



INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN

REPRESENTACIÓN SEMÁNTICA DE DATOS
ESPACIALES RASTER

TESIS

QUE PARA OBTENER EL GRADO DE

DOCTOR EN CIENCIAS DE LA COMPUTACIÓN

PRESENTA

ROLANDO QUINTERO TÉLLEZ

DIRECTOR DE TESIS

DR. SERGUEI LEVACHKINE



MÉXICO D.F.

NOVIEMBRE DE 2007

Agradecimientos

A Dios.

Gracias por el enorme regalo de la vida y por la gracia para poder disfrutarla.

A mi familia.

Gracias por ser mi soporte y mi motor; mi inspiración y mi razón de ser.

A ti Paty, amiga, esposa, compañera y cómplice, gracias por estar conmigo. A mis niños Zyanya y Ronaldo, que son la razón de todo, gracias por existir.

A mis padres, Elías y Socorro, y a mis hermanos, Socorro, Alma, Elías, Gustavo, Rubén, Tacho y Ricardo, por todo su apoyo y sus sabios consejos a lo largo de toda mi vida.

A Marylu y Abundio, complemento ideal de la esposa ideal, gracias por su apoyo.

A mis amigos.

Gracias por su ayuda, por su compañía y, sobre todo, por su amistad.

A los Bonobos, Giovanni, Marco, Miguel, Miguel Alejandro, Rolando, Benjamín, Chadwick y Serguei, por los innumerables buenos momentos que hemos pasado.

A mi asesor, Serguei Levachkine, por su guía y ayuda para lograr este sueño.

A Toño y Pedro, por su amistad y consejos.

Gracias a mi alma mater, el Instituto Politécnico Nacional, por brindarme el medio para alcanzar esta meta.

Y a todos aquellos que de una forma o de otra me han ayudado para desarrollarme, tanto profesional como intelectualmente.

¡A todos, muchas gracias!



Índice

R esumen	i
Abstract	ii
A gradecimientos.....	iii
Í ndice.....	v
Índice de Figuras.....	ix
Índice de Tablas.....	xvii
Índice de Siglas	xviii
O bjetivos.....	xix
Objetivo.....	xix
Objetivos particulares	xix
1. Introducción.....	1
1.1. Semántica espacial	3
1.1.1. Conceptualización	4
1.1.2. Síntesis.....	4
1.1.3. Descripción.....	4
1.2. Justificación.....	5

2.	Estado del arte	9
2.1.	Modelos del terreno	9
2.1.1.	Modelos digitales de elevación	14
2.2.	Enfoques adoptados por el INEGI	25
2.3.	Geomorfometría	28
2.4.	Enfoques que utilizan procesamiento de imágenes	30
2.4.1.	Conversión de <i>raster</i> a vector	33
2.5.	Fractales	37
2.6.	Incertidumbre	39
2.7.	Filosofía, semántica y ontologías	40
2.8.	Discusión	50
3.	Marco teórico	53
3.1.	Ontologías	53
3.1.1.	Compromiso ontológico	57
3.1.2.	Ontologías formales	60
3.1.3.	"Concepto"	61
3.1.3.	Ontologías en los sistemas de información	62
3.2.	Agrupamiento (Clustering)	69
3.2.1.	<i>K-means</i>	70
3.2.2.	<i>Fuzzy C-Means</i>	71
3.3.	Relaciones	72
3.3.1.	Propiedades	73
3.3.2.	Relación de equivalencia	74
3.3.3.	Relación de orden	75

4.] Metodología	77
4.1. Conceptualización	79
4.1.1. Axiomas	82
4.1.2. Relaciones	84
4.1.3. Conceptos.....	85
4.1.4. Propiedades	86
4.1.5. Habilidades.....	86
4.1.6. Instancias	87
4.1.7. Clases	87
4.1.8. Restricciones.....	87
4.1.9. Conceptualización de los dominios geográfico y específico	91
4.2. Síntesis	92
4.4. Descripción.....	98
4.4.1. Formas de clasificación	105
4.4.2. Formas de representación	106
5.] Resultados	109
5.1. Conceptualización	110
5.1.1. Ontología del dominio geográfico.....	111
5.1.2. Ontología del dominio de formas del terreno	127
5.1.3. Ontología de la aplicación.....	142
5.1.4. Integración de las ontologías	145
5.2. Síntesis.....	148
5.2.1. Suavizado	150
5.2.2. Cálculo de la zona llana más extensa.....	155

5.2.3. Etiquetado	157
5.2.4. Segmentación	159
5.2.5. Extracción	167
5.3. Descripción.....	174
5.4. Pruebas adicionales.....	176
5.4.1. Pruebas con el DEM del Valle de México	176
5.4.2. Pruebas con el DEM de Chicago.....	191
6. Conclusiones y trabajo futuro.....	207
6.1. Logros	209
6.2. Alcances y limitaciones.....	210
6.3. Trabajo futuro	211
Bibliografía.....	213
Anexos.....	221
A.1. Código fuente en MATLAB.....	221
A.2. Pruebas con el DEM de Las Vegas.....	232
A.3. Pruebas con el DEM de Salt Lake City	239

Índice de Figuras

Figura 1. Metodología para la representación semántica de datos espaciales <i>raster</i>	5
Figura 2. Curvas de nivel digitales	12
Figura 3. Redes de triángulos irregulares	13
Figura 4. Matrices regulares	13
Figura 5. Matrices de resolución variable	14
Figura 6. Modelos digitales de elevación	14
Figura 7. Proceso de generación de DEM con base en mapas con curvas de nivel.....	26
Figura 8. Correlación de pares estereoscópicos	27
Figura 9. Ejemplo de fractal determinístico	38
Figura 10. Ejemplo de fractal autoafín	38
Figura 11. Descomposición del perfil de un DEM en diez tipos geométricos elementales	47
Figura 12. Tipos de ontologías, según Guarino (1998)	55
Figura 13. Aproximación de una ontología a una conceptualización	60
Figura 14. Ejemplo gráfico de agrupamiento (<i>clustering</i>)	69
Figura 15. Relación entre los elementos de un conjunto	73
Figura 16. Panorama general de la metodología planteada	78
Figura 17. Agrupamiento de objetos utilizando diferentes criterios	93
Figura 18. Modelo básico para la conceptualización de la aplicación.....	94
Figura 19. Ejemplo de conceptualización de los algoritmos de extracción	95
Figura 20. Vinculación de la conceptualización de la aplicación con el resto	96
Figura 21. Relación entre conceptos de la O_D y O_A	97

Figura 22. Estado inicial de la relación entre un extracto en O_A y un concepto en O_D	100
Figura 23. Medición de las propiedades del extracto señaladas por la conceptualización.....	101
Figura 24. Clasificación del extracto con base en valores de referencia	102
Figura 25. Asignación del extracto como instancia del nuevo concepto	103
Figura 26. Proceso de descripción de los extractos.....	104
Figura 27. Grafo difuso de los parámetros utilizados para clasificar diferentes formas del terreno.....	106
Figura 28. Utilización de una descripción textual para la representación.....	108
Figura 29. De la metodología al caso de estudio	110
Figura 30. Clases principales en la ontología kaab.....	112
Figura 31. Clases y subclases de objeto geográfico.....	114
Figura 32. Subclases de "relación geográfica"	119
Figura 33. Definición de ACUEDUCTO en el diccionario de datos topográficos del INEGI	120
Figura 34. Ubicación existencial de ACUEDUCTO en la conceptualización.....	123
Figura 35. Asignación del atributo de dominio fijo "condición de construcción" a "acueducto" como una propiedad	125
Figura 36. Asignación del atributo de dominio variable "identificador numérico" a "acueducto" como una propiedad.....	126
Figura 37. Asignación de la relación geográfica "comparte con túnel" a "acueducto" como una habilidad.....	127
Figura 38. Clases de la ontología del dominio de las formas del terreno	130
Figura 39. Conceptos de la clase "forma de terreno"	136

Figura 40. Conceptos de la clase "sitio de interés"	138
Figura 41. Conceptos de la clase "propiedad fisiográfica"	139
Figura 42. Extractos que genera el algoritmo de extracción	143
Figura 43. Fragmento de la jerarquía de extractos que describe la ontología de la aplicación	144
Figura 44. Ejemplos de instancias en la ontología de la aplicación	145
Figura 45. Integración de la ontología hunxeet con la ontología kaab	146
Figura 46. Integración de la ontología wiinkil con la ontología hunxeet	147
Figura 47. Un ejemplo de jerarquía de extractos	149
Figura 48. Aplicación de análisis de derivadas, iteración 1	151
Figura 49. Aplicación de análisis de derivadas, iteración 30	152
Figura 50. Ejemplo de suavizado y sus efectos en el etiquetado	153
Figura 51. Modelo Digital de Elevación de la Zona del Gran Cañón de Colorado, E.U.	154
Figura 52. Vista tridimensional del Modelo Digital de Elevación de la Zona del Gran Cañón de Colorado, E.U.	155
Figura 53. Imagen de los contornos del resultado de aplicar el filtro Laplaciano al DEM	156
Figura 54. Algoritmo de búsqueda de la Zona Llana más Extensa	157
Figura 55. Algoritmo de etiquetado	158
Figura 56. Ejemplo de etiquetado de un DEM	158
Figura 57. Algoritmo para la segmentación	159
Figura 58. Resultados del proceso de segmentación aplicado al DEM original	160
Figura 59. Resultados por separado del proceso de segmentación aplicado al DEM original.....	161

Figura 60. Resultados del proceso de segmentación aplicado al segmento "elev".....	162
Figura 61. Resultados del proceso de segmentación aplicado al segmento "llan"	163
Figura 62. Resultados del proceso de segmentación aplicado al segmento "depr"	164
Figura 63. Resultados del proceso de segmentación aplicado al segmento "elev-elev"	165
Figura 64. Resultados del proceso de segmentación aplicado al segmento "elev-llan"	166
Figura 65. Resultados del proceso de segmentación aplicado al segmento "elev-depr"	167
Figura 66. Algoritmo para la extracción.....	168
Figura 67. Extractos obtenidos de las firmas "elev-elev-elev", "elev-elev-llan" y "elev- elev-depr"	169
Figura 68. Extractos obtenidos de las firmas "elev-llan-elev", "elev-llan-llan" y "elev-llan- depr"	170
Figura 69. Extractos obtenidos de las firmas "elev-depr-elev", "elev-depr-llan" y "elev- depr-depr"	171
Figura 70. Extractos obtenidos de las firmas "llan-elev", "llan-llan" y "llan-depr"	172
Figura 71. Extractos obtenidos de las firmas "depr-elev", "depr-llan" y "depr-depr"	173
Figura 72. Gráfica tridimensional de un extracto de la firma "elev-elev-llan"	174
Figura 73. Algoritmo para realizar la descripción de un extracto	175
Figura 74. Parte de la descripción obtenida a partir del DEM de ejemplo	176
Figura 75. Modelo Digital de Elevación del Valle de México.....	177
Figura 76. Vista tridimensional del Modelo Digital de Elevación del Valle de México ...	177
Figura 77. Resultados del proceso de segmentación aplicado al DEM del Valle de México.....	178

Figura 78. Resultados del proceso de segmentación aplicado al segmento "elev" DEM del Valle de México.....	179
Figura 79. Resultados del proceso de segmentación aplicado al segmento "llan" DEM del Valle de México.....	180
Figura 80. Resultados del proceso de segmentación aplicado al segmento "depr" DEM del Valle de México.....	181
Figura 81. Resultados del proceso de segmentación aplicado al segmento "elev-elev" DEM del Valle de México	182
Figura 82. Resultados del proceso de segmentación aplicado al segmento "elev-llan" DEM del Valle de México	183
Figura 83. Resultados del proceso de segmentación aplicado al segmento "elev-depr" DEM del Valle de México	184
Figura 84. Extractos obtenidos de las firmas "elev-elev-elev", "elev-elev-llan" y "elev-elev-depr" del DEM del Valle de México	185
Figura 85. Extractos obtenidos de las firmas "elev-llan-elev", "elev-llan-llan" y "elev-llan-depr" del DEM del Valle de México	186
Figura 86. Extractos obtenidos de las firmas "elev-depr-elev", "elev-depr-llan" y "elev-depr-depr" del DEM del Valle de México	187
Figura 87. Extractos obtenidos de las firmas "llan-elev", "llan-llan" y "llan-depr" del DEM del Valle de México.....	188
Figura 88. Extractos obtenidos de las firmas "depr-elev", "depr-llan" y "depr-depr" del DEM del Valle de México	189
Figura 89. Parte de la descripción obtenida a partir del DEM del Valle de México.....	191
Figura 90. Modelo Digital de Elevación de Chicago.....	192
Figura 91. Vista tridimensional del Modelo Digital de Elevación de Chicago	192
Figura 92. Resultados del proceso de segmentación aplicado al DEM de Chicago	193

Figura 93. Resultados del proceso de segmentación aplicado al segmento "elev" del DEM de Chicago.....	194
Figura 94. Resultados del proceso de segmentación aplicado al segmento "llan" del DEM de Chicago.....	195
Figura 95. Resultados del proceso de segmentación aplicado al segmento "depr" del DEM de Chicago.....	196
Figura 96. Resultados del proceso de segmentación aplicado al segmento "elev-elev" del DEM de Chicago	197
Figura 97. Resultados del proceso de segmentación aplicado al segmento "elev-llan" del DEM de Chicago	198
Figura 98. Resultados del proceso de segmentación aplicado al segmento "elev-depr" del DEM de Chicago	199
Figura 99. Extractos obtenidos de las firmas "elev-elev-elev", "elev-elev-llan" y "elev-elev-depr" del DEM de Chicago	200
Figura 100. Extractos obtenidos de las firmas "elev-llan-elev", "elev-llan-llan" y "elev-llan-depr" del DEM de Chicago	201
Figura 101. Extractos obtenidos de las firmas "elev-depr-elev", "elev-depr-llan" y "elev-depr-depr" del DEM de Chicago	202
Figura 102. Extractos obtenidos de las firmas "llan-elev", "llan-llan" y "llan-depr" del DEM de Chicago.....	203
Figura 103. Extractos obtenidos de las firmas "depr-elev", "depr-llan" y "depr-depr" del DEM de Chicago.....	204
Figura 104. Parte de la descripción obtenida a partir del DEM de Chicago.....	205
Figura 105. Código fuente en MATLAB	231
Figura 106. Modelo Digital de Elevación de Las Vegas	232
Figura 107. Vista tridimensional del Modelo Digital de Elevación de Las Vegas.....	232

Figura 108. Resultados del proceso de segmentación aplicado al DEM de Las Vegas	233
Figura 109. Resultados de la segmentación aplicada al segmento "elev" del DEM de Las Vegas.....	233
Figura 110. Resultados de la segmentación aplicada al segmento "llan" del DEM de Las Vegas	234
Figura 111. Resultados de la segmentación aplicada al segmento "depr" del DEM de Las Vegas.....	234
Figura 112. Resultados de la segmentación aplicada al segmento "elev-elev" de Las Vegas	235
Figura 113. Resultados de la segmentación aplicada al segmento "elev-llan" de Las Vegas	235
Figura 114. Resultados de la segmentación aplicada al segmento "elev-depr" de Las Vegas	236
Figura 115. Extractos obtenidos de las firmas "elev-elev" del DEM de Las Vegas	236
Figura 116. Extractos obtenidos de las firmas "elev-llan" del DEM de Las Vegas.....	237
Figura 117. Extractos obtenidos de las firmas "elev-depr" del DEM de Las Vegas	237
Figura 118. Extractos obtenidos de las firmas "llan" del DEM de Las Vegas	238
Figura 119. Extractos obtenidos de las firmas "depr" del DEM de Las Vegas	238
Figura 120. Modelo Digital de Elevación de Salt Lake City.....	239
Figura 121. Vista tridimensional del Modelo Digital de Elevación de Salt Lake City	239
Figura 122. Resultados del proceso de segmentación aplicado al DEM de Salt Lake City	240
Figura 123. Resultados de la segmentación aplicada al segmento "elev" de Salt Lake City.....	240
Figura 124. Resultados de la segmentación aplicada al segmento "llan" de Salt Lake City.....	241

Figura 125. Resultados de la segmentación aplicada al segmento "depr" de Salt Lake City.....	241
Figura 126. Resultados de la segmentación aplicada al segmento "elev-elev" de Salt Lake City	242
Figura 127. Resultados de la segmentación aplicada al segmento "elev-llan" de Salt Lake City	242
Figura 128. . Resultados de la segmentación aplicada al segmento "elev-depr" de Salt Lake City	243
Figura 129. Extractos obtenidos de las firmas "elev-elev" del DEM de Salt Lake City	243
Figura 130. Extractos obtenidos de las firmas "elev-llan" del DEM de Salt Lake City	244
Figura 131. Extractos obtenidos de las firmas "elev-depr" del DEM de Salt Lake City ...	244
Figura 132. Extractos obtenidos de las firmas "llan" del DEM de Salt Lake City.....	245
Figura 133. Extractos obtenidos de las firmas "depr" del DEM de Salt Lake City	245

Índice de Tablas

Tabla 1. Parámetros y modelos topográficos utilizados en los DEM	24
Tabla 2. Algoritmo <i>K-means</i>	71
Tabla 3. Algoritmo <i>Fuzzy C-means</i>	72
Tabla 4. Restricciones para la relación "es" en la forma simple	88
Tabla 5. Restricciones para la relación "tiene" en la forma simple	88
Tabla 6. Restricciones para la relación "hace" en la forma simple	89
Tabla 7. Definición de montaña y colina	98
Tabla 8. Definición de las clases principales en la ontología kaab	111
Tabla 9. Definición de las clases principales descendientes de "objeto geográfico" ...	113
Tabla 10. Definición de las subclases en la ontología kaab	115
Tabla 11. Clases descendientes de "relación geográfica"	118
Tabla 12. Definición de las clases de la ontología del dominio de las formas del terreno	128
Tabla 13. Definición de los conceptos de la clase "forma de terreno"	130
Tabla 14. Conceptos de la clase "sitio de interés"	137
Tabla 15. Conceptos de la clase "propiedad fisiográfica"	138
Tabla 16. Otros conceptos utilizados en la conceptualización	139
Tabla 17. Propiedades de los conceptos de la ontología del dominio de las formas del terreno	141
Tabla 18. Lista de relaciones entre la ontología wiinkil y la ontología hunxeet	147

Índice de Siglas

CDER	Conjunto de Datos Espaciales <i>Raster</i>
DEM	<i>Digital Elevation Model</i> – Modelo Digital de Elevación
DER	Datos Espaciales <i>Raster</i>
INEGI	Instituto Nacional de Estadística, Geografía e Informática
MDT	Modelo Digital del Terreno
OWL	<i>Ontology Web Language</i> – Lenguaje de Ontologías Web
SE	Semántica Espacial
SIG	Sistema de Información Geográfica
SPOT	Satélites de Prueba para la Observación de la Tierra
TIN	<i>Triangulated Irregular Network</i> – Red de Triángulos Irregulares
USGS	United States Geological Survey – Inspección Geológica de los Estados Unidos
ZLE	Zona Llana más Extensa

R esumen

Cuando las personas pensamos espacialmente, no lo hacemos teniendo en mente cosas como coordenadas o proyecciones geográficas; más bien pensamos en cosas que nos sirven como referencias y en las relaciones que esas cosas tienen entre sí. Ante preguntas que tienen una connotación espacial, las personas no contestamos con mapas ni con coordenadas; contestamos con referencias a objetos cuya ubicación es "bien conocida". Por ejemplo, ante la pregunta "¿Dónde está el CIC?", la respuesta de un sistema de información geográfica convencional sería "en 19.50314°N, 99.14759°O", pero una persona nos diría "en Zacatenco" o "cerca del Eje Central". El procesamiento semántico trata de alcanzar un nivel de abstracción parecido al que las personas utilizamos. Este tipo de procesamiento, aplicado a datos espaciales, no depende de escalas, resoluciones, proyecciones ni de ninguna de esas cosas que son importantes en los sistemas convencionales. Consideramos que el primer paso para realizar procesamiento semántico es la descripción semántica de los datos espaciales "crudos", dicha descripción consiste en la identificación de los objetos contenidos en los datos y la ubicación de esos objetos en un marco conceptual dentro del cual obtienen un significado. En este trabajo se presenta una metodología para realizar esta descripción semántica; utilizando como caso de estudio los modelos digitales de elevación. La metodología consta de tres etapas: la conceptualización, en la que se define el marco conceptual para la descripción; la síntesis, en la cual se procesan los datos espaciales "crudos" y se obtienen los objetos espaciales que contienen; y la descripción, en la que se realiza la representación de los resultados de la síntesis en términos del marco conceptual. Los resultados principales del trabajo resuelven dos problemas fundamentales: ¿Qué hay en un conjunto de datos espaciales raster? y ¿Cómo describir un conjunto de este tipo en un formato computacional listo para ser intercambiado, integrado e interoperado con otros similares?

Abstract

When people think spatially, they do not usually consider geographic coordinates nor projections. Instead, they think about things which serve as spatial references and the relations between such things. Facing the questions which have a spatial sense, people do not answer with maps or coordinates, but use some references whose spatial location is "well known". For instance, the answer of a conventional geographic information system to the question "Where is the CIC?" would be "in coordinates 19.50314°N, 99.14759°W". In contrast, a person would answer "in Zacatenco" or "near Eje Central". The semantic processing attempts to enrich an abstraction level similar to the one that people use commonly. This processing, applied to spatial data, does not depend on scales, resolutions, projections or others which are fundamental in conventional systems. We assume that the first step for making semantic processing is the semantic description of "raw" spatial data. Such description is the identification of the objects contained in data and the location of such objects within a conceptual framework, where they get a meaning. In this work, we present a methodology for making this semantic description using as a case study the digital elevation models. The methodology is build up of three stages: conceptualization, to define the conceptual framework of the description; synthesis, to process "raw" spatial data and to obtain the spatial objects contained in data; and description, to generate the representation of results from the synthesis according to the conceptual framework. The main results of this work solve two fundamental problems: Which objects are within a set of raster spatial data?, and How to describe this type of data, in a computational format ready to interchange, integrate and interoperate it with others similar?



Objetivo

Definir un concepto de semántica espacial en el contexto de los datos espaciales *raster*, así como una metodología para obtener una representación semántica de un conjunto de estos datos.

Objetivos particulares

- Definir el concepto de semántica espacial aplicado a los datos espaciales, de tal manera que cualquier conjunto de datos espaciales *raster* pueda ser descrito a través de ella.
- Desarrollar los algoritmos que permitan obtener representación semántica de un conjunto de datos espaciales *raster* mediante la identificación de los rasgos principales que lo componen.
- Utilizar los modelos digitales de elevación como un caso de estudio de los conjuntos de datos espaciales *raster*.

1. Introducción

En este trabajo se describe el desarrollo de una metodología que puede ser utilizada para describir semánticamente los objetos espaciales contenidos en un Conjunto de Datos Espaciales Raster (CDER). El ánimo global es que dicha metodología pueda ser utilizada para describir cualquier tipo de Dato Espacial Raster (DER) semánticamente, aunque como caso de estudio utilizaremos modelos digitales del terreno, particularmente, Modelos Digitales de Elevación (DEM¹). En otras palabras, lo que se pretende hacer es que, dado un CDER, podamos obtener una descripción de los objetos que están representados en él. Por ejemplo, en un DEM se representan las elevaciones del terreno, las cuales a su vez y en su conjunto representan objetos tales como montañas, cerros, valles, llanuras y toda clase de formas de terreno que conforman el paisaje de la región que es representada por el DEM; entonces, lo que debemos obtener es una descripción que enumere los elementos contenidos en el paisaje, algo así como “En este modelo de elevación hay un cerro que tiene una altura de tantos metros y está ubicado en tal sitio; también tiene una llanura de ...”.

Este trabajo tiene su origen en las necesidades planteadas por personal del Instituto Nacional de Estadística, Geografía e Informática (INEGI) a mediados del año 2003 y que tienen que ver con la generación, el procesamiento y, fundamentalmente, con la interpretación de los DEM.

¹ Por las siglas en inglés del término *Digital Elevation Model*, el cual es ampliamente utilizado para describir a este tipo de datos espaciales.

El INEGI produce diversos tipos de datos geográficos (INEGI, 2000b), entre los que destacan los archivos de datos digitales que representan a su vez diversos objetos geográficos. Para fines prácticos, tales datos son agrupados en tres clases:

- **Conjuntos de Datos Vectoriales.** Corresponden a la presentación digital de los mapas que tradicionalmente ha elaborado el INEGI: consignan los rasgos u objetos geográficos mediante una representación de puntos, líneas y polígonos, separados por temas en diferentes capas de información tales como, vías de comunicación, localidades, hidrografía, curvas de nivel, etcétera.
- **Conjuntos de Datos Raster.** Corresponden a datos en formato *raster*², en los que se incluyen las ortofotos y los DEM. La estructura de estos archivos es de un arreglo matricial de valores de un atributo particular; para el caso de imágenes, los valores son de la reflectancia del terreno para cada elemento de imagen, o bien, de valores de altura cuando se trata de los DEM.
- **Conjuntos de Datos Alfanuméricos.** Corresponden a archivos de texto con diferentes atributos considerados de interés, relativos a los diferentes rasgos existentes en los conjuntos de datos espaciales. En esta clase, se incluyen archivos de nombres geográficos (topónimos y localidades), puntos geodésicos, puntos de muestreo para mapas de recursos naturales, descripciones de unidades temáticas, etcétera.

Los diferentes archivos digitales mencionados provienen de información generada para diferentes niveles de generalización o de especificidad, según se quiera ver.

² Los sistemas raster basan su funcionalidad en una concepción implícita de las relaciones de vecindad entre los objetos geográficos. Su forma de proceder es dividir la zona en una retícula o malla regular de pequeñas celdas (llamadas píxeles) y atribuir uno o varios valores numéricos a cada celda como representación de su valor temático. La localización de cada celda es implícita, dependiendo directamente del orden que ocupa en la rejilla, a diferencia de la estructura vectorial en la que se almacena de forma explícita la topología. Dado que la malla es regular (el tamaño del píxel es constante) y que conocemos la posición en coordenadas del centro de una de las celdas, se puede decir que todos los píxeles están geo-referenciados.

La necesidad de que usuarios que no son expertos puedan tener acceso a la interpretación o al análisis de este tipo de datos, es lo que ha dado lugar a que se pueda conocer el contenido de información útil en los datos, es decir su semántica.

1.1. Semántica espacial

Para alcanzar tal objetivo, planteamos que la parte más importante es la definición de la Semántica Espacial (SE) en el contexto de los DER. En el grupo de trabajo del Laboratorio de Procesamiento Inteligente de Información Geoespacial (PIIGLab, antes Geoprocesamiento) del Centro de Investigación en Computación del Instituto Politécnico Nacional hemos desarrollado diferentes aproximaciones al concepto de SE (Quintero y otros, 2002; Quintero y otros, 2003a; Quintero y otros, 2003b; Torres y otros, 2004; Torres y otros, 2005; Moreno, 2007; Torres, 2007). Este concepto ha sido trabajado desde un punto de vista general, es decir, se pretende definir cómo se debe hacer la descripción semántica de cualquier tipo de información geoespacial. Así bien, desprendiendo de los trabajos mencionados, podemos dar una primera aproximación a una definición de SE de la siguiente manera:

La semántica de una colección de objetos espaciales está dada por la definición y/o descripción de estos, de acuerdo a la conceptualización del dominio en el cual se procesan dichos objetos.

Tomando como base la definición anterior, resulta evidente que es necesario especificar una conceptualización detallada para cada caso de estudio particular. Por ejemplo, (Torres, 2007) presenta la aplicación del concepto de SE para el caso de conjuntos de datos espaciales vectoriales. En dicho trabajo se plantea la conceptualización de un conjunto de objetos espaciales mediante la generación de una ontología del dominio geográfico y una ontología de aplicación (turismo en Acapulco) que interactúan para llegar a una descripción semántica de una base de datos espacial.

Otro caso de estudio lo describe (Moreno, 2007), quien plantea el uso de la SE para controlar la calidad del proceso de generalización automática de mapas, en el cual el conjunto de características semánticas lo constituyen las relaciones espaciales entre

objetos que no deben ser cambiadas por el proceso de generalización, por ejemplo que una ciudad está al norte de un río. Conceptualizando las relaciones espaciales entre los objetos generalizados, se controla la calidad del proceso.

De acuerdo con lo anterior, proponemos que para realizar una representación semántica de datos espaciales son necesarias tres etapas: conceptualización, síntesis y descripción.

1.1.1. Conceptualización

El primer paso para realizar una representación semántica de DER es contar con una conceptualización detallada, la cual habrá de constar de tres partes fundamentalmente: una conceptualización del dominio geoespacial (de alto nivel), la conceptualización del dominio particular del CDER con el que se está trabajando³ y, por último, la conceptualización de la aplicación por medio de la cual se realizará la representación. Estas conceptualizaciones son representadas a través de sendas ontologías.

1.1.2. Síntesis

Este proceso es guiado por la misma conceptualización, en la que se cuenta con diferentes algoritmos (como se verá más adelante, dichos algoritmos forman parte de la conceptualización de la aplicación) para extraer las características del conjunto de datos espaciales, de acuerdo al dominio en el que éstos son analizados.

1.1.3. Descripción

Como resultado del proceso de síntesis se obtienen “extractos”, que no son otra cosa que fragmentos de los datos espaciales identificados como de alguna clase particular. El proceso de descripción determina qué son esos extractos y los representa de acuerdo a la conceptualización. Se propone utilizar, además de la descripción textual o codificada de cada objeto, formas geométricas de referencia de cada tipo

³ En nuestro caso, es una conceptualización de las formas del terreno contenidas en modelos digitales de elevación.

de objeto conceptualizado con el fin de construir una representación sintética de los DER basada en su semántica.

En la Figura 1 se ilustran las etapas mencionadas en los párrafos anteriores, las cuales conforman la metodología propuesta para la representación semántica de datos espaciales *raster*.

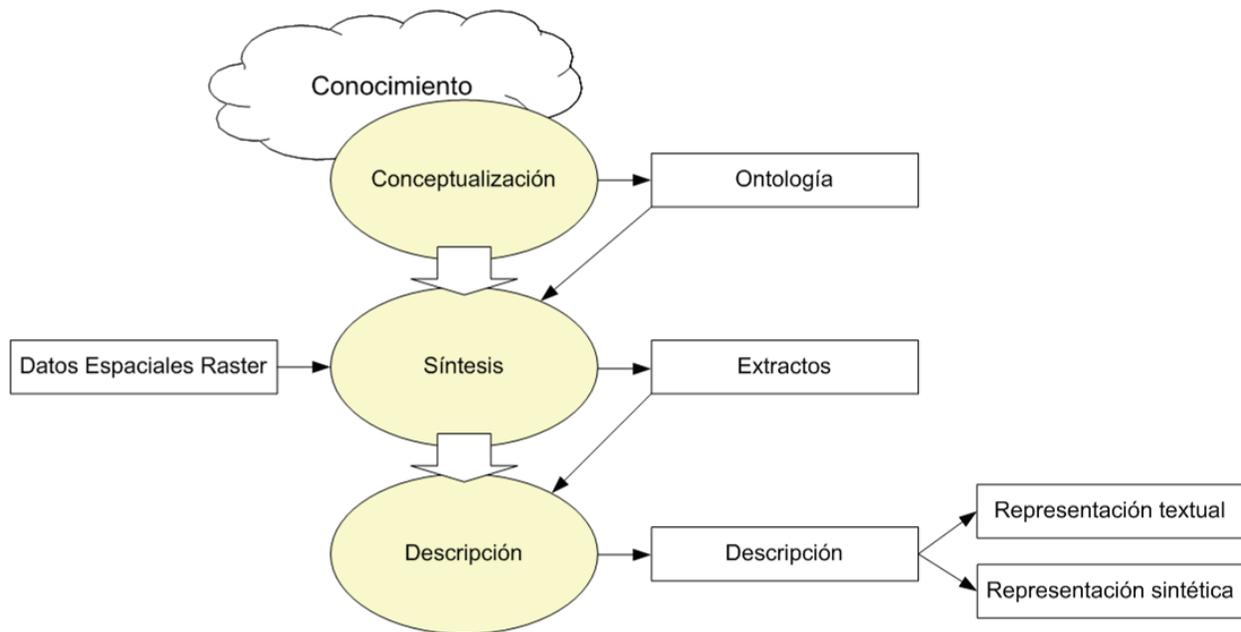


Figura 1. Metodología para la representación semántica de datos espaciales *raster*

1.2. Justificación

Se ha mencionado que este trabajo surge de una necesidad del INEGI: la de generar y procesar los DEM. De la misma fuente, sabemos que no cuentan con una metodología para realizar esta tarea, por lo que todo lo que se haga a este respecto puede resultar valioso.

En el grupo de trabajo de PIIGLab se ha trabajado desde hace algún tiempo sobre la SE. Como ya se mencionó, este concepto está siendo definido desde un punto de vista general. En este trabajo se estudia la aplicación de los conceptos de SE (generados por el grupo de trabajo) a un tipo de datos espacial particular. Los DER que describen la forma de una sección de la superficie terrestre, son ampliamente utilizados

para análisis hidrológicos, prevención de desastres y planeación urbana, entre otras aplicaciones prácticas.

Los conceptos de SE aplicados a los DER tiene varios propósitos. El primero de ellos es servir como una línea de trabajo dentro del marco general que se ha creado en torno a la SE en el PIIGLab. De esta manera, se pueden demostrar los conceptos planteados sobre SE a través de su aplicación para tipos de datos espaciales particulares. Por otro lado, la utilización de modelos de elevación como caso de estudio permite abrir un camino hacia la aplicación de conceptos de SE a otros tipos de DER. Finalmente, el tratamiento semántico de DER podrá ser aplicado para el análisis de zonas de riesgo (inundaciones, incendios, etc.), el cual es objeto de estudio en la investigación aplicada de nuestro grupo de trabajo. El análisis semántico de DER presenta algunos problemas interesantes que pueden ser resueltos mediante la aplicación de algunas metodologías que normalmente no han sido utilizadas en el campo de la geoinformática.

Entre los problemas fundamentales que supone la aplicación de la semántica espacial a los DER, podemos mencionar en primer lugar la definición de la forma que debe tener la representación de la semántica de los DER, problema que es atacado mediante la conceptualización en sus diferentes niveles. Un segundo problema es el proceso de síntesis, que es el proceso que se ha de aplicar para “separar” los objetos contenidos en un CDER. Obviamente, el proceso de síntesis dependerá del dominio de la aplicación específica (caso de estudio), pero es necesario mencionar que tal síntesis involucra múltiples disciplinas del área de la computación (reconocimiento de patrones, procesamiento de imágenes, lógica difusa, por mencionar algunas). Por ejemplo en (Quintero y otros, 2002; Quintero y otros, 2003a; Quintero y otros, 2003b), se presentan algoritmos para procesar y sintetizar datos provenientes de DEM con un enfoque geométrico. Su aplicación en la generación de DEM va desde la conversión de curvas de nivel a un formato *raster*, hasta su aplicación en la minimización de errores locales ajustando puntos de control. En esos mismos trabajos, hemos introducido algunas ideas para realizar análisis semántico de DER, entre las cuales se encuentra el

uso de funciones recursivas⁴ para realizar dicho análisis. Dicha aplicación puede traer consigo algunas ventajas significativas sobre los modos convencionales de descripción de DER. La primera de ellas radica en que es posible alcanzar una granularidad muy pequeña en la descripción y al mismo tiempo tener una descripción compacta (en cuanto a la cantidad de datos se refiere); con lo que sería posible comprimir la enorme cantidad de datos que llegan a tener los CDER.

