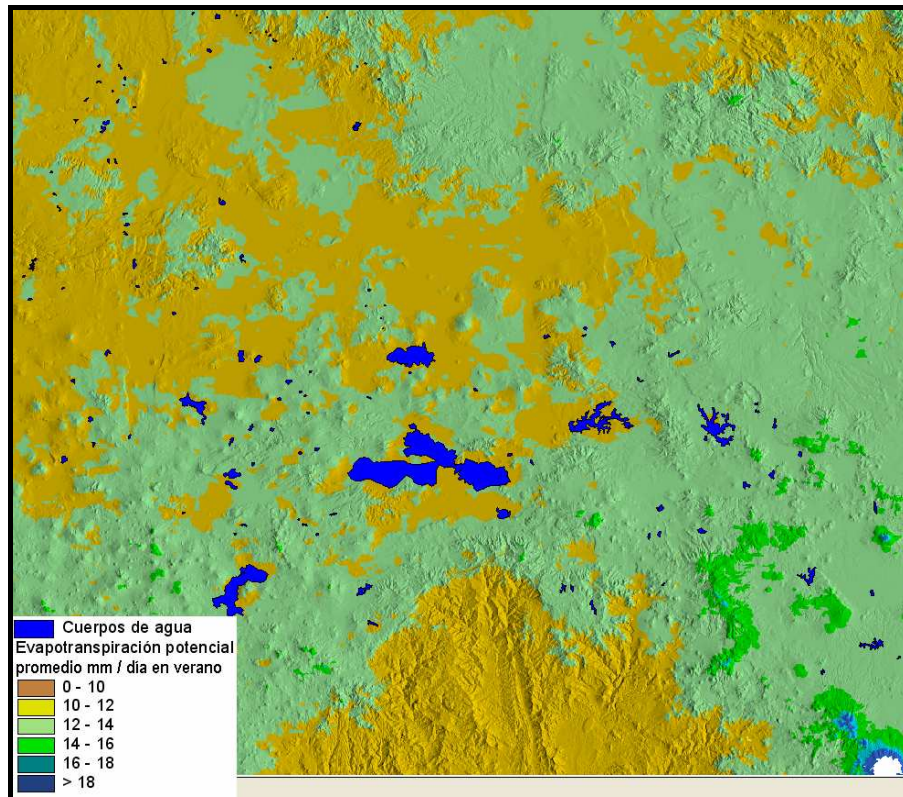
$T_A$  = temperatura media del aire por periodo, °C

$T_M$  = temperatura media normal del aire por periodo, °C

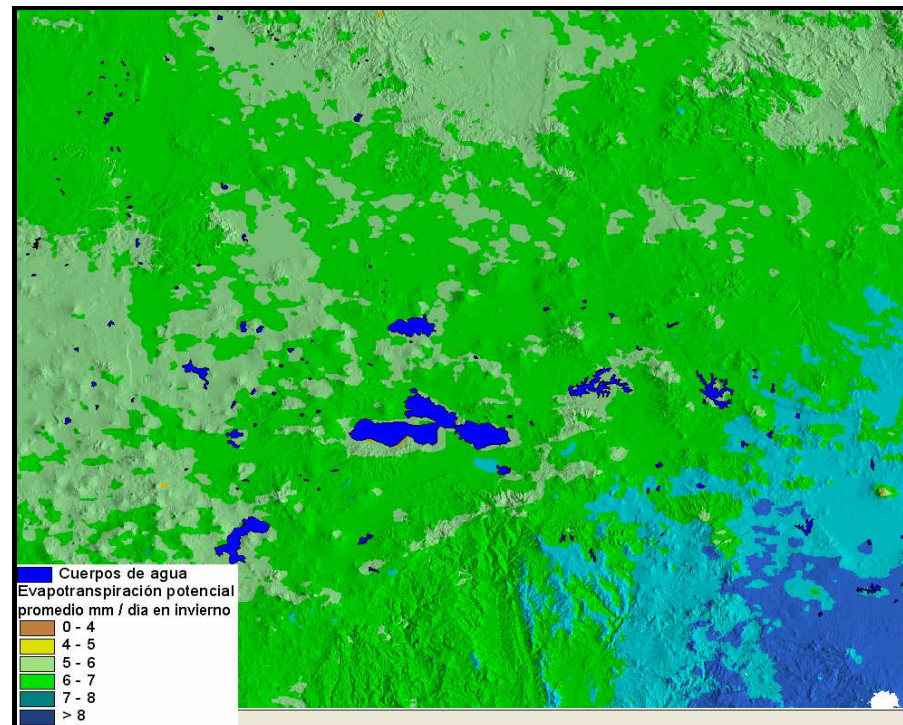
$I$  = índice de temperatura del aire media anual por periodo

Se emplearon estas relaciones para estimar la evaporación a partir de datos interpolados de temperatura del aire promedio diario para cada periodo y estimaciones de  $h$  respecto de la latitud, se emplearon además normales históricas previamente interpoladas por Mora e Iverson (1998:1850-1860). En los mapas de la figura 7.24 se observan las superficies obtenidas para la evaporación en verano e invierno. Se puede distinguir a primera vista diferencias en el dominio de escala espacial de la variable para ambos periodos, la evapotranspiración potencial resulta ser más regional durante el verano. Se observa la mayor evaporación en las zonas altas en un gradiente que va de noroeste a sureste, la cual es una de las direcciones de principal variación.

Fig. 7.24 Mapas de evapotranspiración potencial.



a)



b)

a) Evapotranspiración potencial promedio en verano, b) Evapotranspiración potencial promedio en invierno, mm. / día para periodos de 16 días.

Para completar la identificación del dominio de escala de la evapotranspiración potencial se detallan en la tabla 7.11 los parámetros del ajuste al variograma experimental de la variable.

**Tabla 7.11 Parámetros de variogramas teóricos ajustados a datos de evaporación promedio diaria**

Variable	Modelo	Nugget (C <sub>0</sub> )	Sill Parcial	Sill	Dirección	Rango mayor	Rango menor	Media error
Evap. Verano	Exponencial	0.0464	0.024	0.070	47.900	256130.000	97675.000	0.018
Evap. Invierno	Exponencial	0.0354	0.043	0.079	74.000	154850.000	79551.000	0.004

Estimación en verano e invierno de 2002. Las unidades de los rangos son metros.

Con base a los parámetros del variograma teórico ajustado a los datos estimados de la evapotranspiración potencial, una de las principales direcciones de variación es la noroeste-sureste y en ambas fechas esta dirección presenta el rango menor. Por ello en esta dirección el gradiente es más evidente.

### 7.2.2c. Aportación de la fase de análisis exploratorio o estructural

En esta etapa se presentaron herramientas de análisis exploratorio necesarias para identificar la presencia de estructura espacial en las observaciones a partir de los formatos de representación lógica puntual y raster. Se identificaron propiedades intrínsecas de procesos geográficos como la presencia de vecindarios, la heterogeneidad, la autocorrelación espacial y el dominio de escala de las variables analizadas. Así mismo, se formalizaron asociaciones entre variables, reconociéndose que los periodos de tiempo analizados son instancias de estructuras temporales que se diferencian no sólo por la extensión de la autocorrelación temporal con periodos cercanos, sino por la forma como los procesos explicatorios afectan las variables dependientes. La identificación de dominios de escala en periodos de tiempo diferentes equivale a probar la presencia de configuración espacio-temporal en los procesos estudiados, la cual es un elemento fundamental dentro del modelo jerárquico del paisaje. Las superficies que representan los componentes del balance hídrico junto con sus propiedades intrínsecas constituyen datos con un nivel III de información. La magnitud del error de la estimación determina las limitaciones de la modelación. Como vemos, la forma de representación de los datos es similar a la de la etapa anterior, pero tenemos elementos formales que nos caracterizan el comportamiento de las variables.

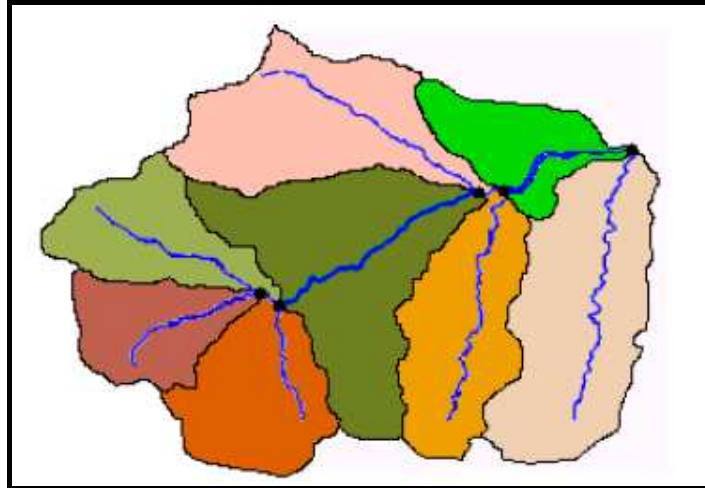
Lo que sigue a continuación, es aplicar los conceptos espaciales a estas superficies para extraer categorías tipo clases y caracterizar los patrones de precipitación y evapotranspiración potencial que resulten válidos y significativos. A partir de aquí, iniciamos con el proceso de extracción de patrones.

#### **7.2.2d. Regionalización hidrológica**

En esta sección nos interesa conformar las unidades de captación en las que identificaremos los patrones espaciales de la precipitación y la evapotranspiración potencial. El método general utilizado para delimitar unidades hidrológicas es el que se plantea en el módulo Arc Hydro de ArcGis (Maidment, 2000:3-59). Consiste en asociar áreas de captación pluvial que contribuyen a puntos de salida o de acumulación de flujos en la red hidrológica. Estos puntos (outlets), pueden seleccionarse manualmente, en estaciones de aforo o bien de forma estandarizada mediante la definición de un área umbral. Este último método es el empleado para la definición de unidades de captación o de drenaje y emplearemos la referencia sobre los dominios de escala de los componentes del balance hídrico para construir regiones hidrológicas. Debido a que se han identificado diferentes dominios de escala para la precipitación y para la evapotranspiración potencial, al menos en verano, se sugiere modelar unidades de captación a dos escalas: unidades a partir de 300 km<sup>2</sup> y unidades a partir de 1000 km<sup>2</sup>.

Como sabemos, las unidades de drenaje se representan como polígonos y conforman en conjunto el paisaje a analizar. En el esquema de la figura 7.25, se observa la forma como se definen las áreas que contribuyen a los puntos de acumulación a lo largo de un escurrimiento o línea de flujo.

**Fig. 7.25 Esquema de unidades de drenaje tipo.**



Se componen de áreas cuyos escurrimientos contribuyen directamente a puntos de acumulación de flujo predeterminados. Tomado de Maidment (2000:4).

En la figura 7.25, los puntos en negro son salidas de acumulación y las áreas diferenciadas en color, las áreas de drenaje que aportan escurrimientos a una misma salida o punto. La red de escurrimientos es un elemento fundamental para la definición de las regiones hidrológicas, sin embargo, su representación no es necesaria para extraer patrones de precipitación y evapotranspiración potencial. Por ello, se omite a partir de aquí la consideración de las redes de drenaje en los análisis.

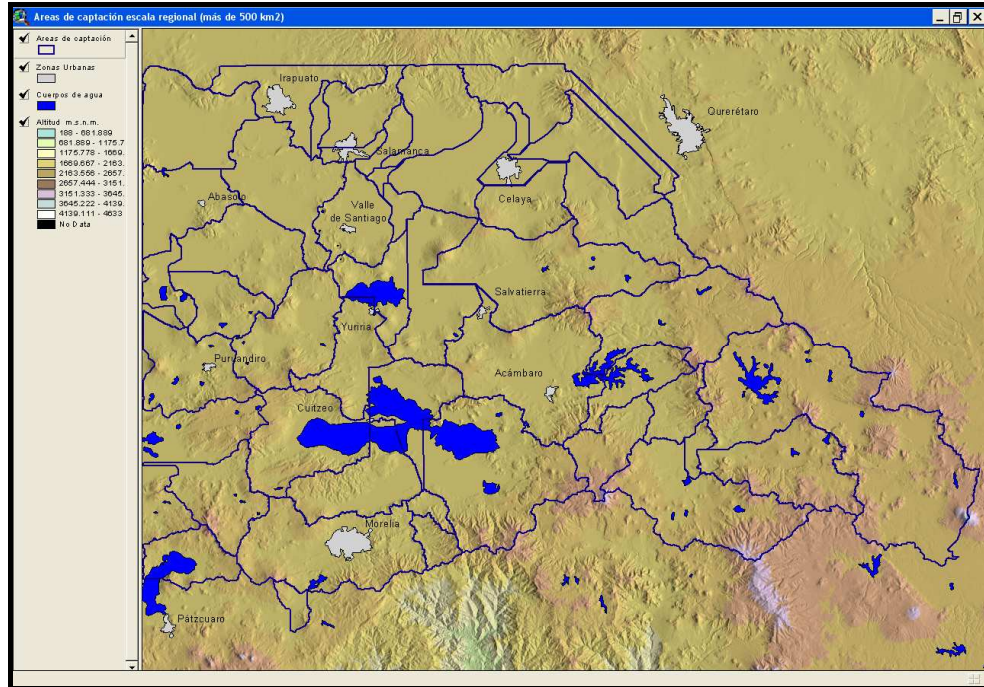
A continuación, se presentan las unidades de drenaje modeladas a dos escalas considerando la cuenca del Lerma y la zona del Bajío (ver mapas de las figuras 7.26a y 7.26b). El modelo de elevación del terreno se incluye como mapa base para destacar el desempeño del algoritmo de ArcHydro en la modelación de las áreas de captación.

En el mapa de la figura 7.26a se delimitan áreas de captación a partir de 300 km<sup>2</sup> con la intención de capturar la heterogeneidad espacio temporal de la precipitación y la evapotranspiración potencial. Al norte de la región se observan 4 regiones que se ubican en el límite septentrional y presentan al menos un segmento recto, estas áreas se añaden para evitar que las unidades interiores presenten bordes rectos de manera errónea, la presencia de bordes rectos en las unidades al interior de la región indican un relieve plano. Se modelaron unidades mayores que comprenden un área más extensa, con el propósito

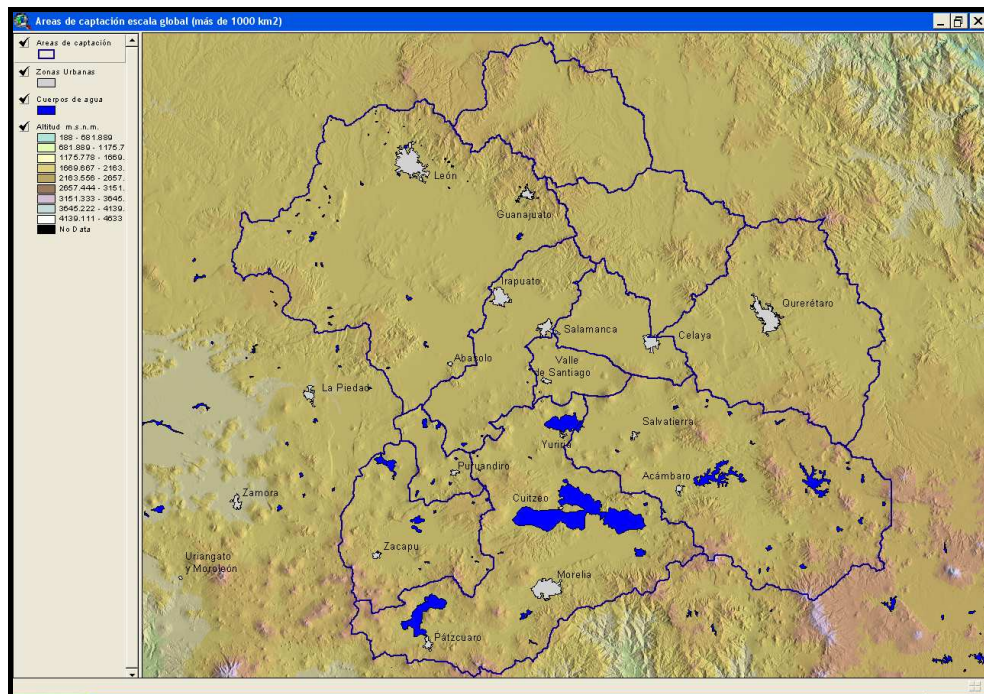


de establecer un umbral de escala en el cual ya no es posible identificar patrones en los componentes del balance debidos a la autocorrelación espacial (ver figura 7.26b).

**Fig. 7.26 Mapas de áreas de captación pluvial.**



a)



b)

a) Unidades hidrológicas a partir de 300 km<sup>2</sup>. b) Unidades hidrológicas a partir de 1000 km<sup>2</sup>. Se incluye la zona del Bajío y algunas otras áreas del Alto Lerma al sureste, así como el lago de Pátzcuaro.

Para identificar regiones de evapotranspiración potencial y precipitación en el área, debemos continuar con el proceso de construcción de clases o clasificación el cual no podrá completarse si no se procede con la caracterización de las unidades respecto de tales componentes. Además, se continuará explorando la asociación de la fracción vegetal de la cobertura terrestre, la altitud y la temperatura de superficie con la precipitación y la evaporación, todo, a nivel de regiones.

### **7.2.3. Construcción de hipótesis**

En esta etapa se requiere un fuerte componente de visualización de los patrones que se conforman al caracterizar las unidades de captación generadas en la etapa previa. Lo que implica esta fase es una inferencia lógica inductiva pues conduce a la generalización de los datos. Esta generalización busca caracterizar a los objetos espaciales estudiados y sus propiedades, en este caso nos referimos a las unidades de captación, su caracterización y la visualización de patrones.

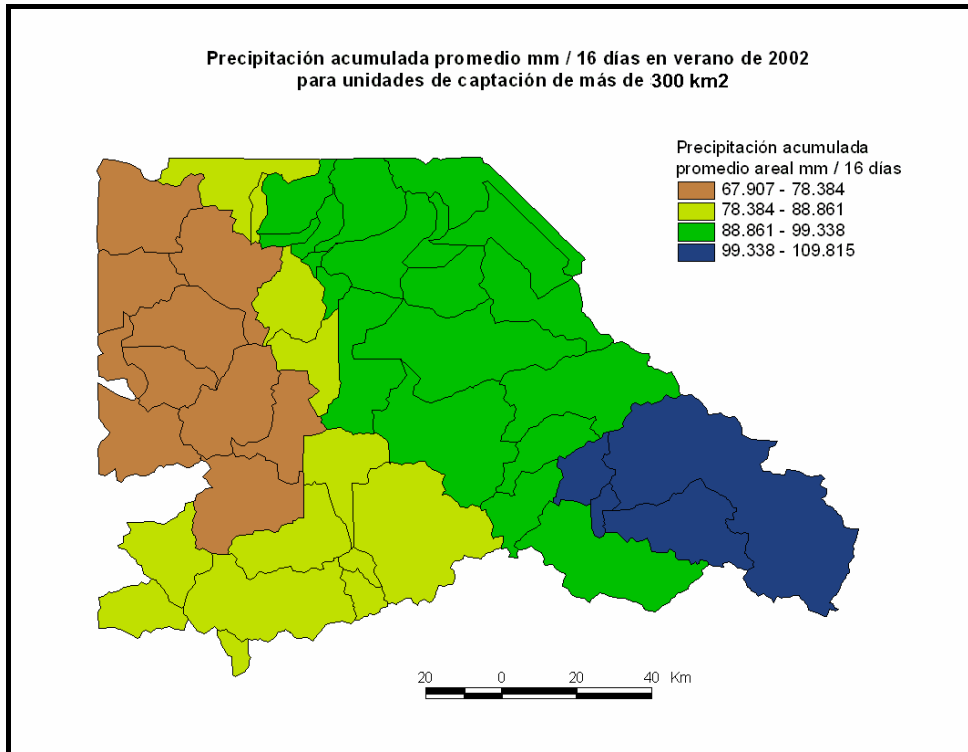
#### **7.2.3a. Geovisualización de patrones**

Los mapas de unidades hidrológicas pueden ser caracterizados mediante promedios areales y nos pueden dar elementos para realizar generalizaciones sobre la heterogeneidad espacial de las variables. Dado el dominio de escala de la precipitación de aproximadamente 400 km. en dirección noroeste-sureste y de 200 km. en dirección noreste-suroeste para la evapotranspiración potencial, esperamos encontrar unidades de captación con un promedio areal tan similar que puedan agruparse en clases conformando regiones en esas direcciones.

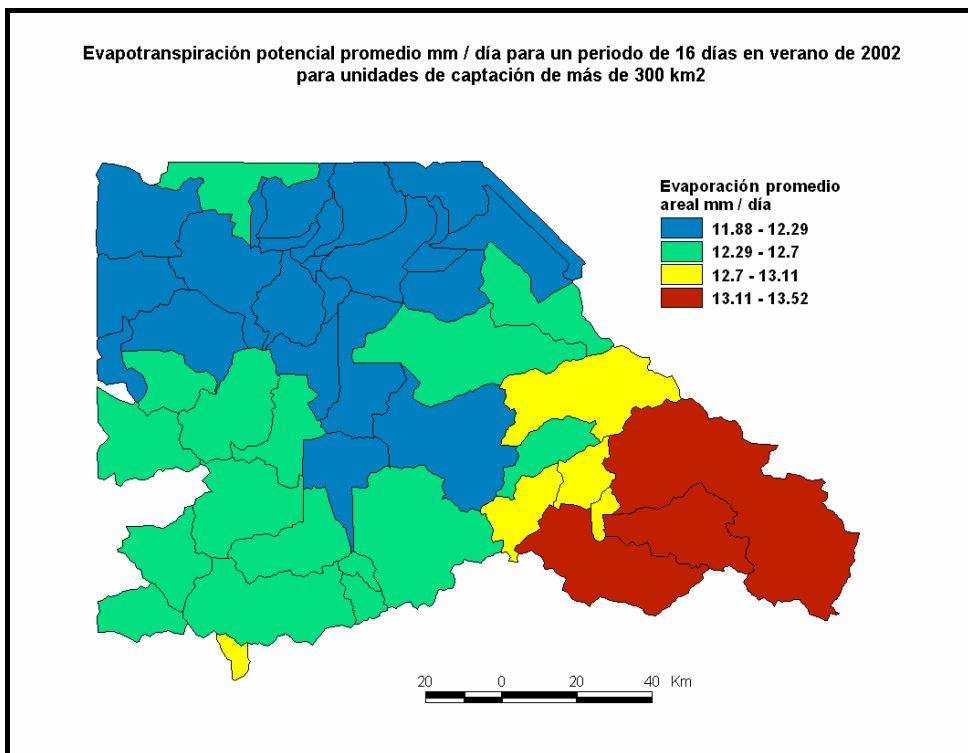
El problema de cuál es la mejor representación de patrones en un mapa tiene que ver con los objetivos del análisis, en este caso buscamos caracterizar la región en cuanto a acumulados promedio de precipitación y promedios diarios de evaporación. Podemos partir de la representación de la variación total de las variables en intervalos iguales. En los mapas de las figuras 7.27a, 7.27b, 7.27c y 7.27d, observamos indicios de la presencia de regiones, debemos probar que los patrones tienen que ver o no con la autocorrelación espacial de las variables.