El resto del trabajo está organizado de la siguiente manera. El capítulo 2 presenta una recopilación y análisis de los trabajos relacionados con el procesamiento y descripción semántica de los datos espaciales, en general, y de los DER en particular. Se presentan temas como: modelos del terreno, geomorfometría y conceptualización, entre otros.

En el capítulo 3 se presentan algunos temas que son utilizados más adelante en la descripción de la metodología, por lo que se requiere describirlos para colocar en contexto las ideas en las que se basa la metodología. Se presentan temas como: ontologías y relaciones matemáticas, entre otros.

En el capítulo 4 presentamos detalladamente cada una de las etapas de la metodología propuesta. En la etapa de conceptualización proponemos una metodología para la conceptualización del dominio geográfico, basada en la premisa de reducir el número de elementos axiomáticos. La etapa de síntesis va muy de la mano con la etapa previa, puesto que involucra la conceptualización de los algoritmos utilizados para segmentar los DER. Esta conceptualización puede entenderse como la definición de cada posible resultado de los algoritmos que se utilizan para segmentar. En la etapa de descripción se establece cómo se deben interpretar los datos arrojados por los algoritmos de segmentación y cómo se lleva a cabo la representación semántica de los DER.

⁴ La aplicación de las funciones recursivas puede ser vista como un modelo de aproximaciones sucesivas, en el que cada iteración es registrada como una firma de operación. Bajo este esquema, las aproximaciones pueden ser refinadas cada vez más adaptándose al contexto del terreno, para así lograr una mejor aproximación a éste.

El capítulo 5 trata sobre los resultados de aplicar la metodología propuesta al caso de estudio, que en este trabajo se trata de los DEM. Se conceptualizan, sintetizan y describen cosas como: cerros, montañas, colinas, etc.

Finalmente, en el capítulo 6 se presentan las conclusiones, los logros, los alcances y las limitaciones de la metodología propuesta. Asimismo, se presentan las tareas pendientes en el desarrollo de la metodología y la visión de los trabajos futuros para la extensión de la misma.

2. Estado del arte

En este capítulo se presentan las revisiones de algunos trabajos relacionados con el presente. Lo que aquí se presenta es un estudio del desarrollo teórico de la representación de las formas de terreno, así como de los problemas relacionados con ésta.

2.1. Modelos del terreno

En este trabajo presentamos una manera de realizar la representación de DER semánticamente, para lo que utilizaremos un tipo particular de DER como caso de estudio. Como mencionaremos después, los modelos digitales de elevación⁵ son una rica fuente de datos espaciales *raster*, pues están disponibles de manera más o menos accesible⁶. Así pues, al utilizar los DEM como caso de estudio de los DER, abordamos en esta sección el estado del arte respecto a los modelos digitales del terreno, en general, y de los DEM en particular.

La superficie de la Tierra no es un caos cuando es vista a través de la conceptualización humana. Las pendientes colgantes son extremadamente raras, lo cual implica que, casi sin pérdida de generalidad, la elevación de la superficie de la Tierra relativa al geode puede ser modelada o conceptualizada científicamente como una función univaluada de la posición horizontal, esto es, como un campo continuo.

⁵ Conocido como DEM por las siglas del inglés *Digital Elevation Model*

⁶ Ejemplo de ello es **USGS**. (2007a). "Digital Elevation Models." en www.usgs.gov.

Generalmente hablando, es esto lo que hacen los científicos para modelar la geometría de la superficie de la Tierra. El hecho de que la superficie de la Tierra pueda ser representada por una función de este tipo ha tenido implicaciones para la cartografía científica; los mapas del tipo con el cual estamos familiarizados son posibles gracias a que la superficie de la Tierra misma se aproxima mucho al estado de la bidimensionalidad. Los primeros mapas topográficos que utilizaron isolíneas llamadas 'curvas de nivel' aparecieron alrededor del año 1800, y una teoría matemática de la topología como superficies continuas y suaves univaluadas fue presentado después en el artículo de Cayley en 1859, llamado "*On contour lines and slope lines*" y en la extensión que en 1870 hizo Maxwell al trabajo de Cayley, en el artículo llamado "*On hills and dales*". La idea de un campo de elevaciones univaluado como una representación de la forma de la Tierra ha sido incorporada implícita o explícitamente en las representaciones de las formas de la tierra desarrolladas por computadora desde los años 50's (Wood, 1997).

Los mapas comúnmente contienen elementos de ambos puntos de vista: tanto basado en campos (al describir la superficie de la tierra con base en elevaciones) y basado en objetos (por medio de una teoría primaria). Esto resulta ser especialmente evidente cuando miramos cómo se representan las montañas en los mapas, lo que típicamente se hace a través de curvas de nivel. Los cartógrafos comúnmente ponen los nombres de la montaña en sus mapas, en la vecindad de la colina o de alguna parte correspondiente a las curvas de nivel que indican la forma de la superficie de la Tierra en el vecindario de la etiqueta. Pero entonces ellos dejan al usuario del mapa inferir la extensión del objeto al cual se refiere el nombre que aparece en la etiqueta (Smith y Mark, 2003).

Cuando las personas ven, aprenden y describen una escena, parece que no piensan en la escena como un campo de elevaciones en el sentido científico. En lugar de ello ven en la escena objetos sueltos, que tienen superficies completamente cerradas, haciéndolos móviles al menos en principio, y objetos anclados, que son los objetos grandes, o forman parte de las capas de la superficie. La superficie percibida de la Tierra parece ser poblada por aquellos objetos muy grandes anclados los cuales

llamamos formas del terreno. Las partes convexas son mayormente salientes, aunque regiones cóncavas (agujeros) como los valles y los cráteres aparecen como objetos en muchos contextos (Mark y Turk, 2003?).

En algunos dominios de aplicación como el modelado ambiental y los sistemas de información geográfica, la principal infraestructura de datos, llamada superficie topográfica o datos de elevación, ha sido obtenida digitalizando mapas que estaban en papel. Esfuerzos tempranos para representar la topografía en las computadoras intentaron utilizar las curvas de nivel, pero hacer cálculos acerca de las superficies con base en datos de curvas de nivel resultó ser ineficiente, por otro lado las rejillas regulares, látrices, y matrices de altitud probaron ser más efectivas (Etzelmüller y Sulebak, 2000). Después, se volvió popular el representar superficies dentro de los sistemas de información geográfica como conjuntos de triángulos no sobrepuestos en redes irregulares de triángulos. Una ontología formal de las superficies de elevación debe estar relacionada con los modelos de datos utilizados para representar la topografía en los sistemas de información geográfica actuales. En este trabajo utilizaremos modelos digitales de elevación como caso de estudio, aunque existen diferentes tipos de CDER en los cuales la metodología que se presenta en este trabajo puede ser aplicable.

Así, podemos mencionar que el Modelo Digital del Terreno (MDT) es una estructura numérica de datos que representa la distribución espacial de una variable cuantitativa y continua. Se utiliza el término MDT, para designar al conjunto de datos numéricos que describe la distribución espacial de una característica del territorio (Felicísimo, 2004). Lo anterior permite que la característica a representar sea diferente de la altitud, aunque es habitual identificar los MDT con los modelos digitales de elevación cuando, en realidad, pueden ser representadas muchas otras propiedades del terreno. En este sentido, parece deseable diferenciar claramente qué variable se está representando en el modelo (Etzelmüller y Sulebak, 2000).

Existen dos clases de MDT, los vectoriales y los *raster*; a continuación presentamos algunos de estos modelos (Felicísimo, 2004).

- **Curvas de nivel digitales.** La estructura básica es el vector, compuesto por un conjunto de pares de coordenadas que describe la trayectoria de líneas isométricas. El número de elementos de cada vector es variable. Una curva de nivel concreta queda definida mediante un vector ordenado de puntos que se sitúan sobre ella a intervalos no necesariamente iguales. La localización espacial de cada elemento es explícita, conservando los valores individuales de coordenadas. El caso más sencillo está constituido por el conjunto de las curvas de nivel que pasan por la zona representada, separadas generalmente por intervalos constantes de altitud.

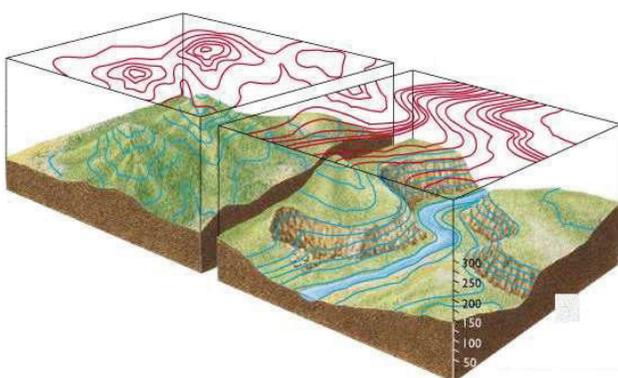
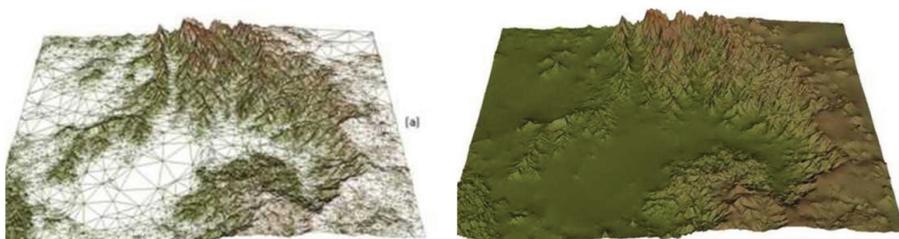
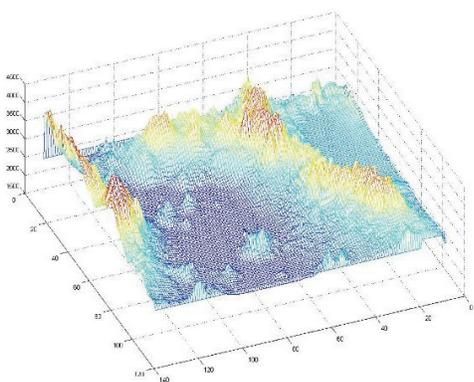


Figura 2. Curvas de nivel digitales

- **Redes de triángulos irregulares.** Modelo vectorial que se compone de un conjunto de triángulos irregulares adosados y que suele identificarse por las siglas en inglés TIN (*Triangulated Irregular Network*), (Peucker, 1978). Los triángulos se construyen ajustando un plano a tres puntos cercanos no colineales, y se aproximan sobre el terreno formando un mosaico que puede adaptarse a la superficie con diferente grado de detalle en función de la complejidad del relieve. Se trata de una estructura en la que el terreno queda representado por el conjunto de superficies planas que se ajustan a una estructura anterior de puntos. Los TIN pueden considerarse como una estructura derivada de otra anterior de puntos o líneas. Aunque la distribución original puede ser cualquiera (incluso puntos distribuidos aleatoriamente), es frecuente partir de una base de curvas de nivel para generar la red de triángulos.

**Figura 3. Redes de triángulos irregulares**

- **Matrices regulares.** Esta estructura es el resultado de superponer una retícula sobre el terreno y extraer la altitud media de cada celda; aunque habitualmente se utiliza un valor puntual asociado a cada nodo de la retícula o punto medio de la celda. La retícula puede adoptar formas variadas, pero la más utilizada es una red regular de malla cuadrada con filas y columnas igualmente espaciadas. En esta estructura, la localización espacial de cada dato está implícitamente determinada por su situación en la matriz, una vez que son definidos su origen y el intervalo entre filas y columnas. Las matrices de altitudes suelen ser generadas por interpolación a partir de un modelo previo de contornos o por métodos fotogramétricos (Ver siguiente sección).

**Figura 4. Matrices regulares**

- **Matrices de resolución variable.** El interés de las matrices de resolución variable reside en la posibilidad de solucionar el principal problema de las matrices regulares: su resolución espacial prefijada, manteniendo, en principio, sus principales ventajas: la sencillez conceptual y operacional. En este tipo de matrices los elementos pueden ser, bien datos elementales (como en las matrices regulares),

bien submatrices con un nivel de resolución diferente. La estructura final es un árbol jerárquico y dinámico de submatrices con una profundidad en principio arbitraria y cuya resolución espacial se multiplica en cada nivel.

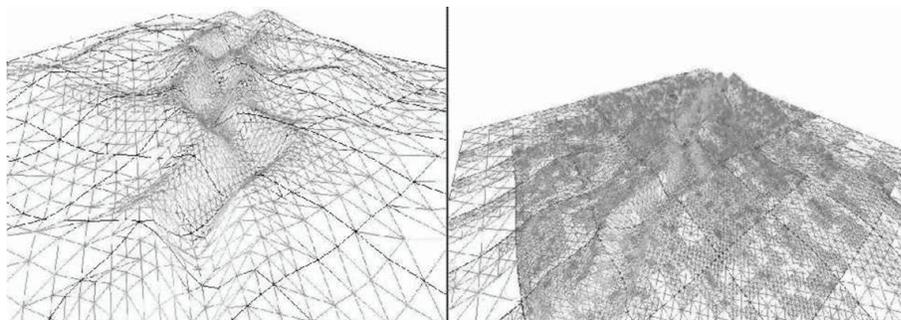


Figura 5. Matrices de resolución variable

2.1.1. Modelos digitales de elevación

Aunque como vemos existen otros tipos de MDT, el más ampliamente utilizado es el DEM. Aunque el término es usado en la literatura de manera inconsistente (Burrough, 1986; Weibel y Heller, 1991), para este trabajo tomaremos la definición de (Burrough, 1986): “una representación de la variación continua del relieve sobre el espacio, por medio de una matriz de valores relacionados con una rejilla regular”. No se proporciona ninguna otra información acerca de la superficie del terreno.

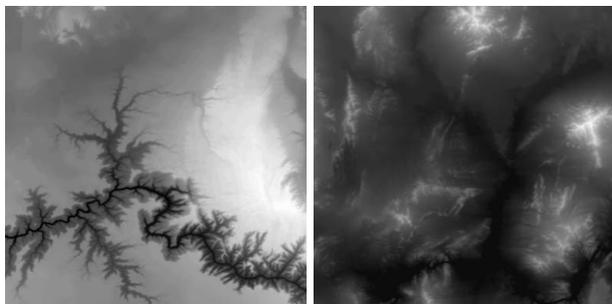


Figura 6. Modelos digitales de elevación

El DEM es un modelo de superficie para representaciones de terrenos. Es una aproximación a la superficie topográfica basada en un conjunto de valores de elevación dados en un conjunto finito de puntos en la superficie que se desea modelar. En una malla rectangular los nodos están impuestos por un criterio fijo de

subdivisión. Un DEM puede describirse de forma genérica como $z = \zeta(x, y)$, donde z es la altitud del punto situado en las coordenadas x y y , y ζ es la función que relaciona la variable con su localización geográfica (Felicísimo, 2004). Los valores de x y y suelen corresponder con las abscisas y ordenadas de un sistema de coordenadas plano, habitualmente un sistema de proyección cartográfica. La ecuación anterior representa una superficie en la que la altitud es una variable continua. Dado que esta superficie está formada por un número infinito de puntos no es posible su modelado sin cierta pérdida de información, proceso equivalente al de la generalización cartográfica en los mapas convencionales. La simplicidad de tener una matriz con valores de elevación fue una de las razones por las cuales los DEM fueron tomados como estándar de facto para el análisis del terreno a principios de los 70's (Evans, 1972). Más recientemente, continúa siendo ampliamente utilizado como modelo de la superficie debido a su fácil integración en los modernos Sistemas de Información Geográfica (SIG) (Weibel y Heller, 1991).

Pero la utilización de los DEM como modelo de la superficie trae consigo algunos problemas. Primero, la *fidelidad* con la cual el DEM describe la superficie del terreno depende tanto de la naturaleza del terreno (rugosidad, accidentes, etc.), como de la resolución espacial del modelo. Lo anterior quiere decir que existe una dependencia de la escala en el análisis que se realice con los DEM. Se han realizado trabajos respecto a esta dependencia de la escala (Ackermann, 1993; Garbrecht y Martz, 1993; Hodgson, 1995), y sin embargo, muchas veces no son considerados cuando se realiza el análisis del terreno con los DEM. Un segundo problema con los DEM, es que no se proporciona información respecto a las relaciones que existen entre los valores contenidos en el modelo; es decir, que intuitivamente podemos considerar que existe una relación lineal entre dos valores adyacentes, pero en la realidad la relación debería ser alguna otra. (Kumler, 1994) sugiere que una interpolación lineal entre celdas adyacentes en un DEM produce residuos más pequeños al compararlos con el modelo TIN de la misma área. Además, reconoce que diferentes procedimientos de interpolación ocasionan diferentes estimaciones de altura, lo que lleva a considerar el

uso de interpoladores cuadráticos y cúbicos para aplicaciones donde se requiere un modelo continuo de la superficie (Quintero y otros, 2002).

Durante las últimas décadas los avances de las tecnologías de información y del poder de cómputo han favorecido el desarrollo de la tecnología de los sistemas de información geográfica en general, y de los DEM en particular. Esto ha hecho que los sistemas de información geográfica sean una herramienta invaluable para modelar, analizar y visualizar fenómenos geomorfológicos en la mayoría de los proyectos al respecto. En este contexto, los DEM juegan un papel importante debido a que los procesos geomorfológicos que dan forma a la superficie de la Tierra, son altamente gobernados por la gravedad y por lo tanto por la topografía. Así, existen relaciones cercanas entre ciertas características topográficas, respuestas a procesos, tipos de sedimentos y su distribución en el espacio. Las características topográficas son elementos importantes para el entendimiento y modelado de las formas del terreno y de su distribución en el espacio. Actualmente, los DEM se encuentran a la mano en una amplia gama de escalas, lo cual abre la posibilidad al modelado espacial de la distribución de las formas (Quintero y otros, 2003a).

Con los DEM se pueden realizar diferentes tipos de análisis topográfico; se distinguen seis tipos básicos de análisis:

- las relaciones entre las características del terreno y los procesos geomorfológicos.
- análisis de la escala.
- análisis de los cambios en la superficie.
- análisis del flujo o movimientos de superficie.
- análisis utilizando técnicas de visualización.
- modelos topográficos.

En las siguientes secciones se detalla cada uno de estos puntos.

Geomorfometría

La idea básica en la geomorfometría es una relación cercana entre procesos superficiales y características topográficas especiales, representados como parámetros topográficos. La combinación de varios parámetros topográficos define regiones topográficas (unidades de relieve) después de un proceso de clasificación (clasificación del relieve). Esas unidades de relieve son consideradas para representar el dominio de ciertos procesos superficiales o formas del terreno. Si se pudiera establecer una relación empírica o física entre un parámetro topográfico (regiones) y un proceso superficial (forma del terreno), la geomorfometría puede ser utilizada como herramienta para el modelado espacial. Si además asumimos que las relaciones son independientes de la escala, al menos en cierto rango, es posible bajar o subir la escala de la información espacial (Wood, 1997).

La altitud de la superficie describe la función de rugosidad de la superficie, la cual es continua en cada punto. La superficie puede ser descrita conceptualmente como una combinación de funciones periódicas. La longitud de onda más pequeña (textura) se refiere a las formas más pequeñas que pueden ser identificadas en un modelo de elevación a esa escala. La longitud de onda más grande se refiere a los mayores componentes de un terreno, tales como valles o montañas (grano). Las relaciones entre las formas pequeñas y grandes en el terreno describen la rugosidad de la superficie. En términos de funciones periódicas, la rugosidad describe qué tanto de la variación del terreno es representada por las diferentes longitudes de onda. La relación entre la distribución de la altitud y el área de la superficie modelada es descrita por la distribución hipsográfica de la masa. Los parámetros relacionados a este concepto son normalmente estimados con base en matrices de altitud (Etzelmüller y Sulebak, 2000).

La clasificación topográfica da como resultado unidades homogéneas del relieve, las cuales tienen una respuesta predecible con respecto a la distribución de las formas del terreno y las tasas del proceso. Se han propuesto diferentes tipos de procesos de clasificación como definiciones subjetivas de jerarquías de las formas del terreno (Hammond, 1964), firmas geométricas (Pike, 1988; Weibel y DeLotto, 1988; Friedrich, 1996; Sulebak y otros, 1997), definiciones de elementos de formas del terreno (Dikau,

1989; Dikau, 1990; Dikau, 1994) y técnicas de agregación de escala (Dikau, 1994; Sulebak y otros, 2000).

La clasificación topológica involucra además la identificación y delineación de ciertas formas del terreno, como ciertas áreas de drenaje (Dikau, 1989), características de valles y colinas (Graff y Usery, 1993), formas glaciales (Evans, 1984) y pendientes (Romstad, 2001). La identificación de las formas del terreno es realizada utilizando cualquier algoritmo de combinación estadística de clases o a través de ciertos algoritmos de emparejamiento (*matching*). Además, las técnicas de clasificación se han utilizado para identificar restricciones topográficas para la distribución de ciertas formas del terreno.

Escala

La forma de la superficie de la Tierra es el producto final de un sistema de procesos que interactúan recíprocamente, que funcionan a la vez en una variedad grande de escalas espaciales y temporales. La parte central del problema de la escala espacial en el proceso geomorfológico es que el sistema geomorfológico se debe ver en su complejo contexto jerárquico. Es decir, cada sistema consiste en un arreglo de sistemas de nivel inferior más pequeños y es al mismo tiempo parte de una secuencia de sistemas más grandes, de alto nivel. (Dikau, 1989) define tal espacio jerárquico en los diversos niveles o tipos de relieve, extendiéndose del pico- hasta el mega- relieve. En un sistema geomorfológico las diferencias ocurren en la sensibilidad de la morfología al cambiar la condición del proceso. En la situación ideal de un escenario geológico uniforme, el dominio de la evolución del paisaje por fenómenos de alta frecuencia y procesos de baja magnitud, daría lugar a un paisaje liso que tiene irregularidades en reducida escala, mientras que la dominación por fenómenos de baja frecuencia y procesos de gran magnitud, produciría un paisaje escabroso que tiene irregularidades de gran escala (De Boer, 1992).

La escala es el término de uso frecuente para describir el nivel de detalle geográfico, pero su significado se confunde generalmente en un mundo geográfico digital. Muchos fenómenos geográficos, como la topografía de la Tierra, son

infinitamente complejos, así que cualquier tentativa de capturar su patrón en la información geográfica debe implicar la aproximación (Goodchild, 1997). Así, la mayoría de la información geográfica debe tener un nivel explícito o implícito del detalle geográfico. Limitando el nivel del detalle de un conjunto de datos nos aseguramos de que la descripción es posible, en costos razonables, y de que el volumen de datos es manejable. La idea de que la superficie de la Tierra revela más detalle conforme se le mira más de cerca, fue formalizada por (Mandelbrot, 1982) con sus conceptos de fractales. Otros han procurado formalizar el proceso inverso –la pérdida de detalles conforme uno mira la superficie menos de cerca– con teorías y métodos de generalización, de circunvolución y de agregación (Buttenfield y McMaster, 1991).

Claramente, el nivel del detalle geográfico de los datos afecta el resultado del análisis de una forma que es difícil de predecir, particularmente en la geomorfología cuantitativa (Openshaw, 1983). Una pregunta recurrente en geomorfología es el grado al cual una generalización se puede decir que es cercana a la realidad (Salge, 1995). La tarea del especialista es representar la topografía en un nivel específico de detalle, utilizando los modelos y las especificaciones para ese nivel, con el cual se debe poder describir los tipos específicos de características de las formas del terreno bajo estudio. En un nivel más tosco, estas características pueden ser descritas, lo que puede conducir a una mala interpretación y errores en los resultados.

En cartografía tradicional el nivel de detalle es especificado por la fracción representativa o el cociente de la distancia en el mapa a la distancia en la Tierra, como la escala, la escala métrica o escala del mapa. Un mapa de escala grande (1:1000, 1:500) implica indirectamente un alto detalle geográfico, mientras que una escala pequeña del mapa (1:100000, 1:1000 000) implica indirectamente un bajo nivel de detalle geográfico. Un mapa de escala grande puede representar generalmente un área más pequeña en la superficie de la Tierra. Esto, sin embargo, es solamente debido al tamaño práctico de una hoja del mapa. En teoría no existe ningún límite para que un mapa cubra un área grande con alto nivel de detalle (Wood, 1997).

Diversos significados o definiciones se relacionan con el concepto de escala o nivel del detalle geográfico. Muchos de éstos han sido aplicados en el análisis geomorfológico de diversas maneras. Como ejemplo, la teoría variable regional dentro de la geoestadística ofrece una herramienta para investigar la escala de superficies topográficas. La dependencia espacial en un conjunto de datos topográficos se puede considerar también como medida de rugosidad. Una dependencia (autocorrelación) espacial fuerte significa rugosidad superficial pequeña y relieve liso y viceversa. En el análisis geomorfométrico, la granularidad topográfica se define como la longitud de onda más grande de la topografía dentro de un área (Evans, 1972; Mark, 1975; Pike y Rozema, 1975; Pike y otros, 1989). Sobre esta distancia, los parámetros topográficos deben ser calculados. Esto se puede definir como una medida de la escala topográfica. Se ha investigado sobre la naturaleza fractal de la topografía y en estimar la dimensión fractal de diferentes paisajes (Mark y Aronson, 1984; Huang y Trucotte, 1990; McClean, 1991; Polidori y otros, 1991; Klinkenberg y Goodchild, 1992; Outcalt y Melton, 1992). (Chase, 1992) utilizó análisis fractal en el estudio de la evolución de paisajes de gran tamaño. (Sulebak, 1999) analizaba el comportamiento fractal de los paisajes para propósitos hidrológicos, mientras que (Evans y McClean, 1995) indicaron que la topografía nunca puede ser autoafín o de comportamiento autosimilar. El análisis de dimensiones fractales de la topografía es de gran interés, especialmente para el desarrollo de la metodología ascendente y descendente.

Análisis de cambios superficiales verticales

Hay una relación dinámica entre las formas del terreno y los procesos que los construyen, modifican o destruyen. Los procesos incluyen el desplazamiento de material superficial de la Tierra, que varía en lo referente a velocidad, dirección, cantidad de material involucrado y frecuencia o período del proceso (Brunsden, 1996). Esto cambia gradualmente la topografía de la forma del terreno. Por lo tanto, el tipo y la magnitud de los cambios superficiales dan información valiosa sobre el tipo, magnitud y frecuencia de los procesos implicados. El análisis de los cambios verticales de la superficie da como resultado las medidas de la magnitud de dichos cambios y cómo se distribuyen espacialmente.

Cambio de la altitud. En un sistema dinámico, el cambio de la altitud es la suma de la masa local y de la componente vertical de la velocidad.

Cambio de pendiente. El cambio de pendiente exhibe el levantamiento o lo plano de una superficie. En una escala local el patrón espacial de la distribución de la pendiente cambia y su magnitud da información valiosa para describir acontecimientos de futuros movimientos.

Cambio de la rugosidad superficial. La cuantificación de los cambios de rugosidad (la curvatura) muchas veces suceden debido a que ocurren acontecimientos geomorfológicos que conducen a los cambios superficiales rápidos o de gran magnitud. Esto es causado por la alta sensibilidad de las medidas de rugosidad a la calidad del DEM.

Propagación de error. Muchos investigadores no han observado el peligro de usar curvas de nivel convertidas a DEM, y magnifican las preocupaciones donde la estadística se utiliza para describir los DEM y sus derivados, además de cuando los DEM se utilizan en simulaciones de proceso. Muchos usuarios aceptan a los DEM sin criticar y hay una necesidad de acentuar de nuevo que los usuarios deben determinar la calidad de los DEM que utilizan. Teniendo en cuenta las leyes de propagación del error, el calcular diferencias es un procedimiento problemático, pues los errores de los conjuntos de datos originales se incrementan durante el cálculo. Esto quiere decir que los errores relativos pueden llegar a ser muy grandes, lo que exige la exactitud de los datos básicos tiene que ser muy buena (Burrough, 1986; Etzelmüller y Sulebak, 2000).

Visualización y mapeo interactivo

El desarrollo de mayores capacidades tanto gráficas como de procesamiento que han tenido las computadoras, abren la alternativa del análisis mediante la visualización interactiva de las formas del terreno en escalas variables. Las facilidades estándar encontradas en muchos sistemas de información geográfica han sido rebasadas rápidamente en términos de flexibilidad y fotorealismo por otras cada vez más disponibles de sistemas tridimensionales realistas (Etzelmüller y Sulebak, 2000).

Las técnicas de visualización de la topografía se han aplicado en una variedad de estudios científicos, como por ejemplo dentro de la silvicultura o del estado meteorológico. Dentro de la geomorfología, los mapas de relieve sombreados fueron utilizados en el mapeo manual y en la identificación de las formas del terreno (Quintero y otros, 2003b). Hay un alto potencial en la utilización de la visualización interactiva ("vuelo virtual"): en la geomorfología estructural de gran escala y en la visualización de los DEM, se puede utilizar esta técnica en vez de fotos aéreas o mapas topográficos para localizar las formas estructurales de la Tierra (Quintero y otros, 2002).

Esto ha conducido a requisitos para el desarrollo y la incorporación de las nuevas técnicas para la visualización geográfica. La información geográfica sigue basándose en mapas de dos dimensiones tradicionales. Aunque es apropiado para la comunicación de muchos tipos de información espacial, el mapa tradicional y las técnicas cartográficas tradicionales tienen limitaciones (Etzel Müller y Sulebak, 2000). Con la nueva visualización geográfica (animación, vídeo y realidad virtual) las restricciones impuestas por los geógrafos al mapa de dos dimensiones deben ser eliminadas (Quintero y otros, 2003a). Aunque técnicas de visualización avanzadas recientemente se integran más completamente dentro de los SIGs de hoy, lo que no se ha incorporado actualmente es la animación y la visualización del cambio dinámico y la representación realista de paisajes reales.

No sólo la visualización es necesaria: la animación es también un componente muy importante para mostrar aspectos dinámicos de modelos geomorfológicos. La animación es una técnica con gran potencial para la visualización en una gran variedad de aplicaciones geográficas. Incorpora el reciente desarrollo del hardware, de los gráficos por computadora, del modelado 3D y de las tecnologías de información geográfica. La animación puede ser definida como la creación de la ilusión del movimiento o del cambio mediante la presentación de una serie de cuadros sencillos, como en las películas o vídeos (Quintero y otros, 2002). En los pasados treinta años han sido pocos los ejemplos de animación geográfica, debido en gran parte a la complejidad de su creación y a la fijación de los geógrafos por los mapas impresos. Recientes contribuciones en animación geográfica han intentado formular una

fundamentación conceptual para el uso de la animación dentro de la geografía (Quintero y otros, 2003a). La animación geográfica se utiliza principalmente en dos aplicaciones importantes: la animación geográfica temporal y la no-temporal. El objetivo del primer tipo de animación es demostrar el cambio en un cierto plazo, como el cambio en la cubierta de la vegetación durante determinados períodos de tiempo. La animación geográfica no-temporal no representa el cambio de las formas del terreno con respecto al tiempo, sino el efecto de la clasificación de los datos donde cada cuadro representa un esquema distinto de clasificación (Etzel Müller y Sulebak, 2000).

Muchos usos de los modelos digitales del terreno requieren el uso de datos del terreno en varios niveles de detalle (Quintero y otros, 2002). Cambiar la resolución del modelo puede cambiar el nivel del detalle de un modelo del terreno. Los diferentes niveles se pueden almacenar en una jerarquía para proporcionar un modelo jerárquico del terreno consistente en una pirámide de niveles de resolución en aumento. Así, los modelos del terreno de múltiple resolución describen una superficie topográfica en diversos niveles de resolución. Además de proporcionar el mecanismo de compresión de datos para datos topográficos densos, tales modelos en general permiten analizar y visualizar diversas superficies en resolución variable. En este contexto, la resolución se ve como una medida del nivel del detalle o granularidad de un conjunto de datos. La alta resolución indica datos detallados y la resolución baja indica datos más toscos. La tecnología de multiresolución tiene la capacidad de almacenar modelos del terreno para áreas grandes en un alto nivel del detalle.

Modelos topográficos

Los modelos basados en topografía son modelos basados en la Física que implican componentes topográficos fuertes. El principio de estos modelos es que los DEM entregan parámetros importantes para los cálculos; algunos de estos modelos se resumen en la Tabla 1 (Wood, 1997; Etzel Müller y Sulebak, 2000). Todos los modelos utilizan una discretización del área en una rejilla regularmente espaciada.

Tabla 1. Parámetros y modelos topográficos utilizados en los DEM

Z es la altitud, Z' y Z'' son la primera y segunda derivada de Z , respectivamente. Z'' plana y de perfil son la curvatura horizontal (el contorno o curva de nivel) y la curvatura vertical, respectivamente. X y Y son las distancias en las direcciones x y y , respectivamente. VAR es la varianza, $SKEW$ es la oblicuidad calculada sobre una vecindad local. E es la tasa de elevación del relieve, Z_{\min} , Z_{\max} y Z_m son el valor mínimo, máximo y medio de Z , respectivamente, calculado sobre una vecindad local. A_s es el área específica de flujo inverso específico calculado en cualquier punto del DEM. IR es el índice de rugosidad, con $X = -\sin(y)\cos(a)$, $Y = \cos(y)\sin(a)$ y $Z = \sin(a)$, con y y a como el aspecto y la pendiente, respectivamente; n es el número de muestras en el cálculo.

Parámetros y modelos topográficos

Nombre	Descripción
Altitud	$Z(m)$
Pendiente	$Z' = \alpha = \sqrt{\left(\frac{\Delta Z}{\Delta x}\right)^2 + \left(\frac{\Delta Z}{\Delta y}\right)^2}$
Aspecto	$\omega = -\frac{\frac{\Delta Z}{\Delta y}}{\frac{\Delta Z}{\Delta x}}$
Curvatura (plana y de perfil)	$Z'' = \frac{\delta^2 Z}{\delta x^2}$
Relieve local	$R_l = Z_{\max} - Z_{\min}$
Varianza	$Z_{\text{var}} = VAR(Z)$
Tasa de elevación de relieve	$E = \frac{Z - Z_{\min}}{Z_{\max} - Z_{\min}}$
Oblicuidad	$Z_{sk} = SKEW(Z)$

Índice de humedad

$$\varpi = \ln \frac{A_s}{\tan \alpha}$$

Índice de erosión

$$\Omega = A_s \tan(\beta) Z''_{plan}$$

Índice de rugosidad

$$IR = 1 - \frac{\sqrt{(\sum X_i)^2 + (\sum Y_i)^2 + (\sum Z_i)^2}}{n}$$

Los modelos de balance energético y de radiación calculan estos valores en cada punto de la rejilla de un área. La radiación solar directa y la cantidad de energía que alcanza la superficie dependen de la pendiente, de la exposición y de los efectos de las sombras topográficas. Se han desarrollado varios algoritmos para la parametrización de los efectos topográficos en la radiación directa.