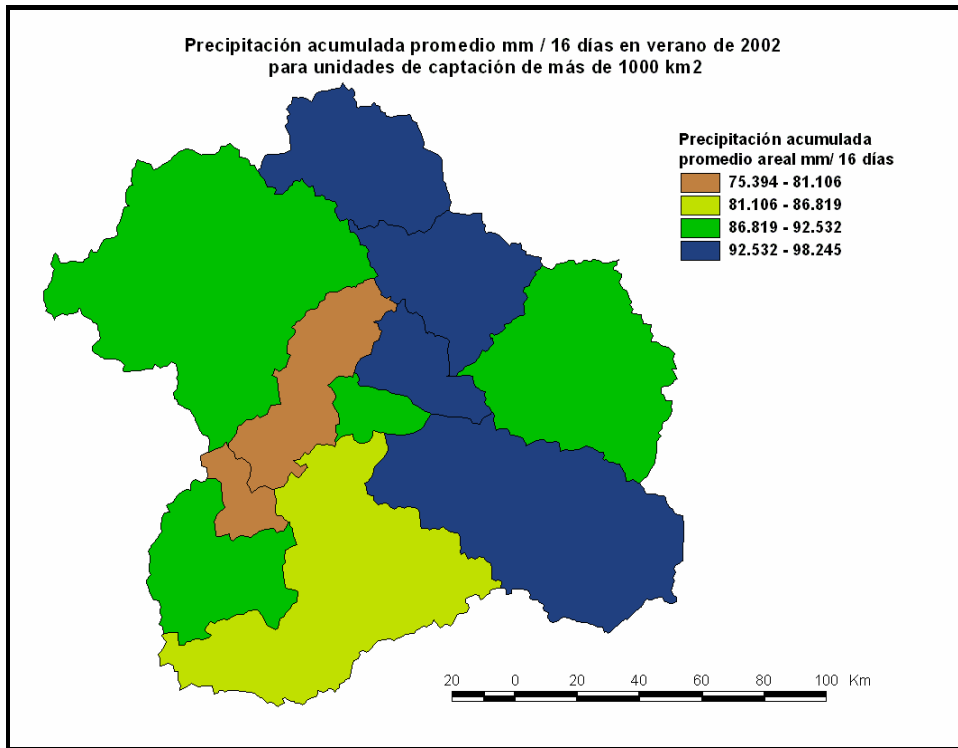
**Fig. 7.27 Mapas de promedios areales para unidades de captación pluvial.**



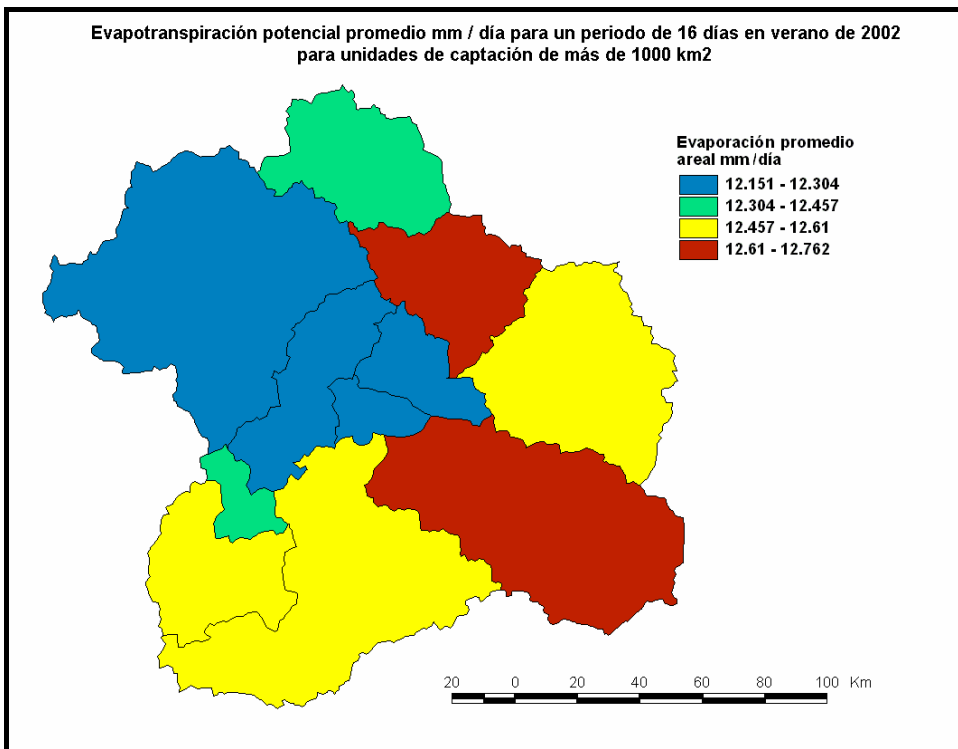
a)



b)



c)



d)

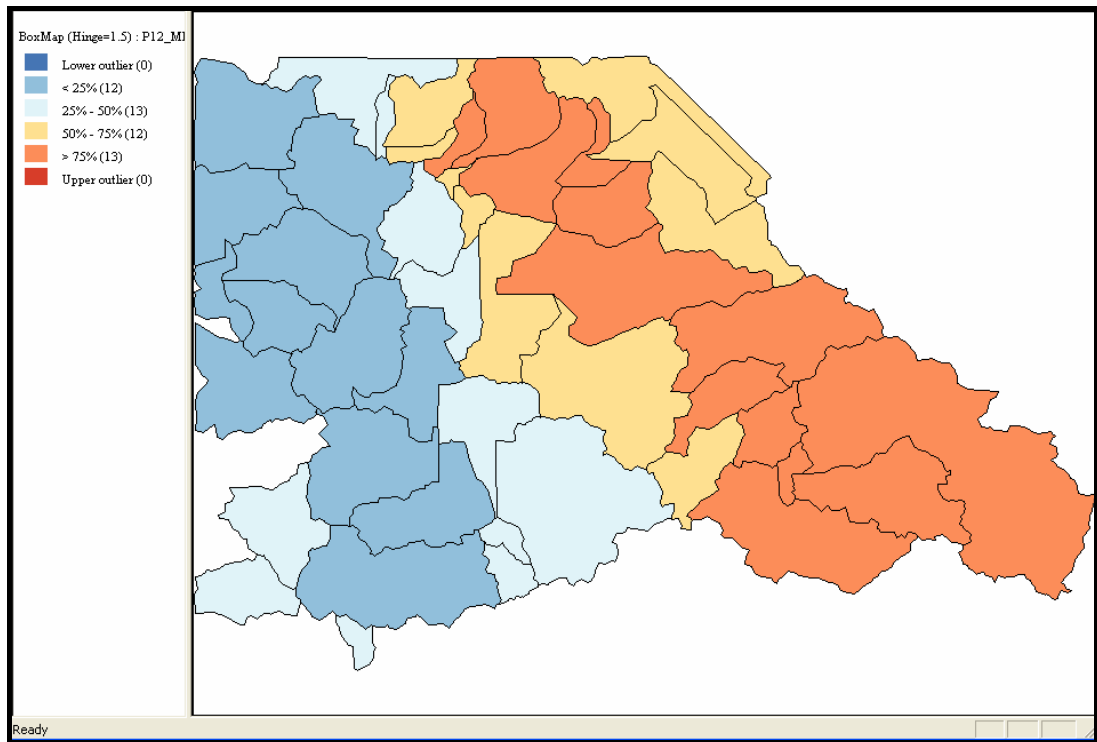
a) Precipitación, verano 2002, b) Evaporación, verano de 2002, c) Precipitación, verano de 2002 y d) Evaporación, verano de 2002. Los incisos a y b representan unidades mayores a 300 km<sup>2</sup> y los incisos c y d unidades mayores a 1000 km<sup>2</sup>.

Lo que nos dicen estos mapas no es comparable entre variables ya que los intervalos de clase difieren entre sí. Esto es debido a que no se está considerando en la segmentación de clases la frecuencia de cada una, lo cual permite estandarizar rangos de variación diversa a una escala de 0 a 100. Esto es lo que se hace de forma corriente en los diagramas de caja y brazo, donde se divide la variación total en clases con igual frecuencia y se representan los casos de tal manera que el primer cuarto o cuartil contiene el 25 % de los datos que presentan los valores más bajos. En orden ascendente se agrupa el siguiente 25 % de los datos y así sucesivamente hasta cubrir el total de los datos y de la variación. La comparación entre fechas y variables se hace válida en cuanto a la búsqueda de patrones ya que efectivamente comenzamos a delimitar posibles clases. Cuando asociamos esta herramienta a las unidades de análisis obtenemos lo que se conoce como mapas de caja.

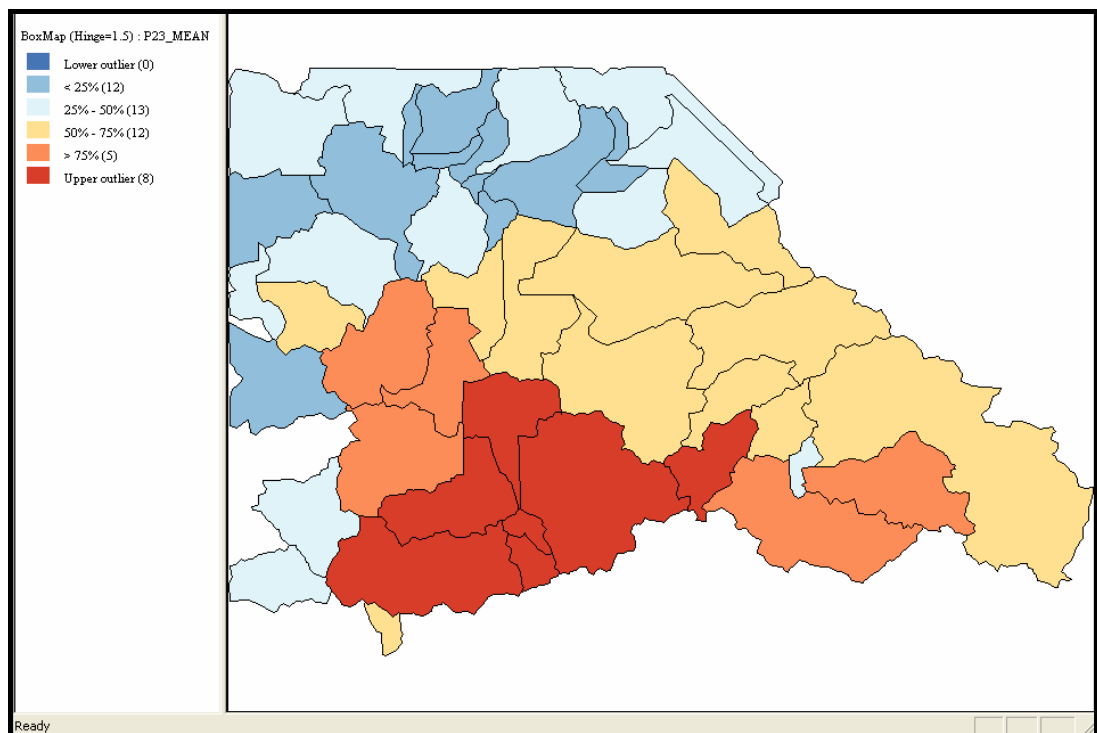
A continuación se presentan las correspondientes representaciones de los mapas anteriores a través de mapas de caja. En los mapas de las figuras 7.28a, 7.28b, 7.28c y 7.28d, se observan patrones mejor definidos ya que se manejan los 4 cuartiles usuales y se identifican outliers o valores extremos. Se pueden apreciar también las direcciones donde la autocorrelación espacial se ha identificado, estos mapas nos permiten construir hipótesis acerca del comportamiento de las variables.

En cuanto a los patrones observados a nivel de las unidades mayores de 300 km<sup>2</sup>, la precipitación en verano (ver mapa de la figura 7.28a) presenta dos grandes franjas en dirección noreste-sureste en los dos extremos del área, mientras que en la porción central no hay patrón bien definido. En invierno (figura 7.28b), el patrón sea hace menos regional y se observan dos grandes aglomerados al sur de la región uno de ellos también en la dirección noreste-sureste. El sur de la región es la zona que presenta mayor acumulación de lluvias. Lo que sigue, es comprobar que efectivamente estos patrones son válidos y significativos y que los clusters representan clases de precipitación discernibles entre sí.

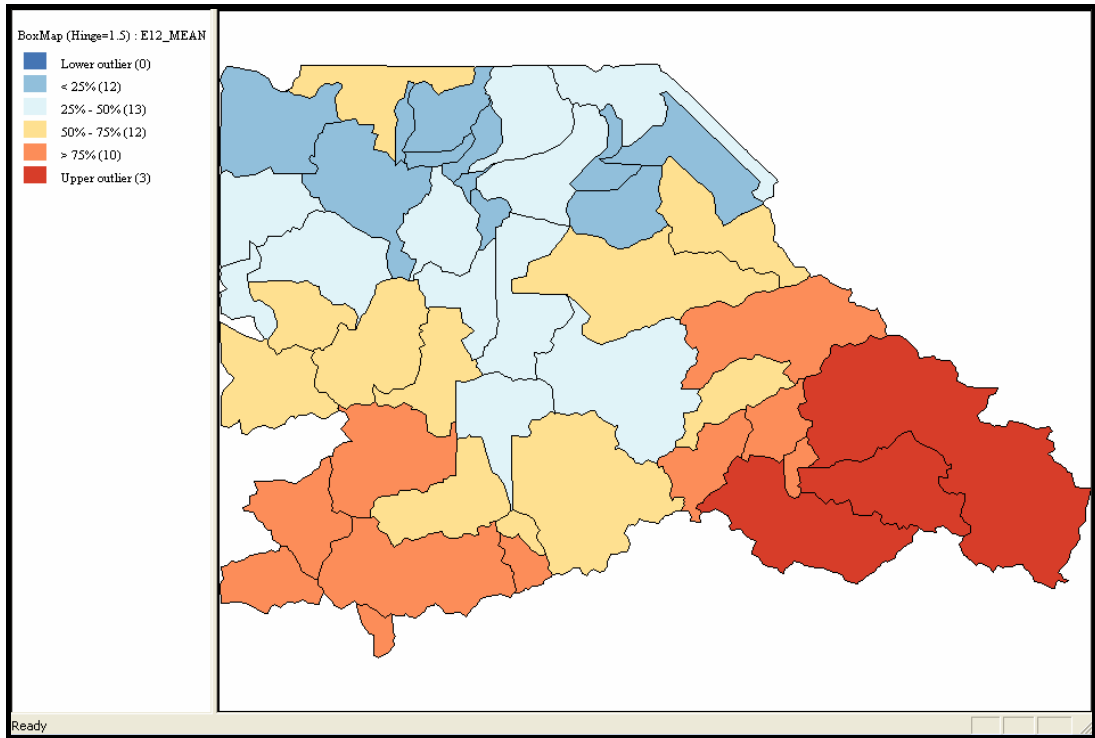
**Fig. 7.28 Mapas de caja de promedios areales para unidades de captación mayores a 300 km<sup>2</sup>.**



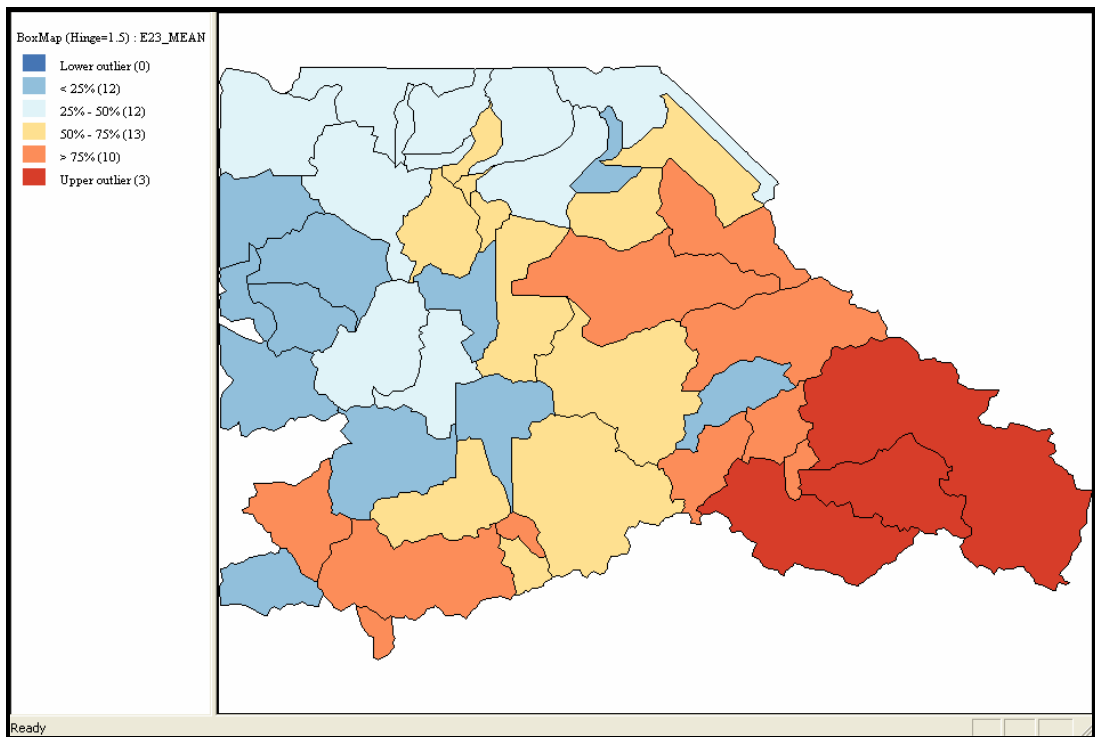
a)



b)



c)



d)

a) Precipitación en mm/16 días, verano 2002, b) Precipitación en mm/16 días, invierno 2002, c) Evapotranspiración potencial en mm/día, verano 2002, d) Evapotranspiración potencial en mm/día, invierno

2002. La simbología representa los cuatro cuartiles tradicionales y además, 2 cuartiles con valores extremos (por lo menos 1.5 veces menores o mayores que los valores extremos del primero y cuarto cuartil respectivamente).

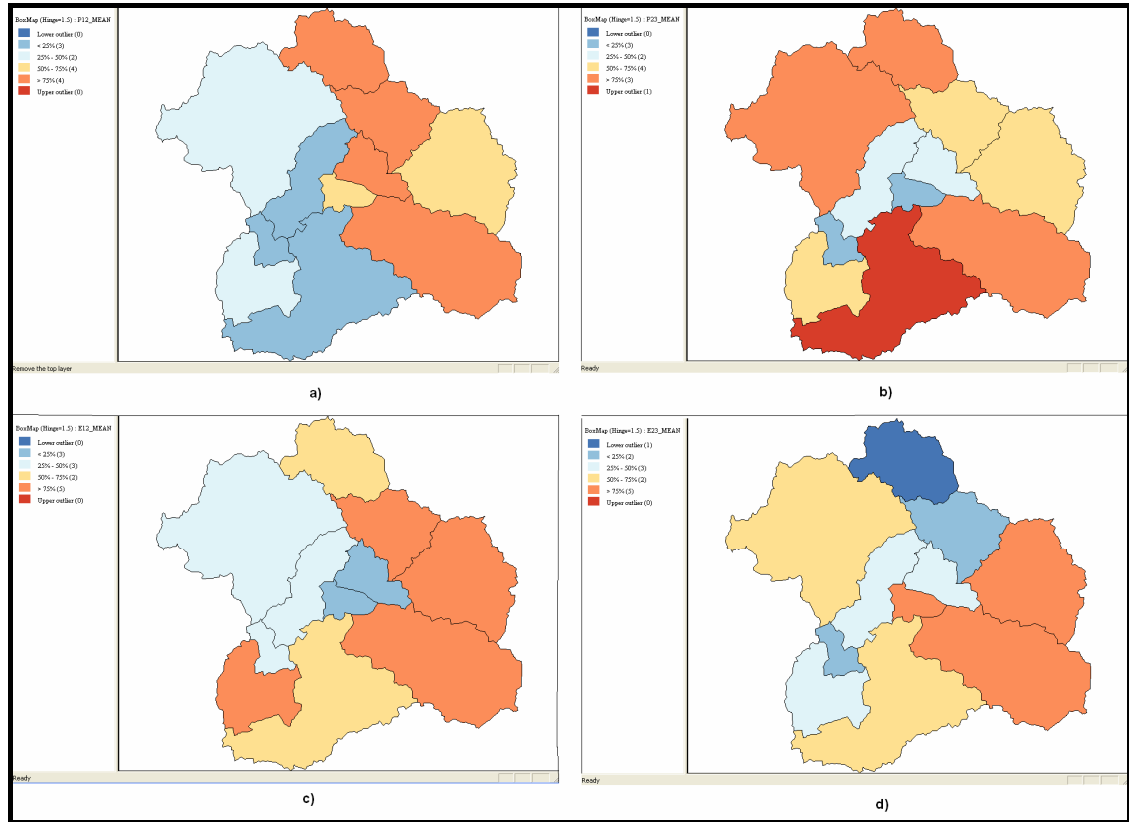
Respecto de la evapotranspiración potencial (ver figuras 7.28c y 7.28d), se observan patrones mucho más locales en dirección norte-sur principalmente, tal como se esperaba, según los parámetros que se obtuvieron con el ajuste a los variogramas experimentales en ambas fechas; aparentemente el efecto de la latitud se manifiesta claramente en los patrones identificados para la evaporación estimada. No hay diferencias discernibles a simple vista entre ambas fechas y se distinguen aglomerados al sur de la región principalmente. En los dos periodos se identifican valores máximos de evaporación promedio en la parte alta de la cuenca del Lerma al sureste de la región. Al igual que en el caso de la precipitación debemos probar si estos patrones son válidos o no.

Partimos del supuesto de que el efecto de la autocorrelación es lo suficientemente fuerte a nivel de las unidades de captación más pequeñas como para extraer patrones tipo clases o regiones discernibles entre sí, en las direcciones identificadas en los variogramas. No obstante parece haber algunas unidades de captación donde la variación es independiente de la estructura espacial en los propios componentes. Paralelamente sugerimos la misma visualización a nivel de unidades más grandes con el objetivo de definir un umbral de escala para la extracción de patrones en la precipitación y la evapotranspiración potencial explicados fundamentalmente por el espacio. Los mapas de caja correspondientes a las dos variables se presentan en la figura 7.29 y efectivamente se observan patrones, sin embargo, debemos probar que no se deben a la estructura espacial de las variables. Básicamente sugerimos que los promedios de precipitación y evaporación a nivel de estas unidades son independientes entre sí y que podemos enfocarnos en formalizar relaciones causa efecto sin la influencia de la autocorrelación espacial.

En otras palabras, para formalizar asociaciones de los componentes analizados, con posibles variables explicatorias, se requiere considerar la estructura o autocorrelación espacial como un factor de peso o diferenciación; esto, a nivel de unidades a partir de 300 km<sup>2</sup>, (escala mayor a 1: 800000, para representar una unidad de 300 km<sup>2</sup> con un área de 4 cm<sup>2</sup> en el mapa como mínimo). En cambio, si se emplean unidades mayores a 1000 km<sup>2</sup> (escala mayor a 1: 1500000 aprox.), no se tendría que considerar la autocorrelación

espacial en el proceso de parametrización de modelos hidrológicos a nivel de promedios areales.

**Fig. 7.29** Plantilla de mapas de caja para unidades de captación mayores a 1000 km<sup>2</sup>.



a) Precipitación en mm/16 días, verano, b) Precipitación en mm/16 días, invierno, c) Evapotranspiración potencial en mm/día, verano, d) Evapotranspiración potencial en mm/día, invierno. La simbología corresponde a la leyenda de los mapas de caja presentados anteriormente.

Además de la autocorrelación espacial de los componentes que configura patrones específicos, debemos considerar los procesos causales que también determinan patrones; estos procesos, como ya lo identificamos, presentan autocorrelación espacial y un dominio de escala propio. Por lo tanto, además de la diferenciación espacial debida a la estructura espacial de los componentes hay otro factor de diferenciación de carácter funcional que se expresa en forma de diversos grados de asociación entre variables en toda el área de estudio. Se propone establecer la diferenciación espacial de los componentes analizados a partir de regiones, las cuales sean resultado de una clasificación funcional, considerando la autocorrelación espacial y las relaciones de asociación entre componentes.

### **7.2.3b. Aportes de la fase de construcción de hipótesis.**

En esta fase se formulan generalizaciones acerca del comportamiento de los componentes del balance y su asociación con variables explicatorias a partir de unidades de captación. Esto, con el propósito de establecer diversos niveles de análisis correspondientes al dominio de escala de las variables dependientes analizadas y establecer cuál sería el tamaño adecuado de las unidades del paisaje. Para el estudio de caso se evaluaron dos escalas 1: 800000 y 1: 1500000, ambas variables parecen conformar patrones del tipo clases o regiones a la escala mayor, pero no hay suficientes evidencias de que se puedan conformar regiones por autocorrelación espacial a la escala más general. No obstante, se propone que debe haber además una diferenciación funcional o de asociación entre variables que contribuya también a la regionalización. Sin duda, se hace más evidente la presencia de patrones mediante los mapas de caja, sin embargo, fue fundamental el análisis previo para definir a priori el detalle de la regionalización hidrológica por áreas de drenaje. Los datos resultantes de los análisis en esta etapa corresponden a un nivel de información IV, el cual implica, la representación de los procesos como entidades espaciales del tipo regiones y la sumariación de sus atributos para comprender mejor su comportamiento.

### **7.2.4. Análisis confirmatorio**

En esta etapa proseguimos con la formalización de relaciones causales o **asociaciones** y la extracción de **regiones**. Definimos herramientas que nos permitan probar la efectividad de los modelos aplicados para describir el comportamiento de los objetos espaciales que estudiamos. Significa que debemos probar las hipótesis derivadas de la etapa anterior y obtener como resultado la regionalización funcional de los componentes del balance hídrico analizados.

En esta fase se proponen herramientas de geovisualización y minería de datos útiles para la extracción de patrones del tipo regiones y se emplea nuevamente el método de regresión parcial para formalizar las asociaciones entre variables explicatorias y los componentes del balance hídrico a nivel de regiones.

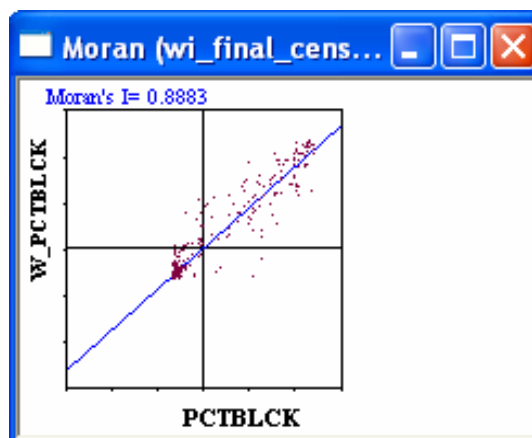


### 7.2.4a. Extracción de patrones explicados por autocorrelación espacial de los componentes del balance hídrico.

El procedimiento general que a continuación se describe es aplicable a cualquier proceso geográfico en el que se requiera extraer patrones debidos a la autocorrelación espacial de dicho fenómeno. Este procedimiento se modificó del propuesto por Anselin, Syabri y Kho (2004:2-18). Como vimos, una primera evidencia de la presencia de configuración espacio-temporal es la observación de clases mediante mapas de caja, herramienta que facilita la comparación entre diferentes periodos de tiempo y variables, sin importar la magnitud y escala de medición de éstas.

El primer paso en la extracción de regiones es la definición de una matriz de pesos basados en la vecindad de cada unidad de observación o bien en la distancia de separación entre registros. En esta matriz se asigna el valor más probable para la variable de acuerdo a los valores de sus vecinos más cercanos o de los vecinos que se encuentren en el rango definido como umbral por el analista. Una vez definida esta matriz de pesos espaciales, se grafican los eigenvalores correspondientes a cada registro respecto de los valores observados (ver figura 7.30). En el mismo gráfico se ajusta una línea, cuya pendiente es el Índice de autocorrelación de Moran que ya conocemos. Recordemos que se generalizó el comportamiento de las variables a nivel de unidades de captación mediante el promedio areal de acumulados de precipitación y evaporación, sin embargo esperamos que la autocorrelación espacial siga siendo significativa al primer nivel de agregación de datos (escala 1:800000).

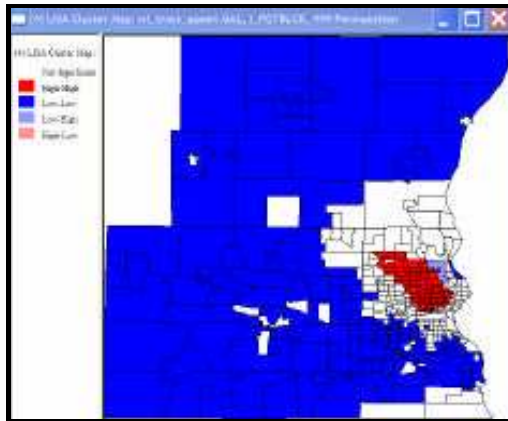
**Fig. 7.30 Diagrama de dispersión de Moran.**



Tomado de Anselin (2005:133).

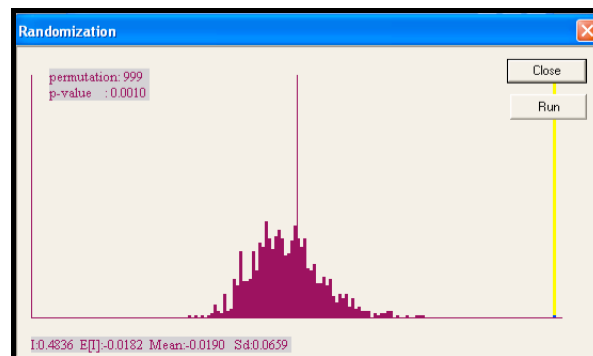
En un segundo paso se incluye la asociación del diagrama de dispersión de Morán a un mapa de autocorrelación espacial donde se resaltan aquellas unidades de observación que están correlacionadas a un nivel de significancia mínimo de 0.05, es decir con un 95% de probabilidad de que el supuesto de autocorrelación espacial es verdadero. De esta forma, se pueden observar porciones del área de estudio en forma de regiones (figura 7.31). Se ofrece además, una prueba de significancia del índice  $I$  en la que se corre un número máximo de permutaciones para obtener una distribución de probabilidad correspondiente a un proceso aleatorio independiente del espacio (figura 7.32). Esta distribución se compara con el parámetro de autocorrelación para observar si el valor del índice se encuentra cerca del valor esperado  $E$  del proceso aleatorio o si se ubica en los rangos correspondientes a las colas del histograma, fuera del 95 % de los valores probables. Lógicamente si ocurre lo último se acepta la hipótesis de autocorrelación espacial.

**Fig. 7.31 Mapa de autocorrelación espacial.**



Tomado de Anselin (2005:142).

**Fig. 7.32 Distribución del índice de Moran respecto de una distribución aleatoria.**



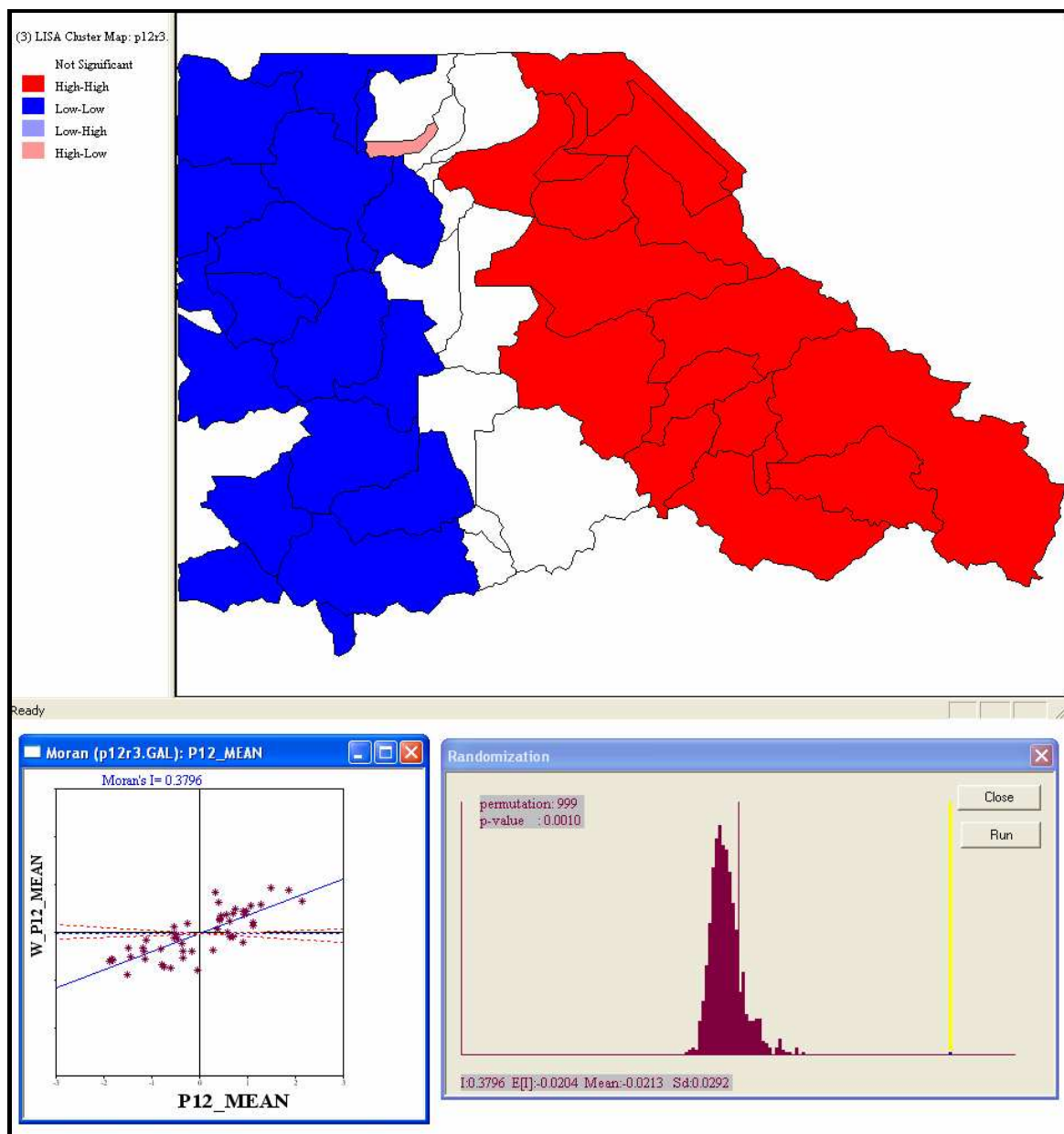
Tomado de Anselin (2005:135).

Los mapas de autocorrelación para la precipitación y la evaporación a nivel de unidades de más de 300 km<sup>2</sup> se obtuvieron empleando el software Geoda v. 0.9.5-i (ver figuras 7.33, 7.34, 7.35 y 7.36). Como se dijo antes, se resaltan aquellas unidades donde el efecto de la autocorrelación es significativo. En las regiones que aparecen en rojo o azul oscuro, la autocorrelación es positiva, asimismo, pueden aparecer patrones de autocorrelación espacial negativa donde algunas unidades se colorean de un color diferente al de sus vecinos, este caso se interpreta como el límite de una clase respecto de otra cuya media es inversamente proporcional a la primera. En los mapas de autocorrelación estos casos se representan con color rosa o azul claro y usualmente corresponden a observaciones aisladas o raras. El resto de las unidades no se resalta y la interpretación que damos es que no hay un patrón explicado por la estructura espacial pero las unidades pueden clasificarse por cualquier método tradicional de clasificación.

Además de los mapas de autocorrelación se incluye en las figuras el diagrama de Morán con sus respectivas pruebas de significancia. Como vemos en la figura 7.33, se observan dos grandes regiones para la precipitación acumulada, que coinciden con la dirección noroeste-sureste y su perpendicular, lo cual demuestra que efectivamente es posible discriminar dos grandes regiones a nivel de las unidades de más de 300 km<sup>2</sup>. El nivel de significancia de estos patrones es 0.01 por lo que en la regionalización se debe considerar la autocorrelación espacial de la precipitación. En el periodo de invierno (ver figura 7.34), también se distinguen dos regiones donde la autocorrelación es positiva y significativa al nivel  $p = 0.01$ , no obstante, aquí es menor el rango por lo que las dos regiones son más pequeñas y se observa además una tercera región con autocorrelación negativa.

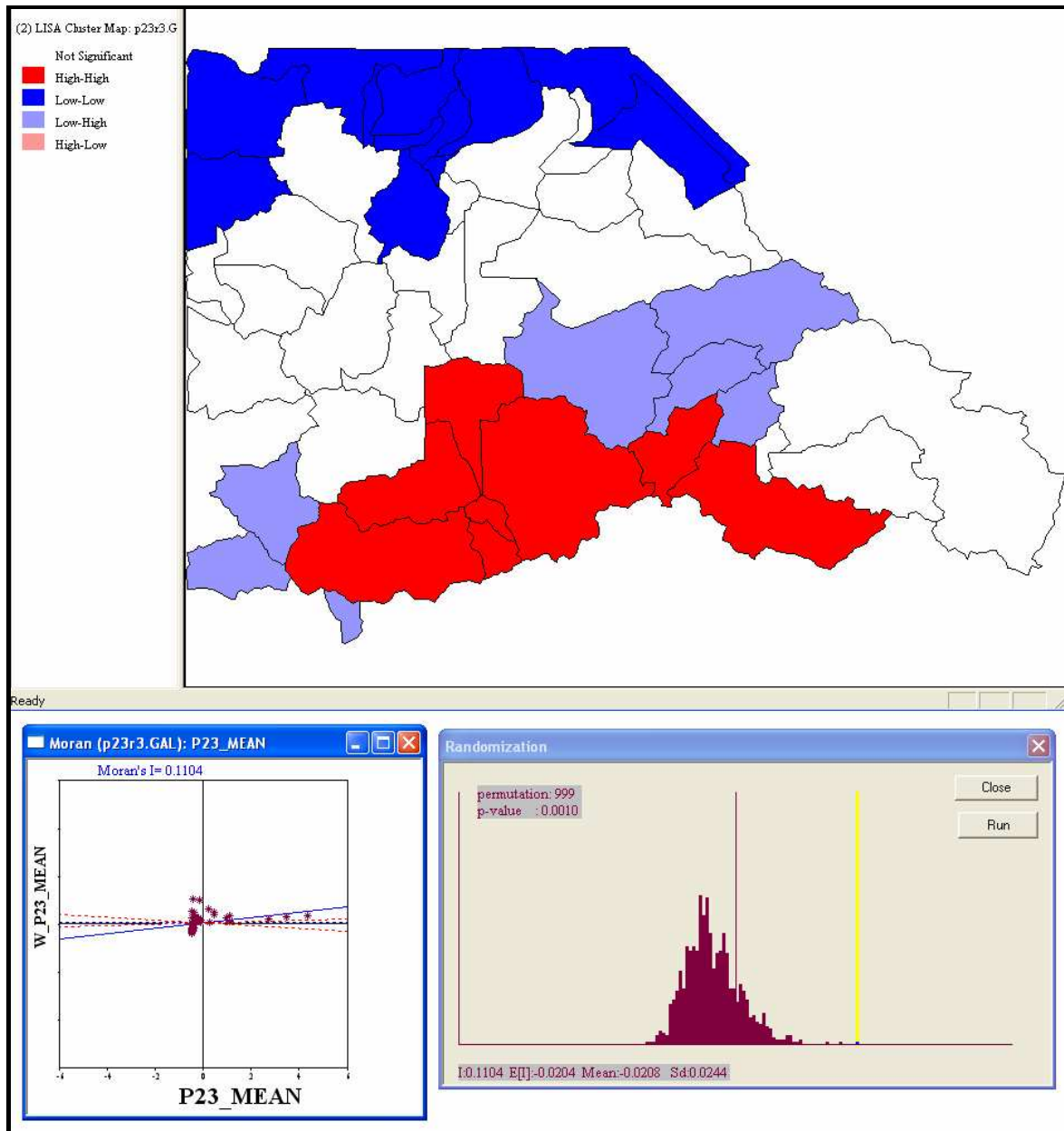
Respecto de la evapotranspiración potencial (ver figuras 7.35 y 7.36), se observan patrones por autocorrelación espacial muy similares en dirección noroeste-sureste principalmente. Estos resultados nos indican que la diferenciación en regiones se verifica principalmente a nivel de la magnitud de la variable entre los dos periodos.

**Fig. 7.33 Mapa de autocorrelación espacial con regiones de precipitación acumulada, unidades mayores a 300 km<sup>2</sup>, verano 2002.**



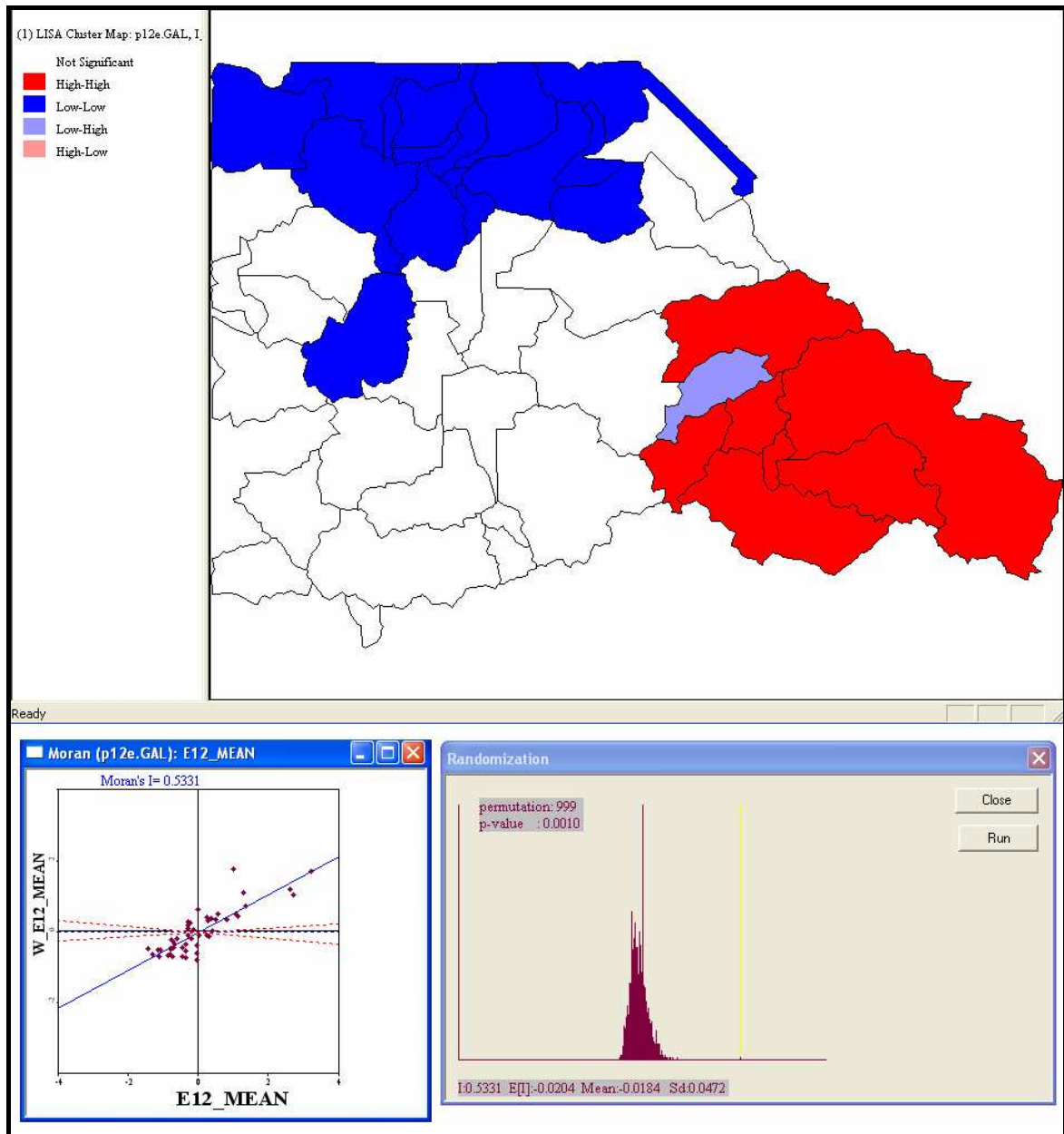
Se anexa el diagrama del índice Moran mostrando la dispersión de la variable respecto a promedios ponderados por distancia de la misma variable. Las líneas punteadas indican los límites de una distribución aleatoria (percentiles 2.5 a 92.5) en la que cualquier punto coincidente tiene un comportamiento aleatorio; claramente, la línea de ajuste (en azul) queda fuera de esta área. Otra forma de representación de la validez del supuesto de autocorrelación es la comparación del parámetro  $I$  con un proceso aleatorio puro donde el valor esperado  $E$  difiere significativamente (a un nivel  $p$  de 0.001) del valor del índice.