2.2. Enfoques adoptados por el INEGI

En el año 2003, personal del INEGI acudió al congreso GEOPRO 2003 organizado por el entonces llamado Laboratorio de Geoprocesamiento (GeoLab). En ese tiempo se platicó acerca de posibles proyectos vinculados en los que el Laboratorio podría colaborar con dicho Instituto. En particular, dos de esos proyectos resultaron ser de gran interés para nosotros, pues se acercaban a la incipiente línea de investigación que se estaba forjando en el GeoLab. Uno de tales proyectos tiene que ver con la instauración de una red de metadatos para la integración de diferentes fuentes gubernamentales de datos espaciales; el otro proyecto de interés trataba del análisis de alto nivel de los datos contenidos en los DEM que el Instituto genera. Como resultado de estas ideas, decidimos trabajar en un proyecto que conjuntara ambas necesidades, puesto que pensamos que ambas tienen solución realizando el procesamiento y el análisis desde un punto de vista semántico. Así pues, surge un caso de estudio real en el que se puede aplicar las ideas que dieron origen al presente trabajo de tesis. Por ello, en esta sección presentamos los enfoques que maneja el INEGI en cuanto a los DEM.

En el INEGI no existe una metodología establecida para la generación de los DEM (INEGI, 2000b; INEGI, 2000a). La manera en la que se realiza la generación de esta

información tiene como base la experiencia de los cartógrafos y en muchos casos dependiente de las habilidades manuales adquiridas a través de los años.

Existe una serie de especificaciones emitidas en (INEGI, 2000b), en las que se presentan las normas bajo las cuales se deben generar los DEM. Sin embargo, no se presenta una metodología para su generación. En dicho documento, en cambio, se presentan dos posibles fuentes para obtener datos de elevación: por conversión de curvas de nivel y por fotogrametría de pares estereoscópicos. Se mencionan, además, consideraciones del tipo geométrico y de exactitud como guía para la generación de los DEM.

En el primer caso los datos de elevación se obtienen al digitalizar mapas que contienen curvas de nivel del terreno. Con estos datos, se aplica una etapa de interpolación para obtener las alturas en las posiciones que estarán representadas en el DEM. Como se ha mencionado, los puntos sobre la superficie del terreno están referenciados por coordenadas que se encuentran equiespaciadas; es por ello que se requiere de la etapa de interpolación, ya que las curvas de nivel no se intersectan necesariamente con los puntos que se han de representar. Ver Figura 7.

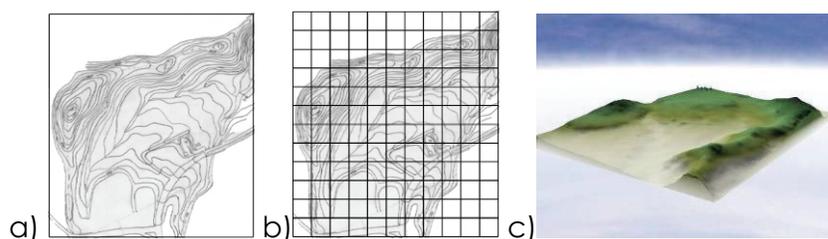


Figura 7. Proceso de generación de DEM con base en mapas con curvas de nivel

a) Mapa con curvas de nivel, b) Se aplica una rejilla (*grid*) a las curvas de nivel vectorizadas y se interpola para obtener la altitud en los puntos de cruce de la rejilla, c) Modelo del terreno generado.

El segundo método de generación está basado en la fotogrametría de pares estereoscópicos. En dicho método se emplean pares de imágenes aéreas del terreno, las cuales son tomadas desde dos perspectivas distintas (izquierda y derecha). Con ello se obtiene un par estereoscópico, de cuyas propiedades se pueden obtener los datos

de elevación del terreno fotografiado. El problema principal de este método reside en encontrar los píxeles (uno en cada imagen del par estereoscópico) que representen la misma ubicación en el terreno. Para resolver este problema se utiliza la técnica de correlación de imágenes.

La correlación de imágenes consiste en extraer subimágenes homólogas de ambas fotografías, las que subsecuentemente se correlacionan por medio de su información radiométrica, para determinar los puntos del DEM a partir de píxeles homólogos de cada imagen. Para cada posición en la imagen se calcula el coeficiente de correlación de la matriz de búsqueda (ver Figura 8). El resultado de la aplicación de la función de correlación cruzada define la posición del mejor "empate" de la matriz de referencia y su correspondiente en la matriz de búsqueda.

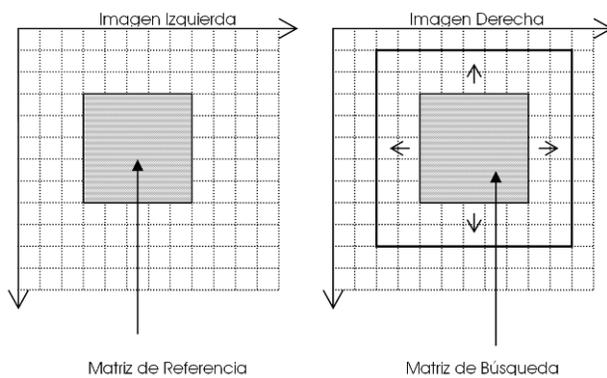


Figura 8. Correlación de pares estereoscópicos

La matriz de referencia se toma de la imagen izquierda y es utilizada para comparar con los tonos de gris de una segunda matriz (llamada de búsqueda) para encontrar el mejor punto con el que se relaciona en la imagen derecha.

Para calcular las coordenadas planimétricas de los puntos del DEM, así como sus elevaciones, se hace uso tanto de las coordenadas de "fotografía" obtenidas por la transformación píxel a foto para todos los pares de píxeles homólogos, así como de los elementos de orientación exterior de las fotos aéreas (INEGI, 2000a).

2.3. Geomorfometría

La morfometría, o la medición de la forma, es un área que ha cobrado gran relevancia en diversas ramas de la ciencia que van desde la medicina y la arqueología hasta la física y el reconocimiento de patrones. En este trabajo proponemos realizar la descripción de los DEM conceptualizando las formas que puede tener un terreno para posteriormente buscar esas formas en los datos. Por ello es importante para nuestro enfoque estudiar la geomorfometría. En este trabajo consideraremos a la geomorfología como *“la ciencia que trata de la geometría o la forma del terreno”*. Como parte en el estudio de la geomorfometría, en el siglo pasado se dio gran importancia a la medición y cuantificación de la forma topológica del terreno. Ejemplo de ello es (Evans, 1972), quien señala una diferencia entre los estudios que miden las características geomorfológicas específicas (*geomorfometría específica*) y la *geomorfología general*: *“la medición y análisis de aquellas características del terreno que son aplicables a cualquier superficie continua rugosa”*. Los inicios de este enfoque están en el análisis cartométrico de mapas y la derivación de índices descriptivos a partir de los datos. Este enfoque es unificado por (Mark, 1975) al intentar medir la *“rugosidad”* de un terreno. (Mark, 1975) sugiere que la variación planimétrica (rugosidad) tiene dos escalas principales a las que les llamó *grano*, que se refiere a la longitud de onda significativa más grande en una superficie, y *textura* que es la más pequeña. Este tipo de mediciones presentan algunos problemas al ser aplicadas. Primeramente, las mediciones propuestas se hacen con base en definiciones arbitrarias o ambiguas, como *‘longitud de onda significativa’*. Por otro lado, existe una superposición entre las diferentes medidas que se proponen, lo que nos lleva a tener definiciones redundantes. Una forma más sistemática de parametrizar la rugosidad del terreno fue presentada en (Evans, 1972; Evans, 1979; Evans, 1980; Evans, 1984). En estos trabajos se sugiere tomar la primera y segunda derivada de la altitud (pendiente, aspecto, convexidad de perfiles y convexidad de los planos) para realizar el análisis de las variables importantes relacionadas con la altitud (media, desviación estándar y oblicuidad) y sus derivadas (pendiente, aspecto, convexidad de perfil y de plano). (Heerdegen y Beran, 1982) y (Pike, 1988) utilizan la pendiente y la

curvatura para caracterizar diferentes formas del terreno mediante la construcción de 'firmas'⁷ a partir de los parámetros mencionados. (Pike, 1988) sugiere que una buena parametrización del terreno es necesaria para su correcta clasificación. La clasificación geomorfométrica del terreno ha sido realizada por medio de 'regiones homogéneas' (Dikau, 1989), así como por la identificación de características geomorfométricas específicas (Tribe, 1990). Asimismo, en (Chorowicz y otros, 1995) se describen algunas aproximaciones al reconocimiento automático de características geomorfológicas a partir de los DEM y de imágenes de satélite. Se presenta un método paramétrico que reconoce patrones de deslave, principalmente enfocado a detectar isolíneas con patrones circulares y los parámetros relacionados. Se presenta, también, un método estocástico de reconocimiento de patrones, el cual calcula la dimensión fractal local, lo que permite encontrar patrones de deslave en los DEMs y en los mapas geológicos digitalizados.

La identificación de redes fluviales es una parte importante del análisis que se realiza con los DEM. Es posible mencionar algunas características importantes que presentan los trabajos realizados relacionados (Wood, 1997). Primero, el enfoque en la forma, que puede ser topológico o geométrico. Con el enfoque topológico se obtiene una descripción multidimensional del terreno, es decir, una caracterización del mismo; en contraste, con el enfoque geométrico se realiza una transformación de los datos del modelo del terreno y se obtiene una nueva descripción del mismo. En segundo lugar, tenemos el ámbito del análisis, que puede ser local o global. Cuando es local se utiliza una vecindad local para calcular los parámetros; si la vecindad es global, cada uno de los elementos del modelo afecta el cálculo de los parámetros o de los nuevos valores. La tercera característica es la precisión de las mediciones, que pueden ser exactas, en donde todos los valores que se generan y que coinciden con un valor conocido dentro del modelo deberán ser exactamente iguales; o aproximadas, en las cuales se permite la existencia de cierta tolerancia. En cuarto lugar, podemos

⁷ Las firmas que menciona Pike son una especie de vector de características con los cuales se pueden comparar, e incluso identificar regiones de drenaje.

mencionar que el procedimiento de análisis puede ser directo (el resultado puede ser identificado y medido directamente del resultado de los datos) o indirecto (los resultados son asociados con algunas otras propiedades relacionadas con la morfometría del terreno). Finalmente, la aplicación del procedimiento de análisis puede ser sistemática o recursiva. En el caso de ser sistemática, la aplicación del procedimiento se realiza de tal manera que es independiente de las características del modelo del terreno. Los métodos recursivos se aplican en un orden que es determinado por la propia naturaleza de los datos contenidos en el modelo.

2.4. Enfoques que utilizan procesamiento de imágenes

Para el procesamiento de información como la contenida en los CDER, resulta interesante estudiar los trabajos realizados en el área de procesamiento de imágenes. Particularmente, sobre las técnicas de procesamiento para el análisis de texturas y sobre todo, para el reconocimiento de patrones.

El análisis de textura dentro del procesamiento de imágenes es un símil del proceso geomorfométrico de medición de rugosidad o variación del terreno. Según (Gonzalez y Woods, 1992), no existe una definición formal de textura; (Tomita y otros, 1982) distinguen entre textura estadística y los elementos estructurales de una imagen. Una textura estadística es la descripción de la frecuencia más alta de la variación dentro de una imagen, mientras que un elemento estructural es el arreglo de 'primitivas de textura que conforman la imagen. (Haralick, 1983) presenta un método de análisis de texturas en el que se utiliza una matriz de dependencias espaciales, que mide las propiedades de autocorrelación espacial de una imagen. Con base en esta información, se calculan medidas estadísticas para obtener propiedades de las texturas, tales como entropía y linealidad. Cabe destacar que esta técnica depende de la resolución de la imagen (de la escala, en términos geográficos). Algunos trabajos estudian el uso de transformadas *wavelet* (Graps, 1995) para realizar análisis que permitan medir y controlar la dependencia de la escala. (Gallant y Hutchinson, 1996) utilizan transformadas *wavelet* para reconstruir los DEM en un rango controlado de escalas.

El reconocimiento de patrones en imágenes puede resultar de gran utilidad para nuestro trabajo. En general, en este contexto, el reconocimiento de patrones consta de dos acciones: la segmentación y la clasificación. La primera tiene como su principal objetivo dividir una imagen en partes que tienen una fuerte correlación entre sí. La segmentación puede ser total, en la cual un conjunto de regiones disjuntas tienen una correspondencia única con objetos de la imagen de entrada, o puede ser parcial, en la que las regiones no tienen una correspondencia con los objetos.

Existen diferentes algoritmos de segmentación. Podemos mencionar los de conocimiento global, en los cuales el conocimiento es usualmente representado por el histograma de la imagen. Ejemplos de algoritmos de este tipo son el de umbralizado (*thresholding*), que consiste en utilizar las líneas de los contornos para establecer los límites de un objeto; mediante el histograma de la imagen se define el límite óptimo conocido como umbral (Sonka y otros, 1999). Otro tipo de algoritmos de segmentación que se utilizan son los de agrupamiento (*clustering*) en los que se considera un vector de medidas multiespectrales (color, desviación estándar, media estadística, etc.) de cada píxel en la imagen (Jain, 1984). Otro tipo de algoritmos son aquellos que realizan una segmentación basada en contornos, los cuales utilizan características de la imagen como textura, brillantez, color, etc. (Sonka y otros, 1999). Como ejemplo, podemos mencionar el algoritmo de trayectoria del borde interno y externo (*tracing*), que se utiliza cuando no se conoce el borde de las regiones pero éstas ya han sido definidas. La segmentación basada en regiones es otro tipo interesante, que al igual que la anterior, utiliza características de la imagen como textura, brillantez, color, etc. Como ejemplo podemos mencionar la segmentación basada en el crecimiento de regiones, en la que los píxeles vecinos que presentan una amplitud similar son agrupados para formar una región segmentada. Un par de píxeles son combinados en grupos llamados regiones atómicas si los píxeles presentan la misma amplitud y están conectados (Castleman, 1996).

La clasificación se realiza normalmente con dos tipos de algoritmos: supervisados y no supervisados. En los supervisados es necesario conocer previamente las clases que

serán manejadas durante la clasificación, mientras que en los no supervisados el objetivo es identificar los grupos naturales o *clusters*.

El reconocimiento de patrones puede ser considerado como un primer paso en el análisis geomorfométrico (Chorowicz y otros, 1995), debido a que es necesario extraer características específicas a partir de las fuentes de datos antes de comenzar con su interpretación cualitativa. Algunos métodos, como el análisis paramétrico, necesitan primero extraer mediciones de las fuentes de datos para entonces producir tanto mapas de patrones específicos, como una descripción de las características cuantitativas relevantes a su propia naturaleza. En (Chorowicz y otros, 1995) se revisan brevemente algunas técnicas de reconocimiento de patrones que los autores han aplicado para detectar características geomórficas. Según el mismo trabajo, en las geociencias las principales aplicaciones del reconocimiento de patrones es el análisis de las estructuras compuestas por líneas, tanto curvas como rectas. En este tipo de análisis, se extraen líneas de los DEM y se utilizan como imágenes binarias. Entonces se definen los parámetros necesarios para describir las líneas; las cuales se miden y se utilizan después para detectar características específicas. Como ejemplo se menciona el método descrito en (Parrot y Taud, 1992) para extraer estructuras circulares en imágenes SPOT. Debido a que estos métodos utilizan parámetros numéricos, tanto la información cuantitativa acerca de las características y su reconocimiento se llevan a cabo simultáneamente. Las clasificaciones pueden ser bastante correctas en dependencia de los límites especificados por los usuarios para diferentes parámetros. Sin embargo, debido a que sólo algunas de las características son extraídas, este método debe ser implementado en combinación con otros. En (Chorowicz y otros, 1995) se presenta otro tipo de aplicación del reconocimiento de patrones, pues se menciona que se pueden utilizar métodos estadísticos aplicados al análisis de texturas para detectar morfología rugosa a un nivel local en superficies de deslave. Cuando el análisis estocástico de la textura no es suficientemente adecuado para detectar confiablemente una característica, éste se combina con otras características peculiares (como tipo de inclinación y/o litología). Como ejemplo de esto se utilizan los modelos presentados en (Mandelbrot, 1982; Pentland, 1984) con los que se calculan

dos índices: 1) la desviación estándar de fractales locales (equivalente a la cuesta) y 2) la dimensión fractal local (equivalente a la rugosidad), de manera que las zonas de deslave son extraídas al empatar los mapas geológicos con las imágenes de la dimensión fractal y la desviación estándar.

2.4.1. Conversión de *raster* a vector

Una de las áreas paralelas a este trabajo es la conversión de *raster* a vector, pues es un área en la cual se requiere hacer un análisis de los datos *raster*, en donde se necesita, para tener una buena conversión, agregar información y/o conocimiento a los datos *raster*. El problema de la conversión automática de *raster* a vector ha tomado mucha atención de los desarrolladores y los investigadores de software durante las últimas dos décadas. Numerosas aproximaciones para resolver el problema se han originado principalmente del área de los SIG. Desafortunadamente, los sistemas de conversión completamente automáticos parecen ser irreales y algunos autores sugieren poner al operador en el centro del sistema de conversión.

Las tecnologías de digitalización de mapas *raster* pueden ser divididas en cuatro grupos intersectables (Levachkine y Polchkov, 2000):

- *Manuales*. En la práctica, los métodos de digitalización manual de mapas *raster* coinciden con los métodos de digitalización de mapas en papel. Por ejemplo, en el caso de objetos puntuales, el operador localiza visualmente los símbolos gráficos y fija sus coordenadas. En el caso de objetos lineales y poligonales, el operador utiliza segmentos de rectas para aproximar contornos curvos. La tasa de digitalización manual es de uno o dos objetos por minuto.
- *Interactivos*. La digitalización interactiva utiliza programas especiales, en los cuales, una vez que el operador indica el punto de inicio de un segmento de línea, el sistema sigue automáticamente el contorno de la línea (trazo). Esos programas son capaces de seguir líneas relativamente simples. Si el programa no puede resolver una ambigüedad en el mapa *raster*, regresa un mensaje al operador.

- *Semiautomáticos*. Recientemente han aparecido editores vectoriales capaces de llevar a cabo el proceso de digitalización, reduciendo el tiempo de procesamiento. Estos editores pueden ser llamados sistemas semiautomáticos (Able-Software, 2000; Levachkine y Polchkov, 2000; Bodansky, 2002; e-Cognition, 2003).
- *Automáticos*. En teoría, los editores de vectores automáticos digitalizan automáticamente todos los objetos de una clase, dejando al operador las capas vectoriales resultantes. Los editores de vectores más populares utilizan esta aproximación. Sin embargo, en la práctica, el alto nivel del error resultante de cualquier pequeña complicación en el mapa *raster* significa que se deben tomar métodos alternativos para reducir el gran volumen de correcciones manuales. Además, el uso de métodos cada vez más complejos y algoritmos para el reconocimiento de objetos cartográficos realmente no mejora los resultados (Levachkine y Polchkov, 2000).

En (Levachkine, 2003) se sugiere el desarrollo de un sistema para la digitalización automática de mapas, el cual utiliza un principio básico: los métodos y el software deben de ser desarrollados dejando al operador solamente aquellas tareas que la computadora no puede llevar a cabo. Esto implica el análisis detallado de todos los procesos existentes de digitalización de mapas y desarrollo de software para automatizar todas las operaciones tecnológicas, desde actualización basada en algoritmos formales, heurísticas, bases de conocimiento e interactivos, los cuales puedan ser efectivamente utilizados por la computadora. Una aplicación de dicha aproximación puede mejorar enormemente no solo el procesamiento (el paso principal de reconocimiento de imágenes cartográficas), sino además el preprocesamiento (preparación de los mapas en papel y sus análogos en *raster*) y el posprocesamiento (procesamiento final de los resultados de la digitalización automática).

Preprocesamiento de los mapas raster

La meta principal del preprocesamiento es preparar las imágenes cartográficas *raster* de una forma que se incremente la probabilidad de su reconocimiento en un

sistema automático. A continuación se presenta una secuencia de operaciones propuestas para la preparación de mapas *raster* para el reconocimiento automático (Levachkine, 2003).

- Preparación de los materiales para su escaneo (Restauración, Copiado, Mejora)
- Escaneo (Pruebas de escaneo, Definición de los parámetros óptimos , Escaneo final, Unión de imágenes en mosaico)
- Preparación del mapa *raster* para el reconocimiento de objetos cartográficos (Edición del mapa *raster*, Separación de los colores básicos, Restauración de la paleta de colores, Estratificación del mapa *raster*)

Procesamiento de los mapas raster

La meta principal de este paso es el reconocimiento de imágenes cartográficas; esto es, la generación de capas vectoriales y de la información atributiva en mapas electrónicos. Desde nuestro punto de vista, la línea más promisoría del desarrollo de software es el desarrollo de métodos, algoritmos y programas que se enfocan en localizar e identificar objetos cartográficos específicos. En los SIG tradicionales, los objetos en mapas vectoriales están divididos en tres tipos: puntos, líneas y polígonos, representando respectivamente objetos puntuales, lineales y de superficies. Esta clasificación puede ser fácilmente extendida al análisis de imágenes cartográficas en mapas *raster*. Los objetos son dibujados en mapas temáticos en la forma de símbolos gráficos, los cuales son los mismos para todos los objetos de un grupo dado.

Un elemento importante en la automatización de la vectorización del mapa *raster* es el desarrollo de una secuencia óptima de pasos para el reconocimiento de imágenes cartográficas, sucesivamente eliminando elementos que ya han sido decodificados del mapa y restaurar partes de la imagen, las cuales hubieran sido ocultadas por los elementos eliminados. En (Levachkine, 2003; González-Gómez y Levachkine, 2004) se sugiere la utilización de mapas vectoriales de pequeña escala para el reconocimiento de las correspondientes imágenes cartográficas en mapas de escala grande. Un mapa de escala pequeña contiene información generalizada acerca de una proporción considerable de los objetos existentes en el mapa a escala

grande. El uso de mapas de pequeña escala resuelve un gran problema en la digitalización automática: la búsqueda de objetos en todo el mapa *raster*. En este caso, un objeto en el mapa de escala grande puede encontrarse muy cerca de su análogo en el mapa de pequeña escala, y no en otro lado. La zona de búsqueda para objetos puntuales puede estar restringida a un círculo con un radio definido por la correlación entre las escalas de los mapas (de escala grande y pequeña).

Posprocesamiento de los mapas raster

La meta principal del post procesamiento de mapas *raster* es la corrección automática de los errores en la vectorización. Para ello se sugieren dos opciones: 1) utilizando las características topológicas de los objetos en las capas vectoriales y 2) la correlación espacial (conectividad) de las capas vectoriales correspondientes.

La primera está basada en el hecho de que muchos objetos cartográficos en el sistema tienen características topológicas bien definidas, las cuales pueden ser utilizadas para la corrección de los errores en la vectorización. Por ejemplo, las isolíneas, cuyas características topológicas son: a) las isolíneas son continuas; b) estas líneas no se pueden interceptar una con otra; c) cada isolínea es cerrada, es decir inicia y finaliza en el dominio del vecindario; y finalmente d) los polígonos que cubren todo el dominio sin intersecciones pueden ser ensamblados mediante los arcos del sistema correcto de isolíneas junto con las fronteras del dominio. Sin embargo, en un mapa *raster* esas características, como una regla, pueden perderse debido a varias razones: a) las líneas están interrumpidas donde el valor para la isolínea está escrito, b) algunas secciones de las isolíneas no están bien dibujadas en regiones de alta densidad, y c) en la imagen *raster*, algunas isolíneas se mezcla debido a defectos de impresión en los mapas en papel (Levachkine, 2003).

A2R2V (*Analog to Raster to Vector*) es un sistema de conversión *raster* a vector que se presenta en (Levachkine, 2003), basado en el análisis semántico de imágenes de color. El análisis semántico de imágenes cartográficas es interpretado como una representación separada de los patrones cartográficos (alfanuméricos, puntuales, lineales y de área). El sistema de interpretación de mapas explora la idea de la síntesis

de invariantes en las imágenes a un nivel bajo de procesamiento (vectorización y segmentación). Esto significa que nosotros aceptamos los sistemas de "vectorización – reconocimiento" y "segmentación – interpretación" simultáneamente. Aunque esos sistemas pueden generar algunos errores, es mucho más útil para los siguientes algoritmos que su salida esté cerca de reconocer los objetos de interés.

Comenzamos el reconocimiento del mapa partiendo de una binarización global, seguido de una identificación clásica de OCR con redes neuronales artificiales, clasificación supervisada, reglas de reconocimiento basado en conocimiento y vectorización basada en morfología. Para enfrentar el problema del intensivo trabajo que representa el entrenamiento, diseñamos imágenes simplificadas. Para este propósito, utilizamos combinaciones lineales de componentes de color o representaciones de la imagen (técnica de falso color) y composición de representaciones binarias (técnica de imagen compuesta). Esas técnicas son independientes de la aplicación. Sin embargo, en el marco de trabajo de nuestra solución, el reconocimiento del mapa puede ser tratado como una tarea común (dependiente de la aplicación) (Doermann, 1998).

2.5. Fractales

La idea de describir un conjunto de datos *raster* utilizando elementos primitivos, nos lleva a la necesidad de la existencia de un conjunto de "partes" elementales mediante las cuales se pueda describir cualquier terreno. Pensamos que esas "partes" se replican una y otra vez en diferentes tamaños, así que creemos que es interesante explorar con respecto a las "partes" que se construyen de otras, muy similares, pero más pequeñas; a esas "partes" se les conoce como fractales. Existen un gran número de ideas importantes que han tenido su origen en este campo de estudio. De acuerdo a (Feder, 1988) uno de los aspectos más persuasivos en el trabajo de Mandelbrot son las imágenes creadas por (Voss, 1985) de terrenos "que lucen naturales". La esencia de un objeto fractal es la autosimilaridad, la cual puede ser definido como "*una forma hecha de partes similares al todo*" (Mandelbrot, 1986). Más específicamente, los fractales pueden ser definidos en dos contextos:

- **Determinísticos o estocásticos.** Un fractal determinístico está definido por una función matemática que es aplicada recursivamente sobre un gran rango de escalas; como resultado se obtiene un objeto que contiene copias escaladas de sí mismo. En un fractal estocástico existe una probabilidad de aplicar diferentes formas en cada paso de la recursión, pero siempre similares al todo.

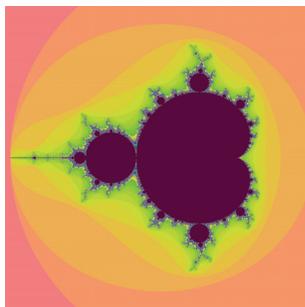


Figura 9. Ejemplo de fractal determinístico

- Autosimilares o autoafines. Si las copias que se hacen son idénticas y la escala es isotrópica, entonces el fractal es llamado autosimilar; de otra manera es llamado autoafín.

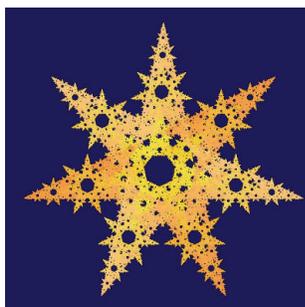


Figura 10. Ejemplo de fractal autoafín

El comportamiento fractal de las superficies puede ser detectado con algunas de las técnicas geoestadísticas utilizadas para la interpolación. En particular, el variograma puede ser utilizado para identificar elementos autosimilares (Rees, 1995). Es este comportamiento lo que representa una de las más importantes implicaciones en las geociencias en general, y en la caracterización de superficies en particular.

En (Mandelbrot, 1977) se define que para un objeto autosimilar verdadero su dimensión fractal será constante sin importar la escala seleccionada para su medición. Dado que la dimensión fractal puede ser medida, es posible preguntarnos si la topografía es fractal. (Wood y Snell, 1957) argumentan que el relieve medido en un rango de escalas demuestra que puede ser utilizado para predecir la caracterización del relieve en otras escalas. (Mandelbrot, 1967) provee alguna evidencia de que la línea costera de Gran Bretaña es fractal; lo que sugiere que los terrenos son fractales (Mandelbrot, 1982). Existen análisis de diferentes partes del mundo (Dietler y Zhang, 1992; Rees, 1995) en los cuales se arrojan resultados que sugieren que los terrenos no poseen sólo una dimensión fractal, sino una variedad de valores que cambian con la escala. Aunque no existe evidencia sobre el comportamiento fractal de la topografía, es importante destacar dos cosas: primero, que la escala a la cual las superficies son modeladas y analizadas determinarán en parte su caracterización; y segundo, que la total dependencia a la escala es un dato útil, ya que nos lleva a considerar este parámetro en la caracterización de la superficie.

Existen trabajos en los que se utilizan las ideas introducidas por (Mandelbrot, 1977) para el análisis de la superficie del terreno, en los que se hace un símil entre la dimensión fractal y la rugosidad o variación de la superficie del terreno (Pentland, 1984; Chorowicz y otros, 1995).

2.6. Incertidumbre

Aunque ya se ha mencionado que el procesamiento semántico busca elevar el nivel en el análisis en los problemas, es inevitablemente necesario tener una representación que sea manejable a través de las computadoras actuales, es decir, una representación sintáctica e incluso léxica (numérica) de los datos. Así pues, la descripción semántica de los DEM debe involucrar no solo la identificación de los objetos presentes en el modelo, sino también involucrar la realización de mediciones y su caracterización para que la descripción sea útil para posteriores análisis. Por ello, es necesario aceptar que los resultados son susceptibles del error intrínseco en el modelo, por lo que se debe pensar con qué grado de incertidumbre puede realizarse la

representación⁸. En (Wood, 1997) se presenta un amplio análisis sobre la incertidumbre en los DEM, en el que se pretende incorporar esta medida al caracterizar los terrenos. Para realizar dicho análisis, utiliza las siguientes medidas de incertidumbre:

- técnicas estadísticas (error cuadrático medio, tasa de precisión, media y desviación estándar, entre otras),
- mediciones espaciales (autocorrelación espacial, variogramas y correlogramas, exactitud de superficie, etc.), y
- análisis hipsométrico.

De igual manera, analiza los métodos comúnmente utilizados para la interpolación de los valores de un DEM y formula cómo se comporta cada una de las medidas mencionadas. Finalmente propone un modelo para la incertidumbre de los DEM. Luego, trata el proceso de parametrización de los modelos. Adopta como base a (Evans, 1979), en donde se toman como parámetros principales la elevación, así como su primera y segunda derivada en cada uno de los ejes. Se presenta cómo modelar la superficie mediante *Parametrización Cuadrática* (parches bicuadráticos).

2.7. Filosofía, semántica y ontologías

Un enfoque actual de procesamiento de modelos de terreno, y en general de datos espaciales, utiliza una modelación del terreno con base en una conceptualización de los elementos o conceptos que existen en este contexto. Por ejemplo, en (Smith y Mark, 2003) se reconoce que *se requiere una ontología, basada en objetos, de montañas y otras formas del terreno para hacer justicia a nuestra concepción cotidiana del entorno que nos rodea, así como para soportar el razonamiento espacial y el procesamiento del lenguaje natural*. Para ello cita que *Montaña es el ejemplo de tipo*

⁸ Aunque no es importante para un análisis semántico, pero ya mencionamos que se debe aterrizar en una representación sintáctica. Incluso, tal vez, los análisis posteriores se podrían realizar con el enfoque tradicional de análisis numérico; por ello, el análisis de la incertidumbre que se maneja, tanto en el modelo de entrada como en los resultados de la descripción, puede resultar de importancia, aunque es algo que queda fuera de los alcances de este trabajo.

de característica (o cosa) geográfica que es más comúnmente citada por las personas, en inglés y en otros idiomas (Mark y otros, 1999; Smith y Mark, 2001).

La gente en general convive con el mundo cotidiano en una variedad de formas: además de actuar dentro del mundo directamente, utilizamos mapas, leemos periódicos, escuchamos el estado del tiempo y del tráfico. Cada una de esas actividades involucra cierta conceptualización, lo cual significa que involucramos un sistema de conceptos o categorías que dividen el dominio en el que nos encontramos en objetos, cualidades, relaciones, etc.

En el sentido filosófico, nos debemos referir a una ontología como un sistema particular de categorías válidas para cierta visión del mundo. Así, este sistema no depende de un lenguaje particular. En inteligencia artificial se utiliza mayormente el término ontología como un artefacto de ingeniería, constituido por un vocabulario específico, utilizado para describir cierta realidad, más un conjunto de reglas para intentar dar significado a las palabras del vocabulario. Estas reglas tienen usualmente la forma de una teoría lógica de primer orden, donde las palabras del vocabulario aparecen como nombres de predicados unarios o binarios, respectivamente llamadas conceptos y relaciones. El término ontología es actualmente utilizado también por científicos de la información, para los cuales una ontología no involucra cuestiones de existencia. Está diseñada para asistir la tarea de especificar y clarificar los conceptos empleados en un dominio dado, y por encima de todo, formaliza los conceptos dentro del marco de cierta teoría formal con una estructura lógica (sintáctica y semántica) bien entendida. (Gruber, 1993) define una ontología en sistemas de información como *una especificación de una conceptualización*. Una meta central del empleo de ontologías en sistemas de información es la resolución de las incompatibilidades que resultan en ciertas circunstancias, idealmente por medio de una descripción concisa y sin ambigüedades de las entidades principales y relevantes de un dominio de aplicación y sus relaciones potenciales entre sí. En el sentido de los sistemas de información una ontología es una cosa pragmática.

Las interpretaciones de "ontología" descritas anteriormente están profundamente relacionadas una con otra; utilizaremos la palabra conceptualización para referirnos a

la interpretación filosófica de la palabra ontología. Así, dos ontologías pueden ser diferentes en el vocabulario (utilizando palabras en inglés y español, por ejemplo), mientras comparten la misma conceptualización.

Así, mientras que el proceso para la creación de una ontología depende de la sintaxis, es decir de la forma de representación, la conceptualización es un proceso de abstracción de la realidad para encontrar la esencia de ésta en un contexto definido; según (Gruber, 1993) *una conceptualización es una visión abstracta y simplificada del mundo que deseamos representar para algún propósito*. Es así entonces que una conceptualización tiene como resultado la semántica de la realidad que se está estudiando⁹. Para (Smith y Mark, 2003), una conceptualización tiene por lo menos dos etapas, llamadas teoría primaria y secundaria. La teoría primaria, según (Horton, 1982), es aquella parte del sentido común que encontramos en todas las culturas y seres humanos. Las teorías secundarias, en contraste, son aquellas colecciones de creencias, populares o científicas, que son peculiares para grupos sociales particulares.