**Fig. 7.34 Mapa de autocorrelación espacial con regiones de precipitación acumulada, unidades mayores a 300 km<sup>2</sup>, invierno de 2002.**



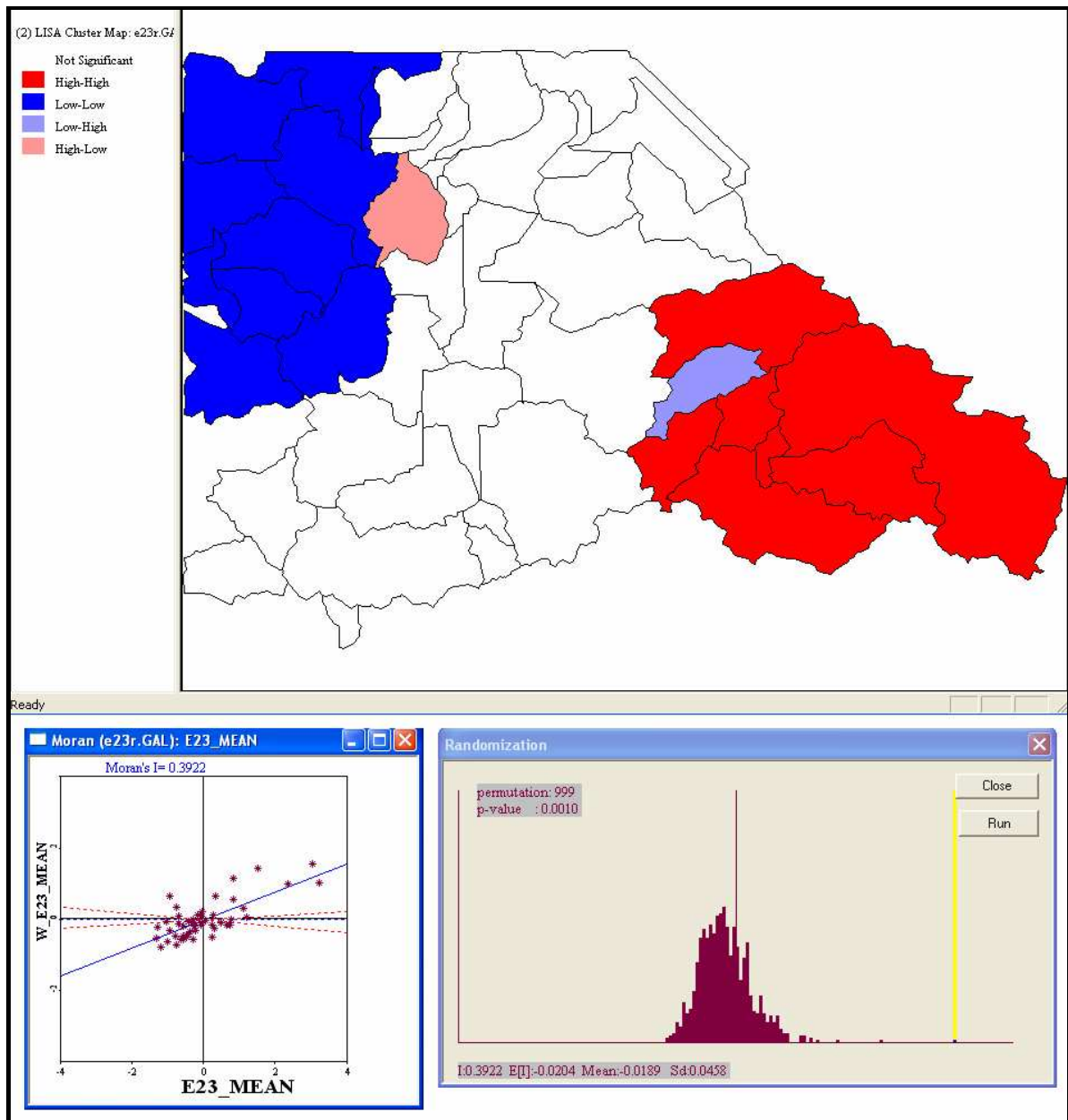
Se anexa el diagrama del índice Moran mostrando la dispersión de la variable respecto a promedios ponderados por distancia de la misma variable. Las líneas punteadas indican los límites de una distribución aleatoria (percentiles 2.5 a 92.5) en la que cualquier punto coincidente tiene un comportamiento aleatorio; claramente, la línea de ajuste queda fuera de esta área. Otra forma de representación de la validez del supuesto de autocorrelación es la comparación del parámetro  $I$  con un proceso aleatorio puro donde el valor esperado  $E$  difiere significativamente (a un nivel  $p$  de 0.001) del valor del índice.

**Fig. 7.35 Mapa de autocorrelación espacial con regiones de evapotranspiración potencial promedio diario, unidades mayores a 300 km<sup>2</sup>, verano de 2002.**



Se anexa el diagrama del índice Moran mostrando la dispersión de la variable respecto a promedios ponderados por distancia de la misma variable. Las líneas punteadas indican los límites de una distribución aleatoria (percentiles 2.5 a 92.5) en la que cualquier punto coincidente tiene un comportamiento aleatorio; claramente, la línea de ajuste (en azul) queda fuera de esta área. Otra forma de representación de la validez del supuesto de autocorrelación es la comparación del parámetro  $I$  con un proceso aleatorio puro donde el valor esperado  $E$  difiere significativamente (a un nivel  $p$  de 0.001) del valor del índice.

**Fig. 7.36 Mapa de autocorrelación espacial con regiones de evapotranspiración potencial promedio diario, unidades mayores a 300 km<sup>2</sup>, invierno de 2002.**



Se anexa el diagrama del índice Moran mostrando la dispersión de la variable respecto a promedios ponderados por distancia de la misma variable. Las líneas punteadas indican los límites de una distribución aleatoria (percentiles 2.5 a 92.5) en la que cualquier punto coincidente tiene un comportamiento aleatorio; claramente, la línea de ajuste (en azul) queda fuera de esta área. Otra forma de representación de la validez del supuesto de autocorrelación es la comparación del parámetro  $I$  con un proceso aleatorio puro donde el valor esperado  $E$  difiere significativamente (a un nivel  $p$  de 0.001) del valor del índice.

Sin duda, el efecto de la latitud sobre la evaporación estimada por el método de Thornthwaite es importante, ya que se homogeneizó la variabilidad intrínseca de la temperatura del aire (empleada en la estimación de la evaporación), de tal manera que no

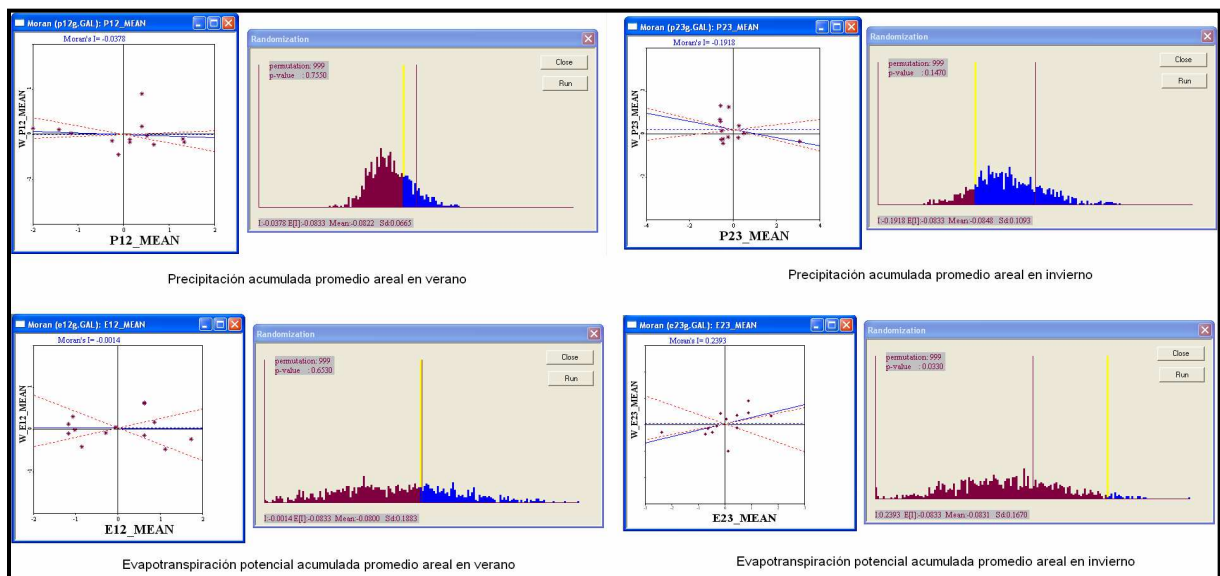


encontramos patrones significativamente diferentes de evapotranspiración potencial entre ambos periodos. Recordemos que la función de estructura o variograma de la temperatura del aire es específica para cada periodo (ver tabla 7.8).

Los resultados confirman la presencia de autocorrelación espacial en la precipitación y la evaporación a nivel de las unidades mayores a 300 km<sup>2</sup>. Por ello, se debe considerar la estructura espacial como una variable más en el proceso de calibración de modelos hidrológicos cuando se trabaje con unidades de hasta 1000 km<sup>2</sup> en el área de estudio, siempre que se identifique que el dominio de escala de las variables supera los 200 km.

En contraparte, en la figura 7.37 se muestran los resultados del análisis para verificar la presencia de estructura espacial en la precipitación y la evaporación a nivel de las unidades mayores. No se obtuvieron regiones por autocorrelación espacial a esta escala de análisis, por esta razón, sólo se presentan los diagramas de dispersión de Morán y la prueba de significancia al nivel  $p = 0.001$ , en la figura 7.37. Según esta última prueba, el valor del índice  $I$ , se acerca al valor de  $E$  o queda dentro de los 2 cuartiles intermedios en el histograma.

**Fig. 7.37 Diagramas de dispersión del índice Moran para precipitación y para evapotranspiración vs. promedios ponderados por distancia.**



Análisis a nivel de unidades mayores a 1000 km<sup>2</sup>. Las líneas punteadas indican los límites de una distribución aleatoria (percentiles 2.5 a 92.5) en la que cualquier punto coincidente tiene un comportamiento aleatorio. Claramente, la línea de ajuste (en azul) queda dentro de esta área. Otra forma de representación de la validez del



supuesto de autocorrelación espacial es la comparación del parámetro  $I$  con un proceso aleatorio puro donde el valor esperado  $E$  debería diferir (a un nivel  $p$  de 0.001) del valor del índice.

Estos resultados apoyan la hipótesis de que la estructura espacial en las variables delimita la extracción de patrones por autocorrelación a un dominio de escala. Sin embargo, el espacio no es el único proceso que define patrones del tipo regiones, sabemos que el comportamiento de la precipitación y la evaporación obedece además a procesos causales de diferente naturaleza, los cuales, igualmente, presentan un dominio de escala. Tales relaciones causales o de asociación entre variables establecen el carácter funcional o las propiedades corológicas en las unidades del paisaje y definen también patrones. Esta particularidad de los procesos geográficos nos obliga a realizar una diferenciación estructural y funcional del área de estudio.

#### **7.2.4b. Conformación de regiones mediante una clasificación restringida espacial y funcionalmente.**

En esta sección se describe el procedimiento general para generar una regionalización que considere la diferenciación funcional y la estructura de los componentes del balance hídrico analizados. Conocemos las regiones que se conforman por autocorrelación espacial para la precipitación y la evaporación, por lo que el primer criterio de aglomeración de unidades es el de autocorrelación espacial, el segundo criterio, igual de importante, es el de maximizar la diferencia entre clases y disminuir la variación dentro de las clases, este criterio es utilizado en el algoritmo denominado K-means, el cual emplea un método aglomerativo para conjuntar los casos en clases.

El principal problema en toda clasificación es la definición de clases a través de funciones discriminantes, firmas o descriptores de clases, los más empleados son la media y la desviación estándar. La determinación del número esperado de clases es fundamental y de ello depende el resultado de la clasificación. Aquí se parte de la conformación de regiones por autocorrelación espacial, los resultados de la sección anterior muestran que se pueden diferenciar para cada componente al menos tres clases: las dos que conforman patrones espaciales y que corresponden a las dos clases extremas de la variación de los procesos y una clase intermedia. Para la evaporación se sugiere que sean cuatro pues las regiones debidas a autocorrelación espacial son más pequeñas y una clase intermedia resultaría muy dispersa, sin embargo, para propósitos de ilustrar el proceso y no complicar

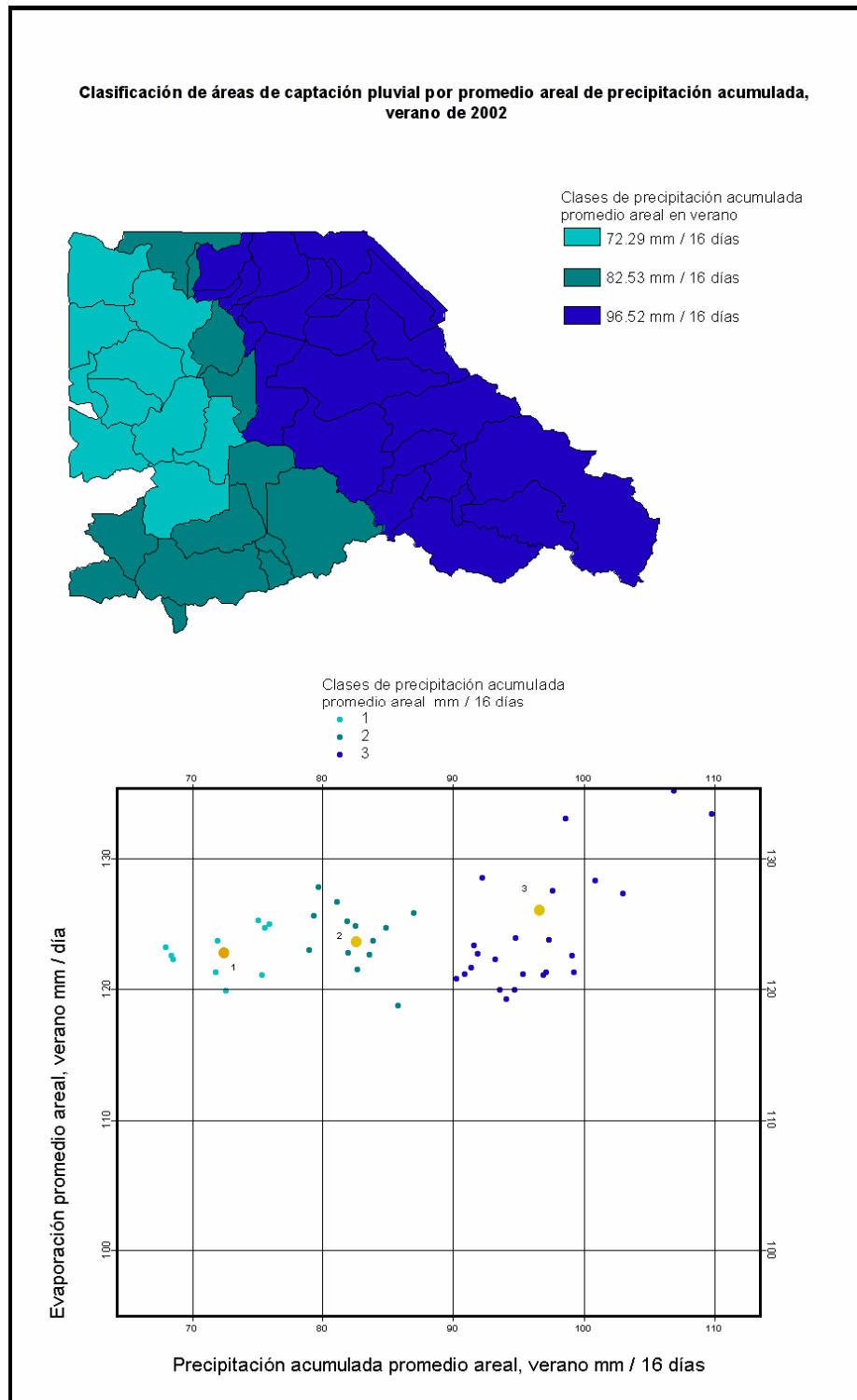
demasiado posteriores análisis, se procede a construir también sólo tres clases de evaporación.

El proceso general para la asignación de clases consiste en conservar juntas aquéllas unidades que exhiben autocorrelación espacial entre sí, pero se separan en caso de que el segundo criterio de maximización de diferencias entre clases no se cumpla. De esta forma, se obtienen las regiones que se ilustran en las figuras 7.38, 7.39, 7.40 y 7.41, se añade además el diagrama de dispersión de las unidades, en el espacio conformado por los promedios areales de precipitación y evaporación. Asimismo, se indica la posición de los centroides de cada clase.

De acuerdo con las figuras 7.38 y 7.39, la precipitación presenta patrones del tipo regiones, específicos para cada periodo analizado, tanto el tamaño como la dirección difieren, así como la magnitud del promedio de clases. Respecto de la evaporación (figuras 7.40 y 7.41), los patrones resultan ser similares para ambos periodos, la magnitud del promedio de clases es la única diferencia. Esto es debido a que se obtuvieron sólo tres clases y las regiones por estructura espacial son bastante parecidas. Además, aún debemos confirmar la presencia de diferenciación funcional.

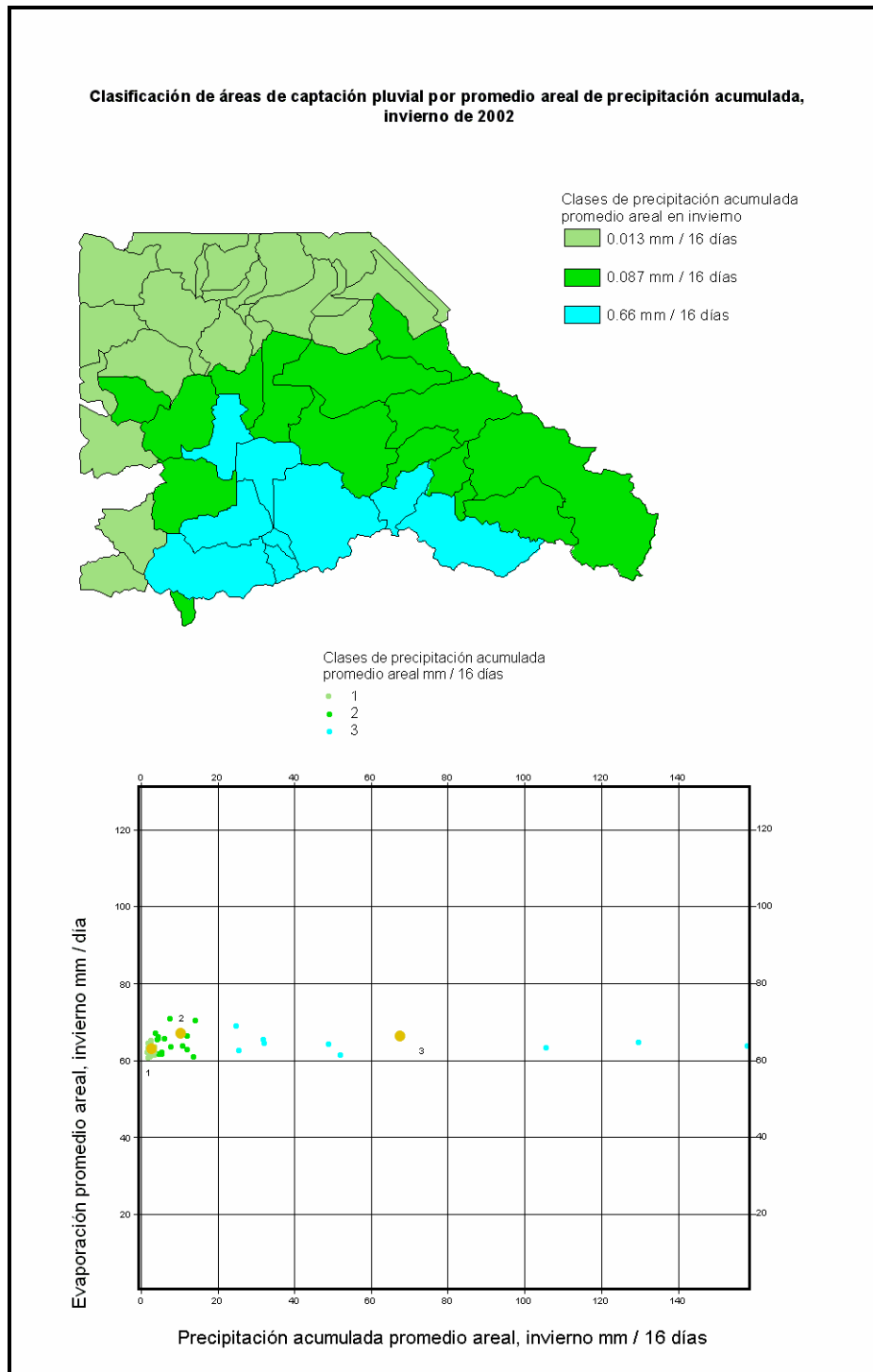
Observamos en los cuatro mapas que los elementos de cada región son contiguos entre sí y se observan gradientes en las direcciones esperadas. La principal razón de este resultado es que se aplicó una regionalización que considera la estructura espacial y los parámetros de variografía. En cambio, una clasificación típica que no considera la autocorrelación espacial o cuyos casos no presentan autocorrelación significativa conformará regiones cuyos elementos no necesariamente cumplen con las características mencionadas.