Una teoría primaria consiste en la física básica, psicología básica, y otras familias de esas creencias básicas que toda la gente necesita para percibir y actuar en situaciones cotidianas. Es importante mencionar que la teoría primaria será en muchos aspectos incompleta. Por ejemplo, la teoría primaria reconoce la existencia de montañas, pero no da explicación de cómo las montañas están demarcadas (o como no lo están) de sus alrededores. Una teoría primaria debe ser compatible con los resultados de la ciencia (Smith, 1995b; Smith, 1995a).

Una teoría secundaria, en contraste, consiste tanto en teorías científicas de fenómenos físicos microscópicos y macroscópicos (átomos y quarks, galaxias y quasars) como en creencias populares (ángeles y demonios, cielos e infiernos). Ambas cosas no son accesibles a nuestra percepción y acción cotidiana debido a que son o muy grandes o muy pequeñas, o tal vez porque no existen completamente.

⁹ La cual puede resultar muchas veces difícil, si no imposible de representar. Por ello, pensamos que toda representación de una conceptualización implica pérdida de semántica. Diferentes tipos de representación tendrán diferentes cantidades de pérdida.

De manera similar, en el sentido de los sistemas de información (Guarino, 1998) identifica varios tipos de ontologías dependiendo de su nivel de generalidad, a saber:

- Ontologías de alto nivel
- Ontologías de dominio y de tarea
- Ontologías de aplicación

En general, los primeros dos tipos de ontologías representan lo mismo que las teorías primaria y secundaria. Como una adición del campo de los sistemas de información, las ontologías de aplicación pueden verse como un puente entre el conocimiento general y el cómo los programas de software lo utilizan en un sistema de información.

La importancia de la utilización de ontologías en los sistemas de información tiene su origen en el hecho de que las aplicaciones convencionales y los sistemas basados en conocimiento están contruidos sobre plataformas de hardware heterogéneas, diferentes lenguajes de programación, diferentes protocolos de red, etcétera (Gruber, 1993). Como ya se ha mencionado, el mismo autor define ontología como *una especificación explícita de una conceptualización*, reconociendo que el término ha sido tomado prestado de la filosofía, donde una ontología es *una afirmación semántica de la existencia*. Para los sistemas de información, lo que "existe" es aquello que puede ser representado en un programa.

Se utilizan ontologías comunes para describir compromisos ontológicos para que un conjunto de programas puedan comunicarse acerca del dominio del discurso sin la necesidad de tener una teoría global compartida.

Según (Smith, 2004), existen también elementos negativos en el empleo de ontologías para el modelado de la realidad, si no en la teoría fundamental, sí en el enfoque que se utiliza para la conceptualización, pues menciona que algunos miembros de la comunidad de representación de conocimiento y de otras comunidades afines, han abrazado una u otra de las siguientes visiones:

- una visión de acuerdo a la cual no existe una realidad objetiva a la que correspondan los conceptos o términos generales de nuestra representación del conocimiento.
- una visión de acuerdo a la cual no podemos saber cómo es la realidad objetiva, así que no hay un beneficio práctico en tratar de encontrar una correspondencia con dicha realidad en nuestras representaciones del conocimiento.
- una visión según la cual el término 'realidad' en cualquier caso significa nada más que una construcción hecha de conceptos, de manera que cada sistema de conceptos, en principio, construye su propia 'realidad' o 'mundo posible'.

Esto lleva a las tesis en las que se piensa que la realidad objetiva desarrollada por las ciencias naturales no es otra cosa que una construcción cultural, comparable con la astrología y la brujería. En el campo de los sistemas de información pasa algo similar, Por ejemplo, podemos leer en el sitio owlseek.com (OwlSeek, 2007) la siguiente definición de ontología:

No podemos conocer la realidad en su forma más pura; podemos solamente interpretarla a través de nuestros sentidos y experiencias. Por lo tanto, cada quien tiene su propia perspectiva de la realidad. Una ontología es una especificación formal de una perspectiva. Si dos personas acuerdan utilizar la misma ontología cuando se comunican, entonces no debe existir ambigüedad en la comunicación. Para esto, una ontología codifica la semántica utilizada para representar y razonar con un conjunto de conocimientos.

En esta definición se asume como un hecho que la representación del conocimiento no tiene que ver con la realidad, sino con los conceptos concebidos como creaciones humanas. El problema de esta definición radica en asumir que la realidad puede ser modelada solamente a través de nuestras propias construcciones mentales; es decir, que la realidad se adapta al modelo y no al revés.

En muchos contextos las personas que trabajan con ontologías tratan con conceptos como analogía de las expresiones lingüísticas con las cuales están

asociadas. Así ellos hablan de "definir" y mapear los conceptos de diferentes ontologías, entendiendo los conceptos efectivamente como herramientas (análogamente a telescopios con microscopios) que pueden ser utilizadas para obtener acceso cognitivo a las entidades correspondientes en la realidad.

Es por ello que resulta interesante el trabajo realizado en este sentido. En este trabajo abrazaremos este enfoque para el procesamiento de modelos del terreno, pues como hemos visto, el procesamiento numérico ha sido bastante estudiado. Así mismo, existen trabajos que utilizan el enfoque conceptual para la solución de problemas; ejemplos de estos son los ya mencionados trabajos de (Guzmán, 2007; Moreno, 2007; Torres, 2007), en los que se utiliza la conceptualización del dominio geográfico como herramienta para el procesamiento semántico.

Otros trabajos presentan aproximaciones cualitativas cuyo objetivo es encontrar una identificación computacional de características del terreno. Por ejemplo, el modelo desarrollado por (Chevriaux y otros, 2005) es derivado de la frontera que se forma entre el cielo y la tierra, es decir, la silueta de la Tierra que se forma en el horizonte. Dicha silueta es modelada por una secuencia de formas del terreno.

Hay varios intentos de hacer esto (Frank, 1996; Montello y otros, 2003; Smith y Mark, 2003), los cuales están basados en propiedades geométricas de las líneas que conforman la silueta. Sin embargo, (Chevriaux y otros, 2005) presenta una idea en la que una silueta está formada por una secuencia de formas de terreno primitivas, las cuales están semánticamente categorizadas. Es decir, las propiedades geométricas de una silueta dada estarán encapsuladas en las propiedades semánticas de las formas del terreno que la constituyen. Para ello establecen tres restricciones importantes:

- Una forma de terreno puede ser detectada a través de diferentes niveles de resolución.
- Una forma de terreno primitiva está asociada a una y solo a una categoría.
- Una forma de terreno puede ser detectada a través de varias representaciones que se refieren a la misma categoría.

La propuesta se basa en el modelado de las diferentes clases de formas de terreno ubicándolas dentro de un cuadro contenedor (*bounding box*), e identificando unos puntos notables que se corresponden con los puntos en los que se intersecta la silueta y el cuadro contenedor. Al mismo tiempo, se identifican en el cuadro ocho puntos característicos, a saber, IA (izquierda-arriba), A (arriba), DA (derecha-arriba), D (derecha), DB (derecha-abajo), B (abajo), IB (izquierda-abajo) e I (izquierda). Así, se describe una forma de terreno mediante una secuencia de estos elementos, los cuales concuerdan (o se hacen concordar) con las intersecciones del cuadro con la silueta.

Posteriormente se establecen criterios para eliminar las secuencias duplicadas, definiendo que dos descripciones están duplicadas cuando comparten una misma parte de la silueta y son caracterizadas por la misma secuencia de vértices.

Otro trabajo en este sentido es el de (Chorowicz y otros, 1995), en el cual se menciona que no se ha desarrollado un reconocimiento directo de patrones tridimensionales, pero se puede presentar una aproximación simplificada que utiliza perfiles topográficos (extraído de un DEM), que podemos entender como las siluetas del trabajo de (Chevriaux y otros, 2005). La descripción que se presenta en (Chorowicz y otros, 1995) sobre la organización de las características, está basada en el análisis de sucesiones (de tipos de elementos) específicas a lo largo de los perfiles. Dicha sucesión de tipos puede ser codificada y así establecer reglas sintácticas, resultando en un análisis estructural sintáctico. Según (Chorowicz y otros, 1995), las características tridimensionales (las formas del terreno) pueden ser analizadas en dos dimensiones utilizando sus perfiles. Las variaciones en la cuesta a lo largo de esos perfiles pueden ser analizadas utilizando la primera y la segunda derivada local. Sucesiones de derivadas a lo largo de un perfil pueden caracterizar diez tipos geométricos elementales, con los cuales se construyen perfiles complejos (ver Figura 11).

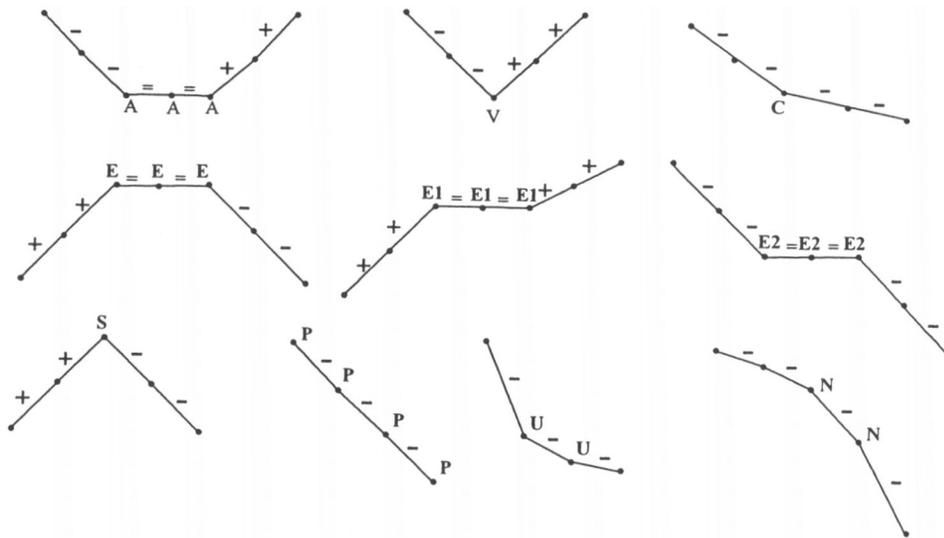


Figura 11. Descomposición del perfil de un DEM en diez tipos geométricos elementales

Tomada de (Chorowicz y otros, 1995).

Los perfiles son descritos en términos de características geomorfológicas. Con tales descripciones se forman secuencias que pueden ser reconocidas (mediante un análisis sintáctico) y mapeadas en el DEM, y subsecuentemente las características geomorfológicas que han sido reconocidas pueden ser desplegadas. Los autores reconocen que el reconocimiento de patrones utilizando análisis sintáctico, por sí solo no puede identificar completamente las características geomorfológicas más complejas.

En otro orden de ideas, en cuanto al desarrollo de técnicas y métodos para desarrollar ontologías, encontramos que (Mizen y otros, 2005) describen el desarrollo de un método sistemático para crear ontologías de dominio. Se ha elegido reconocer explícitamente las necesidades que difieren del dominio experto humano y de la máquina, por medio de la representación de dos tipos de ontologías: *conceptual* y *lógica*. La ontología conceptual está pensada para el entendimiento humano y la lógica está expresada en descripciones lógicas y se deriva de la ontología conceptual; por lo tanto, ésta se orienta para el procesamiento con máquinas. La principal contribución del trabajo es la división de estas dos etapas en el desarrollo de las

ontologías, con un énfasis esencial en el dominio de expertos, para crear la ontología conceptual que representa el conocimiento acerca del dominio.

Ordnance Survey es la organización nacional de mapas de Gran Bretaña, e investiga los beneficios potenciales de introducir un Sistema de Referencia Semántica Topográfico para mejorar la integración de los datos topográficos y temáticos. Una ontología es un componente importante de un sistema de referencia semántico, por lo cual el trabajo de (Mizen y otros, 2005) describe el desarrollo de una metodología para crear ontologías de dominio. Los autores indican que se encuentran investigando si las tecnologías actuales aplicadas al desarrollo de Web Semántica, y particularmente si las ontologías pueden facilitar la captura del dominio de conocimiento en tal forma que se detecten errores en la integración de datos o debido a la naturaleza explícita de la semántica, para prevenir este tipo de ocurrencias. En algunas ocasiones, los procesos y servicios manuales pueden ser automatizados; sin embargo, en aspectos de integración semántica algunas entradas deben realizarse en forma manual.

El Sistema de Referencia Semántica propuesto por (Kuhn, 2003) se describe en términos de ontologías de nivel superior que proporcionan una base para otras ontologías. Estas son ontologías que están pensadas para establecer de facto la semántica para un área particular. (Kuhn, 2003) cita que un sistema de este tipo es más que una ontología, debido a que éste soporta transformaciones entre dominios. (Mizen y otros, 2005) toman como base el sistema de referencia del Reino Unido para poder proporcionar una definición semántica común de los principales conceptos topográficos que pueden aplicar. Con esto se intenta construir un sistema que primeramente permita entender la estructura necesaria para una ontología y como ésta debe ser construida.

La creación de una ontología es usualmente visualizada como una tarea de adquisición de conocimiento, tal como la define (Kidd, 1987), en donde involucra la verdad, el análisis y la interpretación del conocimiento del humano experto, así como la transferencia de este conocimiento en una representación legible para la máquina.

Muchas metodologías para ontologías están basadas alrededor de una estructura similar o contienen criterios de diseño similares; sin embargo, todos difieren ligeramente

en un estándar formal de facto. La metodología propuesta por (Uschold y King, 1995) y la propuesta por (Fernández-López y otros, 1997) son los casos más representativos. Ambas proponen una fase de modelado inicial que desarrolla un entendimiento compartido implícito y glosarios del ser humano legible y explícito, pero informal, antes de estructurar la información en una ontología lógica. (Uschold y King, 1995) definen primero sus clases en forma precisa y sin ambigüedad, utilizando lenguaje natural, con lo cual las clases son estructuradas como jerarquías semiformales antes de construir una ontología lógica. En el caso de (Fernández-López y otros, 1997) desarrolla un método más sistemático para la conceptualización del dominio; éste proporciona un conjunto de tareas para asistir al modelador de la ontología en capturar y estructurar la información requerida para una ontología lógica, utilizando una serie de tablas, un diccionario de datos y una serie de árboles de conceptos. Algunas de estas representaciones son especificadas claramente asociadas a diversos dominios tales como la química y podrían ser adecuados para una ontología geográfica.

Las metodologías más populares (Uschold y King, 1995; Gómez-Pérez y otros, 1996) promueven la creación de árboles de conceptos y subgrupos de clases similares. Estas promueven una dependencia de las estructuras de lenguajes formales y motivan a los modeladores de ontologías a agrupar clases bajo encabezados familiares que en algunos casos no representan la lógica verdadera de las relaciones fundamentales. Esto es particularmente cierto para relaciones de subordinación, por ejemplo: en una ontología topográfica los conceptos pueden no necesariamente dividirse bajo las ramas "natural" y "artificial" en una jerarquía. Por lo tanto, una ontología debe ser más que una taxonomía, puesto que éstas minimizan el potencial para inferir y reutilizar la dependencia de conceptos. Bajo este argumento, es posible que no todos los dominios tengan una estructura de clasificación natural o clara y en algunas ocasiones no pueden ser divididos en módulos más pequeños. En este mismo sentido, se puede decir que la representación del conocimiento es procedural, ya que la gente encuentra dificultades para describir exactamente cómo los seres humanos llevan a cabo estos procedimientos o tareas.

2.8. Discusión

En este capítulo hemos tratado sobre algunos temas que consideramos de utilidad para el desarrollo del presente trabajo. En mayor o menor medida, cada tópico tratado ha aportado algo significativo. Los enfoques del INEGI respecto a los modelos de elevación fueron analizados, pues como se mencionó, el presente trabajo de tesis tiene su origen en la respuesta a las inquietudes externadas por el personal del Instituto; por ello consideramos no sólo necesario sino también justo mencionar cómo se trabajan los modelos de elevación en el INEGI. En el desarrollo del presente trabajo, fundamentalmente en el Capítulo 5, se utilizan los DEM como caso de estudio para probar la viabilidad y utilidad de la metodología que se presenta en este trabajo; así pues nos pareció lógico presentar algunas definiciones acerca de los modelos del terreno, sus orígenes y su evolución hasta llegar a ser modelos de gran utilidad. Presentamos, igualmente, que gran parte de los trabajos que existen están guiados desde un punto de vista numérico y que la gran mayoría de ellos están encaminados al análisis de los flujos o las líneas de drenaje existentes en el modelo. Encontramos que otros esfuerzos para caracterizar a los DEM siguen por la misma senda que ya hemos mencionado: análisis numérico. Todo ello nos da una pista de que lo que planteamos en este trabajo puede ser útil, si en un futuro avanzamos hacia un análisis semántico, no numérico, de los problemas. Por otro lado, debemos recordar que el enfoque que proponemos es general para los datos *raster* y que el modelo de elevación es el caso de estudio que hemos seleccionado; otro tipo de datos espaciales *raster* han de encontrar utilidad en otros campos y otras aplicaciones. Sin embargo, hemos seleccionado este caso de estudio, pero no nos enfocaremos a encontrar redes de drenaje, ni a caracterizar paramétricamente el terreno; lo que vamos a realizar es identificar las formas presentes en el terreno para posteriormente describirlo con base en una conceptualización del dominio de las formas del terreno.

En este capítulo exploramos también algunos otros temas que consideramos útiles para el procesamiento de datos de elevación y de las formas del terreno. En este sentido, la geomorfometría es un tema obligado. Al realizar el estudio sobre los trabajos realizados encontramos que, nuevamente, los estudios que se realizan tienen una visión

basada en el análisis numérico; sólo en algunos trabajos se manejan “categorías” o “clases” para realizar el análisis, sin llegar a expresar la conceptualización en otros términos que no sean los numéricos.

Finalmente, hemos analizado los enfoques sobre conceptualizaciones y ontologías que se manejan en la literatura. Encontramos trabajos meramente filosóficos que nos hablan de “existencia” y “realidad”, que nos sirven para encontrar las bases de la necesidad de organizar el conocimiento. De manera análoga, vimos trabajos que nos presentan el qué, el cómo, el porqué y el para qué de las ontologías, en los sentidos filosófico y práctico, como formas de modelar y entender la realidad y como herramientas para diseñar e implementar sistemas computacionales. Esto último, es una meta, si bien no final, sí intermedia para cumplir con los objetivos de este trabajo.

3. Marco teórico

En este capítulo presentamos algunos de los fundamentos teóricos inherentes a los tópicos del conocimiento que utilizaremos más tarde en este trabajo. Comenzaremos por explorar la teoría actual sobre ontologías, siempre desde el punto de vista de los sistemas de información.

3.1. Ontologías

La gente en general convive con el mundo cotidiano en una variedad de formas: además de actuar dentro del mundo directamente, utilizamos mapas, leemos periódicos, escuchamos el estado del tiempo y del tráfico. Cada una de esas actividades involucra cierta conceptualización, lo cual significa que involucramos un sistema de conceptos o categorías que dividen el dominio en el que nos encontramos en objetos, cualidades, relaciones, etc. Una ontología es una representación formal del conocimiento donde los conceptos, las relaciones y las restricciones conceptuales son explicados explícitamente mediante formalismos en un determinado dominio. Algunas definiciones de ontología son:

- *Una ontología es una entidad computacional, y no ha de ser considerada como una entidad natural que se descubre, sino como recurso artificial que se crea (Mahesh, 1996).*
- *Una ontología es un entendimiento común y compartido de un dominio, que puede comunicarse entre científicos y sistemas computacionales (Gangemi y otros, 1998).* Esta última característica, el hecho de que puedan compartirse y

reutilizarse en aplicaciones diferentes, explica en parte el gran interés suscitado en los últimos años en la creación e integración de ontologías.

El término ontología es actualmente utilizado por científicos de la información, en un sentido distinto al que lo utilizan los filósofos. Una ontología en el sentido de los sistemas de información no involucra cuestiones de existencia. Está diseñada para asistir la tarea de especificar y clarificar los conceptos empleados en un dominio dado, y por encima de todo, formaliza los conceptos dentro del marco de cierta teoría formal con una estructura lógica (sintáctica y semántica) bien entendida. Como hemos visto en el capítulo anterior, suele utilizarse ontología y conceptualización como sinónimos, aunque esto no sea del todo exacto, pues, según la definición de (Gruber, 1993), una ontología constituye *una especificación formal y explícita de una conceptualización*. En esta definición, conceptualización se refiere a un modelo abstracto y conceptual de algún fenómeno del mundo del que se identifican los conceptos que son relevantes; explícito hace referencia a la necesidad de especificar de forma consciente los distintos conceptos que conforman una ontología.

Desde un punto de vista más práctico, una ontología es una base de datos que describe los conceptos en el mundo o en algún dominio, algunas de sus propiedades y cómo los conceptos se relacionan con cada uno de los demás.

(Gangemi y otros, 1998) distinguen tres tipos de ontologías:

- **Ontologías de un dominio**, en las que se representa el conocimiento especializado pertinente de un dominio o subdominio, como la medicina, las aplicaciones militares, la cardiología o, en nuestro caso particular, el dominio geográfico.
- **Ontologías genéricas**, en las que se representan conceptos generales y fundacionales del conocimiento como las estructuras parte/todo, la cuantificación, los procesos o los tipos de objetos.
- **Ontologías representacionales**, en las que se especifican las conceptualizaciones que subyacen a los formalismos de representación del conocimiento, por lo que también se denominan *metaontologías*.

(Guarino, 1998) añade las ontologías que han sido creadas para una actividad o tarea específica (denominadas *task ontologies*). (Guarino, 1998) puntualiza que la denominación "conceptualización" se acerca más al entendimiento filosófico de una ontología como sistema de categorías que dan cuenta de una visión del mundo. Su clasificación de ontologías es la siguiente (se ilustra en la Figura 12).

- **Ontologías de alto nivel** que describen conceptos muy generales tales como, tiempo, materia, objeto, evento, acción, etc., que son independientes de un problema o dominio particular: parece entonces razonable, al menos en teoría, tener Ontologías de alto nivel unificadas para grandes comunidades de usuarios.
- **Ontologías de dominio** que describen el vocabulario relacionado al dominio genérico (como medicina, o automóviles) mediante la especialización de los términos introducidos en las ontologías de alto nivel.
- **Ontologías de tarea** que describen el vocabulario relacionado a una tarea o actividad genérica (como diagnóstico o venta), mediante la especialización de los términos introducidos en las ontologías de alto nivel.
- **Ontologías de aplicación** que describen conceptos tanto de un dominio como de una tarea particular, que son muchas veces especializaciones de ambas ontologías relacionadas (de dominio y de tarea).

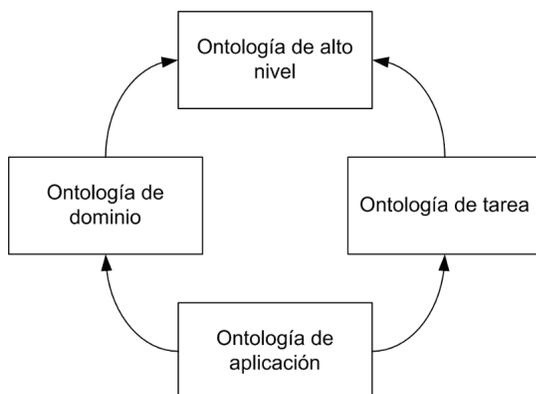


Figura 12. Tipos de ontologías, según Guarino (1998)

Tipos de ontologías, de acuerdo al nivel de dependencia a una tarea o punto de vista particular.

Una ontología necesariamente incluirá un vocabulario de términos y una especificación de su significado (definiciones e interrelaciones entre conceptos) que impone estructura al dominio y restringe las posibles interpretaciones. Según (Gruber, 1993), las ontologías se componen de:

- **conceptos:** son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc.
- **relaciones:** representan la interacción y enlace entre los conceptos de un dominio. Suelen formar la taxonomía del dominio. Por ejemplo: subclase-de, parte-de, parte-exhaustiva-de, conectado-a, etc.
- **funciones:** son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología. Por ejemplo, pueden aparecer funciones como: asignar-fecha, categorizar-clase, etc.
- **instancias:** se utilizan para representar objetos determinados de un concepto.
- **reglas de restricción o axiomas:** son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Por ejemplo: "Si A y B son de la clase C, entonces A no es subclase de B", "Para todo A que cumpla la condición K, A es C", etc. Los axiomas, junto con la herencia de conceptos, permiten inferir conocimiento que no esté indicado explícitamente en la taxonomía de conceptos.

Los objetos de una ontología se organizan en dos niveles: el nivel de individuos y el nivel de tipos. Nuestro conocimiento de individuos es muchas veces marcado por el uso de nombres propios ('Cerro del Chiquihuite') y por expresiones léxicas ('ese cerro'). Nuestro conocimiento de tipos es marcado por el uso de nombres comunes tales como 'cerro', 'montaña', etc.

La utilización de ontologías trae consigo algunas ventajas; por ejemplo, pueden existir ontologías múltiples: si el propósito de una ontología es hacer explícito algún punto de vista, en algunos casos, necesitamos combinar dos o más ontologías, donde

cada ontología introduce conceptualizaciones específicas. Por otro lado, se pueden identificar distintos niveles de abstracción estableciendo una topología de ontologías: se puede caracterizar una red de ontologías usando multiplicidad y abstracción. Al no poder realizar una descripción completa del mundo, se puede pensar una estrategia de construcción gradual que vaya de abajo hacia arriba o viceversa. Al mismo tiempo se puede manejar ambigüedad (que los seres humanos manejamos muy bien), a través de la multiplicidad de la representación, pues un concepto puede ser representado de muchas formas, por lo que pueden coexistir múltiples representaciones del mismo concepto.

3.1.1. Compromiso ontológico

En el sentido filosófico, nos debemos referir a una ontología como un sistema particular de categorías válidas para cierta visión del mundo. Así, este sistema no depende de un lenguaje particular. En inteligencia artificial se utiliza mayormente el término ontología como un artefacto de ingeniería, constituido por un vocabulario específico utilizados para describir cierta realidad, más un conjunto de características para intentar dar significado a las palabras del vocabulario. Este conjunto de reglas tiene usualmente la forma de una teoría lógica el primer orden, donde las palabras del vocabulario aparecen como nombres de predicados unarios o binarios, respectivamente llamadas conceptos y relaciones.

(Gruber, 1993) menciona que el cuerpo del conocimiento formalmente representado está basado en una conceptualización: los objetos, conceptos, y otras entidades que se asume que existen en algún área de interés, en las relaciones establecidas entre ellos. Una conceptualización es una visión abstracta y simplificada del mundo que deseamos representar para algún propósito. Todas las bases de conocimiento, sistemas basados en conocimiento, o agentes de nivel de conocimiento se apegan a una conceptualización, ya sea explícita o implícitamente. Una ontología es una especificación explícita de una conceptualización. Se utilizan ontologías comunes para describir compromisos ontológicos para un conjunto de entidades de tal forma que éstos puedan comunicarse acerca del dominio del discurso sin la necesidad

de tener una teoría global compartida. La idea de compromisos ontológicos está basada en la perspectiva de nivel de conocimiento. Las entidades que comparten un vocabulario no necesitan compartir una base de conocimiento. El compromiso a una ontología común garantiza la consistencia, pero no la completitud.

(Guarino, 1998) ha definido una conceptualización como una estructura $\langle D, R \rangle$, donde D es un dominio y R es un conjunto de relaciones relevantes en D . Esta definición ha sido utilizada también por (Gruber, 1993). El conjunto de relaciones que se define con la estructura anterior no refleja la intención se puede tener una relación. Así, se necesita hablar de relaciones intencionales, a las que llamaremos relaciones conceptuales, dejando el término "relación" para las relaciones no intencionales o relaciones extensionales.

Se define ahora el espacio de dominio como una estructura $\langle D, \mathcal{W} \rangle$, donde D es el dominio y \mathcal{W} es el conjunto maximal de estados de ese dominio (también llamados mundos posibles). Dado un espacio de dominio $\langle D, \mathcal{W} \rangle$, se define una relación conceptual ρ^n de aridad n en $\langle D, \mathcal{W} \rangle$ como una función total $\rho^n: \mathcal{W} \rightarrow 2^{D^n}$ desde \mathcal{W} hacia el conjunto de todas las relaciones n -arias en D . Para una relación conceptual genérica ρ , el conjunto $E_\rho = \{\rho(w) \mid w \in \mathcal{W}\}$ contendrá las extensiones admisibles de ρ . Una conceptualización para D puede ser ahora definida como una tripleta ordenada $C = \langle D, \mathcal{W}, \mathfrak{R} \rangle$, donde \mathfrak{R} es un conjunto de relaciones conceptuales en el espacio del dominio $\langle D, \mathcal{W} \rangle$. Así, se puede decir entonces que una conceptualización es un conjunto de relaciones conceptuales definidas en un espacio del dominio.

Consideremos ahora estructura $\langle D, R \rangle$. Ya que se refiere a un mundo particular (o estado de las cosas), se le llama estructura del mundo. Se puede ver que una conceptualización contiene muchas de esas estructuras del mundo, una para cada mundo: esas estructuras serán llamadas estructuras de mundos previstos de acuerdo a la conceptualización.

Sea $C = \langle D, W, \mathfrak{R} \rangle$ una conceptualización. Para cada mundo posible w que pertenece a W , la estructura prevista de w de acuerdo a C es la estructura $S_{wC} = \langle D, R_{wC} \rangle$, donde $R_{wC} = \{\rho(w) \mid \rho \in \mathfrak{R}\}$ es el conjunto de extensiones (relativas a w) de elementos de \mathfrak{R} . Se denota con S_C el conjunto $\{S_{wC} \mid w \in W\}$ de todas las estructuras de mundos previstos de C .

Consideremos ahora un lenguaje lógico L , con vocabulario V . Se define un modelo para L , la estructura $\langle S, I \rangle$, donde $S = \langle D, R \rangle$ es una estructura del mundo e $I: V \rightarrow D \cup R$ es una función de interpretación que asigna los elementos de D a símbolos constantes de V , y elementos de R a símbolos de predicados de V . Análogamente, se puede componer una interpretación intencional por medio de la estructura $\langle C, \mathfrak{I} \rangle$, donde $C = \langle D, W, \mathfrak{R} \rangle$ es una conceptualización y $\mathfrak{I}: V \rightarrow D \cup \mathfrak{R}$ es una función que asigna elementos de D a símbolos constantes de V , y elementos de \mathfrak{R} a símbolos de predicados de V . Se le llama a esta interpretación intencional un "compromiso ontológico" para L . Si $K = \langle C, \mathfrak{I} \rangle$ es un compromiso ontológico para L , decimos que L cumple C por medio de K , mientras C es la conceptualización intrínseca de K .

Dado un lenguaje L con vocabulario V , y un compromiso ontológico $K = \langle C, \mathfrak{I} \rangle$ para L , un modelo $\langle S, I \rangle$ será compatible con K si:

- $S \in S_C$;
- para cada constante c , $I(c) = \mathfrak{I}(c)$;
- existe un mundo w tal que, para cada símbolo de predicado p , I mapea a una extensión admisible de $\mathfrak{I}(p)$, esto es, que existe una relación conceptual ρ tal que $\mathfrak{I}(p) = \rho \wedge \rho(w) = I(p)$.

El conjunto $I_K(L)$ de todos los modelos de L que son compatibles con K es llamado el conjunto de modelos previstos de L de acuerdo con K . El conjunto de modelos previstos es solamente una débil caracterización de una conceptualización:

este solamente excluye algunas interpretaciones absurdas, sin que realmente se describa el "significado" del vocabulario.

3.1.2. Ontologías formales

Se puede ahora establecer el papel de una ontología, considerada como el conjunto de axiomas lógicos designados para definir el significado previsto de un vocabulario. Dado un lenguaje L con compromiso ontológico K , una ontología para L es el conjunto de axiomas designados de tal manera que el conjunto de sus modelos se aproxima lo mejor posible al conjunto de modelos previstos de L de acuerdo con K (Figura 13). Diremos que una ontología O para lenguaje L aproxima una conceptualización C existe un compromiso ontológico $K = \langle C, \mathfrak{S} \rangle$ tal que los modelos previstos de L de acuerdo con K están incluidos en los modelos de O . Una ontología cumple con C si k:

- ha sido diseñada con el propósito de caracterizar C , y
- la ontología aproxima C

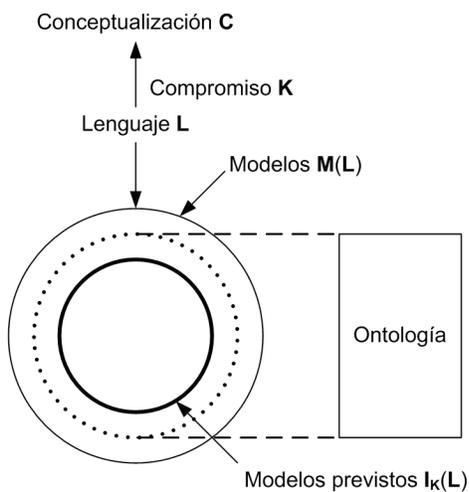


Figura 13. Aproximación de una ontología a una conceptualización

Los modelos posibles de un lenguaje lógico reflejan su compromiso a una conceptualización. Una ontología refleja indirectamente este compromiso (y conceptualización asociada) al aproximarse a este conjunto de modelos posibles.

Un lenguaje L cumple con ontología O siempre cumple con alguna conceptualización C tal que O está acuerdo con (commits) C .

Como hemos mencionado, una ontología es una teoría lógica que da significado posible de un vocabulario formal. Los modelos posibles de un lenguaje lógico utilizando dicho vocabulario están restringidos por su compromiso ontológico. Una ontología indirectamente refleja su compromiso (y la conceptualización intrínseca) al aproximarse a sus modelos posibles.

3.1.3. "Concepto"

Según (Smith, 2004) hay unos pocos intentos convincentes para definir el término "concepto", los cuales son marcados por la característica de la circularidad; ésta puede ser ilustrada utilizando como ejemplo la red semántica del sistema del lenguaje médico unificado (McCray, 1993), que define una idea o concepto como "un concepto abstracto, por ejemplo social, religioso o filosófico".

Los conceptos no son claramente distinguidos de las entidades en la realidad, o de los nombres o descripciones por el lado del lenguaje. La tendencia de manejar juntos conceptos y entidades se encuentra también en la lingüística. Sin embargo, dicha distinción es indispensable cuando nos embarcamos en el desarrollo de ontologías.