**Fig. 7.38 Mapa de clases de precipitación acumulada, verano de 2002.**



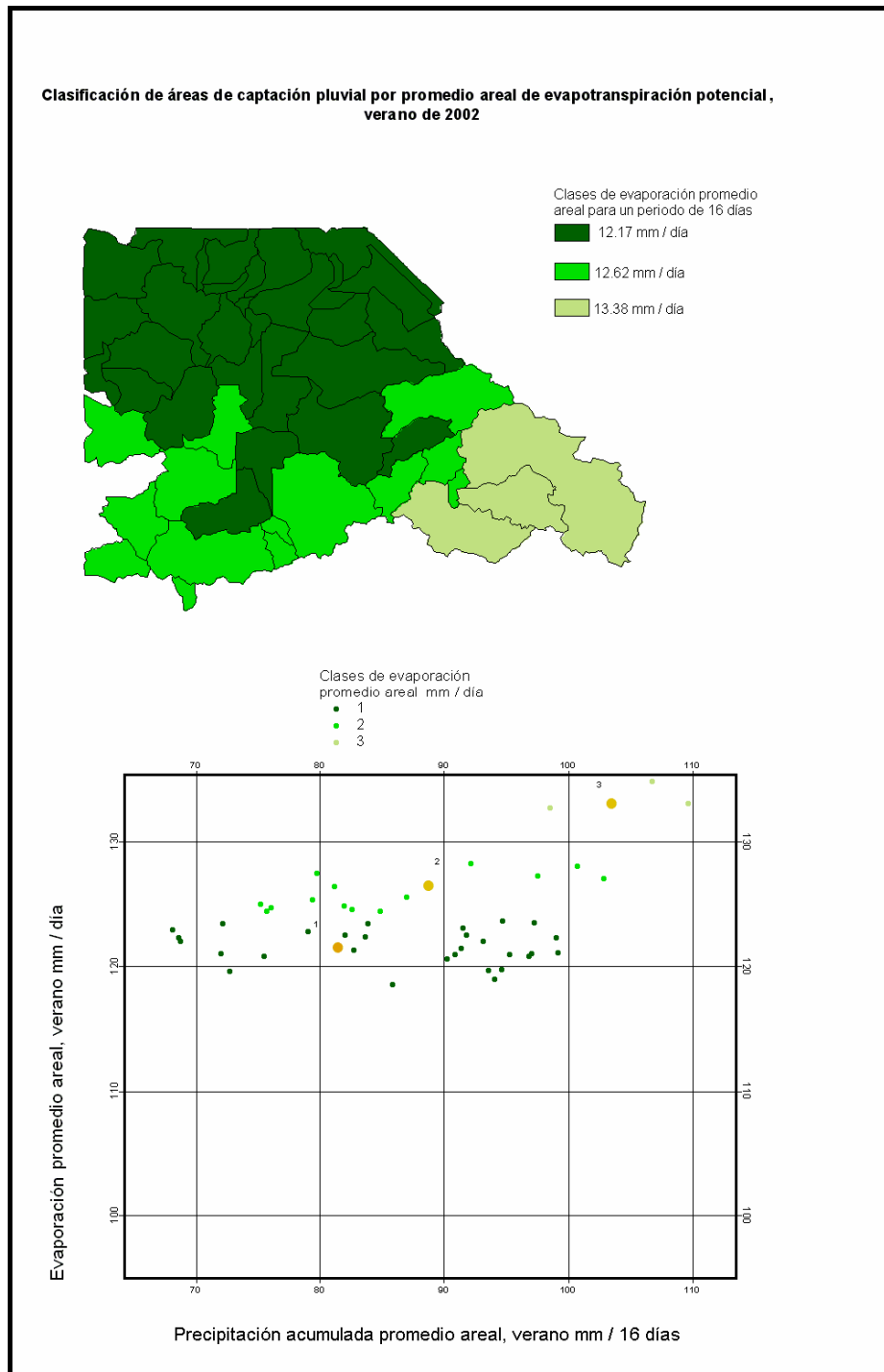
El diagrama de dispersión muestra la posición de las unidades correspondientes a cada clase en el espacio precipitación- evaporación, en naranja se presentan los centroides de las clases correspondientes al mapa. Las unidades de evaporación se escalan a 10X para hacer comparables las escalas. Las unidades hidrológicas son mayores a 300 km<sup>2</sup>.

**Fig. 7.39 Mapa de clases de precipitación acumulada promedio areal, invierno de 2002.**



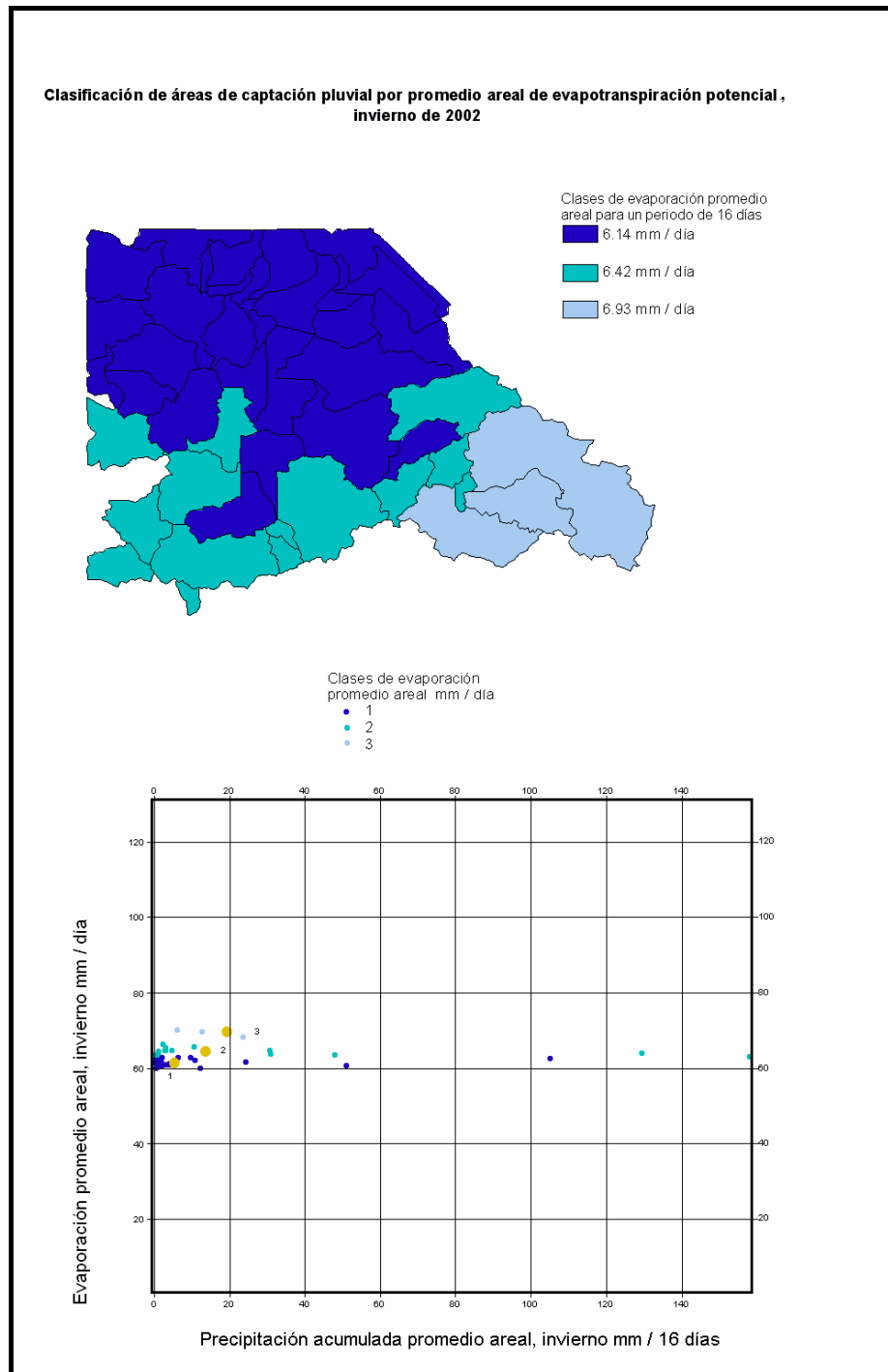
El diagrama de dispersión muestra la posición de las unidades correspondientes a cada clase en el espacio precipitación-evaporación, en naranja se presentan los centroides de las clases correspondientes al mapa. Las unidades de evaporación se escalan a 10X y las de precipitación a 100X para hacerlas comparables. Las unidades hidrológicas son mayores a 300 km<sup>2</sup>.

**Fig. 7.40 Mapa de clases de evaporación promedio areal, verano de 2002.**



El diagrama de dispersión muestra la posición de las unidades correspondientes a cada clase en el espacio precipitación-evaporación, en naranja se presentan los centroides de las clases correspondientes al mapa. Las unidades de evaporación se escalan a 10X para hacer comparables las escalas. Las unidades hidrológicas son mayores a 300 km<sup>2</sup>.

**Fig. 7.41 Mapa de clases de evaporación promedio areal, invierno de 2002.**

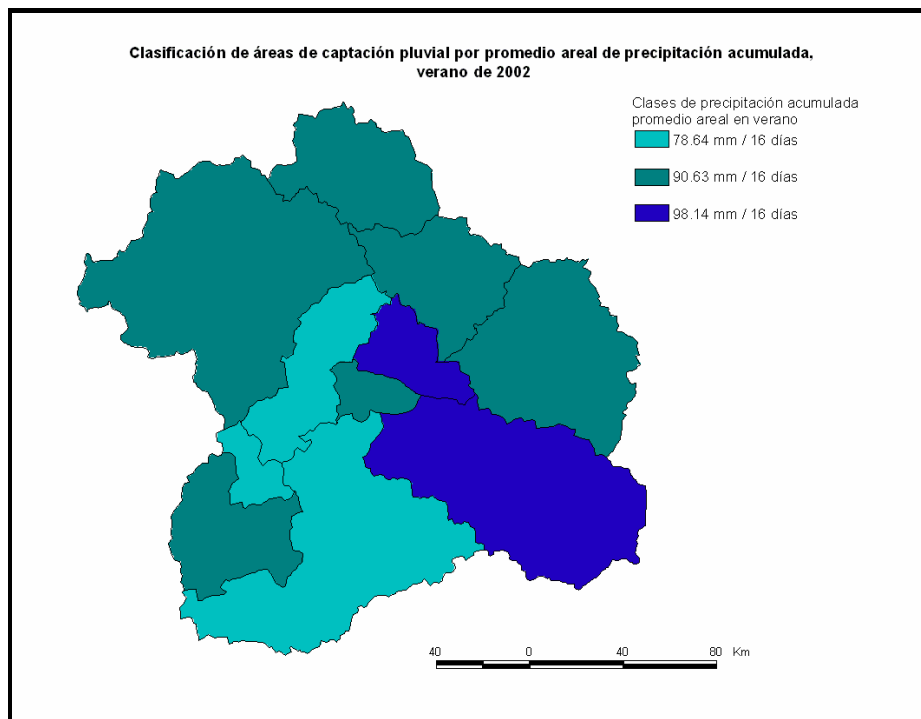


El diagrama de dispersión muestra la posición de las unidades correspondientes a cada clase en el espacio precipitación evaporación, en naranja se presentan los centroides de las clases correspondientes al mapa. Las unidades de evaporación se escalan a 10X y las de precipitación a 100X para hacerlas comparables. Las unidades hidrológicas son mayores a 300 km<sup>2</sup>.

En las unidades hidrológicas mayores se regionaliza únicamente con el criterio del algoritmo K-means. Se obtienen también 3 clases para cada componente y periodo (ver figuras 7.42, 7.43, 7.44, 7.45).

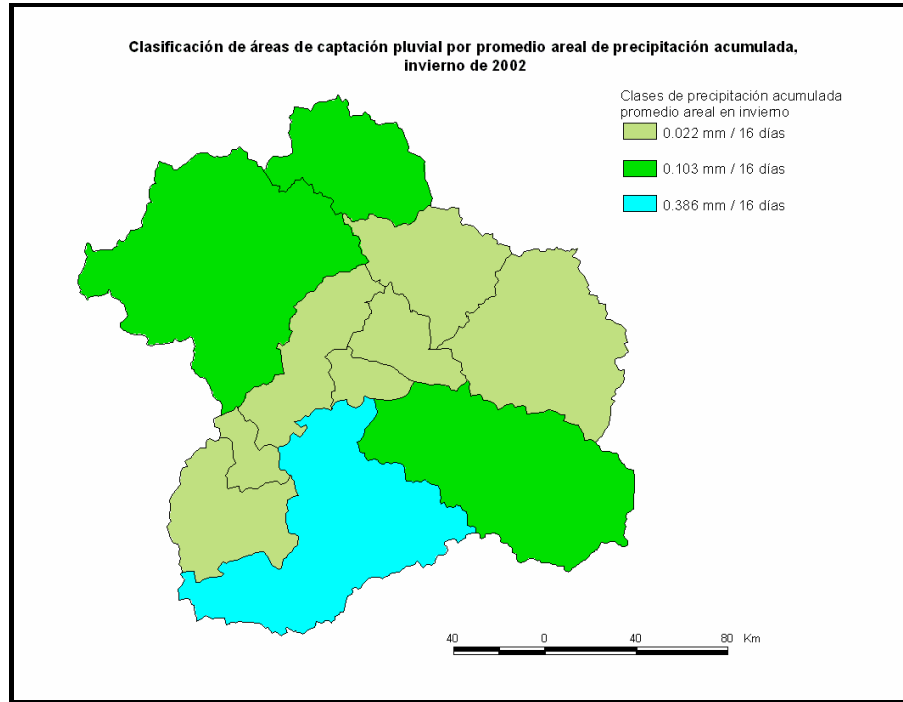
Se añaden las escalas gráficas en estos mapas para comparar las dimensiones de clases con el dominio de escala de los componentes. Los resultados reflejan lo que esperábamos en términos generales, algunas regiones no están formadas por elementos contiguos y las clases no se arreglan en gradientes. La diferencia entre ambos niveles, es el efecto de la autocorrelación espacial en la conformación de regiones. Debido a que se utilizan los mismos datos, las clases están caracterizadas por promedios muy similares a los promedios de las unidades mayores a 300 km<sup>2</sup>.

**Fig. 7.42 Mapa de clases de precipitación acumulada promedio areal, verano de 2002, unidades mayores de 1000 km<sup>2</sup>.**

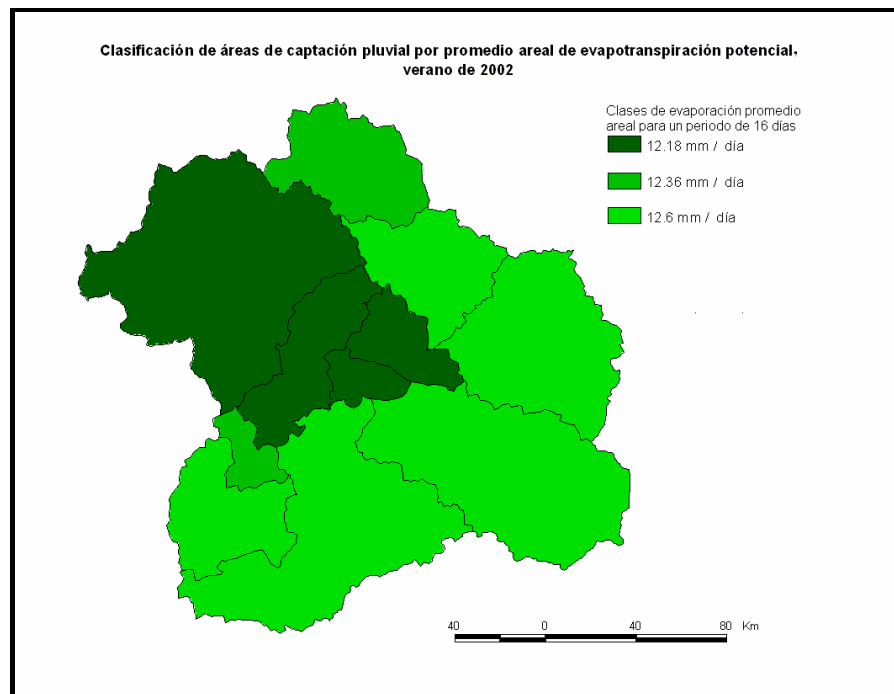


A este nivel de agregación, la autocorrelación espacial de la precipitación no es significativa en la clasificación.

**Fig. 7.43 Mapa de clases de precipitación acumulada promedio areal, invierno de 2002, unidades mayores de 1000 km<sup>2</sup>.**

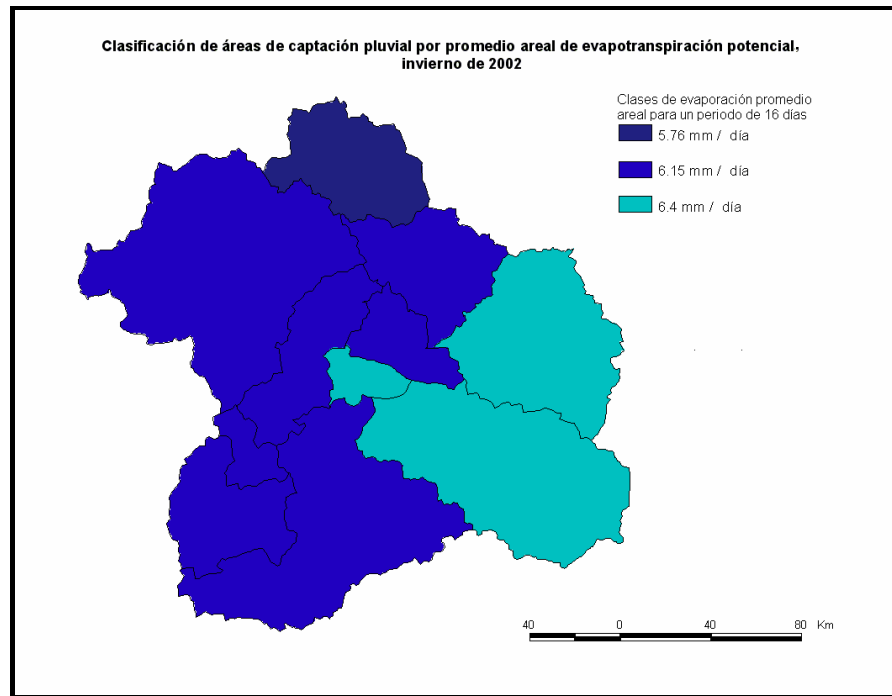


**Fig. 7.44 Mapa de clases de evaporación promedio areal, verano de 2002, unidades mayores de 1000 km<sup>2</sup>.**





**Fig. 7.45 Mapa de clases de evaporación promedio areal, invierno de 2002, para unidades mayores de 1000 km<sup>2</sup>.**



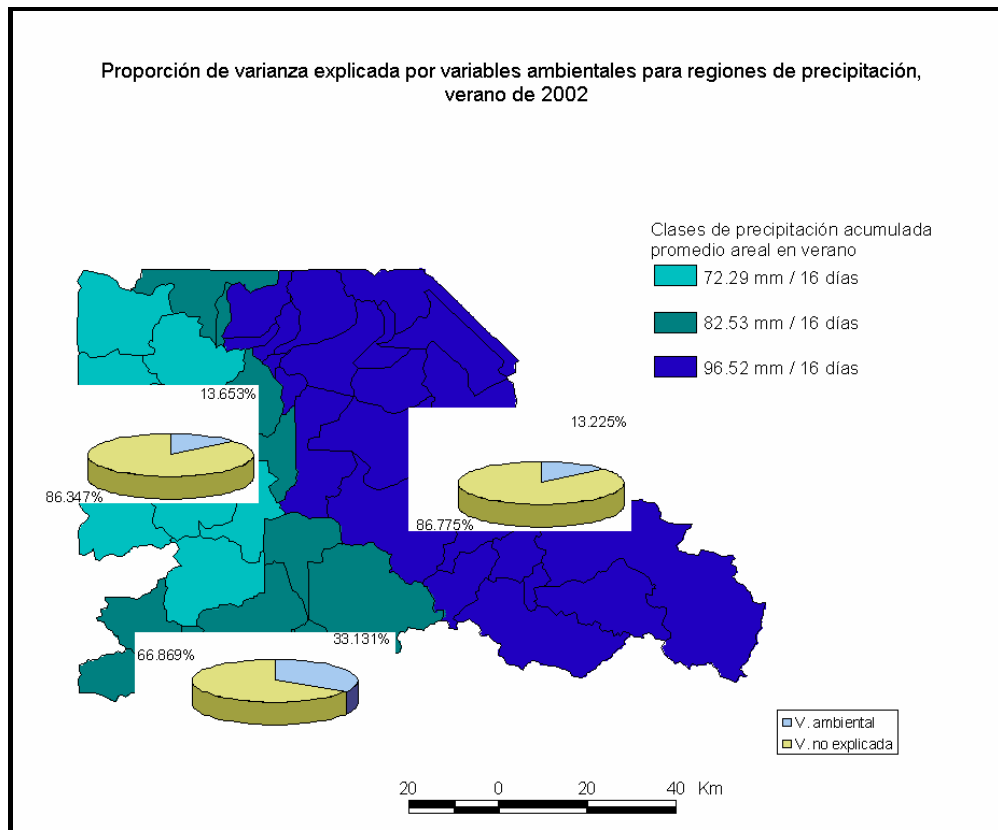
En cuanto a la diferenciación funcional, partimos del supuesto de que la autocorrelación espacial de las variables es importante. De esta forma, aquellas regiones donde la autocorrelación espacial es significativa, diferirán en forma importante en cuanto al grado de asociación entre los componentes analizados y las variables causales, respecto de aquellas regiones donde la autocorrelación espacial no es significativa.

Otro aspecto a considerar es la forma como se estiman las regresiones, si se emplean promedios areales para formalizar relaciones, tal como se hace en los estudios hidrológicos tradicionales, la diferenciación funcional disminuye drásticamente por el efecto de agregación o sumarización. Esta simplificación de la variabilidad de los componentes y sus formas de asociación es cuestionable, especialmente en estudios de diagnóstico y monitoreo. Por esta razón, el análisis de regresión para formalizar relaciones se realiza con variables distribuidas y no con promedios areales. El dominio de escala es otro factor de diferenciación funcional que aplica sólo cuando se realizan análisis distribuidos, esto significa que aún a nivel de unidades mayores de 1000 km<sup>2</sup> será posible observar la diferenciación funcional, ya que, como sabemos, los dominios de escala de las

variables difieren entre sí. Sin embargo, si existen coincidencias en los parámetros de variografía entre variables, la diferenciación funcional disminuirá.