Lectura lingüística de "concepto"

La lectura básica del término "concepto" en la representación del conocimiento, comienza en el reconocimiento que diferentes términos, por ejemplo término en diferentes lenguas como "perro", "can" o "dog", pueden tener el mismo significado. "Concepto" es entonces utilizado en lugar de "nombre" o "palabra" como un dispositivo que nos permite abstraer de las diferencias sintácticas y enfocarnos hacia aquellas relaciones entre los términos que son importantes para el razonamiento. Algunas veces los conceptos parecen como una idea compartida entre aquellas personas utilizando los mismos términos. Algunas otras veces el concepto es visto como una construcción lógica. Una buena ontología es aquella que corresponde a la realidad que existe más allá de nuestros propios conceptos. Sin embargo, el

conocimiento por sí mismo es identificado con el conocimiento de nuestros propios conceptos, y si una ontología es solamente la especificación de una conceptualización, entonces la diferencia entre buenas y malas ontologías es inexistente.

“Concepto” es utilizado de tal manera que se asume que tiene connotaciones normalmente asociadas con términos como “propiedad”, “tipo (*kind*)” o “universo”, términos que utilizados en su forma normal no denotan entidades que sean productos del conocimiento humano. En la lectura lingüística, una frase como “edad humana es_un edad”, no es una aseveración acerca de las edades; en lugar de ello es una aseveración acerca del uso del lenguaje. Esto nos dice que el significado asociado con el nombre “edad humana” es más cercano o específico que el significado asociado con el nombre “edad”. Esta interpretación es especialmente común en trabajos de terminologías y diccionarios.

Lectura de “concepto” en sistemas de información

Podemos notar que el término “información”, así como término “modelo” (y como término “semántica”), es en sí mismo sujeto de muchas de las mismas confusiones que han estado asociadas con el término “concepto” en el área de representación del conocimiento y literatura asociadas (Smith, 2004). Una inspección de la práctica actual revela que los modeladores están preocupados en construir modelos de entidades de la realidad, por ejemplo con la construcción de modelos de la organización del genoma. El término “concepto” en sí mismo se refiere a entidades que son creadas por los modeladores. Los conceptos son criaturas del ámbito computacional que existen (en cierto sentido difícil explicar) a través de sus representaciones en software, en diagramas UML, en representaciones XML, en sistemas axiomáticos, etc.

3.1.3. Ontologías en los sistemas de información

Según (Gruber, 1993) las ontologías formales se diseñan (en el campo de los sistemas de información, por supuesto), pues cuando se selecciona el cómo será representado algo en una ontología, se están tomando decisiones de diseño. El mismo autor propone un conjunto preliminar de criterios de diseño para ontologías cuyo propósito es

compartir conocimiento e interoperar con otros programas basados en la conceptualización compartida.

- **Claridad.** Una ontología debe comunicar efectivamente el significado previsto de los términos definidos. Las definiciones deben ser objetivas. Las definiciones deben ser independientes del contexto social o computacional. Cuando una definición puede ser dada a través de axiomas lógicos, esto debe ser hecho. Cuando sea posible, una definición completa es preferida sobre una definición parcial. Todas las definiciones deben ser documentadas en lenguaje natural.
- **Coherencia.** Una ontología debe ser coherente, esto es, debe sancionar inferencias que son consistentes con las definiciones. Si una expresión que puede ser inferida a partir de los axiomas contradice una definición o ejemplo dado informalmente, la ontología es incoherente.
- **Extensibilidad.** Una ontología debe estar diseñada para explicar los usos del vocabulario compartido. Se debe poder definir nuevos términos para usos especiales basándose en el vocabulario en, de tal manera que no se requiera la revisión de las definiciones existentes.
- **Dependencia mínima de la codificación.** La conceptualización debe ser especificada a nivel del conocimiento, sin depender de una codificación particular a nivel de símbolos.
- **Compromiso ontológico mínimo.** Una ontología debe tener tan pocas excepciones acerca del mundo que estamos hablando como sea posible, permitiendo a quienes utilizan la ontología la libertad de especializar e instanciar la ontología como se necesite.

Por otro lado, (Smith, 2004) menciona que en el diseño de ontologías podríamos hablar no de conceptos como artefactos computacionales o lingüísticos, en lugar de ello hablaremos de universos (*universals*), concebidos tal como son en la realidad. El término "universo" significa lo que las instancias correspondientes (por ejemplo todas las ballenas, todas las bacterias) tienen en común. Los universos son invariantes en la realidad.

Los universos son instanciados por instancias en espacio y tiempo. Esto muestra una diferencia primitiva entre instancias particulares y no particulares. Entre las particulares (individuos, *tokens*) tenemos: usted y yo, el planeta Tierra, etc. Ejemplos de no particulares son universos como: raza humana, mariposa, muerte, etc. Un universo es definido como cualquier cosa que es instanciada, y una instancia como cualquier cosa que instancia algún universo.

(Mizen y otros, 2005) describe el desarrollo de un método sistemático para crear ontologías de dominio. Se ha elegido reconocer explícitamente las necesidades que difieren del dominio experto humano y de la máquina, por medio de la representación de dos tipos de ontologías: *conceptual* y *lógica*. La ontología conceptual está pensada para el entendimiento humano y la lógica está expresada en descripciones lógicas, y se deriva de la ontología conceptual, por lo tanto, ésta se orienta para el procesamiento de máquina.

La creación de una ontología es usualmente visualizada como una tarea de adquisición de conocimiento, tal como la define (Kidd, 1987), en donde involucra la verdad, el análisis y la interpretación del conocimiento del humano experto, así como la transferencia de este conocimiento en una representación legible para la máquina.

(Mizen y otros, 2005) se enfoca en el desarrollo de ontologías de dominio. En este sentido una ontología de dominio es una formalización del conocimiento en un área particular (dominio) como por ejemplo: topografía, ecología, biología, etc., y difieren de otros tipos de ontologías como las ontologías de tareas (las cuales pueden definirse como una formalización del conocimiento necesario para solucionar un problema específico, abstraída de un nivel superior de una situación específica o un contexto organizacional).

Cada ontología puede ser pensada como un par de dos ontologías enlazadas: una conceptual y una lógica. La conceptual está pensada para una concepción humana e intenta balancear la necesidad para una formalidad máxima de la ontología mientras retiene una clara comprensión humana. Esto significa que expertos en un dominio capturan conocimiento del mismo, lo cual sirve para registrar y describir sus ideas explícitamente en una estructura estándar. Esta estructura debe ser libre de las

restricciones de la ontología lógica y no debe estar influenciada por las estructuras o reglas que la descripción lógica presenta. La ontología lógica proporciona una representación legible para la máquina, típicamente utilizando un derivativo de lógica de primer orden tal como una descripción lógica y producida por un lenguaje como OWL. Ésta es generada a partir de la ontología conceptual. Sin embargo se puede asumir que se pierde información al momento durante el proceso de traducción debido a la incapacidad de las descripciones lógicas para representar la complejidad verdadera de una ontología.

La técnica propuesta por (Mizen y otros, 2005) está orientada a proporcionar expertos de dominio con un conjunto sistemático y entendible de criterios y directivas para asistir a estos expertos a través del ciclo de vida completo de la ontología. Se describe el esqueleto básico de tareas para construir una ontología de dominio conceptual, soportada por ejemplos de una ontología para el manejo de riesgos de inundación. La metodología está compuesta por cuatro principales etapas: decidir sobre los requerimientos y el contenido de la ontología, poblar un glosario de conocimiento y construir un conjunto de tripletas (relaciones entre conceptos), evaluar las ontologías y documentar la ontología conceptual.

Etapas 1, Preparatoria. La primera tarea a realizar en esta etapa es la identificación de los requerimientos. En el dominio del conocimiento, los expertos de dominio formulan un conjunto de requerimientos para la ontología. Los criterios para identificar los requerimientos son similares a los utilizados por (Uschold y King, 1995). El modelador registra su definición de una ontología, el propósito para construirla, el ámbito o alcance de la ontología pensada y un conjunto de preguntas de competencia.

El ámbito o alcance de la ontología debe estar contenido y restringido en tamaño, para que las ontologías producidas sean manejables y consistentes. Si el ámbito es muy grande (por ejemplo, el dominio topográfico), entonces el modelador puede desear subdividir el dominio en ontologías de dominio adicionales (hidrología, áreas urbanas, etc.), e integrar los módulos en conjunto cuando todos estén completos.

Para ontologías de dominio, las preguntas de competencia son formuladas para que éstas puedan utilizarse para checar en cada etapa de la construcción de la

ontología si las relaciones adecuadas han sido creadas entre los conceptos, y si éstas describen suficientemente el dominio.

En segundo lugar, en esta etapa se deben coleccionar los datos. Se debe adquirir la base de conocimiento de entrada necesaria para construir el modelo conceptual, basado en el propósito, ámbito y preguntas de competencia. El modelador debe reutilizar otras ontologías que puedan adecuarse al propósito de la ontología a construir. De igual forma, éste debe identificar cualquier documento que capture el conocimiento que se requiera en la ontología. La información debe ser adaptada al propósito, al ámbito y a la representación de los dominios. Asimismo, cuando la información no está disponible o no es suficiente, la ontología debe ser construida utilizando el conocimiento de los expertos en los dominios, ya sea en forma manual o semiautomática. En este caso, el modelador debe extraer sentencias semiestructuradas que contengan información requerida por la ontología.

Etapa 2, Poblar un glosario de conocimiento. El primer paso en capturar y estructurar el conocimiento del dominio es poblar un glosario de conocimiento. Algunas comparaciones con este término pueden ser "diccionario de datos" y las "tablas de atributos"; sin embargo, el glosario es más adaptable para una audiencia menos familiar con "clases" y "atributos". Los autores utilizan lenguaje natural para los encabezados del glosario y proporcionan las directrices para asistir los expertos de dominio, identificando la información correcta. La información requerida para el glosario es extraída de las sentencias semiestructuradas y es mejorada por el experto del dominio. El modelador registra la definición lingüística de un término (i.e. sustantivo, verbo) como un paso intermedio para identificar que términos son conceptos en la ontología y que términos son relaciones o características (atributos). En este caso los sustantivos tienden a ser conceptos y los verbos relaciones. Definiendo los términos y registrando éstos en un significado, se puede clarificar su definición e interpretación del término y su uso dentro de la ontología.

Los conceptos principales que son claves para describir el dominio, son distinguidos a partir de "conceptos secundarios", los cuales describen aspectos de los conceptos principales o tienen relaciones distintas con los mismos. Posteriormente, los conceptos

no son miembros del dominio en consideración, pero son necesarios para habilitar conceptos en el dominio a ser relacionado con otros dominios. Por ejemplo, el en caso de la hidrología, un concepto principal "río" podría definirse una relación a un concepto secundario "campo" que pertenecería a un dominio diferente. Los conceptos principales son vitales para la ontología y cuentan con más relaciones que el resto de los conceptos, éstos pueden ser descritos dentro de la ontología no solo por sus relaciones a otros conceptos, sino por su relación a sus atributos (i.e. tiene tamaño, tiene localización), a lo que se denomina como "características" en la metodología de la ontología conceptual. El experto del dominio identifica éstos utilizando sentencias semiestructuradas y su propio conocimiento, y usará esta información para describir explícitamente los conceptos principales por sus enteros y partes en la ontología conceptual. Las características de los conceptos secundarios no son requeridas en la ontología conceptual.

Etapas 3, Crear una red semántica de tripletas. La siguiente etapa es utilizar la información capturada en el glosario de conocimiento para construir una red de conceptos que describa el dominio. Una red de conceptos visualiza una ontología como nodos (conceptos) y enlaces (relaciones entre conceptos). La técnica propuesta en este trabajo, limita el uso de relaciones jerárquicas que pueden originar la creación de falsos agrupamientos de conceptos o divisiones innecesarias entre grupos de conceptos (i.e. la división de los conceptos "natural" y "artificial" en una clasificación típica de objetos topográficos). Además, se argumenta que una inferencia rica puede lograrse si los conceptos son definidos dentro de sí mismos y a través de un rango de relaciones a otros conceptos (i.e. relaciones de concepto a concepto y relaciones de concepto – característica). Se han adoptado los criterios de diseño de Gruber para reflejar las interpretaciones, los cuales son: claridad, coherencia, extensibilidad, codificación bias mínima y compromiso ontológico mínimo. Los autores especifican un número de reglas para crear una red de conceptos para forzar a la consistencia de la ontología.

El modelador debe trabajar con una técnica *bottom-up*, para construir la ontología con los conceptos más específicos, los cuales pueden entonces generalizarse cuando

sea necesario (identificando super conceptos) para prevenir conceptos que estén siendo agrupados bajo jerarquías o una semántica falsa.

La herencia múltiple debe solo crearse cuando el concepto puede heredar todas las características de ambos (conceptos particulares y "super-conceptos").

Se recomienda crear jerarquías cuando sea necesario describir el dominio en donde algunos conceptos hereden todas las características de sus "super-conceptos" más otras características o cuando la ontología necesite moverse entre diferentes niveles de granularidad.

Si los términos del concepto o de la relación nueva son necesarios cuando se construye la red de conceptos, éstos deben validarse en contra del ámbito o alcance, objetivo o propósito, y deben agregarse al glosario antes de ser colocados en la ontología conceptual. Esto garantizará que el término utilizado sea consistente con su definición.

Los conceptos o grupos pequeños de conceptos que no tengan enlaces con el resto de la red de conceptos deben descartarse de ser posible.

Los autores han utilizado dos métodos de visualización de la red de conceptos: utilizando diagramas de red para desplegar gráficamente enlaces entre conceptos y el otro es creando una lista de "tripletas de la ontología conceptual", en donde los conceptos y relaciones son registradas como sujeto-predicado-objeto.

Etapas 4, Documentación del modelo conceptual. La documentación de la ontología conceptual debe incluir el glosario de conocimiento, las redes de conceptos y relaciones, pérdidas de información y cualquier regla y suposiciones definidas en el proceso de modelado.

Por otra parte, los autores citan que existen tres problemas en esta metodología: 1) Ésta no proporciona directrices para limitar la ontología a pequeños ámbitos o alcances que produzcan ontologías más pequeñas y manejables. 2) Las directrices para separar conceptos dentro de formas más adecuadas en ontologías de tarea o de dominio no son claras, y 3) No existen guías para modularizar la ontología, con el

objeto de que ésta pueda producirse por varias personas al mismo tiempo o para dividirse en subdominios que puedan reutilizarse posteriormente.

En muchos casos, el nivel explícito de detalle requiere de una ontología conceptual que no fue capturada en tripletas de la ontología conceptual, por lo cual los tipos de relaciones registrados son genéricos y ambiguos. Por ejemplo: “La precipitación causa inundación”, de la cual la ontología lógica no inferiría que es una relación lógica verdadera. En otro sentido cualquier información que no pudiera modelarse como tripletas se denominaría entonces como una “restricción”.

3.2. Agrupamiento (Clustering)

El agrupamiento puede ser considerado el problema más importante del aprendizaje no supervisado; como cualquier otro problema de este tipo, trata de encontrar una estructura en una colección de datos sin etiquetas. Una definición de agrupamiento puede ser más o menos: el proceso de organizar objetos en grupos cuyos miembros son de alguna manera similares. Un grupo o *cluster* es una colección de objetos que son “similares” entre ellos y son “disimilares” a los objetos que pertenecen a otros grupos.

En la Figura 14 se muestra un ejemplo gráfico simple, en tal caso se puede identificar fácilmente cuatro grupos en los que los datos pueden ser divididos; el criterio de similitud es la distancia: dos o mas objetos pertenecen a un grupo si se encuentran “cerca” de acuerdo a cierto criterio de distancia (en el caso de la figura, es una distancia geométrica). Esta forma de agrupamiento se llama basada en distancia.

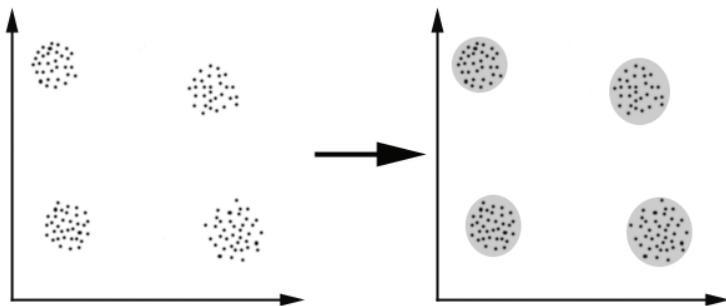


Figura 14. Ejemplo gráfico de agrupamiento (*clustering*)

Otro tipo de agrupamiento es el conceptual: dos o más objetos pertenecen al mismo grupo si este define un concepto común para todos los objetos que pertenecen a él. En otras palabras los objetos son agrupados de acuerdo a su adaptación a conceptos descriptivos, no solo a medidas simples de similitud.

La meta del agrupamiento es determinar el los elementos comunes intrínsecos en un conjunto de datos sin etiquetar. Pero el decidir qué constituye un buen agrupamiento es muy difícil, no existe un “mejor” criterio absoluto el cual sea independiente del objetivo final del agrupamiento. En consecuencia, es el usuario que debe proveer los criterios de agrupamiento, de tal manera que éste cumpla con sus necesidades.

3.2.1. K-means

K-means (MacQueen, 1967) es uno de los algoritmos de aprendizaje no supervisado más simples que resuelven el problema del agrupamiento. El procedimiento sigue una forma simple y fácil para clasificar un conjunto de datos en cierto número preestablecido de grupos (se asumen k grupos). La idea principal es definir k centroides, uno para cada grupo. Esos centroides deben ser colocados astutamente, pues con diferentes posiciones de estos se obtienen diferentes resultados. La mejor decisión es colocarlos tan lejos uno de otro como sea posible. El siguiente paso es tomar cada punto perteneciente al conjunto de datos y asociarlo al centroide más cercano. Cuando se han asociado todos los puntos, el primer paso se ha completado y se cuenta ya con un primer agrupamiento. En este punto necesitamos recalcular k nuevos centroides como los “centros” de cada grupo resultante del paso anterior. Después de calcular esos k nuevos centroides, se debe realizar una nueva asociación entre el conjunto de datos original y los nuevos centroides. En este momento hemos generado un ciclo en el cual, paso a paso, se cambian de posición los k centroides. El ciclo termina hasta que no hay cambios en los centroides. Finalmente, podemos decir que este algoritmo trata de minimizar una función objetivo, en este caso una función de error cuadrático $J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$, donde $\|x_i^{(j)} - c_j\|^2$ es la medida de distancia

seleccionada entre el punto $x_i^{(j)}$ y el centroide c_j . El algoritmo K-means se resume en la Tabla 2.

Tabla 2. Algoritmo K-means

1. Colocar k puntos en el espacio representado por los objetos que están siendo agrupados. Esos puntos representan el grupo inicial de centroides.
2. Asignar cada objeto al grupo cuyo centroide sea el más cercano.
3. Cuando todos los objetos hayan sido asignados, recalculan las posiciones de los k centroides.
4. Repetir los pasos 2 y 3 hasta que los centroides no se muevan al recalcularlos.

3.2.2. Fuzzy C-Means

Fuzzy C-means es un método de agrupamiento que permite que cada elemento en el conjunto de datos a clasificar pertenezca a dos o más clusters. Este método, desarrollado por (Dunn, 1973) y mejorado por (Bezdek, 1981), es frecuentemente utilizado en reconocimiento de patrones. Está basado en la minimización de la función

objetivo $J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2$, donde $m \geq 1$ es cualquier número real, u_{ij} es el grado de

membresía al grupo j del dato x_i , el cual es el i -ésimo dato del conjunto. Así mismo, c_j es el centroide del j -ésimo grupo y $\|x_i - c_j\|$ es alguna medida de similitud entre x_i y c_j . La partición difusa de los datos es llevada a cabo a través de una optimización interactiva de la función objetivo mediante la actualización de las membresías (u_{ij}) y

los centroides (c_j), esto se realiza por $u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$ y $c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m}$

respectivamente. La actualización se detiene cuando $\max_{ij} \left\{ |u_{ij}^{(k+1)} - u_{ij}^{(k)}| \right\} < \varepsilon$, donde

$0 \leq \varepsilon \leq 1$ es el criterio de convergencia, (k) y $(k+1)$ son el k -ésimo y el $(k+1)$ -ésimo pasos en la iteración. En la Tabla 3 se resume el algoritmo Fuzzy C-means.

Tabla 3. Algoritmo Fuzzy C-means

1. Inicializar la matriz $U = [u_{ij}]$, como $U^{(k)}$ con $k = 0$.

2. En el paso (k) calcular los centroides $C^{(k)} = [c_j]$, $c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m}$.

3. Actualizar $U^{(k)}$ y $U^{(k+1)}$, $u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$.

4. Si $|U^{(k+1)} - U^{(k)}| < \varepsilon$ entonces terminamos, de otra forma regresar al paso 2.

3.3. Relaciones

En esta sección presentamos algunas nociones básicas de relaciones entre elementos de un conjunto, pues estas nos servirán para la formalización de la metodología de conceptualización que se presentará en el siguiente capítulo. Particularmente nos interesan las relaciones binarias y sus propiedades.

Sea A un conjunto cualquiera; se dice que R es una relación binaria en A si $R \subseteq A \times A$, es decir, si R es un subconjunto del producto cartesiano citado (Rosen, 1995).

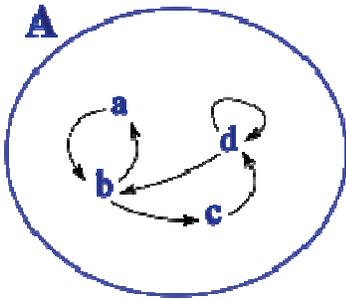


Figura 15. Relación entre los elementos de un conjunto

Dado el conjunto $A = \{a, b, c, d\}$, y la relación entre los elementos de este conjunto, representada en la Figura 15, se puede ver que solo hay un conjunto (A) y que la relación entre los elementos es interior al conjunto, en este caso representado por las flechas.

En este caso podemos decir: $aRb, bRc, cRd, dRd, dRb, bRa$ Como enumeración de las relaciones entre los elementos del conjunto A .

3.3.1. Propiedades

Una relación binaria R sobre un conjunto A es **reflexiva** si se cumple que para todo a perteneciente a A , a está relacionado consigo mismo. Si está escrita en forma de pares, deben figurar tantos pares (a, a) como elementos tenga el conjunto. Si algunos pares (a, a) figuran y otros no, la relación es no-reflexiva. Si ningún par (a, a) figura, la relación es Antireflexiva.

- Reflexiva: $\forall a \in A, aRa$
- No-reflexiva: $\exists a \in A \ni a \neg Ra$
- Antireflexiva: $\forall a \in A, a \neg Ra$

Una relación binaria R sobre un Conjunto A es **simétrica** si se cumple que para todo a y b pertenecientes a A , si a está relacionado con b entonces b está relacionado con a .

- Simétrica: $\forall a, b \in X, aRb \Rightarrow bRa$

- Antisimétrica: $\forall a, b \in X, aRb \wedge bRa \Rightarrow a = b$

Nótese que la antisimetría no es lo opuesto de la simetría. Existen relaciones que son simétricas y antisimétricas al mismo tiempo (como la relación de igualdad), relaciones que no son simétricas ni antisimétricas (como la relación de divisibilidad), relaciones que son simétricas pero no antisimétricas (como la relación de congruencia módulo n), y relaciones que son antisimétricas pero no simétricas (la relación menor-que).

La relación ser menor o igual también es antisimétrica dado que si a es menor o igual que b y b es menor o igual que a es porque $a = b$.

Una relación binaria R sobre un conjunto A es **transitiva** si se cumple que para todo $a, b, y c$ pertenecientes a A , si a está relacionado con b y b está relacionado con c , entonces a está relacionado con c .

- Transitiva: $\forall a, b, c \in A, aRb \wedge bRc \Rightarrow aRc$

Una relación que es reflexiva y transitiva es llamada un preorden. Un preorden que es antisimétrico es un orden parcial. Un preorden que también es simétrico es una relación de equivalencia.

3.3.2. Relación de equivalencia

Sea A un conjunto y supóngase que existe una relación binaria R definida sobre A . Se dice que R es una relación de equivalencia, si es reflexiva, simétrica y transitiva. Una relación de equivalencia R define subconjuntos disjuntos en A llamados clases de equivalencia. Una clase de equivalencia se define como el conjunto de todos los elementos relacionados con un elemento $a \in A$, esto es $[a] = \{b \in A \mid bRa\}$, se le llama la clase de equivalencia asociada al elemento a . Al elemento a se le llama representante de la clase.

Se llama orden al número de clases que genera una relación de equivalencia; si éste es finito, se dice que la relación es de orden finito. Finalmente, el conjunto de todas las clases de equivalencia se denomina conjunto cociente y se le denota como A/R .

3.3.3. Relación de orden

Se dice que una relación R en un conjunto A es de orden si es reflexiva, antisimétrica y transitiva. Si tenemos una relación de orden R en el conjunto A , podremos utilizar el símbolo \leq para su notación. Un conjunto ordenado se puede denotar por (A, \leq) .

Tipos de relaciones de orden.

Dada una relación de orden \leq , diremos que se trata de una **relación de orden total** si dados dos elementos cualesquiera a y b pertenecientes al conjunto A , o bien a es menor o igual que b o bien b es menor o igual que a : $a \leq b, b \leq a \forall a, b \in A$.

Si por el contrario, si esta propiedad no se verifica, diremos que se trata de un conjunto con una relación de orden parcial.

4. Metodología

En secciones anteriores se han descrito, a grandes rasgos, algunos aspectos a desarrollar en este trabajo. Así mismo, se han presentado diferentes enfoques para dar solución a los problemas relacionados. En esta sección, abordaremos nuevamente esos temas con mayor detalle.

Como ya se mencionó, uno de los objetivos de este trabajo es definir la SE aplicada a los DER mediante la identificación de un conjunto de elementos, de tal manera que cualquier DER pueda ser descrito a través de la combinación de éstos. A tal proceso lo denominamos *representación semántica*.

Comenzaremos la descripción de la metodología, retomando la visión general propuesta para realizar la representación semántica de DER, que se muestra en la Figura 16, en la cual se presentan las tres etapas, y los productos e insumos de estas, que conforman la metodología para realizar la representación.

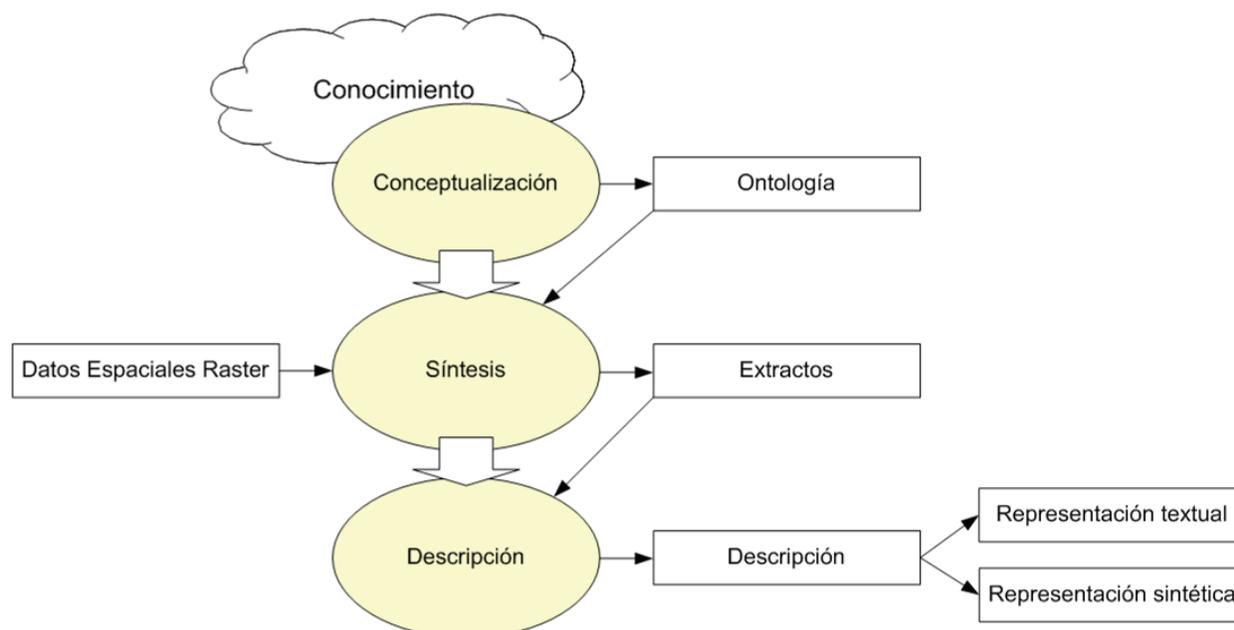


Figura 16. Panorama general de la metodología planteada

- **Conceptualización.** El primer paso para realizar una representación semántica de DER es contar con una conceptualización detallada, la cual consta de tres partes, representadas a través de sendas ontologías, llamadas: de alto-nivel (*top-ontology*), de dominio (*domain-ontology*) y de aplicación (*application-ontology*), respectivamente:
 - La conceptualización del dominio geoespacial (de alto nivel)
 - La conceptualización del dominio particular del CDER con el que se está trabajando y,
 - La conceptualización de la aplicación en la cual se realizará la representación.
- **Síntesis.** Este proceso es guiado por la conceptualización, en la que se tienen algoritmos para la extracción de características específicas del CDER, de acuerdo al concepto que se quiere describir.
- **Descripción.** Como resultado del proceso de síntesis se obtienen “extractos”, que no son otra cosa que segmentos del CDER identificados con algún concepto

particular. El proceso de descripción determina qué son esos extractos y los describe de acuerdo a la conceptualización.

Cada una de las etapas mencionadas tiene asociados algunos productos y/o insumos. A continuación se describen brevemente.

- **Conocimiento.** Se refiere a lo que se sabe sobre el dominio; el conocimiento es poseído por los expertos en el tema.
- **Ontología.** Es una representación que se hace del conocimiento, la cual consta de algunos elementos que la hacen útil para este fin, y para el posterior procesamiento del conocimiento.
- **Datos espaciales raster.** Es la fuente de información en formato *raster*, la cual es el objeto de análisis.
- **Extractos.** Son elementos que han sido identificados y separados a partir de los DER.
- **Descripción.** Es la representación semántica que se hace de los objetos que han sido encontrados en el CDER, y que es realizada con base en la conceptualización del dominio que se ha hecho. Tal descripción puede ser de manera textual (mediante algún lenguaje) o de manera sintética (utilizando formas primitivas que representan los objetos).

En las siguientes secciones se desarrolla más detalladamente cada una de las etapas propuestas para la descripción semántica.

4.1. Conceptualización

En esta sección se describe la metodología para realizar la conceptualización del dominio geográfico. Como se verá, la parte fundamental de esta metodología radica en la filosofía de mantener al mínimo la cantidad de elementos axiomáticas que se utilizan para la conceptualización. Proponemos un conjunto de tan sólo tres relaciones axiomáticas fundamentales (es, tiene y hace) y un conjunto de relaciones auxiliares

(preposiciones), por medio de las cuales creemos que es posible conceptualizar el dominio geográfico.

Podríamos pensar que la reducción del número de relaciones axiomáticas es una limitante para la riqueza de expresividad que puede tener la conceptualización. Por el contrario, el universo de relaciones posibles entre conceptos no está delimitado de facto, pues las “relaciones” en el sentido clásico no están predefinidas. Esto tiene a priori dos ventajas importantes: primero, se pueden definir tantas relaciones “clásicas” como se desee, pues estas son relaciones-concepto, es decir, las relaciones clásicas forman parte de la conceptualización no como axiomas, sino como conceptos. Esto trae consigo la segunda ventaja: las relaciones tienen una semántica asociada, dada por la misma conceptualización, y no a través de axiomas, sino por medio del contexto de cada relación.

Para ilustrar esto consideremos una relación axiomática ampliamente utilizada, la relación “parte_de”; la cual implica que un concepto es elemento constituyente de otro. Utilizando la metodología que proponemos es posible crear esta relación como una relación-concepto, en donde podemos definir “parte” como un concepto y utilizar las relaciones axiomáticas para crear una relación equivalente. Por ejemplo, supongamos la siguiente relación “corazón parte_de cuerpo”; en esta aseveración se involucran dos conceptos (corazón y cuerpo) y una relación axiomática (“parte_de”). Utilizando la metodología que proponemos, la relación se expresaría como “corazón es parte de cuerpo”, en donde utilizamos tres conceptos (corazón, parte y cuerpo) y dos relaciones axiomáticas, una fundamental (es) y la otra auxiliar (de). La ganancia de esto es que la semántica se mantiene y la relación “parte_de” no hubo de ser definida de facto, sino que se define el concepto “parte” (teniendo la riqueza semántica de un concepto) y se utiliza para crear una nueva relación.

De acuerdo a la metodología propuesta, como parte de la conceptualización se deben definir los siguientes elementos:

- **Axiomas.** Etimológicamente hablando, los axiomas son elementos de verdad evidente sobre los cuales descansa el resto del conocimiento o sobre el cual se construyen otros conocimientos. Para fines de la metodología, los axiomas son

conceptos y relaciones cuya semántica es bien conocida y no requieren ser definidos con base en otros conceptos.

- **Relaciones.** Son elementos de vínculo entre conceptos. Las relaciones axiomáticas forman parte de los axiomas y por ende no requieren definición. Para expresar relaciones no-axiomáticas se utilizan conceptos que toman el rol de una relación en el sentido clásico, y se organizan de acuerdo a un modelo de relación no-axiomática.
- **Propiedades.** Son conceptos agregados o asociados a través de una relación de pertenencia a otro concepto. Existen propiedades abstractas y propiedades concretas.
- **Habilidades.** Son conceptos que definen acciones y/u operaciones asociadas a otro concepto, incluso describen la manera que un concepto interactúa con otros. Al igual que las propiedades, existen habilidades abstractas y habilidades concretas.
- **Instancias.** Son conceptos cuyas propiedades y habilidades son concretas. Desde un punto de vista pragmático, el conjunto de todas las instancias conforman los datos originales (el CDER o la base de datos geográfica).
- **Conceptos.** Son una colección de habilidades y propiedades, con una existencia única. Existen tres tipos de conceptos: conceptos relación, conceptos estándar y conceptos clase.
- **Clases.** Son conceptos que no tienen instancias y forman entre sí una partición completa.
- **Restricciones.** Son sentencias que permiten evitar inconsistencias en los elementos que definen a la conceptualización.

A continuación describiremos los elementos mencionados. Cabe mencionar que esta metodología ha sido desarrollada en conjunto con el autor del trabajo de tesis doctoral (Torres, 2007).

4.1.1. Axiomas

La filosofía central de esta metodología de conceptualización está en minimizar el número de axiomas a través de la minimización de la cantidad de relaciones axiomáticas. Para ello, definiremos un conjunto pequeño de relaciones binarias axiomáticas divididas en dos subconjuntos. El primer subconjunto consta de tres relaciones que llamaremos relaciones fundamentales, el cual lo denotamos por A_1 , a saber:

$$A_1 = \{es, tiene, hace\}$$

La relación "es" es una relación de existencia o identidad por medio de la cual podremos caracterizar los conceptos. Esta relación implica la herencia de propiedades y habilidades, y nos permite jerarquizar los conceptos, es decir, que podemos formar una jerarquía de conceptos a través de esta relación. Algunos otros nombres para esta relación pueden ser "hijo de", "es_un".

Como relación binaria, "es" tiene las siguientes propiedades. Sea C un conjunto de conceptos:

- **Es antisimétrica:** $\forall a, b \in C, a(es)b \wedge b(es)a \Rightarrow a = b$, si el concepto A "es" B y B "es" A, necesariamente A y B son el mismo concepto., por ejemplo lago "es" cuerpo_de_agua pero cuerpo_de_agua no("es") lago, puesto que existen otros objetos geográficos que también son cuerpos de agua, tales como presa, laguna, etc.
- **Es reflexiva:** $\forall a \in C, a(es)a$, todo concepto A "es" él mismo, por ejemplo para cualquier lago, podemos decir que lago "es" lago, puesto que un objeto geográfico tiene identidad propia.
- **Es transitiva:** $\forall a, b, c \in C, a(es)b \wedge b(es)c \Rightarrow a(es)c$, si A "es" B y B "es" C, entonces A "es" C, por ejemplo si acueducto "es" og_artificial y og_artificial "es" objeto_geografico entonces acueducto "es" objeto_geográfico.

La relación "tiene" es una relación de pertenencia (agregación) o uso (asociación), mediante la cual se definen las propiedades de las cuales un concepto está constituido.

Como relación binaria, "tiene" tiene las siguientes propiedades:

- **Es no-simétrica:** $\exists a, b \in C, a(\text{tiene})b \wedge b(\text{tiene})a \Rightarrow a = b$, existen los conceptos A y B de tal manera que si A "tiene" B y B "tiene" A, entonces A y B son el mismo concepto. Por ejemplo estado "tiene" municipio pero municipio no("tiene") estado.
- **Es no-reflexiva:** $\exists a \in C \exists a \neg (a(\text{tiene})a)$, existe algún concepto A que no se tiene a sí mismo como propiedad; por ejemplo `area_urbana` no("tiene") `area_urbana`.
- **Es transitiva:** $\forall a, b, c \in C, a(\text{tiene})b \wedge b(\text{tiene})c \Rightarrow a(\text{tiene})c$, si A "tiene" B y B "tiene" C, entonces A "tiene" C, por ejemplo si país "tiene" estado y estado "tiene" municipio entonces país "tiene" municipio.

La relación "hace" es una relación de poder o de acción, mediante la cual se definen las habilidades, acciones u operaciones que están asociadas a un concepto.

Como relación binaria, "hace" tiene las siguientes propiedades:

- **Es no-simétrica:** $\exists a, b \in C, a(\text{hace})b \wedge b(\text{hace})a \Rightarrow a = b$, existen los conceptos A y B de tal manera que si A "hace" B y B "hace" A, entonces A y B son el mismo concepto, por ejemplo `area_urbana` "hace" `compartir` pero `compartir` no("hace") `area_urbana` o bien, `gato` "hace" `comer` pero `comer` no("hace") `gato`.
- **Es no-reflexiva:** $\exists a \in C \exists a \neg (a(\text{hace})a)$, existe algún concepto A que no se tiene a sí mismo como habilidad, por ejemplo, `acueducto` no("hace") `acueducto`.
- **Es no-transitiva:** $\exists a, b, c \in C \exists a(\text{hace})b \wedge b(\text{hace})c \Rightarrow \neg a(\text{hace})c$, si A "hace" B y B "hace" C, no implica que necesariamente A "hace" C. Por ejemplo, `area_urbana` "hace" `compartir` con `aeropuerto` y `aeropuerto` "hace" `colindar` con `panteón` no implica que `area_urbana` "hace" `colindar` con `panteón`.

El segundo subconjunto de relaciones axiomáticas (las relaciones auxiliares o preposiciones), denotado por A_2 , está conformado por las preposiciones¹⁰:

$$A_2 = \left\{ \begin{array}{l} a, \text{ ante, bajo, con, contra, de, desde, en, entre, hacia, hasta,} \\ \text{para, por, según, sín, so, sobre, tras, junto_a, a_través_de} \end{array} \right\}$$

A todas estas relaciones las definiremos como no-simétricas, no-reflexivas y no-transitivas, aunque lingüísticamente algunas de ellas no cumplan estas propiedades¹¹, por conveniencia, y como se verá en las restricciones, para este trabajo las consideraremos de esta manera. Esto es, Sea $r \in A_2$,

- $\forall a, b \in C, arb \wedge bra \Rightarrow a = b$
- $\exists a \in C \ni a \neg ra$
- $\exists a, b, c \in C \ni arb \wedge brc \Rightarrow a \neg rc$

4.1.2. Relaciones

Las relaciones son de dos tipos:

- **Simples:** tienen la forma: $a\rho b \in R_S$, donde $a, b \in C$ y $\rho \in A_1$.
- **Compuestas:** tienen la forma: $a\rho b\pi c \in R_C$, donde $a, b, c \in C$ y $\rho \in A_1$ y $\pi \in A_2$.

Como se ha mencionado previamente, la relación “es” es una relación jerárquica que provee la capacidad de “herencia”, a través de esta relación se puede crear una jerarquía de conceptos (existencia). La relación “tiene” provee la capacidad de “agregación” o “composición” de conceptos (pertenencia). La relación “hace” permite describir las capacidades de actuación de un concepto (operaciones).

¹⁰ La preposición es la clase de palabra invariable que introduce el llamado sintagma preposicional. Constituye un tipo de nexos en tanto que liga palabras, sintagmas e incluso proposiciones, pero subordina una de estas unidades (el elemento regido) a la anterior (elemento regente), de la cual depende a través de la preposición. Su significado es sumamente abstracto y gramatical.

¹¹ Por ejemplo, junto_a es una relación simétrica.

Las relaciones A_2 permiten describir y dar causalidad o intensionalidad a las relaciones A_1 . La forma compuesta permite que tengamos el mínimo de *relaciones axiomáticas* y permiten dar *semántica* a las relaciones que no lo son.

Denotaremos con R_V al conjunto de relaciones válidas, el cual es el conjunto de todas las relaciones tales que cumplen con las restricciones que se presentarán más adelante. Asimismo, denotaremos con R_R al conjunto de relaciones que existen en una conceptualización real o concreta.

4.1.3. Conceptos

Concepto es una colección de habilidades y propiedades, con una existencia única.

Hay cuatro tipos de conceptos:

- **Conceptos relación** (verbos), los cuales se definen como elementos que denotan una acción u operación sobre conceptos. Al conjunto de los conceptos relación le denominaremos C_R
- **Conceptos estándar** (sustantivos), los cuales se definen como elementos que pertenecen a una clase y todas sus habilidades y propiedades son abstractas. Estos se denotan por el conjunto C_E
- **Conceptos clase**, son conceptos estándar, denotados por C_L , los cuales sirven para hacer particiones de C_R y C_E
- **Conceptos instancia** o simplemente instancias, son los que todas sus habilidades y propiedades son concretas.

De esta manera, $C = C_R \cup C_E \cup C_L$, tal que $C_E \cap C_R = \emptyset$, $C_E \cap C_L = \emptyset$ y $C_L \cap C_R = \emptyset$, es decir que C_L , C_R y C_E son conjuntos disjuntos.

Un *concepto abstracto* es un concepto estándar que no tiene como "hijos" instancias.

4.1.4. Propiedades

Una propiedad es un concepto agregado a otro por medio de una relación de pertenencia que define características de este último. Se puede definir de la siguiente manera:

Sean $a, b \in C$, se dice entonces que b es una propiedad de a si $a(\text{tiene})b \in R_R$. b es una propiedad concreta de a si b es una **instancia**. Las propiedades que no son concretas las llamaremos propiedades abstractas.

Por ejemplo Si Lago "tiene" profundidad; en donde profundidad puede ser: profundidad_baja, profundidad_media o profundidad_alta, entonces diremos que profundidad es una propiedad de Lago y profundidad_baja es una propiedad concreta de Lago.

Denotaremos como Π al conjunto de propiedades, Π_C al conjunto de propiedades concretas y Π_A al conjunto de propiedades abstractas. Así podemos decir que, $\Pi = \Pi_A \cup \Pi_C$ y $\Pi_A \cap \Pi_C = \emptyset$

4.1.5. Habilidades

Las habilidades son conceptos que definen acciones y/u operaciones asociadas a otro concepto, incluso describen la manera que un concepto interactúa con otros. Se definen como:

Sean $a, b \in C$, se dice entonces que b es una habilidad de a si $a(\text{hace})b \in R_R$. b es una habilidad concreta de a si b es una instancia. De igual forma, las habilidades que no son concretas las llamaremos abstractas.

Denotaremos como H al conjunto de habilidades, H_C al conjunto de habilidades concretas y H_A al conjunto de propiedades abstractas. Así podemos decir que, $H = H_A \cup H_C$ y $H_A \cap H_C = \emptyset$

4.1.6. Instancias

Son una colección de habilidades y propiedades concretas, con una existencia única. En otras palabras, es un concepto contenido en la jerarquía formada por la relación “es” cuyas propiedades y habilidades son concretas. También es el concepto que no tiene hijos en la jerarquía, no es una clase y todas sus habilidades y propiedades son instancias. Denotaremos con I , al conjunto de todas las instancias¹².

4.1.7. Clases

Una clase se puede definir de acuerdo con las siguientes características relativas a un concepto:

- Tienen como hijos a través de una relación de existencia (“es”) conceptos que no son instancias.
- Forman una partición completa.

En este caso la relación “es” es equivalente a una relación de “subclase de”. Denotaremos como C_L al conjunto de todos los conceptos que son clase.

4.1.8. Restricciones

Una restricción se describe como una sentencia que permite evitar inconsistencias en todos los elementos que definen a una conceptualización.

Restricciones para las relaciones en la forma simple

En la Tabla 4 se presenta una visión compacta de las restricciones que tiene las relaciones “es” en su forma simple de uso. En la misma tabla se muestra, por ejemplo, que una clase no puede tener como hijos instancias; o que una instancia no puede tener hijos.

¹² Desde un punto de vista pragmático, I es la base de datos.

Tabla 4. Restricciones para la relación “es” en la forma simple

a	b	Clase	Concepto	Instancia
Clase	si	si	no	no
Concepto	si	si	si	no
Instancia	no	si	si	no

Formalmente, podemos definir estas restricciones de la siguiente manera. Sean $a, b \in C$ entonces $a(es)b \in R_V$ si,

- $a(es)b \in R_S \Rightarrow b(es)a \notin R_V$, para evitar las herencias redundantes, ya que “es” es reflexiva.
- $a(es)b, b(es)c \in R_S \Rightarrow a(es)c \notin R_V$, para evitar las herencias redundantes, ya que “es” es transitiva.
- $a(es)b \in R_S \Rightarrow b \notin I$, las instancias no tienen hijos.
- $a(es)b \in R_S, b \in C_L \Rightarrow a \notin I$, las instancias no heredan de clases.

De manera análoga, en la Tabla 5 se presentan las restricciones que tiene las relaciones “tiene” en su forma simple, donde se muestra que una clase no tiene propiedades; o que una instancia no puede tener como propiedad una clase.

Tabla 5. Restricciones para la relación “tiene” en la forma simple

a	b	Clase	Concepto	Instancia
Clase	no	no	no	no
Concepto	si	si	si	no
Instancia	no	si	si	si

Formalmente, podemos definir estas restricciones de la siguiente manera. Sean $a, b \in C$ entonces $a(tiene)b \in R_V$ si,

- $a(tiene)b \in R_S \Rightarrow a \notin C_L$, las clases no tienen propiedades.

- $a(\text{tiene})b \in R_S, a \in I \Rightarrow b \notin C_L$, las instancias no tienen clases como propiedades.
- $a(\text{tiene})b \in R_S \Rightarrow b \notin I$, los conceptos no tienen instancias como propiedades.

Al igual que las propiedades, en la Tabla 6 se presentan las restricciones que tiene la relación “hace” en su forma simple, donde se muestra por ejemplo, que una clase no tiene habilidades.

Tabla 6. Restricciones para la relación “hace” en la forma simple

a	b	Clase	Concepto	Instancia
Clase		no	no	no
Concepto		si	si	no
Instancia		no	si	si

Formalmente, podemos definir las restricciones de la relación “hace” de la siguiente manera. Sean $a, b \in C$ entonces $a(\text{hace})b \in R_V$ si,

- $a(\text{hace})b \in R_S \Rightarrow a \notin C_L$, las clases no tienen habilidades. Esta restricción elimina $n_i n$ relaciones posibles, pues se eliminan n relaciones posibles por cada instancia en la conceptualización.
- $a(\text{hace})b \in R_S, a \in I \Rightarrow b \notin C_L$, las instancias no tienen clases como habilidades. Esta restricción elimina $n_i n_i$ relaciones posibles, ya que por cada clase se eliminan todas las relaciones de este tipo con las instancias, es decir, se eliminan n_i relaciones posibles.
- $a(\text{hace})b \in R_S \Rightarrow b \notin I$, los conceptos no tienen instancias como habilidades. Esta restricción elimina $n_i n$ relaciones posibles, pues se eliminan n relaciones posibles por cada instancia en la conceptualización.

Restricciones para las relaciones en la forma compuesta

En el caso de las relaciones compuestas, definimos las siguientes restricciones. Sean $a, b, c \in C$, entonces:

- $a(es)b\pi c \in R_V$ si,
 - $a(es)b\pi c \in R_C, a \in C_L, \pi \in P, b \notin I$, las clases no pueden ser hijos de instancias.
 - $a(es)b\pi c \in R_C, a \in C_E, \pi \in P \Rightarrow b \notin I$, los conceptos no pueden ser hijos de instancias.
 - $a(es)b\pi c \in R_C, a \in I, \pi \in P \Rightarrow b \in C_E$, las instancias solo pueden ser hijos de conceptos.
 - $a(es)b\pi c \in R_C, \pi \in P \Rightarrow b \notin I$, esto indica que la relación "es" puede describir relaciones ternarias si y solo si el concepto-relación no es instancia.
 - $a(es)b\pi c \in R_C, \pi \in P, b \in C_L \Rightarrow a \notin I$, las clases no pueden tener relación de existencia con instancias (las instancias no tienen hijos).
- $a(tiene)b\pi c \in R_V$ si,
 - $a(tiene)b\pi c \in R_C, \pi \in P \Rightarrow b \in C_E$, esto indica que la relación "tiene" puede describir relaciones ternarias ssi son conceptos estándar.
 - $a(tiene)b\pi c \in R_C, \pi \in P \Rightarrow a \notin C_L$, las clases no tienen propiedades.
 - $a(tiene)b\pi c \in R_C, \pi \in P \Rightarrow c \notin I$, los conceptos no tienen instancias como propiedades.
 - $a(tiene)b\pi c \in R_C, a \in I, \pi \in P \Rightarrow c \notin C_L$, las instancias no tienen clases como propiedades.
- $a(hace)b\pi c \in R_V$ si,

- $a(\text{hace})b\pi \in R_C, \pi \in P \Rightarrow b \in C_R$, esto indica que la relación "hace" puede describir relaciones ternarias si y solo si son conceptos relación.
- $a(\text{hace})b\pi \in R_C, \pi \in P \Rightarrow a \notin C_L$, las clases no tienen habilidades.
- $a(\text{hace})b\pi \in R_C, \pi \in P \Rightarrow c \notin I$, los conceptos no tienen instancias como habilidades.
- $a(\text{hace})b\pi \in R_C, a \in I, \pi \in P \Rightarrow c \notin C_L$, las instancias no tienen clases como habilidades.

4.1.9. Conceptualización de los dominios geográfico y específico

Utilizando lo que hemos presentado hasta este punto referente a la metodología de conceptualización, se debe conceptualizar el dominio general y el dominio específico. En nuestro caso, el primero de ellos se trata del dominio geográfico. En el siguiente capítulo presentamos el resultado de la conceptualización del dominio geográfico utilizando como base los diccionarios de datos del INEGI (INEGI, 1996), donde además proponemos una estructura de ordenamiento de los datos del diccionario en lo que hemos denominado como Ontología Kaab. A esta ontología le denominaremos en este trabajo como O_G .

La conceptualización del dominio específico está completamente relacionada con el caso de estudio; en este trabajo utilizaremos el dominio de las formas del terreno como dominio específico. A esta ontología le denominaremos O_D .

Como es de esperarse, ambas ontologías deben estar vinculadas entre sí. Así que deben existir relaciones entre los conceptos de O_D con conceptos de O_G . Para establecer estas relaciones, definiremos la genealogía¹³ de un concepto $a \in C$ como el conjunto de conceptos que tienen una relación de existencia; en otras palabras, la

¹³ La genealogía es el proceso de estudiar y rastrear los lazos familiares. Dicho proceso involucra recabar los nombres de parientes tanto vivos como difuntos, y establecer la relación entre ellos y de esta forma construir un árbol genealógico.

genealogía de a es el conjunto de los padres, abuelos, bisabuelos, etc. del concepto a , a través de los cuales se define la existencia del mismo. Esto puede expresarse de la siguiente forma: sea $a, b \in C$, entonces la genealogía $G(a)$ está dada de manera recursiva por $G(a) = \{b \mid a(\text{es})b \in R_r\} \cup G(b)$. De igual manera, definimos que la genealogía de un conjunto de conceptos $X = \{x_1, x_2, \dots, x_n\}$ es la unión de las genealogías de cada concepto, es decir $G(X) = G(\{x_1, x_2, \dots, x_n\}) = \bigcup_{j=1}^n G(x_j)$.

Esto es, sea $a \in O_D$, $b \in G(a)$ y $c \in O_G$ entonces $\forall a \exists c \ni b(\text{es})c \in R_r$. Es decir, que todos los conceptos de O_D están relacionados, ya sea directamente o por herencia, con algún concepto en O_G . Recordemos que $G(a)$ es la función de genealogía de a y que R_r es el conjunto de relaciones reales o concretas en una conceptualización específica.

4.2. Síntesis

Hasta ahora hemos hablado sólo a nivel conceptual y que el proceso siguiente es extraer conceptos de un CDER. Debemos mencionar que los píxeles, como elementos mínimos del CDER, se tienen que agrupar para ser vinculados a los conceptos mencionados. Esto se debe a que no es posible realizar una clasificación de manera directa a nivel de píxel, ya que hay objetos que requieren tener información que no se puede obtener directamente de éstos. Por ejemplo, si uno de los rasgos en la conceptualización es el área, no se puede diferenciar entre los píxeles por esta característica, ya que el área de todos los píxeles es igual¹⁴. En cambio, solo se pueden obtener rasgos locales (la altura y sus derivadas, comportamiento de los vecinos, etc.). Así es que se requiere primero unir los píxeles en grupos, a los cuales les llamaremos extractos, para posteriormente realizar una clasificación de ellos. A la obtención de extractos es lo que denominamos el proceso de síntesis.

¹⁴ Por lo menos muy parecida, pues dependiendo de la proyección el área en un píxel puede variar.

La forma que proponemos para obtener extractos tiene que ver con la manera natural en la que las personas agrupamos objetos. Por ejemplo, si tenemos un conjunto de muebles, primero separamos las sillas por un lado, las mesas por otro y los pizarrones en otro. Hasta este punto no nos interesa clasificar elemento a elemento, es decir, nos importa nada más que es una silla, una mesa o un pizarrón. No interesa si la silla es roja, verde o azul; o si el pizarrón está en buen o mal estado. En un siguiente nivel separaremos cada grupo por alguna característica de esa clase. Por ejemplo, por el color de la sillas (azules, verdes, rojas, etc.), las mesas por su uso (de computadora, de trabajo, de laboratorio) y los pizarrones por su estado (bueno, regular y malo).

Para un primer agrupamiento de objetos utilizamos un conjunto de características o criterios, para un segundo nivel de agrupamiento utilizaremos otro criterio (o el mismo), pero solo sobre los elementos del grupo. Vamos a aplicar diferentes criterios de agrupamiento a un mismo conjunto de objetos (C1, C1.1, C1.2 y C1.3 en la Figura 17). Para cada grupo de objetos emplearemos un criterio de agrupamiento (no necesariamente un algoritmo diferente). En cada nivel podemos utilizar diferentes algoritmos o criterios. Los algoritmos se pueden repetir, los criterios cambiar.

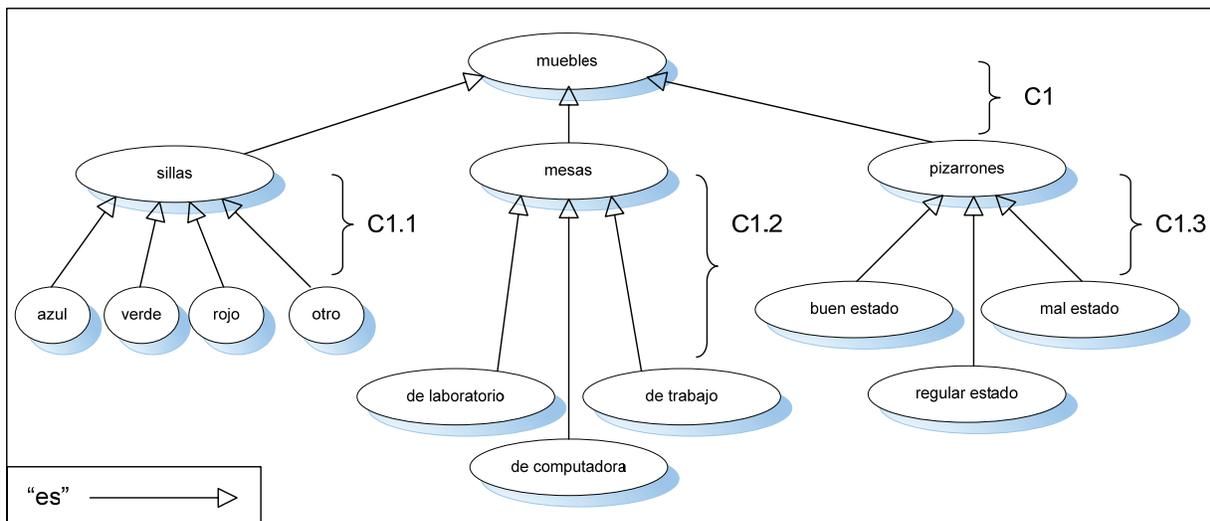


Figura 17. Agrupamiento de objetos utilizando diferentes criterios

La conceptualización de los criterios y algoritmos para agrupar nos da como resultado una conceptualización de los extractos que se obtienen al aplicar esos

algoritmos al CDER; es esta conceptualización la que denominamos conceptualización de la aplicación. Como ya se mencionó, la etapa de síntesis consiste en descomponer el CDER de acuerdo a esta conceptualización.

En la conceptualización del dominio geográfico y del dominio de las formas del terreno, las ontologías resultantes son meramente el modelado del conocimiento que se tiene acerca de tales dominios. Sin embargo, en el caso de la conceptualización de la aplicación no resulta evidente cómo la metodología de conceptualización puede ser aplicada, por lo que en los siguientes párrafos abordamos el tópico de la conceptualización de la aplicación.

Como parte de la metodología de conceptualización, se establece aquí el modelo básico sobre el cual se debe realizar el modelado de la aplicación. El cual se realiza utilizando una sentencia básica: “un algoritmo produce extractos”; lo cual llevado a la conceptualización de acuerdo a la metodología se puede expresar como “Algoritmo(hace)Producir(a)Extracto”, lo cual se ilustra en la Figura 18.

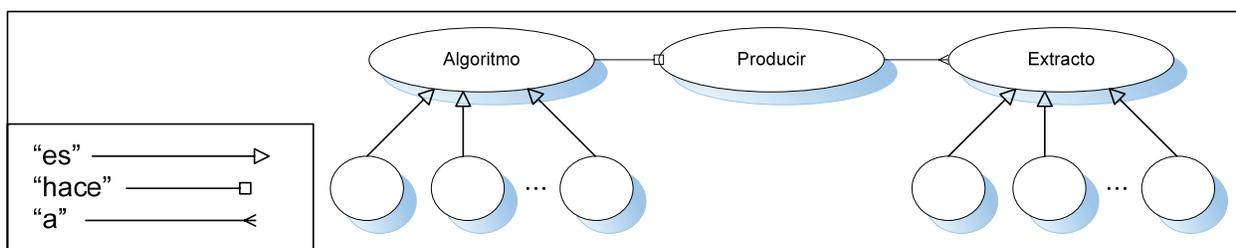


Figura 18. Modelo básico para la conceptualización de la aplicación

La conceptualización arroja una ontología en la cual se encuentran jerarquizados los resultados posibles de la aplicación de cada uno de los algoritmos de extracción que se han de utilizar en dicha aplicación. Esto significa que lo que se debe modelar no es el algoritmo¹⁵, sino más bien los resultados que este arroja. Esto significa que cada algoritmo siempre debe arrojar un número fijo de extractos. En caso que un algoritmo, en dependencia de algún parámetro arrojase un número variable de extractos,

¹⁵ La conceptualización del algoritmo como proceso, se refiere a ubicarlo existencialmente para alcanzar un entendimiento respecto a lo que hace y a lo que produce. No se tiene que modelar, necesariamente, su funcionamiento.

entonces se debe modelar cada posible resultado como si se tratase de un algoritmo diferente. Por otro lado, los extractos se deben modelar en una forma jerárquica (jerarquía que denotaremos como O_A); debido a que, como se presentará más adelante, el esquema de la síntesis involucra una ejecución iterativa de los algoritmos de extracción. En la Figura 19 se presenta una conceptualización de dos algoritmos a_1 y a_2 , en dos niveles de ejecución; así como de los extractos que arroja cada uno. En este ejemplo, el algoritmo a_1 arroja dos tipos de extracto, mientras que el algoritmo a_2 arroja tres tipos. Así por ejemplo, el concepto $ex_a_1_t_2_a_1_t_3$ se refiere al extracto t_3 resultado de aplicar el algoritmo a_1 al extracto t_2 resultado de aplicar a_1 al CDER original.

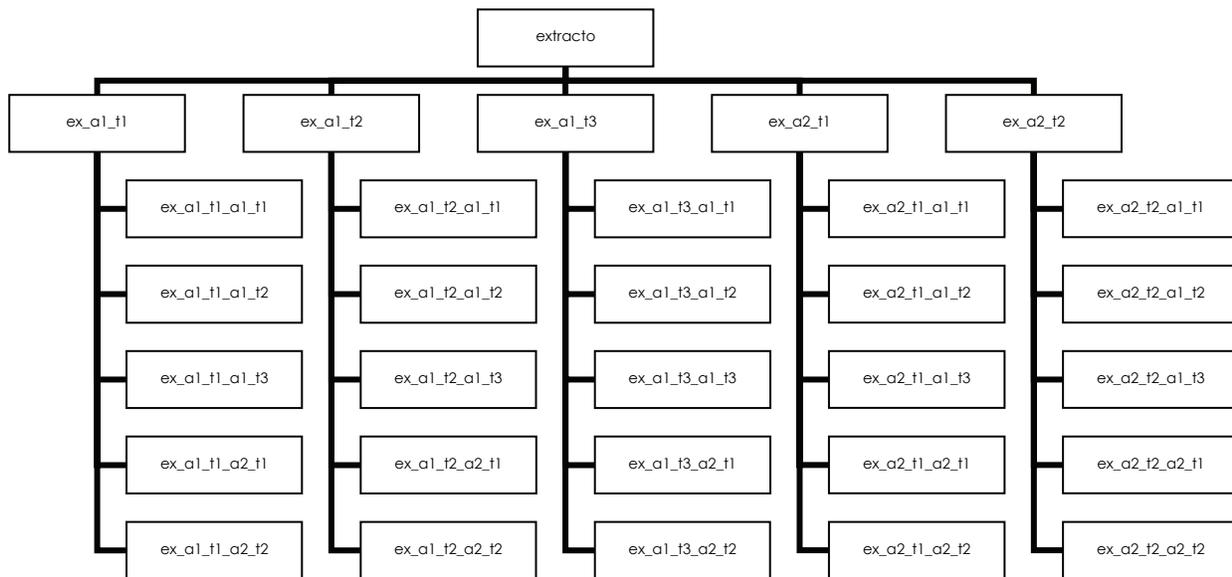


Figura 19. Ejemplo de conceptualización de los algoritmos de extracción

En esta figura se presenta un ejemplo de conceptualización de los algoritmos de extracción, mediante una jerarquía de extractos. En un primer nivel, se presentan los extractos obtenidos de aplicar los algoritmos de extracción al extracto base (que en el primer nivel se trata de todo el CDER). En un segundo nivel se obtienen nuevos extractos de cada uno de los extractos que se encontraron en el primer nivel.

Así como la conceptualización del dominio específico se vincula con la del dominio geográfico, la conceptualización de la aplicación no está aislada del resto de la

conceptualización; por el contrario, cada extracto está relacionado con un concepto en la conceptualización del dominio geográfico o del dominio específico. En la Figura 20 se presenta gráficamente el vínculo que existe entre la conceptualización de la aplicación y la conceptualización del dominio.

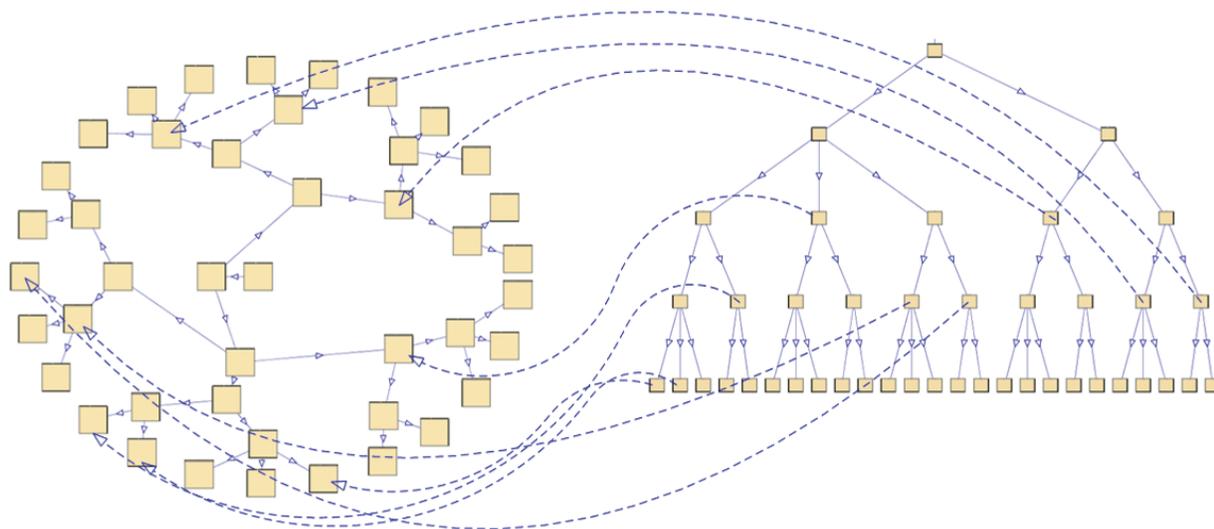


Figura 20. Vinculación de la conceptualización de la aplicación con el resto

En la parte de la izquierda ilustramos la ontología que describe los dominios geográfico y de las formas del terreno. Del lado derecho ilustramos la jerarquía que se forma al conceptualizar los algoritmos de extracción, a lo que llamamos conceptualización de la aplicación. Con flechas punteadas representamos las relaciones que se establecen entre los conceptos en la jerarquía que representan algún extracto y los conceptos en la ontología del dominio geográfico y de las formas del terreno.

Estos vínculos con el resto de la conceptualización permiten identificar qué son los extractos y de esta manera imprimirles la riqueza semántica que aporta la conceptualización completa. En otras palabras, por este medio se asigna a un extracto “anónimo”, como por ejemplo el extracto $ex_a_1_t_2_a_1_t_3$, una semántica al conceptualizar que este extracto “es” es un concepto del dominio geográfico, por ejemplo una “colina”.

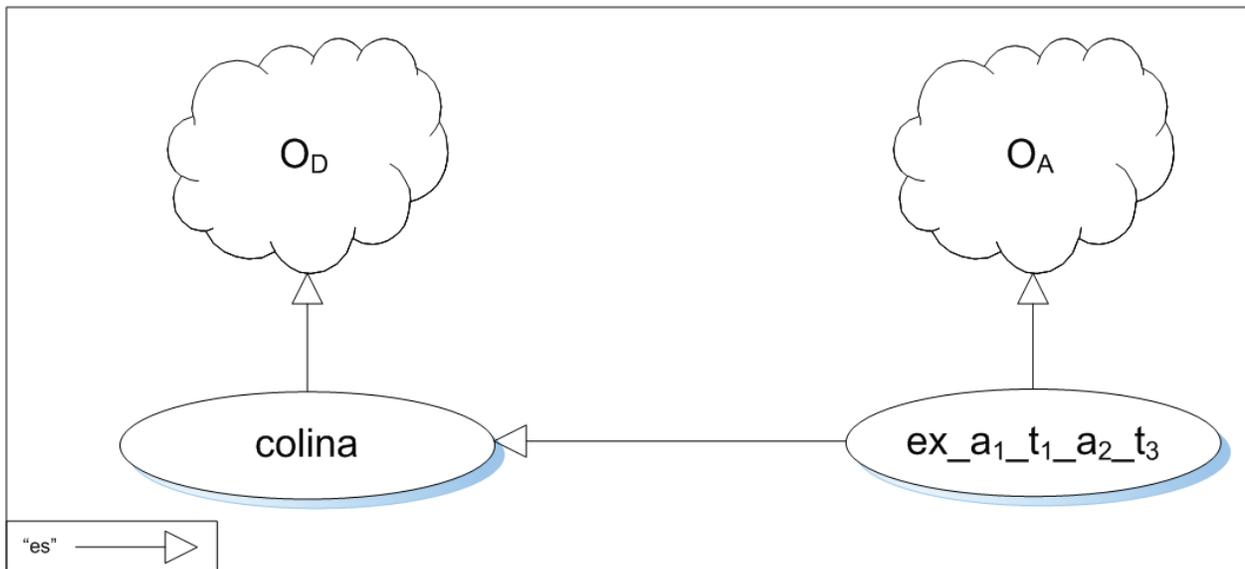


Figura 21. Relación entre conceptos de la O_D y O_A

Al igual que con los conceptos de O_D , como parte de la conceptualización de la aplicación se deben establecer las relaciones de existencia entre los conceptos “anónimos” que obtenemos al conceptualizar los algoritmos de extracción y los conceptos de las ontologías O_G y O_D .

Lo anterior lo podemos expresar como en el caso de las relaciones entre los conceptos de O_D con los conceptos O_G , de la siguiente manera: Sea un concepto $a \in O_A$, $b \in G(a)$ y $c \in O_G \cup O_D$ entonces $\forall a \exists c \ni b(es)c \in R_R$. Lo que quiere decir que todos los conceptos de la ontología de aplicación deben estar relacionados, ya sea directamente o a través de la herencia, con algún concepto en las ontologías del dominio geográfico o del específico. Recordemos que $G(a)$ es la función de genealogía de a y que R_R es el conjunto de relaciones reales o concretas en una conceptualización específica.