A continuación, se ilustran las proporciones de variación explicada por variables causales para los componentes analizados. Se empleó un análisis de regresión múltiple para eliminar efectos entre variables explicatorias. En la figura 7.46 se observa la diferenciación funcional entre las regiones de precipitación acumulada con autocorrelación espacial significativa (clases 1 y 3) y la región central.

**Fig. 7.46 Mapa de proporción de varianza explicada por variación ambiental para regiones de precipitación acumulada en verano.**

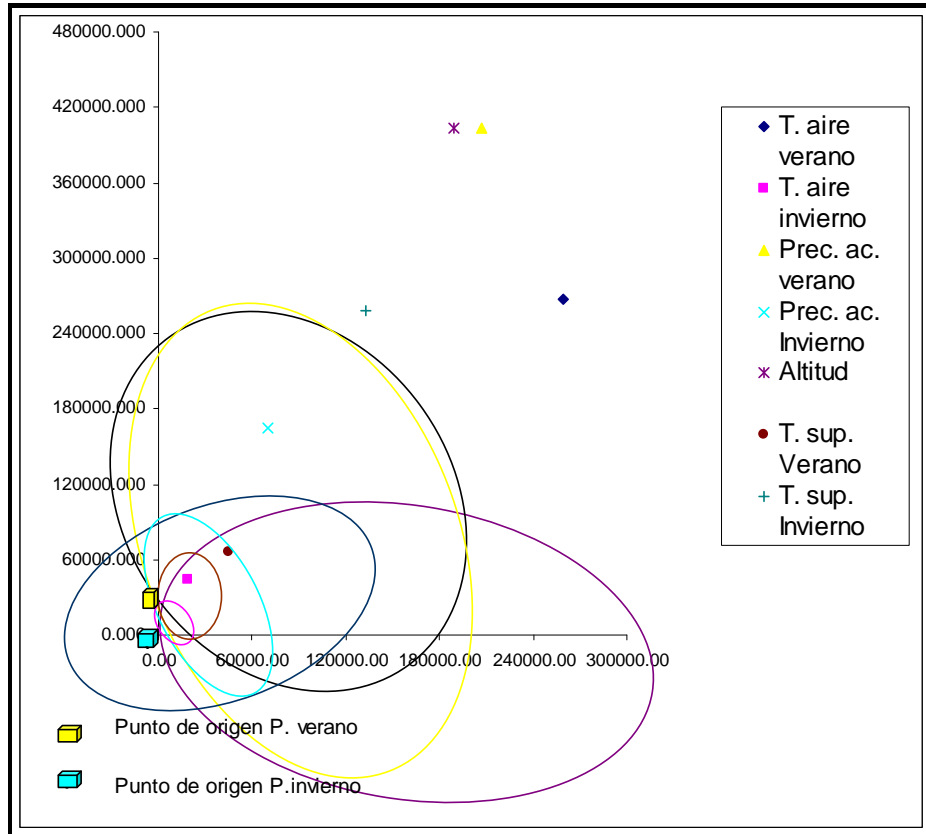


Las variables independientes utilizadas fueron la temperatura del aire promedio, la temperatura de superficie promedio y la altitud.

Para explicar los resultados obtenidos observemos la figura 7.47, que representa el dominio de escala de las variables empleadas en el análisis de regresión para las regiones de precipitación. Enfocándonos en el punto de origen para la precipitación en verano (en amarillo), los dominios de escala de la temperatura del aire (en negro), de la precipitación

y de la altitud (en violeta) tienen rangos bastante similares; sin embargo, la dirección de los ejes difiere lo suficiente como para no contribuir a una fuerte asociación entre estas variables. Aún la temperatura de superficie (en café) presenta ejes en direcciones diferentes a los de la precipitación.

**Fig. 7.47 Dominio de escala de variables empleadas en el análisis de regresión de datos de precipitación a nivel de regiones.**

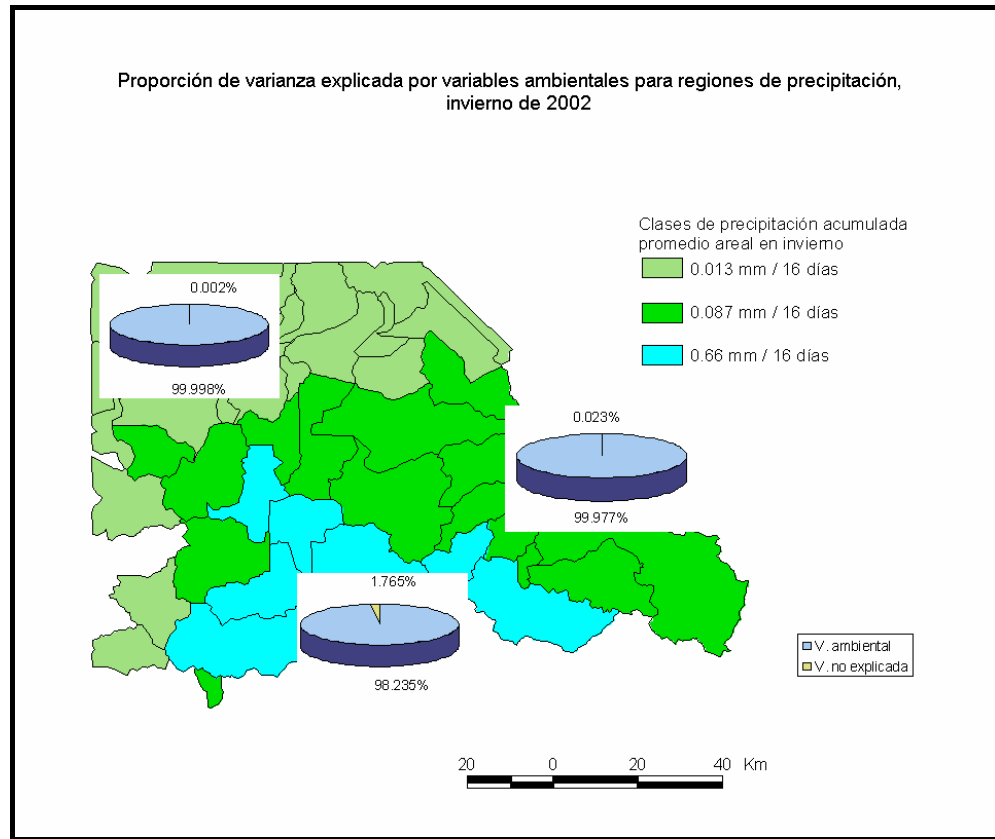


Los ejes mayor y menor de las elipses representan los rangos de las variables obtenidos por variografía. La dirección de estos ejes corresponde a la direcciones de variación de los procesos involucrados. Los puntos de origen se refieren al punto de partida de alguno de los ejes de las elipses y se emplea aquí como elemento de diferenciación entre las variables del periodo de verano y las de invierno. Los ejes x, y se emplean para representar los dominios de escala en forma de puntos en el plano.

En el mapa de la figura 7.48 se presenta la diferenciación funcional entre regiones de precipitación en el periodo de invierno. No obstante, aquí no hay una diferenciación funcional evidente. El primer factor que contribuye a la homogeneización es que el tamaño de las clases debería ser menor o por defecto, debería haber más de 3 regiones para que las clases extremas coincidieran mejor con las unidades que presentaron

autocorrelación espacial y las clases intermedias se comportaran independientemente de la estructura espacial.

**Fig. 7.48 Mapa de proporción de varianza explicada por variación ambiental para regiones de precipitación acumulada en invierno.**

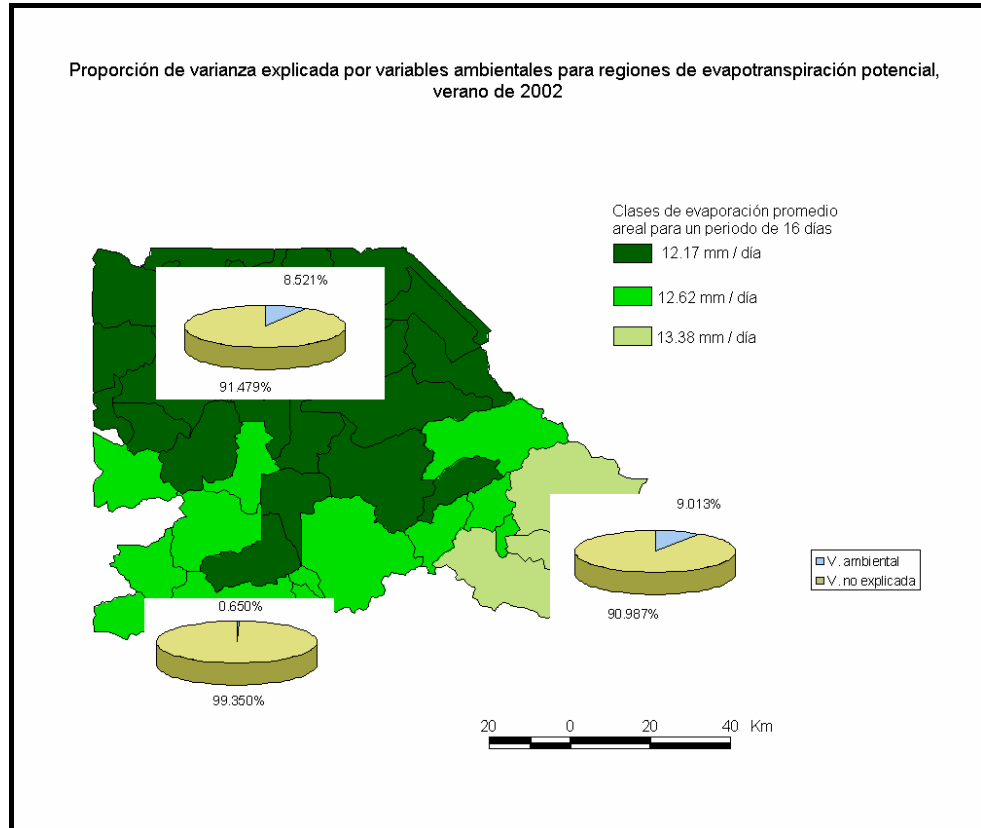


Las variables independientes utilizadas fueron la temperatura del aire promedio, la temperatura de superficie promedio y la altitud.

Respecto a la alta proporción de varianza explicada recurramos al mismo diagrama de la figura 7.47, aquí vemos que las direcciones de variación entre la precipitación (azul turquesa) y la temperatura del aire (rosa), son muy similares, pero el rango de estos procesos es significativamente diferente. Aparentemente el tamaño de las regiones (no mayor a 80 km. en la dirección noroeste sureste y su perpendicular), tiene que ver. Esto significa que las diferencias de rango entre la precipitación y la temperatura del aire no tienen efectos dentro de las regiones y por ello la variación total es prácticamente idéntica. Recordemos que la altitud se empleó en la estimación de la temperatura del aire por lo que la correlación de aquella con la precipitación resulta ser también muy alta. Por último, los rangos de la temperatura de superficie y de la precipitación son casi idénticos en la

dirección noroeste-sureste, sin duda, esto explica porque la correlación de la temperatura de superficie y la precipitación es también cercana a 1. Se observó, en análisis exploratorios previos, que el índice de correlación entre estas variables es muy bajo a nivel de toda la región.

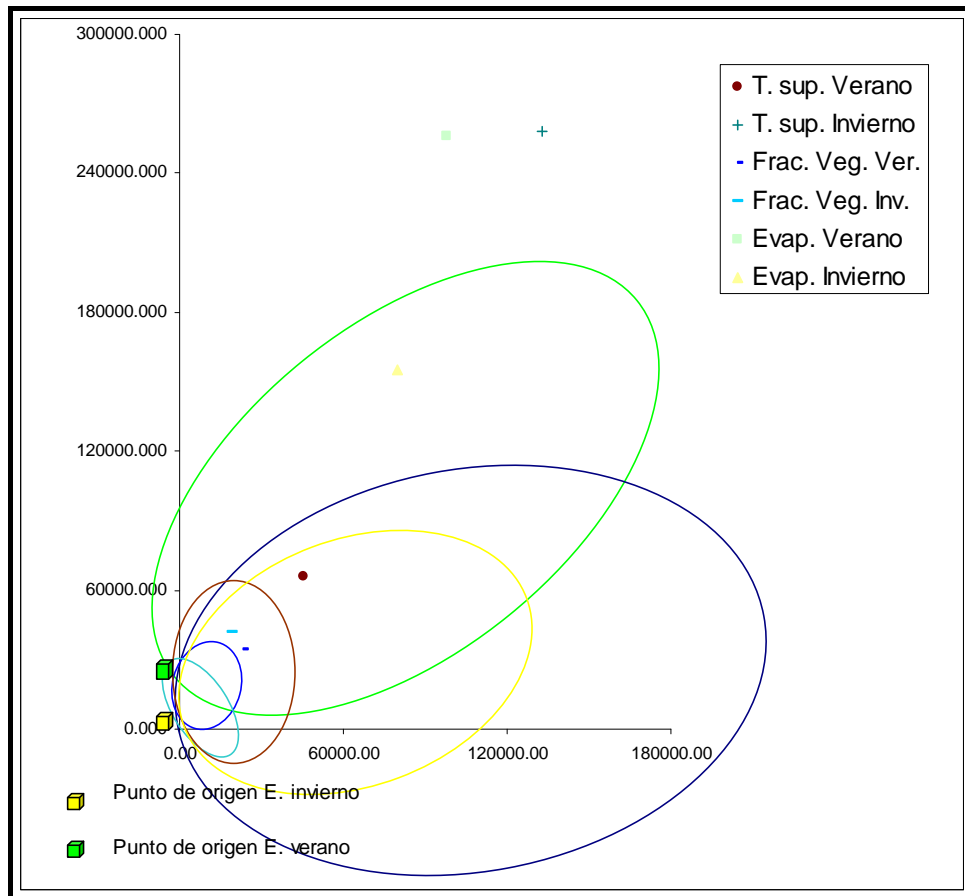
**Fig. 7.49 Mapa de proporción de varianza explicada por variación ambiental para regiones de evaporación promedio en verano.**



Las variables independientes utilizadas fueron la temperatura de superficie promedio y la fracción vegetal de la cobertura.

En cuanto a la evaporación en verano, en la figura 7.49 se resaltan las diferencias entre las regiones conformadas por autocorrelación espacial (evaporación media baja y media alta) de la región intermedia. Observando la figura 7.50, el dominio de escala de la evaporación (en verde) presenta diferencias importantes respecto del rango y dirección de los ejes de las elipses de la fracción vegetal (azul rey) y de la temperatura de superficie (en café). Esto explica la baja proporción de varianza explicada por estas variables para las regiones de evapotranspiración potencial.

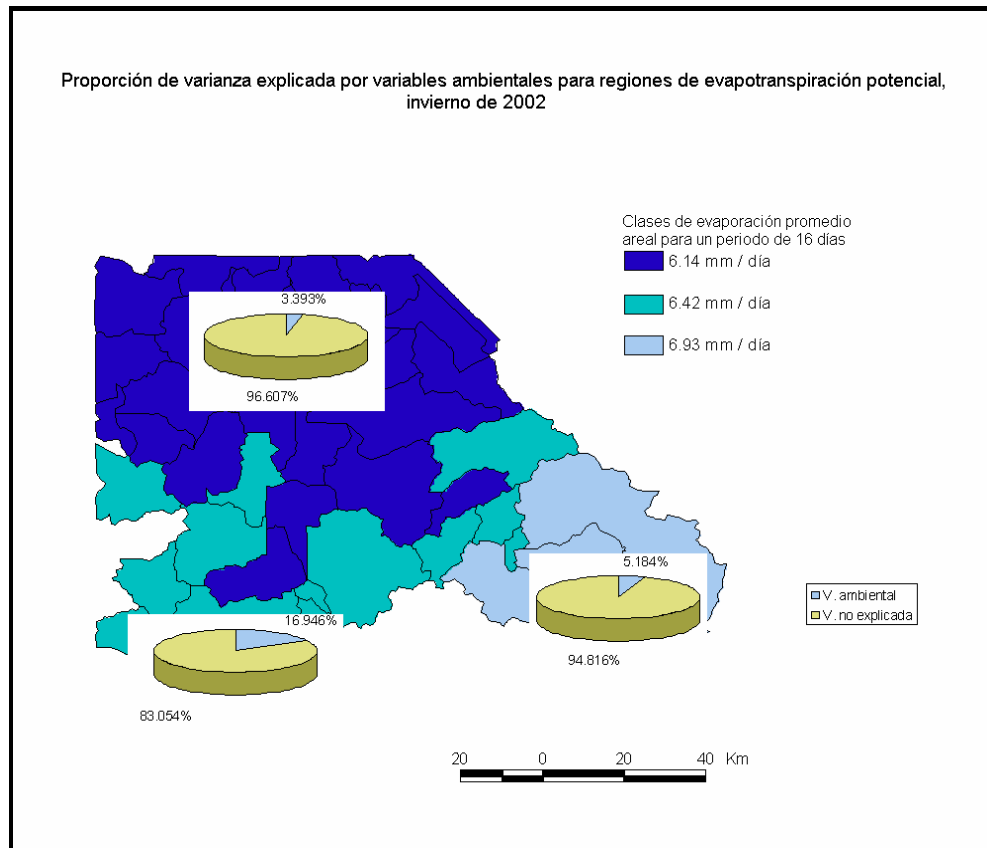
**Fig. 7.50 Dominio de escala de variables empleadas en el análisis de regresión de datos de evaporación a nivel de regiones.**



Los ejes mayor y menor de las elipses representan los rangos de las variables obtenidos por variografía. La dirección de estos ejes corresponde a la direcciones de variación de los procesos involucrados. Los puntos de origen se refieren al punto de partida de alguno de los ejes de las elipses y se emplea aquí como elemento de diferenciación entre las variables del periodo de verano y las de invierno. Los ejes  $x$ ,  $y$  se emplean para representar los dominios de escala en forma de puntos en el plano.

Para la evaporación en invierno (ver mapa de la figura 7.51), se encontró que la región donde la autocorrelación no es significativa (clase intermedia) tiene la mayor proporción de variación explicada. Según la figura 7.50, la temperatura de superficie (en azul marino) y la evaporación (en amarillo), presentan coincidencias en cuanto a la dirección de variación del eje mayor. Esta dirección coincide con el arreglo de la región intermedia (con unos 90 km. a su largo). Sin embargo, la distancia en la que se sobreponen las elipses (más o menos 160 km.), determina diferencias importantes entre las variables y por lo tanto resulta un coeficiente de correlación medio. Las regiones extremas, en cambio, muestran una orientación norte-sur principalmente.

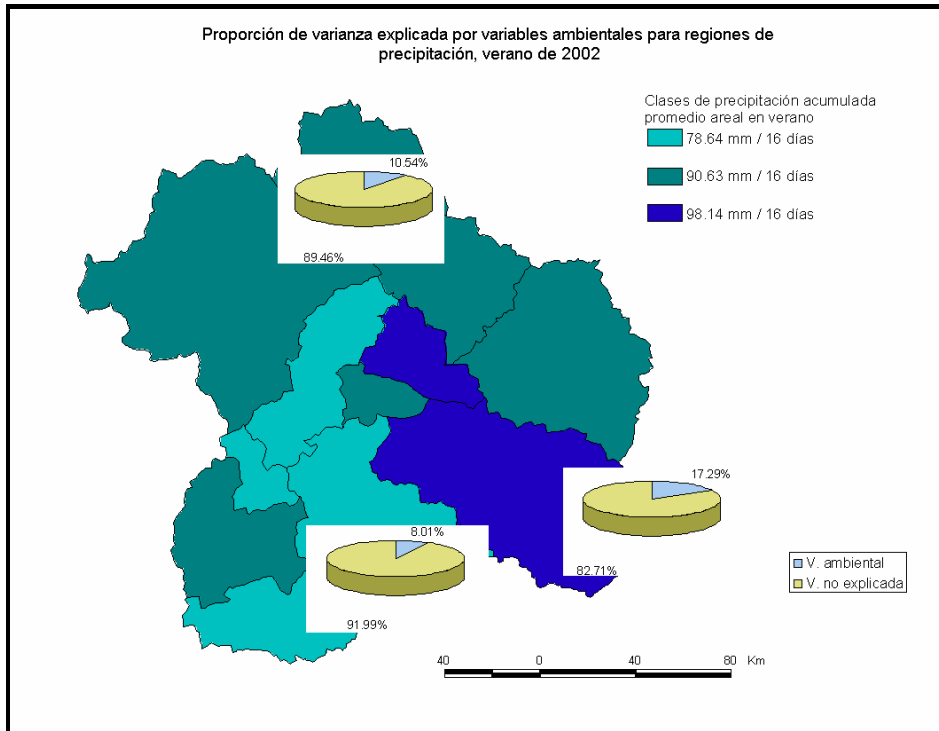
**Fig. 7.51 Mapa de proporción de varianza explicada por variación ambiental para regiones de evaporación promedio en invierno.**



Las variables independientes utilizadas fueron la temperatura de superficie promedio y la fracción vegetal de la cobertura.