Es importante resaltar que en el proceso de síntesis se sabe de antemano los valores posibles de los algoritmos de extracción, puesto que como se ha comentado, los valores arrojados deben ser previamente conceptualizados. Por ello, si alguno de los

algoritmos de extracción requiere una etapa previa (entrenamiento, cálculo de vectores característicos, etc.), esta no forma parte sustantiva del proceso de síntesis.

El proceso de síntesis que presentamos está basado en un algoritmo de segmentación (que presentaremos más adelante). Sin embargo, la metodología propuesta no depende de este ni de ningún otro algoritmo de segmentación o de procesamiento de imágenes; puesto que la propuesta importante es la conceptualización de los resultados mediante una ontología de aplicación.

4.4. Descripción

La etapa de descripción consiste en representar, de acuerdo a la conceptualización, los extractos que se obtienen de la etapa de síntesis. Tal etapa, como ya se ha visto, entrega como resultado una ontología de segmentos O_A , los cuales están vinculados con el resto de la conceptualización. Sin embargo, no es siempre posible vincular un extracto con el concepto exacto que le debe corresponder, pues dependiendo del algoritmo de extracción, un extracto podría vincularse con más de un concepto (generalmente conceptos "hermanos" o "primos"), donde solo se pueden diferenciar entre ellos por rasgos que no es posible obtener mediante los algoritmos de extracción. En ese caso, en la conceptualización de la aplicación se establece una relación con el ascendente común más cercano. Un ejemplo de esto se puede encontrar en la diferencia entre montaña y colina. Tomando la definición de estos dos conceptos (ver Tabla 7), podemos apreciar que la única diferencia es la altura que alcanza cada forma del terreno.

Tabla 7. Definición de montaña y colina

Montaña **Gran elevación natural del terreno.**

Una elevación natural del terreno de una altitud que es impresionante o notable con relación a las elevaciones adyacentes.

Colina **Elevación natural de terreno, menor que una montaña.**

Montaña suave.

Así, en la etapa de descripción se deben analizar las propiedades de los conceptos a los cuales está vinculado cada extracto y así poder hacer una mejor clasificación del extracto para relacionarlo con el concepto que mejor lo describe. Es decir, y retomando el ejemplo anterior, una montaña es de mayor elevación que una colina, por lo que ambos conceptos tienen la propiedad común "elevación", la cual nos servirá para diferenciarlos. Por lo anterior, se debe obtener el valor de la propiedad "elevación" al extracto en cuestión, y mediante este valor definir si se trata de una montaña o una colina. De igual manera, se analizan el resto de las propiedades del concepto, que es más particular que aquel al cual estaba vinculado. Obviamente, es esta la parte más complicada de la diferenciación de los extractos por características comunes.

La metodología para determinar una mejor especialización de la conceptualización de un extracto se basa en realizar una identificación del mejor concepto que se puede vincular con el extracto, de acuerdo a las propiedades que señala la conceptualización y que son medidas en el extracto. Por ejemplo, en la Figura 22 se muestra un extracto perteneciente a la ontología de aplicación O_A , que está relacionado (mediante una relación "es") con cierto concepto en la ontología de dominio específico O_D , el cual tiene asociados una serie de propiedades que han sido dictadas por la conceptualización (prop1, prop2 y prop3). A su vez dicho concepto tiene tres conceptos hijos (A, B y C).

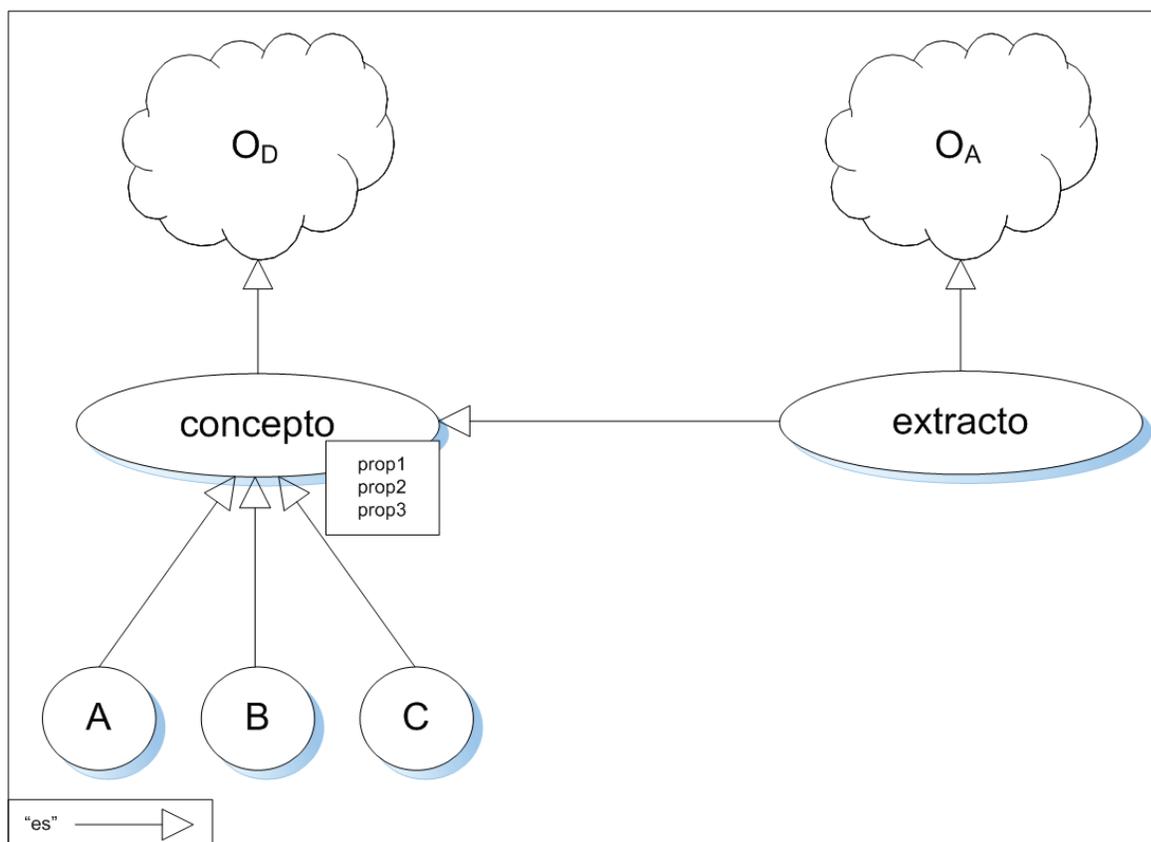


Figura 22. Estado inicial de la relación entre un extracto en O_A y un concepto en O_D

En la Figura 23 se muestra la segunda etapa del proceso. Se deben obtener las mediciones de las propiedades señaladas por la conceptualización (1); obviamente las mediciones son tomadas del extracto (2). Como resultado, se tienen las propiedades del extracto con sus valores medidos (3).

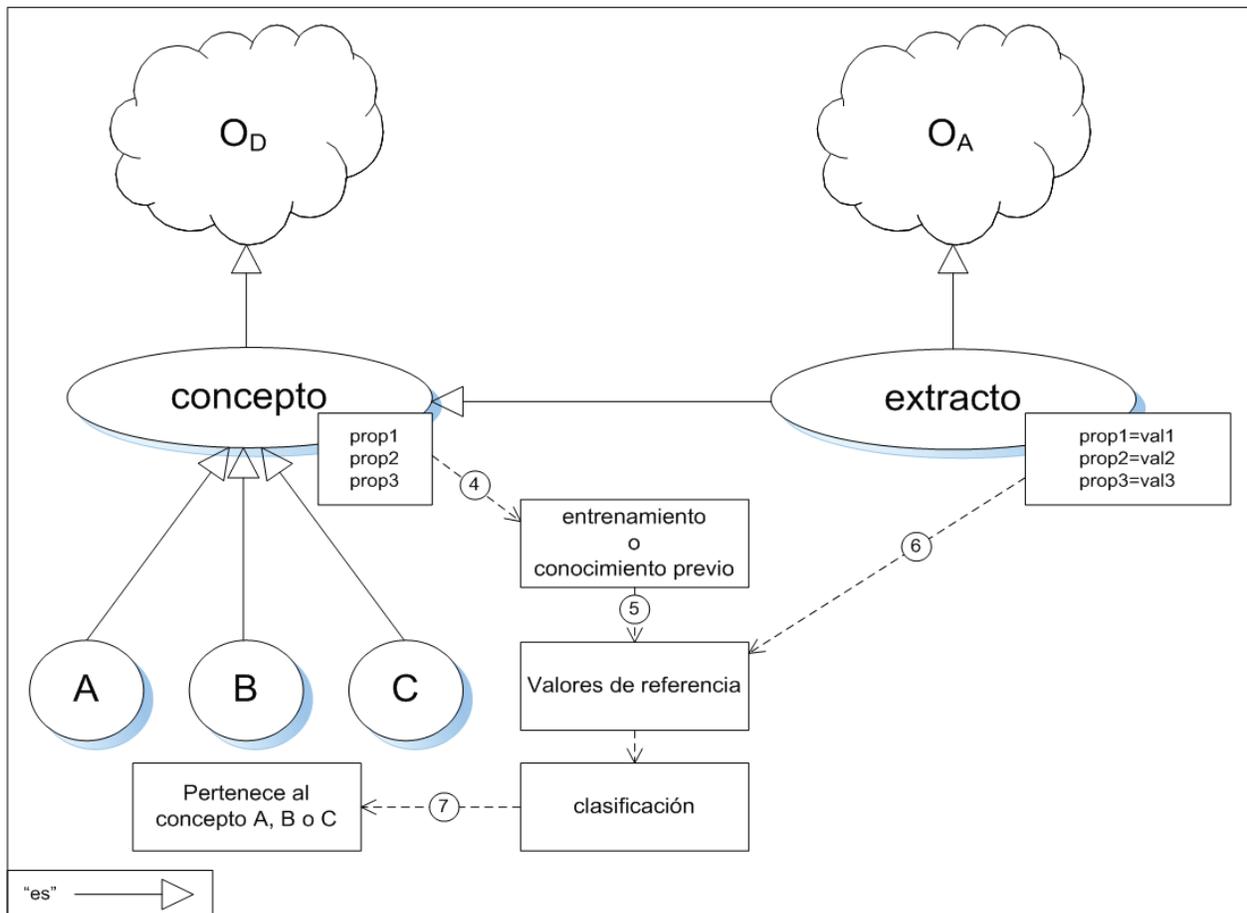


Figura 24. Clasificación del extracto con base en valores de referencia

En la siguiente etapa se asigna la relación "es" (8) entre el extracto y el concepto que ha resultado en la etapa anterior (por ejemplo B). Ahora se realiza la medición de las propiedades adicionales que el nuevo concepto pudiera tener (9). Al igual que con las primeras mediciones, éstas se realizan del extracto (10), el cual toma los valores de las nuevas medidas en conjunto con las anteriores (11).

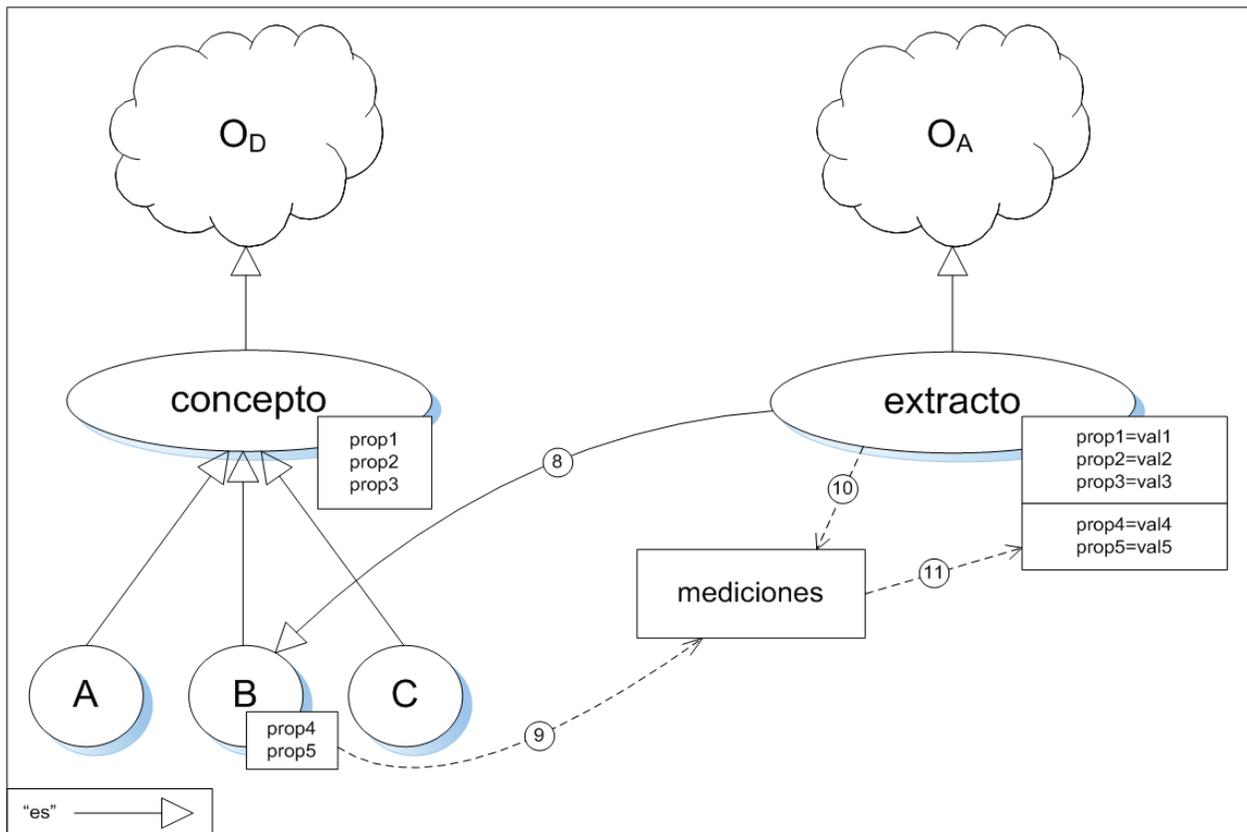


Figura 25. Asignación del extracto como instancia del nuevo concepto

A partir de este punto el extracto pasa a ser una instancia del concepto asignado (B, en el ejemplo). Una vez que el extracto ha sido identificado con el concepto más cercano, se debe proceder a representarlo sintéticamente, ya sea mediante un modelo sintético o textualmente. Este proceso se ilustra en la Figura 26.

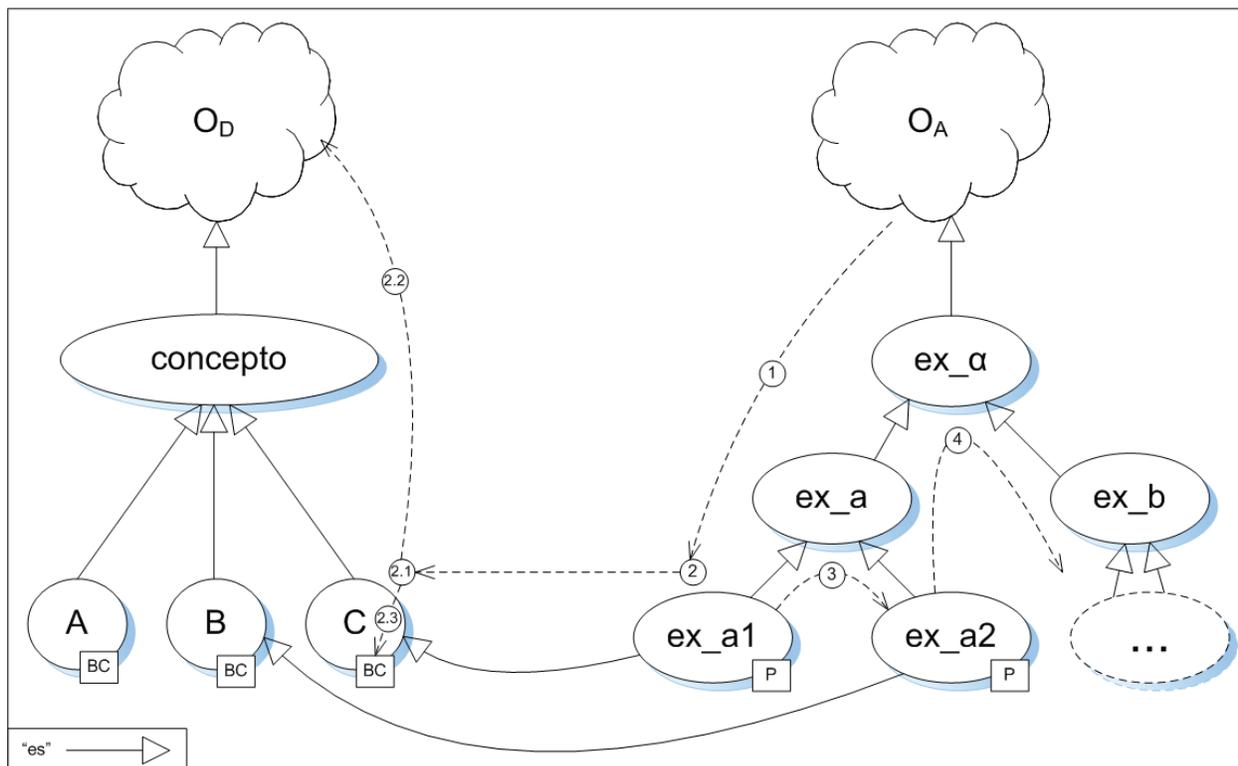


Figura 26. Proceso de descripción de los extractos

Para realizar la representación se procede a buscar el primer extracto identificado como instancia de un concepto (1), esto se realiza con algún recorrido del árbol que forma la jerarquía de extractos (normal, izquierdo, derecho, etc.). Enseguida comienza el proceso de descripción de un solo extracto (2), el cual depende de la forma de representación. En cualquiera de los casos, seguimos la relación que tiene con O_D para obtener la información del concepto del cual el extracto es instancia (2.1). En caso de una representación textual, se obtiene la etiqueta del concepto, así como su información existencial, de pertenencia y de habilidades (2.2). En caso de una representación mediante un modelo sintético, se obtiene el bloque constructivo que representa al concepto (2.3).

Una vez realizada la representación del extracto, se busca el siguiente extracto-instancia para realizar su representación (3), con el cual se repite el proceso de representación mencionado (2). Sucesivamente se continúa con el recorrido y descripción de la jerarquía de extractos (4).

4.4.1. Formas de clasificación

Como se mencionó, la asignación de los extractos al concepto que mejor los describe se realiza por medio de una clasificación, para lo cual se requiere tener valores de referencia. Tales valores pueden ser obtenidos a partir de la conceptualización. Esto significa que la forma de discernir si un extracto, que mediante la conceptualización lo relacionamos a un concepto (elevación, retomando el ejemplo de secciones anteriores), se trata más precisamente de algún otro concepto hijo (cerro, colina, loma, montaña, monte, etc.) dependerá de los valores de referencia que se establezcan en la O_D . Para ello, entonces, en la conceptualización del dominio específico se debe considerar tener los valores de referencia correspondientes a cada concepto. Con los valores de referencia, se puede aplicar alguno de los algoritmos de agrupamiento presentados en el Capítulo 3.

Para obtener los valores de referencia se puede realizar un consenso entre los expertos, realizar entrenamientos utilizando extractos bien identificados o definir conjuntos y grafos difusos con base en las definiciones de los propios conceptos. En la Figura 27 se muestra el grafo difuso de dos de las propiedades más importantes de las elevaciones: el área y la altitud, obtenidos a partir de las definiciones de los conceptos. En la Figura 27 se puede apreciar que, por ejemplo, una montaña tiene una gran altitud (de alta a muy alta) y un área extensa (de grande a muy grande), mientras que un pico tiene, al igual que la montaña, una gran altitud, pero su área es más pequeña que la de la montaña (de pequeña a media).

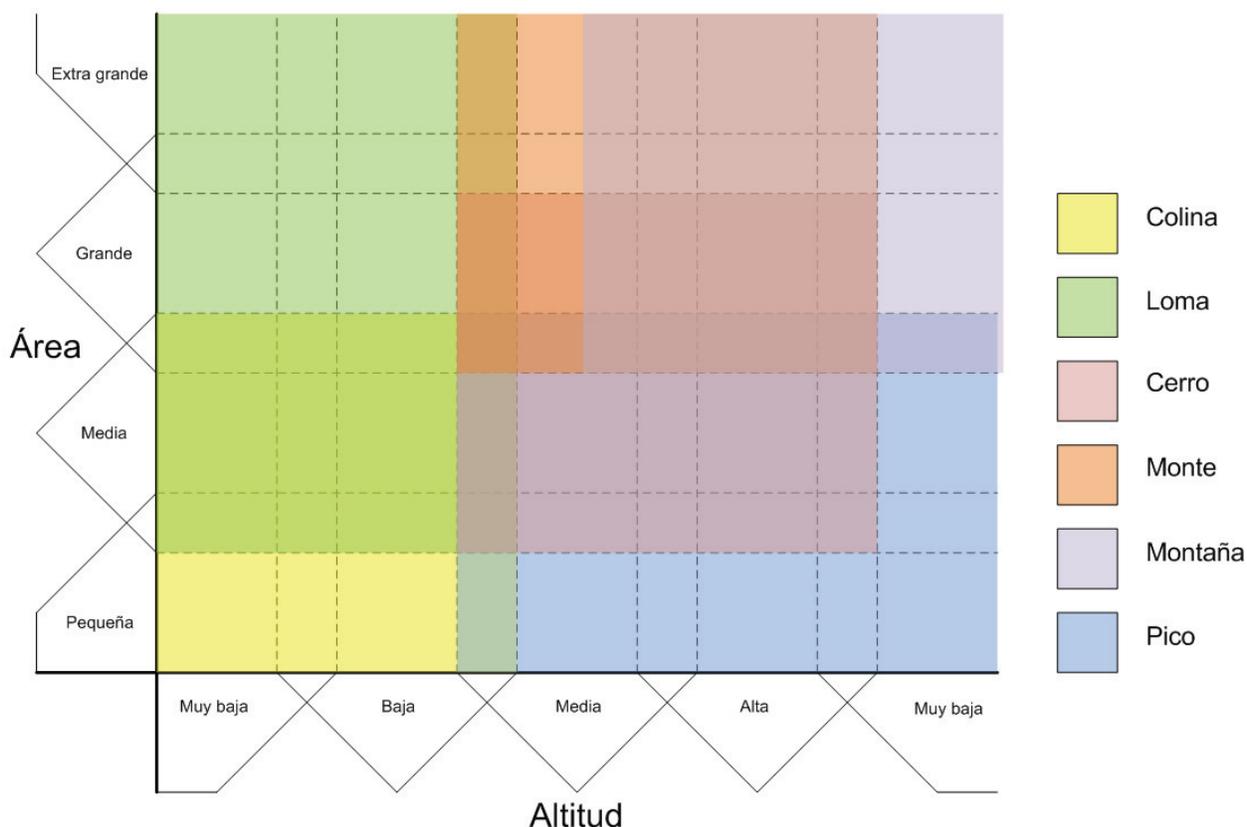


Figura 27. Grafo difuso de los parámetros utilizados para clasificar diferentes formas del terreno

Otro ejemplo se puede ver en la diferenciación entre colina y loma; ambos conceptos tienen altitud baja (de muy baja a media), pero mientras la colina tiene un área pequeña (de pequeña a media) la loma tiene un área extensa (de grande a muy grande).

4.4.2. Formas de representación

La manera final de representar el CDER puede variar dependiendo del sentido que se le desee dar a tal representación. En este trabajo proponemos dos formas de realizar la representación (aunque con solo cambiar el formato de salida, podríamos cambiar la forma de representación); la primera de ellas consiste en el empleo de lo que llamamos bloques constructivos, que se utilizan para crear modelos sintéticos del CDER, es decir, los bloques constructivos nos permiten crear "maquetas" de los datos representados. La segunda forma consiste en hacer una representación textual de los

datos en las que se describe en español (o en cualquier otro idioma) el contenido semántico del CDER que se está describiendo. En los siguientes párrafos se describen más detalladamente las dos formas mencionadas.

Comenzaremos por analizar la primera forma para realizar la representación final del CDER; para ello se debe definir un conjunto de elementos llamados bloques constructivos (BC), esto es, elementos que existen en todos los CDER (o al menos en los de un mismo tipo¹⁷) y que permiten, mediante su combinación, crear cualquier CDER de ese tipo.

Por ejemplo, en el caso de los DEM, se pueden establecer algunas primitivas mediante las cuales se puede construir un modelo sintético del DEM original combinando estas primitivas. El conjunto de BCs está conformado por los conceptos presentes en la ontología del dominio de las formas del terreno, es decir, se tienen elementos tales como planos, cerros, montañas, valles, cuencas, etc. Esta es una forma de incluir el conocimiento previo que se tiene acerca de los DEM para realizar una representación semántica de éstos. Así, la representación consistirá en una combinación de objetos de interés los cuales tienen un valor semántico por sí mismos, y al mismo tiempo permiten, mediante su combinación, la reconstrucción del modelo de elevación en su conjunto. Para utilizar esta forma de representación, es necesario que desde la conceptualización se tengan los BC correspondientes a cada concepto, o por lo menos, que algún concepto en la genealogía de cada concepto tenga un BC representativo.

La representación sintética es una representación visual del contenido semántico de un CDER; sin embargo, no permite apreciar las propiedades de los extractos de una forma más exacta. Por ello, es necesario contar con una alternativa de tener una representación textual que permita apreciar todas las propiedades de los extractos. En la Figura 28 se muestra un ejemplo de representación textual, en el cual se utiliza el lenguaje natural y XML para realizar la descripción del mismo DEM.

¹⁷ Ya que podemos tener modelos digitales de elevación, mapas uso de suelo, imágenes satelitales, mapas escaneados, etc.

CONJUNTO DE DATOS RASTER NÚMERO 1 DEL MODELO DIGITAL DE ELEVACIÓN DE MÉXICO QUE TIENE UNA RESOLUCIÓN ESPACIAL DE 100 M CON PLANO MEDIO EN 1953 M SU ALTITUD MÍNIMA ES DE 1823 M Y SU ALTITUD MÁXIMA ES DE 2315 M LAS COORDENADAS EXTREMAS SON DE 20.0225 GRADOS LATITUD NORTE, 119.0252 GRADOS LONGITUD OESTE HASTA 20.5580 GRADOS LATITUD NORTE, 119.5266 GRADOS LONGITUD OESTE.

CONTIENE:

UNA MONTAÑA CON ÁREA DE 2252 HA, ALTITUD MÍNIMA DE 1950 M, ALTITUD MÁXIMA DE 2301 M Y SU CIMA ESTÁ UBICADA EN 20.1523 GRADOS LATITUD NORTE Y 119.1515 GRADOS LATITUD OESTE A UNA ALTITUD DE 2301 M.

...

```
<CDER ID=1>
<TYPE NAME="DEM">
<LABEL VALUE="MÉXICO">
<PARAMS RES=100 MED PLAN=1953 SYSTEM="GEO">
<ALT MIN=1823 MAX=2315>
<EXTREME>
<XMIN VALUE=-119.0252>
<XMAX VALUE=-119.5266>
<YMIN VALUE=20.0225>
<YMAX VALUE=20.5580>
</EXTREME>
</PARAMS>
<INSTANCE ID=1>
<ES CONCEPTO=OD.MONTANA>
<LABEL VALUE="MONTAÑA">
<TIENE>
<PROPIEDAD ID=OG.AREA>
<VALOR VALUE=2252>
<UNIT VALUE="HA">
<LABEL VALUE="ÁREA">
</PROPIEDAD>
<PROPIEDAD ID=OG.ALT_MIN>
<VALOR VALUE=1950>
<UNIT VALUE="MTS.">
<LABEL VALUE="ALTITUD MÍNIMA">
</PROPIEDAD>
<PROPIEDAD ID=OG.ALT_MAX>
<VALOR VALUE=2301>
<UNIT VALUE="MTS.">
<LABEL VALUE="ALTITUD MÁXIMA">
</PROPIEDAD>
<PROPIEDAD ID=OG.CIMA>
<TIENE>
<PROPIEDAD ID=OG.COORDENADA>
<LATITUD VALUE=20.1523>
<LONGITUD VALUE=-119.1515>
<UNIT VALUE="GRADOS">
</PROPIEDAD>
<PROPIEDAD ID=OG.ALTITUD>
<VALOR VALUE=2301>
<UNIT VALUE="MTS.">
<LABEL VALUE="ALTITUD">
</PROPIEDAD>
</TIENE>
</PROPIEDAD>
</TIENE>
</INSTANCE>
<INSTANCE ID=2>
...
</INSTANCE>
...
</CDER>
```

Figura 28. Utilización de una descripción textual para la representación

En este capítulo hemos presentado desde un punto de vista teórico cómo realizar la representación semántica de los datos espaciales *raster*. Hemos propuesto tres etapas fundamentales para realizarlo: la conceptualización, la síntesis y la descripción. En el siguiente capítulo presentaremos estas mismas etapas, pero ahora desde un punto de vista práctico, utilizando como caso de estudio los DEM.

5. Resultados

Como ya se ha mencionado, utilizaremos los DEM como un caso de estudio, pues estos han adquirido un importante papel en muchos campos de los SIG, incluyendo las ciencias ambientales, la prevención de riesgos, ingeniería, entre algunos otros. Además, en el estudio de los DEM ha sido incorporado el análisis de las formas geométricas del terreno para analizar su distribución y la concentración de ciertos objetos espaciales. En este capítulo se presentan los resultados de las pruebas hechas para el caso de estudio seleccionado, que se trata de los Modelos Digitales de Elevación. Siguiendo la metodología planteada en el capítulo anterior, en la Figura 29 se presentan las etapas adaptadas para el caso de estudio en cuestión.

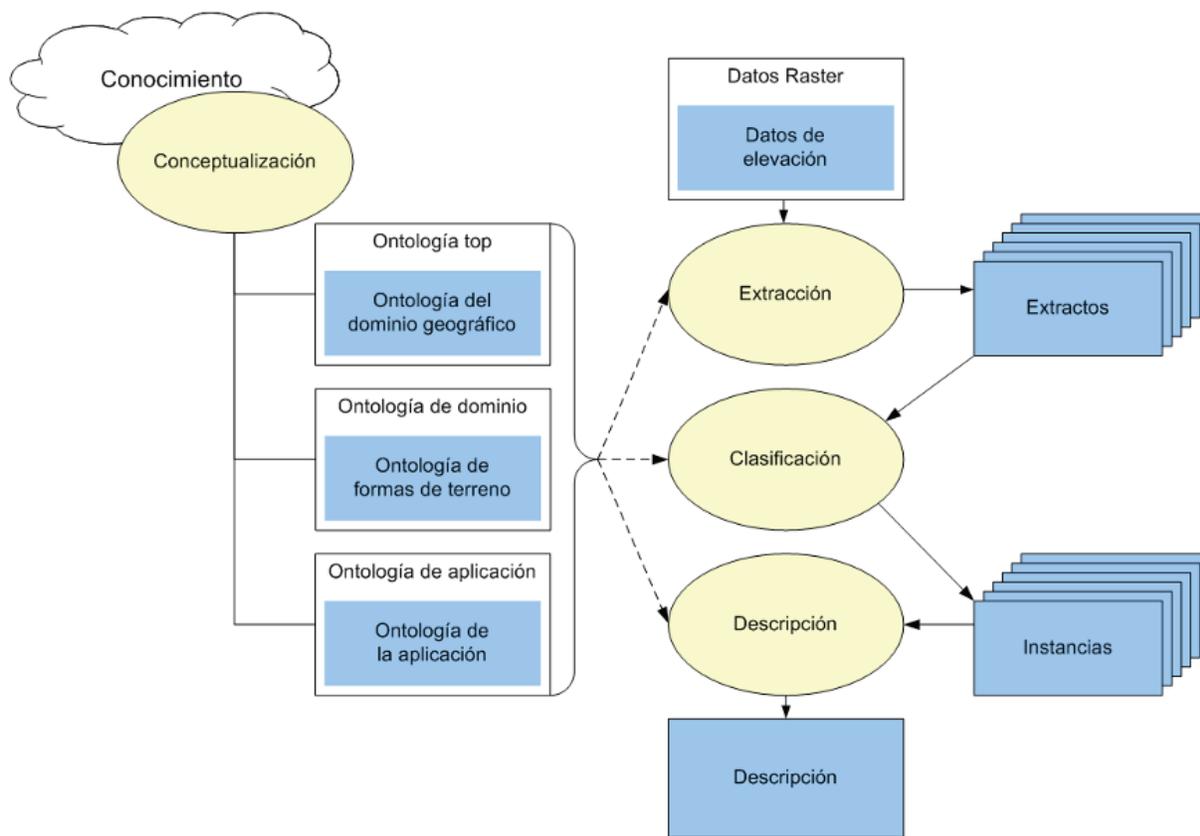


Figura 29. De la metodología al caso de estudio

Se muestra el proceso completo para realizar la descripción semántica de datos *raster*, en general; y en particular se presenta dónde cazan los datos que utilizaremos en el caso de estudio.

5.1. Conceptualización

Como se presenta en la Figura 29, la etapa de conceptualización toma el conocimiento sobre el tema de las formas del terreno y brinda como resultados las ontologías del dominio geográfico (O_G), del dominio de las formas del terreno (O_D) y de la aplicación (O_A). Así, empleando la metodología de conceptualización descrita en el capítulo anterior, presentamos ahora el desarrollo de cada una las ontologías mencionadas.

5.1.1. Ontología del dominio geográfico

En el capítulo anterior, como parte de la metodología presentada, se establece que se deben definir nueve elementos para realizar la conceptualización (axiomas, relaciones, propiedades, habilidades, instancias, conceptos, clases, restricciones y contexto). En la propia metodología de conceptualización ya se han definido algunos de ellos (axiomas, restricciones y contexto) aplicables para todas las conceptualizaciones basadas en la metodología. Los otros elementos (relaciones, propiedades, habilidades, instancias, conceptos y clases) fueron definidos en su forma, mas no en su fondo; así que se deben definir para que el desarrollo de la conceptualización esté completo y obtener como resultado la ontología del dominio geográfico, a la cual vamos a llamar ontología Kaab¹⁸.

Clases de la ontología del dominio geográfico

Las clases en la ontología del dominio geográfico ocupan las posiciones "iniciales" en la conceptualización; por ende, se trata de abstracciones de grandes colecciones de posibles instancias de elementos geográficos que existen. En la Tabla 8 se presenta la definición de las clases más abstractas de la conceptualización del dominio geográfico, como son "entidad geográfica", "objeto geográfico", "fenómeno geográfico" y "relación geográfica". Asimismo, en la Figura 30, se ilustran las relaciones que guardan estas clases entre sí.

Tabla 8. Definición de las clases principales en la ontología kaab

Clase	Definición
Entidad geográfica	Esta clase es la más general del dominio geográfico; es a través de esta clase que iniciamos la conceptualización de dicho dominio. Para el presente trabajo, esta clase es la raíz, descendiendo solo de una clase llamada "cosa".