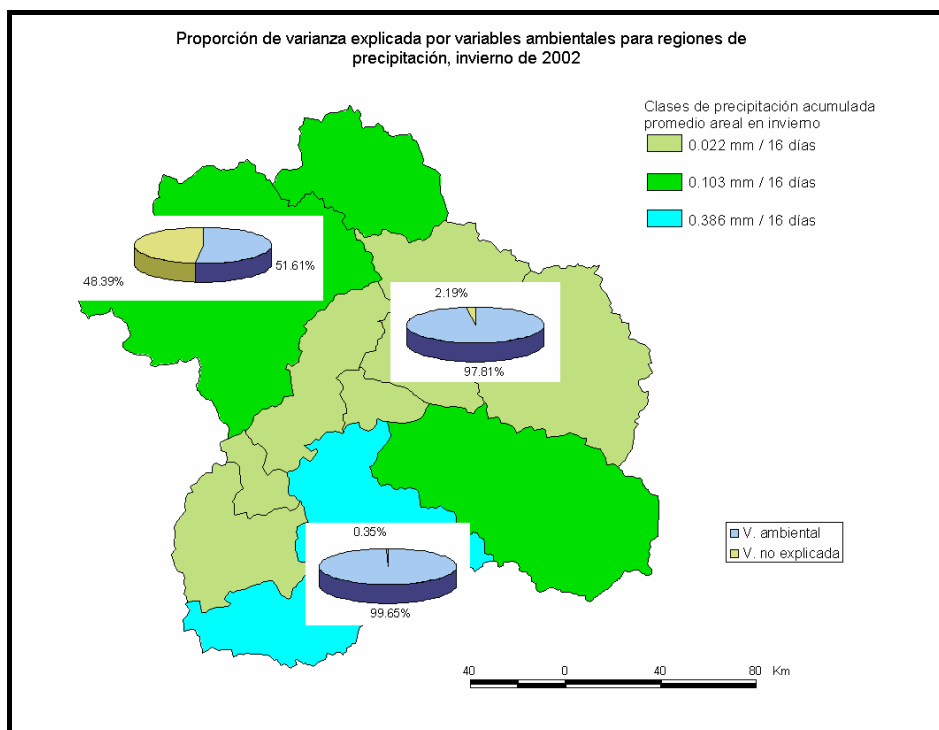
Por último, se presentan los resultados del análisis de regresión a la escala de las regiones mayores a 1000 km<sup>2</sup>. Se analizó la correlación de los componentes con las mismas variables independientes (ver figuras 7.52, 7.53, 7.54 y 7.55). Los mapas en las figuras 7.52 y 7.53 muestran que el efecto del dominio de escala sigue siendo importante aún cuando la autocorrelación espacial ya no es significativa en la conformación de regiones, la diferenciación funcional sigue las mismas reglas generales ya enunciadas. Cuando existe coincidencia en la dirección de variación de las variables analizadas, la asociación entre variables depende principalmente de la diferenciación entre rangos, de tal forma que entre más similares sean, la autocorrelación será alta. El arreglo y tamaño de las regiones es otro factor de diferenciación, si las regiones tienen un tamaño menor o igual al rango de la variable dependiente, el arreglo no es un factor importante; en cambio para regiones mayores, el arreglo en la dirección de los ejes determina fuertemente la correlación.

**Fig. 7.52 Mapa de proporción de varianza explicada por variación ambiental para regiones de precipitación acumulada en verano, unidades mayores a 1000 km<sup>2</sup>.**



Regiones obtenidas por clasificación no restringida espacialmente.

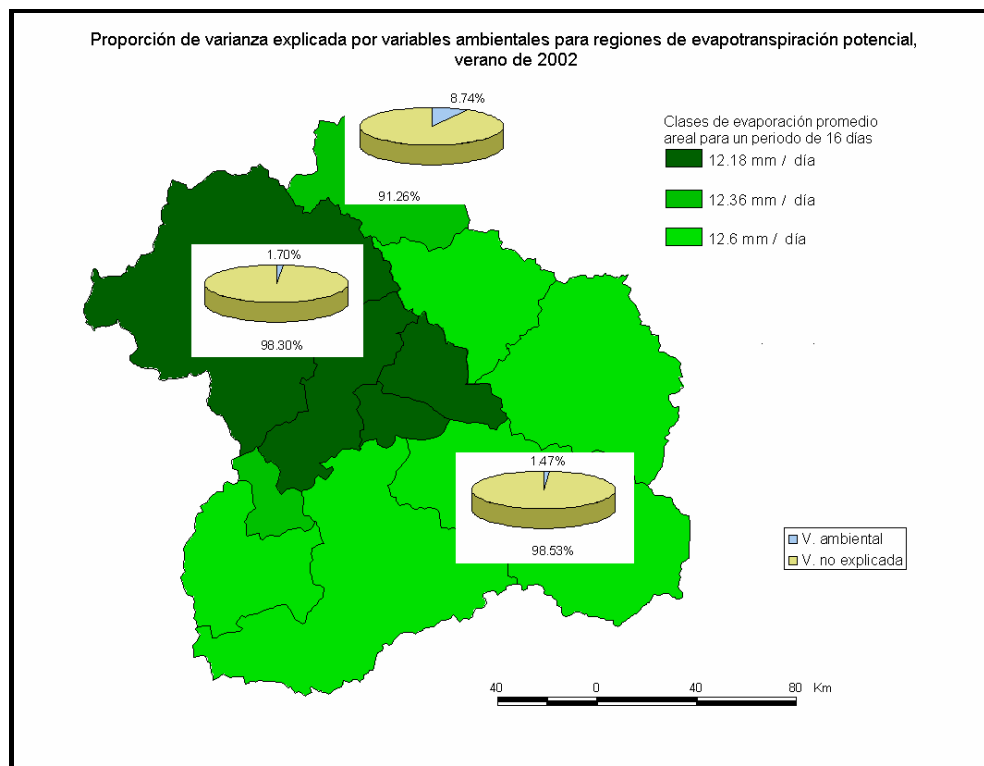
**Fig. 7.53 Mapa de proporción de varianza explicada por variación ambiental para regiones de precipitación acumulada en invierno, unidades mayores a 1000 km<sup>2</sup>.**





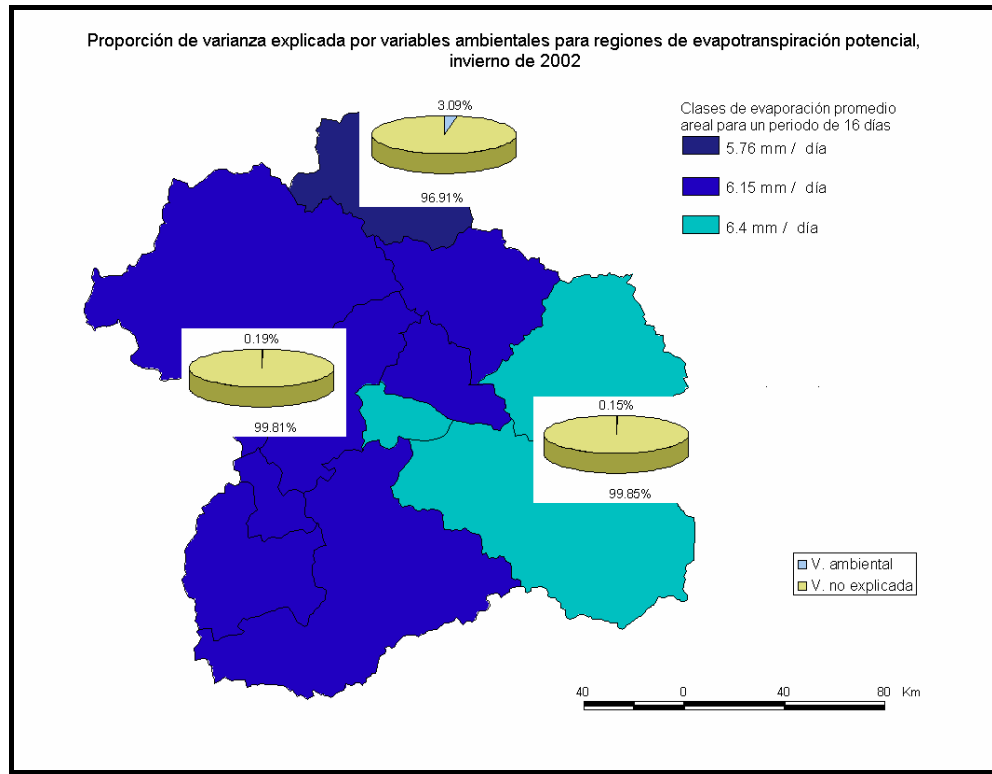
Lo anterior es especialmente cierto en el mapa de la figura 7.52, donde la región 3 está arreglada en la dirección de los ejes del elipse de precipitación y presenta el mayor % de variación explicada por la variación ambiental de la temperatura del aire, la altitud y la temperatura de superficie. En el mapa de la figura 7.53, la clase intermedia (P. muy baja) presenta la más baja correlación entre variables por presentar elementos no contiguos, mientras que las otras 2 clases presentan correlaciones altas por presentar un arreglo en la dirección noroeste-sureste.

**Fig. 7.54 Mapa de proporción de varianza explicada por variación ambiental para regiones de evaporación promedio en verano, unidades mayores a 1000 km<sup>2</sup>.**



En los mapas de las figuras 7.54 y 7.55, resaltan los efectos del tamaño y arreglo de las regiones, pues en ambos periodos las regiones más pequeñas son las que presentan la más alta correlación entre variables. Tales efectos rebasan la diferenciación funcional producida por las variaciones en el dominio de escala de las variables consideradas. Sin duda, la conformación de un mayor número de regiones podría conducir a la diferenciación funcional a esta escala de análisis.

**Fig. 7.55 Mapa de proporción de varianza explicada por variación ambiental para regiones de evaporación promedio en invierno, unidades mayores a 1000 km<sup>2</sup>.**



#### 7.2.4c. Aportes de la fase de confirmación de hipótesis

Esta es, sin duda, una de las fases determinantes en el proceso de descubrimiento de patrones geospaciales, en ella, las generalizaciones propuestas sobre la conformación de regiones de los componentes analizados se confirma. Se establece la importancia de la estructura espacial en la configuración de la precipitación y de la evaporación a una escala determinada. Asimismo, se formaliza la asociación entre variables explicatorias y los componentes, encontrándose que diferentes procesos tienen diferentes efectos sobre la misma variable ya que presentan dominios de escala específicos. El dominio de escala y la autocorrelación espacial resultan ser factores de peso para la diferenciación funcional y la regionalización. Se construyen regiones mediante una clasificación restringida espacialmente y se caracterizan funcionalmente. Estos datos presentan un nivel de información  $V$  que corresponde a la extracción de patrones de los datos, su caracterización por agregación y por la formalización de relaciones causa-efecto entre procesos.

## **8. Consideraciones finales y Conclusiones**

Como se estableció en capítulos anteriores partimos de un problema que puede ser modelado desde la perspectiva sistémica con objeto de abordarlo integralmente, además, se requiere comunicar los procesos y relaciones que conforman el fenómeno del balance hídrico. Para hacerlo, recurrimos al recurso metodológico del DPG, el cual nos permite representar de múltiples formas las características y atributos de los modelos u objetos espaciales que resultan del análisis sistémico del problema.

Desde el enfoque de análisis espacial y la geocomputación, el problema fundamental sobre el balance hídrico consiste en la conformación de modelos o representaciones generalizadas de los datos geográficos u observaciones para destacar las propiedades y atributos de los componentes del balance, profundizar en su comportamiento y ofrecer elementos al proceso de generación de información geoespacial para el manejo y administración del recurso agua.

La extracción de patrones implica un fuerte componente de visualización para la caracterización de los atributos de las variables del balance hídrico. En la tabla 8.1 se listan los procedimientos, métodos y operaciones que involucró la incorporación de cada fase del proceso de descubrimiento de patrones geoespaciales en el estudio de componentes del balance hídrico. Cada fase tiene sus propios métodos de inferencia y herramientas de exploración de datos y nos permite avanzar en la formalización de atributos de las variables del balance hídrico por lo que las salidas de datos presentan diferentes niveles de información. A medida que las propiedades de las variables se van evidenciando el contenido de información es mayor (Rodríguez, 2001:43 y Boisot y Canals, 2004).

Iniciamos en un nivel I que corresponde a la conjunción de los registros sobre el proceso estudiado, donde se define la forma de representación computacional de las observaciones, se diferencia en dos subniveles, respecto de la remoción de ruido o inconsistencia de los datos (nivel IA) y acerca de la precisión de los registros, lo cual tiene que ver con aspectos sobre la capacidad de los formatos lógicos para capturar la realidad del proceso (nivel IB).

**Tabla 8.1 Resumen de los métodos, operaciones y herramientas, empleadas en la incorporación del proceso de DGP en el estudio del balance hídrico.**

FASE	METODOS DE INFERENCIA	HERRAMIENTAS DE EXPLORACIÓN	METAOPERACIONES DE GEOVISUALIZACION	METAOPERACIONES DE DPG
Organización, manejo y visualización de datos de entrada: definición de formatos de representación lógica (nivel IB de información)	Abductivo y deductivo	Herramientas básicas de acceso: Querys, selección y descripción, acceso a data warehouses	Identificación de rasgos	Definición de categorías y formas de representación de los datos
Organización, manejo y visualización de datos de entrada: análisis exploratorio de datos en formatos de representación lógica (nivel II de información)	Abductivo	Herramientas básicas de acceso, herramientas para identificación de rasgos: Diagramas de caja, diagramas de dispersión, diagramas de coordenadas paralelas, plantillas de firmas, animación y perfiles de tiempo, mapas	Identificación, comparación e interpretación de rasgos	Definición de categorías y formas de representación de los datos, identificación de patrones.
Análisis exploratorio o estructural (nivel III de información)	Abductivo, inductivo	Herramientas básicas de acceso, herramientas de identificación y comparación de rasgos: correlograma, variograma, comparación de test de autocorrelación global (Moran y Geary), regresión lineal o no lineal, Kriging, correlograma temporal, mapas, animaciones, modelación de unidades hidrológicas	Identificación, comparación e interpretación de rasgos	Definición de categorías y patrones, identificación de patrones y su caracterización: identificación de vecindarios, definición del dominio de escala y estimación de superficies a partir de observaciones puntuales, formalización de asociaciones, aplicación de modelos de representación conceptual a datos (conformación de regiones o unidades hidrológicas)
Construcción de hipótesis (nivel IV de información)	Inductivo	Herramientas básicas de acceso, herramientas de geovisualización: sumarización o agregación, mapas de percentiles, mapas de caja	Identificación e interpretación de rasgos	Representación de procesos y datos en patrones espaciales, caracterización de patrones o rasgos.
Análisis confirmatorio (nivel V de información)	Deductivo, Inductivo	Herramientas básicas de acceso, herramientas de extracción de fenómenos: clasificación estructural y funcional, regresión múltiple	Extracción e interpretación de rasgos	Extracción de patrones y su caracterización: descriptores de clases, formalización de asociaciones

El nivel II se refiere a la aplicación de herramientas de exploración básica en la muestra de datos que nos sirvan para evidenciar la presencia de rasgos o patrones en las variables. En el caso de las estructuras temporales nos referimos a la autocorrelación temporal y a la escala temporal que los datos capturan sobre el proceso estudiado. En cuanto a las asociaciones nos referimos a la

inclusión del tiempo y el espacio como variables explicatorias de fundamental importancia para entender el comportamiento de los componentes del balance hídrico.

En el nivel III inicia el proceso de identificación de los patrones espaciales que se establecieron como metamodelos a nivel conceptual, en este nivel los formatos lógicos son los mismos de la etapa anterior pero se formalizan las propiedades intrínsecas de las variables: la conformación de vecindarios, la heterogeneidad espacial y temporal, la autocorrelación espacial, el dominio de escala.

En el siguiente nivel se conforman los patrones espaciales del tipo regiones con el fin de caracterizar tales entidades y generalizar el comportamiento de las variables en modelos hipotéticos. Las representaciones de promedios areales de unidades hidrológicas a diferentes escalas espaciales se refieren a abstracciones del nivel IV.

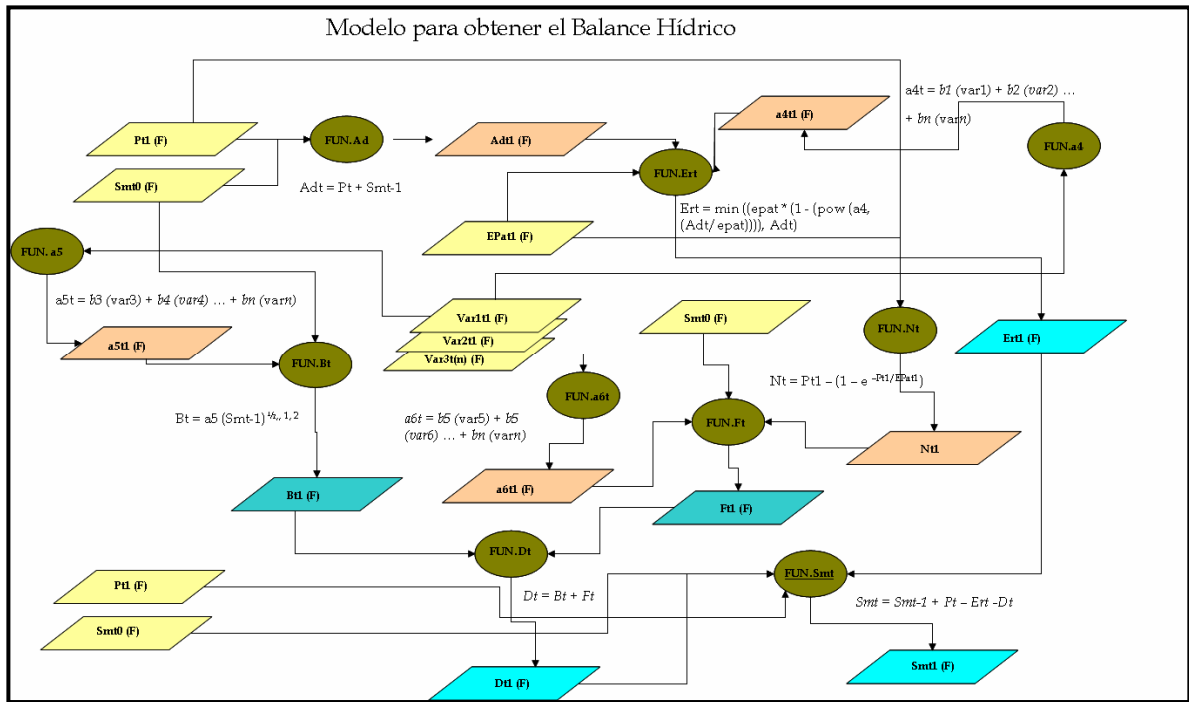
En el nivel V se formalizan los procesos causales o asociaciones y / o se confirma que las variables pueden ser representadas como los patrones espaciales definidos conceptualmente y por tanto los datos en este nivel pasan a ser rasgos geográficos funcionales y concretos (ejemplo: regiones caracterizadas respecto de los componentes del balance hídrico). Aquí debe establecerse el grado de aproximación de los metamodelos a muestras de datos empíricos, en nuestro caso empleamos el método de regresión múltiple para definir de forma explícita el grado de asociación de las variables independientes con los componentes analizados a nivel de regiones.

### **8.1. Detalles de implementación del modelo hidrológico NOPEX**

Uno de los aspectos a cubrir en este trabajo es la implementación computacional de las operaciones a realizar con los datos. Se realizaron algunas tareas que requerían repetirse una y otra vez, para ello, se diseñaron modelos cartográficos que representan de forma lógica los procesos generales para llevar a cabo una tarea particular.

A continuación se muestra un diagrama de procedimiento o modelo cartográfico que bosqueja el proceso general para la estimación de componentes del balance hídrico según el modelo conceptual NOPEX de Xu, Seibert y Halldin (*op cit.*:212), para más detalles véase la figura 8.1.

**Fig. 8.1 Diagrama de procedimientos para la estimación del balance hídrico.**



Este diagrama fue implementado utilizando el lenguaje de programación Arc Macro Language de ArcInfo v. 9.0 el cual desempeña operaciones sobre el formato raster del mismo software. Se emplearon los datos reportados en el artículo de Xu (1999:357), para subcuencas en Suecia. Se utilizaron estimaciones de evapotranspiración real, asimismo se parte de un almacenamiento en suelo de 80mm y una evaporación mensual de 40mm que reporta el trabajo de Xu, Seibert y Halldin (*op cit.*:215), ver apéndice para más detalles.

### 8.2. Limitaciones en la adecuación de modelos conceptuales al estudio del balance hídrico

Al intentar abordar el proceso del balance hídrico, el primer problema a enfrentar es la carencia de observaciones respecto de sus componentes y de variables causales críticas, por ejemplo, datos sobre la permeabilidad del subsuelo, evapotranspiración real, escurrimientos (ya que las observaciones están sesgadas a presas o vasos donde las salidas son controladas o desviadas). Aunque la forma como se ha compensado la falta de datos ha sido la modelación, la mayoría de los modelos más aceptados tienen requerimientos de datos que escapan a la realidad de la generación de datos hidrológicos en México.

La estrategia que se plantea es emplear modelos que tengan la suficiente apertura en los requerimientos de datos, de tal forma que sea posible calibrarlos aún con las deficiencias de información señaladas. Este es el caso de los modelos conceptuales semiempíricos. Se sugiere continuar los esfuerzos hacia la implementación de este tipo de modelos y adecuarlos a las características geográficas del país.

La incorporación del tiempo y el espacio en la estimación del balance hídrico es fundamental y a pesar de que se realizan estudios en forma de series de tiempo, la estructura espacial de las variables se ha omitido en la mayoría de los análisis. La formalización o abstracción de las relaciones se ha llevado a cabo con las restricciones mencionadas anteriormente, lo cual, determina la capacidad actual de establecer modelos aplicables a todas las realidades específicas. En la medida en que se considere la estructura espacio-temporal de los componentes del balance hídrico y de las variables asociadas, estaremos más en posición de fundamentar generalizaciones y aplicar modelos a diferentes condiciones.

### **8.3. Limitaciones en la incorporación del proceso de descubrimiento de patrones geospaciales**

Nos enfrentamos con un problema de enfoque al incorporar el proceso de DPG al estudio del balance hídrico, en este proceso se parte del supuesto de que hay una cantidad inmensurable de datos u observaciones sobre los procesos estudiados y que los datos nos pueden hablar sobre la presencia de patrones, rasgos o entidades bien definidas o conceptuadas. Como vemos, esta no es la realidad sobre algunos fenómenos geospaciales, particularmente del Balance Hídrico, no contamos con grandes volúmenes de datos y estamos en el camino de construir conceptos espaciales claros y válidos.

La estrategia que empleamos, es partir de un conjunto de patrones o modelos conceptuales geospaciales, establecidos desde el enfoque teórico de la ecología del paisaje, que representan adecuadamente el proceso del Balance hídrico y dirigimos el análisis de forma que nos condujera a la formalización de los rasgos identificados. Ciertamente, la principal aportación del proceso de DPG estriba en el énfasis sobre la conformación de patrones y su caracterización mediante técnicas bien definidas y aplicables a la naturaleza de las variables del Balance hídrico.

No hay aportaciones en este trabajo en cuanto a la creación de modelos hidrológicos más precisos o en la construcción de generalizaciones plausibles a escalas más generales, pero la metodología propuesta nos conduce claramente a la formalización de entidades o patrones espaciales en los componentes del Balance hídrico.

Efectivamente se conformaron regiones que representan la diferenciación espacio temporal de la precipitación acumulada y de la evaporación promedio, se pudo establecer que el poder explicatorio del espacio o autocorrelación espacial de estos componentes varía según la época del año. Asimismo, la asociación de los componentes con la fracción de cobertura vegetal, la temperatura del aire, la temperatura de superficie y la altitud, está determinada en cierto grado por la estructura espacial y el dominio de escala de las variables involucradas en el análisis. Se encontraron casos donde hay coincidencias en la dirección y rango de la autocorrelación espacial entre variables, donde las correlaciones lineales a nivel de regiones fueron cercanas a 1 (precipitación acumulada en invierno, ver figuras 7.48 y 7.43). Se estableció que el tamaño de las regiones es otro factor que contribuye a la diferenciación espacio temporal del área de estudio.

Ciertamente se encontrarán limitaciones en la aplicación de la metodología propuesta según la disponibilidad de datos y de herramientas de exploración, de análisis y de validación, no obstante, el énfasis en la geovisualización de las propiedades de las variables del balance hídrico, siempre ofrecerá elementos para avanzar en el descubrimiento de patrones sobre este proceso.

En cuanto a la caracterización, se identificó el dominio de escala espacial de las variables y con base en ello, se definió el tamaño de unidades de captación pluvial que se requieren para capturar la heterogeneidad espacial de la precipitación y de la evaporación (en este caso unidades mayores a 300 km<sup>2</sup> como máximo), realizando de esta forma, la importancia de parametrizar los modelos por región y por temporada no con datos globales o promedios anuales. La generación de información sobre componentes del Balance hídrico a este nivel de detalle es con fines de monitoreo, especialmente porque se propone una metodología replicable y consistente para definir la escala de análisis y de calibración de modelos hidrológicos.



#### **8.4. Líneas de investigación a futuro**

Deben enfocarse algunos esfuerzos de trabajo a análisis exhaustivos que permitan capturar estructuras temporales de los componentes del Balance Hídrico y establecer cuál es la contribución del tiempo en la diferenciación espacio- temporal de este proceso.

Continuar con la implementación de modelos conceptuales semiempíricos, aplicables a la disponibilidad de información sobre los componentes del Balance Hídrico en México, para realizar un análisis más integral del proceso. Particularmente, se sugiere implementar el modelo NOPEX con los datos del Bajío en los dos niveles de agregación analizados (unidades mayores a 300 km<sup>2</sup> y mayores a 1000 km<sup>2</sup>).

Avanzar en la conformación de un marco teórico robusto y multidisciplinario sobre las entidades, objetos o patrones espaciales. Este esfuerzo debe ser paralelo a estudios particulares en los que se busque representar procesos geográficos mediante patrones espaciales.

Continuar con análisis exploratorios para evaluar la relación de otras variables causales con los componentes del Balance Hídrico o bien modelos de regresión diferentes al lineal.

Establecer más claramente cómo el proceso de descubrimiento de patrones geoespaciales incide sobre la administración y planeación del recurso agua y en última instancia sobre la generación de información geoespacial sobre el Balance Hídrico. En este sentido, se sugiere el uso de indicadores del Balance Hídrico (por ejemplo % en volumen de los diferentes componentes), a nivel de unidades de captación pluvial para identificar tendencias generales en el tiempo, incluso de ser necesario se pueden agregar las unidades de acuerdo a la subregión o región hidrológica correspondiente según lo defina el Instituto Nacional de Ecología.

## 9. Bibliografía

- Adams, 2001, *El Octavo Día. La Evolución Social como Autoorganización de la Energía*. México, D.F. Universidad Autónoma Metropolitana Iztapalapa. 200 p.
- Ahern, 1999, Spatial Concepts, Planning Strategies and future Scenarios: a Framework Method for Integrating Landscape Ecology and Landscape Planning. En: *Landscape Ecological Analysis. Issues and Applications*. Klopatek y Gardner (Eds.). New York. Springer-Verlag.175-201.
- Andrienko, Andrienko, Savinov, Voss y Wettschereck, 2001, Exploratory Analysis of Spatial Data Using Interactive Maps and Data Mining, *Cartographic and Geographic Information Science*, **28**(3), 151–65.
- Anselin, Syabri y Kho, 2004, *GeoDa: An Introduction to Spatial Data Analysis*. White paper. Spatial Analysis Laboratory. Department of Agricultural and Consumer Economics. University of Illinois, Urbana – Champaign. 18 p.
- Anselin, 2005, *Exploring Spatial Data with GeodaTM: A Workbook*. Spatial Analysis Laboratory. Department of Geography. University of Illinois, Urbana – Champaign. 244 p.
- Bernier y Caron, 2002, Analyse des méthodes structurées de Développement de systèmes d'information, *Geomática*, **56**(2), 107-119.
- Bockstael, Costanza, Strand, Boynton, Bell y Wainger, 1995, Ecological Economic Modeling and Valuation of Ecosystems, *Ecological Economics*, **14**, 143–159.
- Boisot y Canals, 2004, Data, Information and Knowledge: have we got it right? IN3: UOC (*Working Paper Series*; DP04-002). <http://www.uoc.edu/in3/dt/20388/index.html>.
- Cartwright, 1999, Extending the map metaphor using web delivering multimedia, *Int. J. Geographical Information Science*, **13**(4), 335-353.
- Cook, Symanzik, Majure y Cressie, 1997, Dynamic graphics in a GIS: More examples using linked software, *Computers & Geosciences*, **23**(4), 371–385.
- Cooper e Istok, 1988, Geostatistics applied to groundwater contamination. I: Methodology, *Journal of Environmental Engineering*, **114**(2), 270-299.
- Coopin, Jonckheere, Nackaerts, Muys y Lambin, 2000, Digital change detection methods in ecosystem monitoring: a review, *Int. J. Remote Sensing*, **25**(9), 1565-1596.
- Coronel y Gutiérrez, 2004, *Sistema de Apoyo al Modelo de Gestión Científica del CentroGeo*. Proyecto final del curso de Geomática Aplicada. Documento Interno. 46 p.
- Coronel y Mora, 2006, Dinámica de la fracción vegetal en la cobertura terrestre: Análisis subpixel en imágenes MODIS, *Memorias del XII Simposio Internacional en Percepción Remota y Sistemas de Información Geográfica*. SELPER - Capítulo Colombia Septiembre de 2006. 79-85.
- Costanza, Sklar y White, 1990, Modeling Coastal Landscape Dynamics, *Bioscience*, **40**(2), 91–107.
- Chuvieco, 1995, *Fundamentos de teledetección espacial*. 2a. Ed. Madrid, España. Ediciones Rialp. 453 p.
- DiBiase, 1990, Visualization in the earth sciences. *Earth and mineral Sciences. Bulletin of Earth and Mineral Sciences*, Penn State University, **59**, 13–18.

- Ester, Kriegel, Sander, 1999, *Knowledge Discovery in Spatial Databases*. Invited Paper at 23rd German Conf. on Artificial Intelligence (KI'99). Bonn, Germany.
- Fayyad, Piatetsky-Shapiro, y Smyth, 1996, From Data Mining to Knowledge Discovery in Databases, *American Association for Artificial Intelligence. AAAI*. 37–54.
- Fazal, Imaizumi, Ishida, Kawachi y Tsuchihara, 2005, Estimating groundwater recharge using the SMAR conceptual model calibrated by genetic algorithm, *Journal of Hydrology*, **303**(1/4), 56-78.
- Fortín, 1999, Spatial Statistics in Landscape Ecology. En: *Landscape Ecological Analysis. Issues and Applications*. Klopatec y Gardner (Eds.). New York. Springer-Verlag. 253-279.
- Fuhrmann, 2000, Designing a visualization system for hydrological data, *Computers & Geosciences*, **26**, 11-19.
- Gahegan, Wachowicz, Harrower y Rhyne, ICA, The integration of Geographic Visualization with Knowledge Discovery in Databases and Geocomputation. *ICA Commission on Visualization: Working Group on Database-Visualization Links*. 20 p.
- García y Muñoz, 2002, *El paisaje en el ámbito de la Geografía*. III. Métodos y Técnicas para el estudio del Territorio. México, D.F. Instituto de Geografía. UNAM. 140 p.
- Govaerts, 1999, *Performance comparison of geostatistical algorithms for incorporating elevation into mapping of precipitation*. [http://www.geocomputation.org/1999/023/gc\\_023.htm](http://www.geocomputation.org/1999/023/gc_023.htm). 17 p.
- Haggett, 1965, *Locational Analysis in Human Geography*. London, UK. London: Edward Arnold. 369 p.
- Hall, Townshend y Engman, 1995, Status of Remote Sensing Algorithms for Estimation of Land Surface State Parameters, *Remote Sensing of Environment*, **51**, 138-156.
- Hengl, Heuvelink y Stein, 2003, *Comparison of kriging with external drift and regression-kriging. Technical note, ITC*. Available on-line at [http://www.itc.nl/library/Academic\\_output/](http://www.itc.nl/library/Academic_output/). 17 p.
- Hijmans, Cameron, Parra, Jones y Jarvis, 2005, Very high resolution interpolated climate surfaces for global land areas, *International Journal of Climatology*, **25**, 1965-1978.
- Hoffine, Hurd y Civco, 2002, Development of a Model to Quantify and Map Urban Growth. *ASPRS-ACSM Annual Conference and FIG XXII Congress*. Connecticut University.
- Hurd, Civco, Hoffine, Prisloe y Arnold, 2003, Temporal characterization of Connecticut's Landscape: Methods, Results and applications, *ASPRS 2003 Annual Conference Proceedings*. Anchorage, Alaska.
- INE-SEMARNAT, 2003, *Diagnóstico biofísico y socioeconómico de la cuenca Lerma-Chapala*. Resumen Ejecutivo. Dirección de manejo Integral de Cuencas Hídricas. 20 p.
- Kraak, 1998, The cartographic visualization process: from presentation to exploration, *The cartographic Journal*, **35**, 11–15.
- Kreuseler, 2000, Visualization of geographically related multidimensional data in virtual 3D scenes, *Computers & Geosciences*, **26**(1), 101–108.
- Legendre, 1993, Spatial autocorrelation: Trouble or New Paradigm?, *Ecology*, **74**(6), 1659–1673.

- Legendre y Fortín, 1989, Spatial pattern and ecological analysis, *Vegetatio*, **80**, 107 – 138.
- Lettenmaier, 2004, Terrestrial Hydrologic Aspects of Climate and Global Change Observations Systems. *Joint Meeting of NAS Coordinating Committee on Global Change and Climate Research Committee*. Washington University.
- Maceachren, y Kraak, 1997, Exploratory Cartographic Visualization: Advancing the Agenda, *Computers & Geosciences*, **23**(4), 335–343.
- Maceachren y Kraak, 2001, Research Challenges in Geovisualization, *Cartography and Geographic Information Science*, **28**(1), 1-11.
- Maceachren, Wachowicz, Edsall, Haug y Masters, 1999, Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in database methods, *Int. J. Geographical Information Science*, **13**(4), 311–334.
- Maidment (Eds), 2000, ArcGis Hydro Data Model. Draft Data Model and Manuscript. *20th Annual ESRI User Conference*, San Diego, California. 59 p.
- Mazi, Koussis, Restrepo y Koutsoyiannis, 2004, A groundwater-based, objective-heuristic parameter optimization method for a precipitation-runoff model and its application to a semi-arid basin, *Journal of Hydrology*, **290**(3/4), 243-258.
- Maciek, Lubczynski y Jacek, 2005, Integration of various data sources for transient groundwater modeling with spatio-temporally variable fluxes—Sardon study case, Spain, *Journal of Hydrology*, **306**(1/4), 71-96.
- Mendoza, Bocco, López y Bravo, 2002, Implicaciones hidrológicas del cambio de la cobertura vegetal y uso del suelo: una propuesta de análisis espacial a nivel regional en la cuenca cerrada del lago de Cuitzeo, Michoacán. *Investigaciones Geográficas*, **49**, Boletín del Instituto de Geografía, UNAM. 35-50.
- Mennis y Peuquet, 2003, The Role of Knowledge Representation in Geographic Knowledge Discovery: A Case Study, *Transactions in GIS*, **7**(3), 371-391.
- Miguel, 2002, Región. Complejidad y Caos, *Ciencia y Desarrollo*, **28**(165), 13–19.
- Miller (in press), Geographic Data Mining and Knowledge Discovery. En: *Handbook of Geographic Information Science*. London, UK. Wilson, J.P & Fotheringham, A.S. (eds.). 36-42
- Mitas y Mitasova, 1997, Role of Dynamic Cartography in Simulations of Landscape Processes Based on Multivariate Fields, *Computers & Geosciences*, **23**(4), 437–446.
- Mora e Iverson, 1997, Dynamic Stratification of the Landscape of Mexico: Analysis of Vegetation Patterns Observed with Multitemporal Remotely Sensed Images, *Geocarto International*, **12**(2), 73–88.
- Mora e Iverson, 1998, On the sources of vegetation activity variation and their relation with water balance in México, *Int. J. Remote Sensing*, **19**(10), 1843–1871.
- Mora e Iverson, 2002, A Spatially Constrained Ecological Classification: Rationale, Methodology and Implementation, *Plant Ecology*, **158**, 153-169.

- Myers, 1994, Statistical Methods For Interpolating Spatial Data, *Journal of Applied Science & Computations*, **1**(2), 283–318.
- Naveh, 2001, Ten major premises for a holistic conception of multifunctional landscapes, *Landscape and Urban Planning*, **57**, 269-284.
- Nolan, Healy, Taber, Perkins, Hitt y Wolock, 2006, Factors influencing ground – water recharge in the eastern United States, *Journal of Hydrology*. doi:10.1016/j.jhydrol.2006.06.029.
- Peñaloza y López, 2001, El riego y sus cultivos en Michoacán. En: *Entre campos de esmeralda. La agricultura de Riego en Michoacán*. Sánchez, R. M. (ed). Michoacán, Méx. El Colegio de Michoacán A.C. y Gobierno del Estado de Michoacán. 31-46.
- Pickett, Cadenasso, Grove, Nilon, Pouyat, Zipperer y Costanza, 2001, Urban Ecological Systems: Linking Terrestrial Ecological, Physical and Socioeconomic Components of Metropolitan Areas, *Annual Review of Ecological Systems*, **32**, 127-157.
- Pickett, Cadenasso y Grove, 2004, Resilient cities: meaning, models and metaphor for integrating the ecological, socio-economic and planning reales, *Landscape and urban planning*, **69**, 369–384.
- Portoghese, Uricchio y Vurro, 2005, A GIS tool for hydrogeological water balance evaluation on a regional scale in semi-arid environments, *Computers & Geosciences*, **31**(1), 15-27.
- Ritters, Wickham, O'Neill, Jones y Smith, 2000, Global-scale Patterns of Forest Fragmentation, *Conservation Ecology*, **4**(2), 3–20.
- Rodríguez, 2001, Los servicios de información para la gestión ambiental en México: información, conocimiento y comunicación. En: *Ecología de la Información. Escenarios y actores para la participación ciudadana en asuntos ambientales*. Araya Dujisin (Eds.) Chile. FLACSO. 70 p.
- Rossi, Mulla, Journel, y Franz, 1992, Geostatistical Tools for Modeling and Interpreting Ecological Spatial Dependence, *Ecological Monographs*, **62**(2), 277–314.
- Saraf y Choudhury, 1998, Integrated remote sensing and GIS for groundwater exploration and identification of artificial recharge sites, *International J. Remote Sensing*, **19**(10), 1825–1841.
- Sanders, 1986, Urban vegetation impacts on the Hydrology of Dayton, Ohio, *Urban Ecology*, **9**, 361-376.
- Serafini y Mintz, 1992, A global monthly climatology of soil moisture and water balance, *Climate Dynamics*, **8**(3), 13-27.
- Servenay y Prat, 2003, Erosion Extension of Indurate Volcanic Soils of Mexico by Aerial Photographs and Remote Sensing Análisis, *Geoderma*, **117**, 367–375.
- Singh (ed.), 1995, *Computer Models of Watershed Hydrology*. Water Resources Publications. Department of Civil and Environmental Engineering. Baton Rouge, LA. Louisiana State University. 315 p.
- Soto y Soto, 1990, Cambios recientes en el uso del suelo agrícola en la región del Bajío, Gto. *Geografía y Desarrollo*, **2**(5), 3–11.
- Thomas y Edwards, 2001, Towards an Ontology for Geomatics, *Geomática*, **55**(4), 475–490.

- Tyrtonia, 1999, *Termodinámica de la supervivencia para las ciencias sociales*. México, D.F. Universidad Autónoma Metropolitana Iztapalapa. 90 p.
- Xu, 1999, Estimation of Parameters of a Conceptual Water Balance Model for Ungauged Catchments, *Water Resources Management*, **13**, 353–368.
- Xu, 2000, Climate Change and Hydrological models: a Review of Existing Gaps and Recent Research Developments, *Water Resources Management*, **13**, 369-382.
- Xu, Seibert y Halldin, 1996, Regional water balance modeling in the NOPEX area: development and application of monthly water balance models, *Journal of Hydrology*, **180**, 211–236.
- Xu y Singh, 1998, A Review on Monthly Water Balance Models for Water Resources Investigations, *Water Resources Management*, **12**, 31-50.
- Xu y Vandewiele, 1995, Parsimonious monthly rainfall-runoff models for humid basins with different input requirements, *Advances in Water Resources*, **18**, 39–48.
- Willmott, 1977, Watbug: A Fortran IV algorithm for calculating the climatic water budget, *Publications in Climatology*, **30**(2), 1-54.

## 10. Apéndice

### 10.1. Implementación del diagrama de procedimientos para estimar balance hídrico mediante el modelo NOPEX

A partir de los diagramas de procedimientos o modelos cartográficos se definen los pasos básicos a seguir para la ejecución de una tarea particular, el modelo NOPEX se implementó en el módulo GRID de ArcInfo y se realizaron corridas de prueba que fueron satisfactorias. Primero se detalla la forma como se implementó el algoritmo para la estimación de componentes del ciclo hidrológico y después se hacen observaciones sobre el procedimiento para estimar la evapotranspiración potencial.

Se vaciaron los datos de la tabla 1 del artículo de Xu (1999:357), en 26 unidades hidrológicas tipo, se obtuvieron 6 capas en formato raster cada una con los valores promedio de precipitación mensual, evapotranspiración mensual, % de cuerpos de agua, % de bosque y % de suelo desnudo para cada unidad, respectivamente. Las últimas 3 capas se emplearon para calcular los valores de los parámetros  $a_5$  y  $a_6$  de cada unidad de acuerdo a las siguientes ecuaciones:

$$a_5 = 0.0949 - 0.0350 (\% \text{ agua}) + 0.000999 (\% \text{ suelo})$$

El parámetro  $a_5$  favorece los escurrimientos lentos o base que son los primeros en formarse cuando se rebasa la capacidad de retención de humedad por parte del suelo. En este proceso contribuye la proporción de superficie permeable o suelo descubierto, como vemos en la ecuación lineal del parámetro  $a_5$ , mientras que cantidad de superficie cubierta con agua disminuye la cantidad de escurrimiento lento.

En la ecuación para calcular  $b_t$  la potencia  $b_1$  se representa como una constante y puede tomar valores de 1,2 o  $\frac{1}{2}$  de acuerdo al trabajo de Xu, Seibert y Halldin (1996:218). Este valor varía para ajustar el volumen de escurrimientos estimados con los observados.

$$a_6 = 0.085 - 0.0116 (\% \text{ agua}) + 0.000889 (\% \text{ Bosque})$$

En este caso la proporción de bosque en una subcuenca incrementa el valor del parámetro aunque muy débilmente comparado con el efecto de la presencia de superficies cubiertas con agua que merman el volumen de escurrimientos rápidos.

Empleando como entrada la precipitación y la evapotranspiración real y partiendo de datos observados para la cuenca sobre el almacenamiento disponible en el suelo y sobre la tasa de evaporación mensual se diseñó el programa nopex.aml que corre en la plataforma de GRID en ArcInfo.

```
/*Programa para el cálculo del balance hídrico  
/*Xu, Seibert y Halldin, 1996  
/*Autor: Claudia Coronel Enríquez  
/*Fecha 22 de julio de 2005  
/*CentroGeo 26152508  
  
/*Variables de entrada  
/* Contadores de tiempo  
&sv m = 0  
&sv n = %m% + 1  
/* rt%n% = precm, para este ejemplo  
/* ct%n% = temperatura para el periodo, no se utiliza en el ejemplo  
/* ept%n% = evaporación o evapotranspiración potencial, 40 mm./mensuales para el  
ejemplo  
/* a1, a2, a3, a4, a5, y a6 parámetros citados en el artículo de Xu, 1996  
  
/*variables a calcular  
/*wt%n% = agua disponible, no se emplea para este ejemplo  
/*wt%n% = (smt%m% + rt%n%)  
/*et%n% = evapotranspiración real, evapm para este ejemplo  
/*et%n% = min ((ept%n% * (1 - (pow (a4, (wt%n%/ept%n%))))), wt%n%)  
/*nt%n% = precipitación activa  
/*smt%m% = almacenamiento disponible periodo anterior  
/*bt%n% = flujo lento
```



```

/*ft%n% = flujo rápido
/*dt%n% = escurrimiento total
/*smt%n% = almacenamiento disponible periodo actual

```

```

&do &while %m% lt 2
  &if %m% = 0 &then
    &do
      &type 'Procesando tiempo %n%
      nt%n% = (precm - (40 * (1 - exp((-1 * precm) / 40))))
      bt%n% = (a5t%n% * (sqrt (smt%m%)))
      ft%n% = (a6t%n% * (sqrt (smt%m%)) * nt%n%)
      dt%n% = (bt%n% + ft%n%)
      smt%n% = ((smt%m% + precm) - (evapm + dt%n%))
      kill smt%m% all
      &sv n = %n% + 1
      &sv m = %m% + 1
    &end
  &end &else &do
    &type 'Procesando tiempo %n%
    nt%n% = (precm - (40 * (1 - exp((-1 * precm) / 40))))
    smt%m%b = max (smt%m%, 0)
    bt%n% = (a5t%n% * (sqrt (smt%m%b)))
    ft%n% = (a6 * (sqrt (smt%m%b)) * nt%n%)
    dt%n% = (bt%n% + ft%n%)
    smt%n% = ((smt%m%b + precm) - (evapm + dt%n%))
    kill smt%m%b all
    &sv n = %n% + 1
    &sv m = %m% + 1
  &end
End

```

El programa no presenta ningún problema para correr en GRID y calcula los escurrimientos y el balance hídrico para las 26 unidades en un lapso de 40 a 60 segundos.