¹⁸ Del maya kaab que significa Tierra. Para la definición de las clases utilizaremos algunos vocablos del lenguaje maya.

Objeto geográfico	Esta clase se refiere a las entidades que tienen una presencia geográfica propia, es decir, que tienen una posición espacial que no depende de ninguna otra entidad. Es la representación existencial de todas aquellas “entidades geográficas” que ocupan un lugar en el espacio. Ejemplos: áreas urbanas, aeropuertos, montañas, lagos, etc.
Fenómeno geográfico	Esta clase se refiere también a entidades que tienen una presencia geográfica propia, pero la cual es dinámica, muy cambiante en periodos de tiempo muy pequeños ¹⁹ . Ejemplos: huracanes, corrientes acuáticas, vientos, frentes, etc.
Relación geográfica	Esta clase describe a aquellas entidades que establecen algún tipo de vínculo entre los objetos geográficos. Por medio de esta clase se definen las “entidades geográficas” que no tienen una presencia espacial propia, sino siempre es a través de los “objetos geográficos” que relacionan. Ejemplos: conecta, cruza, al sur, al norte, etc.



Figura 30. Clases principales en la ontología kaab

Continuando con la definición de las clases de la ontología Kaab, en la Tabla 9 se presenta la definición de las clases principales descendientes de “objeto geográfico”. Las clases que se presentan permiten, según nuestra propuesta, agrupar y dar sentido a los datos geográficos presentes en el diccionario de datos del INEGI. Estas clases dan un primer significado a los conceptos (que se presentarán más adelante), ya que

¹⁹ Obviamente, los términos temporales son muy dependientes del contexto; por tanto, al referirnos a periodos muy pequeños de tiempo, en el contexto geográfico se refiere a periodos del orden de meses.

definen siete aspectos fundamentales: su origen, su frontera, su ubicación superficial, su naturaleza existencial, su perceptibilidad, su magnitud y su posición en el mundo.

Tabla 9. Definición de las clases principales descendientes de “objeto geográfico”

Clase	Definición
Moots <i>Origen</i>	Esta clase describe a los objetos geográficos por su tipo o composición intrínseca dentro del mundo real, es decir, indica la forma en la cual éstos fueron contruidos o concebidos en el sentido cognitivo de los seres humanos.
Xuul <i>Frontera</i>	Esta clase indica el tipo de frontera o límite que pueden tener los objetos geográficos de acuerdo con su naturaleza en el mundo real. Además, esta clase permite categorizar a los objetos geográficos de acuerdo con el tipo implícito de sus fronteras.
Ixco <i>Superficialidad</i>	Esta clase describe cómo pueden ser localizados los objetos geográficos dentro del mundo real, es decir, cuál es el ambiente natural en el cual se presentan físicamente los objetos geográficos.
Sayab <i>Naturaleza</i>	Esta clase permite observar si los objetos geográficos pueden ser vistos dentro del mundo real o simplemente existen como parte de un proceso cognitivo de abstracción, en el cual se sabe que ciertos objetos geográficos existen, pero debido a sus características intrínsecas y por el conocimiento del ser humano pueden clasificarse como reales o imaginarios.
Paakat <i>Sentido</i>	Esta clase indica si los objetos geográficos que componen el mundo real son perceptibles por los sentidos del ser humano, considerando como tales la visión, audición, olfato, gusto y tacto, con lo cual se puede determinar cómo el ser humano puede percibir de una u otra forma la presencia de los objetos geográficos.
Chuuk <i>Alcance</i>	Esta clase determina la magnitud con respecto al tamaño que puede presentar un objeto, es decir, el alcance en cuanto a dimensiones que determina el tamaño de un objeto geográfico.
Utskiin <i>Disposición</i>	Esta clase permite identificar la composición sobre la cual se encuentra localizado un objeto geográfico, haciendo referencia a la disposición del mismo sobre los elementos esenciales que componen al mundo real.

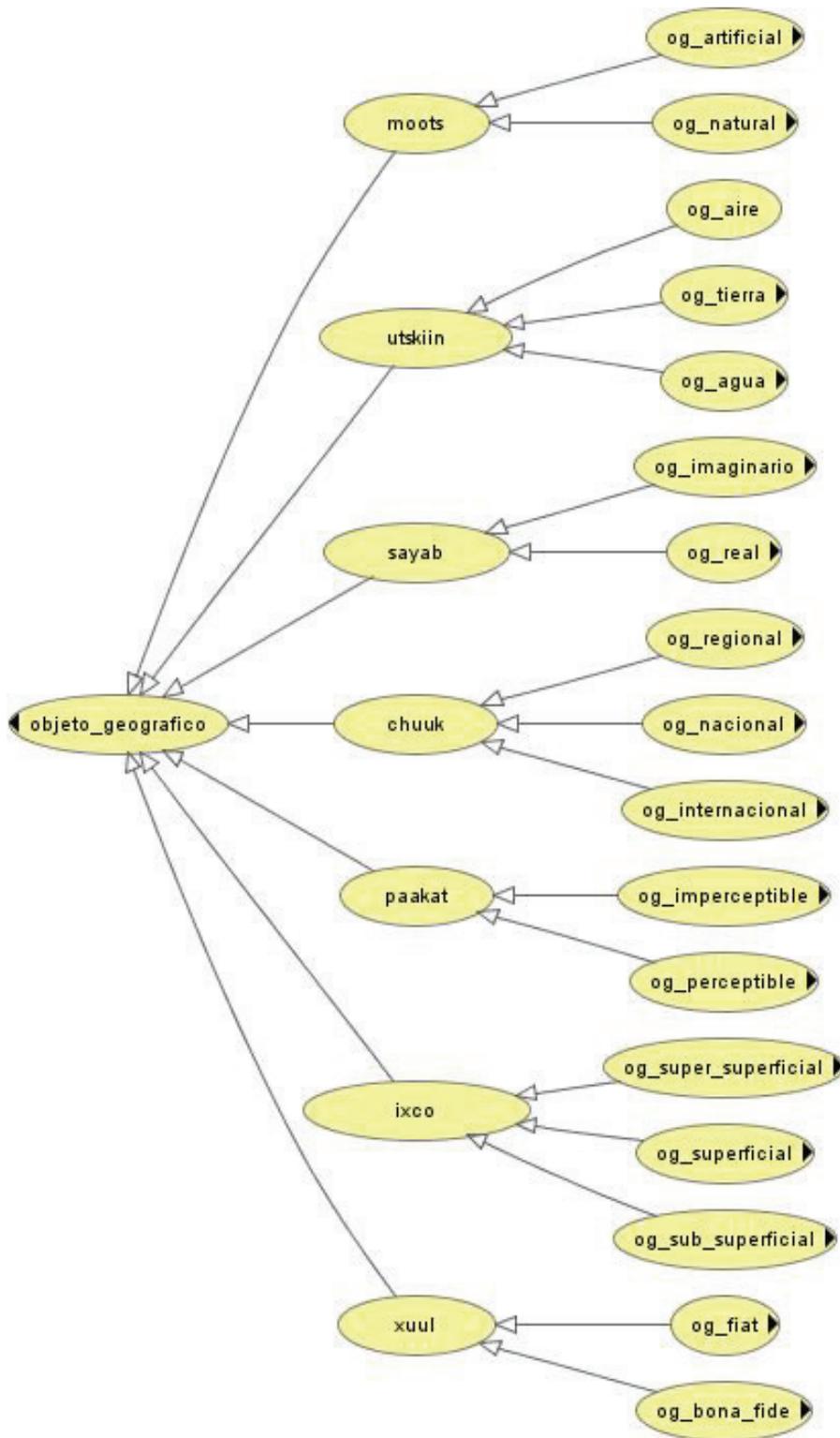


Figura 31. Clases y subclases de objeto geográfico

En la Tabla 10 se presenta la definición de las principales subclases en la ontología Kaab. Estas subclases representan particiones de las clases presentadas en la Tabla 9, lo que significa que todos los objetos del dominio geográfico pertenecen a alguna de las subclases que se presentan. Por ejemplo, la clase *moots* describe el origen de la creación de un concepto geográfico; tiene dos subclases: Objeto geográfico natural y Objeto geográfico artificial. Todos los conceptos del dominio geográfico pertenecen a alguna de las dos subclases mencionadas, es decir, que cualquier objeto geográfico o es de origen natural o es de origen artificial. Lo mismo aplica para las demás clases y sus respectivas subclases.

Tabla 10. Definición de las subclases en la ontología kaab

Clase	Subclase	Significado
Moots	<i>Objeto geográfico natural</i>	Un objeto geográfico natural es todo aquel objeto geográfico que forma parte del mundo real y en donde su creación, no depende de la interacción del ser humano. Por ejemplo: áreas selváticas, corrientes de agua, curvas de nivel, pantanos, zonas arenosas, etc.
	<i>Objeto geográfico artificial</i>	Un objeto geográfico artificial, es aquel objeto que su creación es generada por la intervención humana, en donde diversas características o factores determinan su origen. Por ejemplo: carreteras, conductos, líneas de comunicación, puentes, presas, entre otros.
Xuul	<i>Objeto geográfico bona fide</i>	Un objeto <i>bona fide</i> ²⁰ es un objeto geográfico que su frontera o límite se deriva directamente de la composición implícita y natural del objeto geográfico, en donde los límites o fronteras son determinados de acuerdo con su escala de presentación. Son objetos <i>bona fide</i> : lagos, ríos, áreas verdes, pantanos, etc.
	<i>Objeto geográfico fiat</i>	Los objetos <i>fiat</i> ²¹ son todos aquellos objetos geográficos que su límite o frontera está determinada por el consenso de un grupo de

²⁰ Objeto *bona fide*, proviene del latín y significa buena fe.

²¹ Objeto *fiat* significa frontera física y determinística.

		seres humanos, de acuerdo con un arreglo que es producto de diversas convenciones sociales establecidas. Por ejemplo: límites entre países, cementerios, presas, calles, mojoneras, entre otros.
Sayab	<i>Objeto geográfico real</i>	Son aquellos objetos geográficos visibles en el mundo real, los cuales existen físicamente y pueden ser percibidos por los seres humanos a simple vista. Por ejemplo: arrecifes, salinas, terrenos sujetos a inundación, zonas arenosas, áreas naturales protegidas, cuerpos de agua, etc.
	<i>Objeto geográfico imaginario</i>	Son los objetos geográficos que no son visibles por el ser humano; sin embargo, se sabe de su existencia en el mundo real. Estos objetos pueden dar origen a la construcción de otros objetos geográficos. Ejemplo de objetos geográficos imaginarios: las curvas de nivel, los puntos acotados y los límites.
Ixco	<i>Objeto geográfico superficial</i>	Son todos aquellos objetos geográficos que se encuentran localizados físicamente sobre la superficie de la Tierra. Por ejemplo: aeropuertos, áreas de cultivo, áreas urbanas, instalaciones portuarias, corrientes de agua, etc.
	<i>Objeto geográfico sub-superficial</i>	Son todos aquellos objetos geográficos que se encuentran localizados físicamente bajo la superficie de la Tierra, tales como: acueductos subterráneos, arrecifes bajos, bordos, minas, entre otros.
	<i>Objeto geográfico super-superficial</i>	Son los objetos geográficos que se encuentran localizados y suspendidos físicamente sobre la superficie de la Tierra. Entre los ejemplos de estos objetos tenemos: estructuras elevadas, líneas de transmisión, puentes, rutas de teleféricos, etc.
Paakat	<i>Objeto geográfico perceptible</i>	Son aquellos objetos geográficos sensibles y perceptibles a los seres humanos. Esta percepción está basada en los sentidos. Entre este tipo de objetos geográficos se tienen: canales, caminos, instalaciones, presas, sitios arqueológicos, entre otros.

	<i>Objeto geográfico imperceptible</i>	Esta categoría de objetos hace referencia a todos aquellos objetos geográficos que no pueden ser sensibles al ser humano; es decir, que ningún tipo de sentido permite a las personas percibir este objeto. Por ejemplo: las curvas de nivel, los puntos acotados, los límites y fronteras en su forma natural ²² .
Chuuk	<i>Objeto geográfico regional</i>	Son todos aquellos objetos geográficos que en el contexto de su tamaño, están determinados por una magnitud de cubrimiento, pero en este caso, el "término" <i>regional</i> hace referencia a objetos relativamente pequeños a su alcance y contexto. Ejemplos de objetos con alcance regional: aeropuertos locales, acueductos, áreas urbanas, estanques, instalaciones de bombeo, calles, conductos, cuerpos de agua, linderos, entre otros.
	<i>Objeto geográfico nacional</i>	Son todos aquellos objetos geográficos que en el contexto de su tamaño, está determinado por una magnitud de cubrimiento o extensión, pero en este caso, el "término" <i>nacional</i> hace referencia a objetos relativamente medianos a su alcance y contexto. Ejemplos de este tipo de objetos: aeropuertos nacionales, límites estatales, vías férreas, autopistas federales, etc.
	<i>Objeto geográfico internacional</i>	Son todos aquellos objetos geográficos que en el contexto de su tamaño, está determinado por una magnitud de cubrimiento o extensión, pero en este caso, el "término" <i>internacional</i> hace referencia a objetos relativamente grandes a su alcance y contexto. Ejemplos de este tipo de objetos: aeropuertos internacionales, límites internacionales, mojoneras, entre otros.

²² El hecho de decir que estos objetos son imperceptibles en su forma natural, como es el caso de los límites y fronteras, se refiere particularmente a que este tipo de objetos geográficos por sus características implícitas no pueden ser percibidos, aunque exista un puente divisor entre la frontera de dos países. Este hecho implica que el ser humano utiliza su aparato cognitivo para inferir la existencia de este tipo de objeto geográfico.

Utskiin	<i>Objeto geográfico tierra</i>	Son todos aquellos objetos geográficos que la composición de su localización y la forma en que el ser humano los percibe, esencialmente es en la tierra. Como ejemplo se tienen: bordos, áreas naturales protegidas, depósitos de desechos, faros, límites, malpaís, lumbreras, etc.
	<i>Objeto geográfico agua</i>	Son todos aquellos objetos geográficos que la composición de su localización y la forma en que el ser humano los percibe, esencialmente es en el agua. Por ejemplo: arrecifes bajos, corrientes de agua, corrientes que desaparecen, estanques, cuerpos de agua, instalaciones portuarias, pantanos, rápidos, saltos de agua, entre otros.
	<i>Objeto geográfico aire</i>	Son todos aquellos objetos geográficos que la composición de su localización y la forma en que el ser humano los percibe, esencialmente es en el aire. Por ejemplo: líneas de transmisión, puentes, rutas de teleféricos, etc.

Finalmente, en la Tabla 11 se presentan las clases descendientes de la clase relación geográfica; asimismo en la Figura 32 se presenta la estructuración existencial de tales clases.

Tabla 11. Clases descendientes de “relación geográfica”

Clase	Definición
Relación topológica	Esta clase permite definir al tipo de “relación geográfica” que tiene que ver con la distribución geométrica de los objetos. Este tipo de relaciones permanecen sin variación a través de un proceso de transformación de los “objetos geográficos” o “fenómenos geográficos” que relacionan. Ejemplos: conectar, compartir.
Relación compartir	Esta clase se refiere a la “relación topológica” que establece que dos “objetos geográficos” o “relaciones geográficas” tienen un área en común dentro del espacio que los define. Ejemplo: un aeropuerto comparte con un área urbana, un municipio comparte con un estado.

Relación conectar Esta clase se refiere a la “relación topológica” que establece que dos “objetos geográficos” o “relaciones geográficas” tienen un punto en común dentro del espacio que los define. Ejemplo: una carretera conecta con un área urbana, un río conecta con un lago.

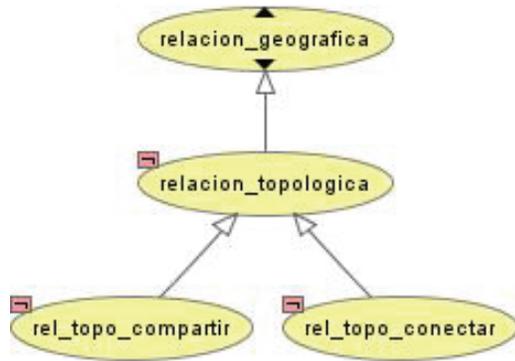


Figura 32. Subclases de “relación geográfica”

Conceptos de la ontología del dominio geográfico

Como se ha mencionado, para la conceptualización del dominio geográfico utilizaremos los diccionarios de datos del INEGI. Estos diccionarios de datos cubren en diversas escalas temas tales como Topografía, Geología, Hidrología, Edafología, Uso del Suelo, etc. Para nuestro caso de estudio hemos utilizado solamente el diccionario de datos topográficos vectoriales de escala 1:50,000 (INEGI, 1996), en donde se definen los rasgos topográficos que actualmente están incluidos en los mapas. Existen rasgos que anteriormente se representaban y que actualmente no es así, por lo que no están considerados en este diccionario. Los nombres, definiciones y la caracterización mediante la asignación de un número determinado de atributos para cada entidad, han sido ajustados con base en estas consideraciones. Algunos rasgos han sido agrupados en una sola entidad y algunos otros han sido desagregados en varias entidades por razones prácticas del manejo digital de la información.

Como ejemplo, en la Figura 33 se muestra la definición de ACUEDUCTO tal y como se presenta en el diccionario de datos del INEGI que mencionamos anteriormente.

ACUEDUCTO	(1)Nombre
Conducto artificial empleado para transportar agua potable.	(2) Definición
ATRIBUTOS	
<p><u>DOMINIO FIJO</u></p> <p>RELACIÓN DEL ACUEDUCTO CON EL SUELO DOMINIO DE VALORES : Subterráneo : Bajo la superficie del terreno. Superficial : Sobre la superficie del terreno.</p> <p>CONDICIÓN DEL ACUEDUCTO DOMINIO DE VALORES : En construcción : Que está en proceso de construcción. En operación : Que está en servicio o puede usarse. Fuera de uso : Que no está en uso.</p> <p><u>DOMINIO VARIABLE</u></p> <p>IDENTIFICADOR DE ACUEDUCTO.: Un número secuencial que se incrementa con cada ocurrencia. DOMINIO DE VALORES : En el intervalo de 1 a N.</p>	(3)Atributos
RESTRICCIONES DE INTEGRIDAD	
<p>COMBINACIÓN(ES) AUTORIZADA(S) DE VALORES DE ATRIBUTOS Subterráneo, en construcción. Subterráneo, en operación. Subterráneo, fuera de uso. Superficial, en construcción. Superficial, en operación. Superficial, fuera de uso.</p> <p>CALIFICADOR AUTORIZADO DE REPRESENTACIÓN(ES) GEOMÉTRICA(S) Aproximada. Definida.</p>	(4)Combinaciones
RELACIONES	
Acueducto (L) Conecta Instalación de bombeo (P) Acueducto (L) Conecta Manantial (P) Acueducto (L) Conecta Tanque de agua (P) Acueducto (L) Conecta Acueducto (L) Acueducto (L) Conecta Canal (L) ...	(5)Relaciones

Figura 33. Definición de ACUEDUCTO en el diccionario de datos topográficos del INEGI

En dicha figura podemos apreciar que la definición de un concepto tiene:

1. **Nombre.** Es la etiqueta mediante la cual se identifica al concepto geográfico.
2. **Definición.** Es una descripción textual en lenguaje natural del significado y alcances del concepto geográfico.
3. **Atributos.** Es una colección de las propiedades del concepto geográfico. Cada atributo tiene un nombre (en el ejemplo de la Figura 33 tenemos como atributos: RELACION DEL ACUEDUCTO CON EL SUELO, CONDICIÓN DEL ATRIBUTO e IDENTIFICADOR DEL ACUEDUCTO), un significado de cada atributo (en el ejemplo, IDENTIFICADOR DEL ACUEDUCTO es "Un número secuencial que se incrementa con cada ocurrencia."), un rango de valores. Este último dependerá del dominio del atributo.
 - a. **Dominio Fijo.** Significa que el atributo puede tomar un valor de entre una colección fija de posibilidades. En el ejemplo, RELACIÓN DEL ACUEDUCTO CON EL SUELO solo puede tomar uno de dos valores: Subterráneo o Superficial. En términos de una base de datos convencional, un atributo de dominio fijo significa que su valor proviene de un catálogo. Para los fines de la conceptualización los atributos de dominio fijo pueden representar especializaciones del concepto si este dominio coincide con alguna de las clases definidas en la sección anterior. Es decir, además del concepto ACUEDUCTO debemos prever la existencia del concepto ACUEDUCTO SUBTERRÁNEO y del concepto ACUEDUCTO SUPERFICIAL, pues subterráneo y superficial coinciden con sendas clases en nuestra conceptualización. Esto no ocurre en el caso de CONDICIÓN DEL ACUEDUCTO, pues sus valores posibles no coinciden con los de clase alguna.
 - b. **Dominio Variable.** Significa que el valor del atributo puede ser alguno de entre una amplia gama de éstos. En el ejemplo, IDENTIFICADOR DEL ACUEDUCTO puede tomar como valor cualquier número natural.
4. **Combinaciones.** Se refiere a qué combinaciones de atributos del dominio fijo son válidas y cuáles no. Por ejemplo, en una carretera un atributo de dominio fijo es

TIPO DE CARRETERA y el rango de valores es pavimentada o terracería, mientras que otro atributo es NÚMERO DE CARRILES con valores de uno a seis. Las combinaciones válidas restringen que no se puede aplicar el número de carriles cuando la carretera es de terracería.

5. **Relaciones.** Describe las relaciones que un concepto geográfico tiene con otros conceptos geográficos.

El diccionario de datos topográficos vectorial 1:50,000 considera los siguientes conceptos: Acueducto, Aeropuerto, Área de cultivo, Área natural protegida, Área urbana, Área verde urbana, Arrecife/bajo, Banco de material, Bordo, Calle, Camino, Canal, Carretera, Caseta de peaje, Cementerio, Conducto, Corriente de agua, Corriente que desaparece, Cuerpo de agua, Curva de nivel, Depósito de desechos, Edificación, Entrada a gruta, Estanque, Estructura elevada, Fango, Faro/radiofaro/vor, Instalación de bombeo, Instalación de comunicación, Instalación deportiva o recreativa, Instalación diversa, Instalación industrial, Instalación portuaria, Límite, Lindero, Línea de comunicación, Línea de transmisión, Lumbrera, Malpaís, Manantial, Mina, Mojonera, Muro de contención, Nieve perpetua, Pantano, Pista de aviación, Pista de carreras, Planta generadora, Pozo de explotación, Presa, Puente, Punto acotado, Rápido, Rasgo arqueológico, Roca, Ruta de embarcación, Ruta de funicular/teleférico, Salina, Salto de agua, Separador, Subestación eléctrica, Tanque, Tanque de agua, Terreno sujeto a inundación, Túnel, Vado, Vegetación densa, Vía férrea, Zona arenosa.

Relaciones, propiedades y habilidades de la ontología del dominio geográfico

En la metodología de conceptualización se establecen solo tres relaciones básicas ("es", "tiene" y "hace"). Utilizando la relación "es" establecemos la existencia del concepto geográfico. Con la relación "tiene" podemos conceptualizar las propiedades que un concepto geográfico posee. Finalmente, con la relación "hace"

se pueden establecer las habilidades de un concepto²³. Así al describir las relaciones que tiene un concepto geográfico estamos definiendo su existencia, sus propiedades y sus habilidades.

La existencia de un concepto geográfico la definiremos creando el concepto con su definición (en lenguaje natural), asignando una etiqueta y ubicándolo mediante la relación “es” como miembro de alguna subclase, para todas las clases presentadas anteriormente, como se ilustra en la Figura 34. En ella se ilustran las relaciones existenciales de acueducto con las subclases de objeto geográfico, así como la creación de conceptos más especializados (acueducto superficial y acueducto subterráneo).

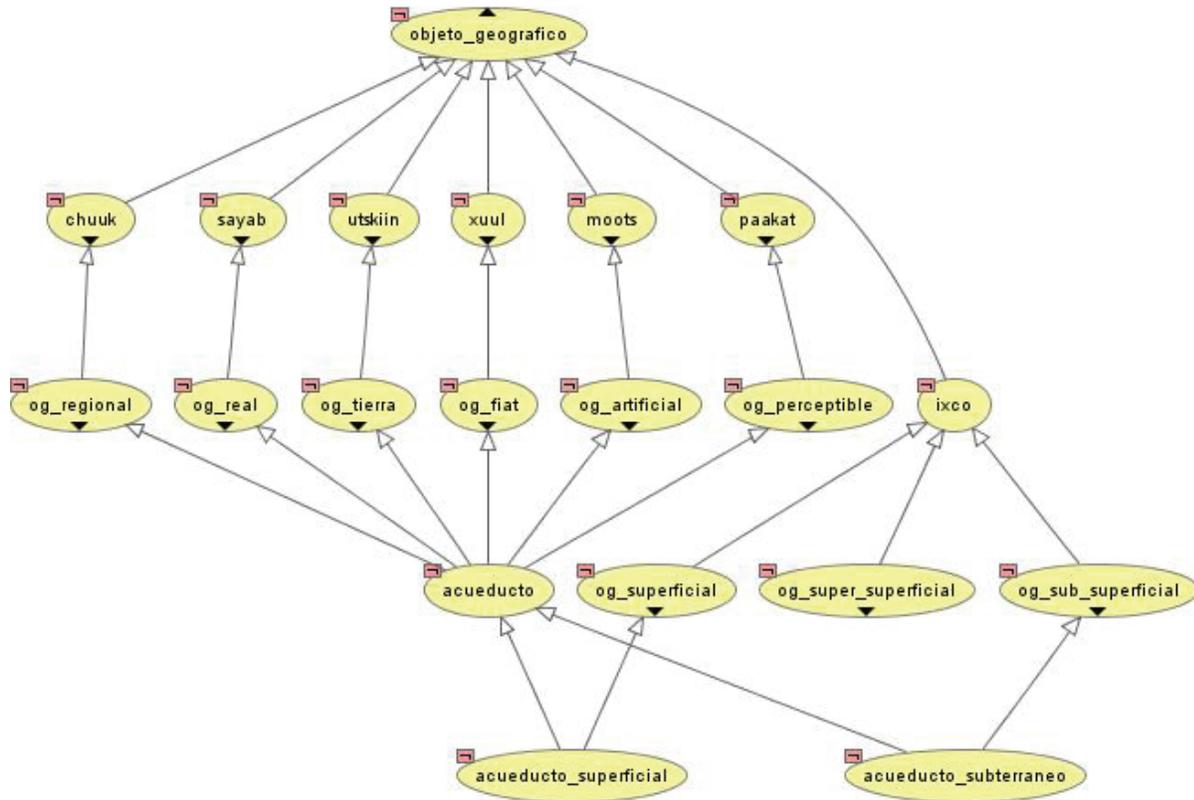


Figura 34. Ubicación existencial de ACUEDUCTO en la conceptualización

²³ En el caso de la conceptualización del dominio geográfico, la principal habilidad que tienen los conceptos es la de relacionarse geográficamente con otros objetos.

En la Figura 33 se presenta la forma en la cual se definen los conceptos geográficos en el diccionario de datos topográficos del INEGI, en el cual se describen las relaciones que un concepto geográfico tiene con otros conceptos (en la sección creada para tal efecto). Asimismo, se describen los atributos que el concepto tiene. También, se ha mencionado que aquellos atributos de dominio fijo que se corresponden con clases de la conceptualización, imponen la creación de conceptos más especializados. Sin embargo, no hemos mencionado qué debe hacerse con aquellos atributos que no están en tal situación (los de dominio fijo que no se corresponden con clases y los de dominio variable). Lo que se hace en esta situación es tomar estos atributos como propiedades de los conceptos. Es decir, y de acuerdo con la metodología de conceptualización, se deben crear conceptos que describen estos atributos y agregarlos al concepto geográfico por medio de una relación "tiene". En la Figura 35 se muestra gráficamente la asignación del atributo "condición de construcción" al concepto geográfico "acueducto" como una propiedad, utilizando una relación "tiene" para tal efecto.

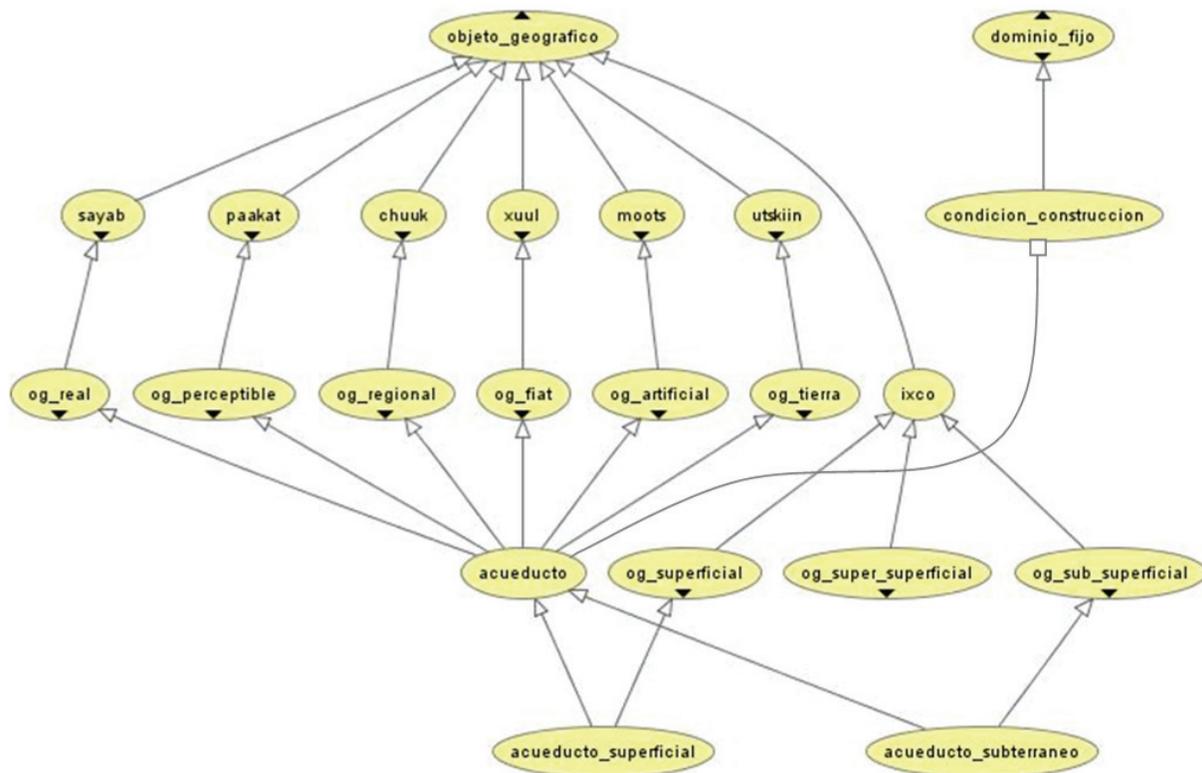


Figura 35. Asignación del atributo de dominio fijo “condición de construcción” a “acueducto” como una propiedad

En la misma figura se puede apreciar la existencia de un concepto llamado “dominio fijo”, el cual es una clase que nos sirve para agrupar precisamente los conceptos de dominio fijo (o catálogos) que presenta el INEGI en su diccionario de datos.

De manera similar, para agregar como propiedad un atributo de dominio variable, es necesario crear la clase “dominio variable” para agrupar aquellos atributos que solo son valores numéricos simples. En la Figura 36 se ilustra la asignación del atributo del dominio variable “identificador numérico” al concepto atributo.

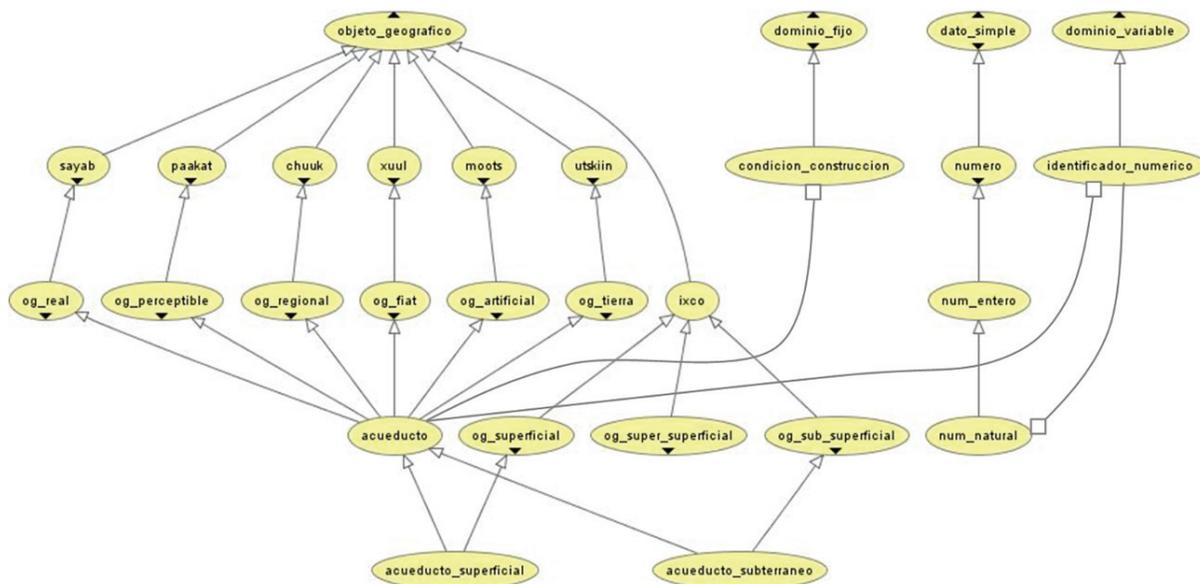


Figura 36. Asignación del atributo de dominio variable “identificador numérico” a “acueducto” como una propiedad

Hasta este punto, hemos mencionado cómo se utilizan las relaciones “es” y “tiene” para definir la existencia y las propiedades de los conceptos geográficos. Ahora utilizaremos la relación “tiene” para definir las habilidades de los conceptos geográficos que, en general, son las relaciones geográficas definidas en el diccionario de datos. En el ejemplo de la Figura 33 se pueden apreciar que las relaciones que están definidas son “conecta” y “comparte”. Aunque en esa figura solo se presenta la definición de “acueducto”, que no tiene definidas relaciones “comparte”, podemos mencionar que esas dos relaciones geográficas son las únicas que se utilizan en el diccionario de datos de INEGI.

Ahora bien, para conceptualizar estas habilidades utilizaremos la forma compuesta de las relaciones, empleando las relaciones axiomáticas definidas por el conjunto de preposiciones; en concreto utilizaremos la preposición “con”. En la Figura 37 se muestra la asignación de la habilidad “compartir con túnel” al concepto “acueducto”; debemos recordar que esta asignación se realiza con la relación “hace” utilizando la forma compuesta, por lo que la lectura de esto es “acueducto (hace) compartir (con) túnel”.

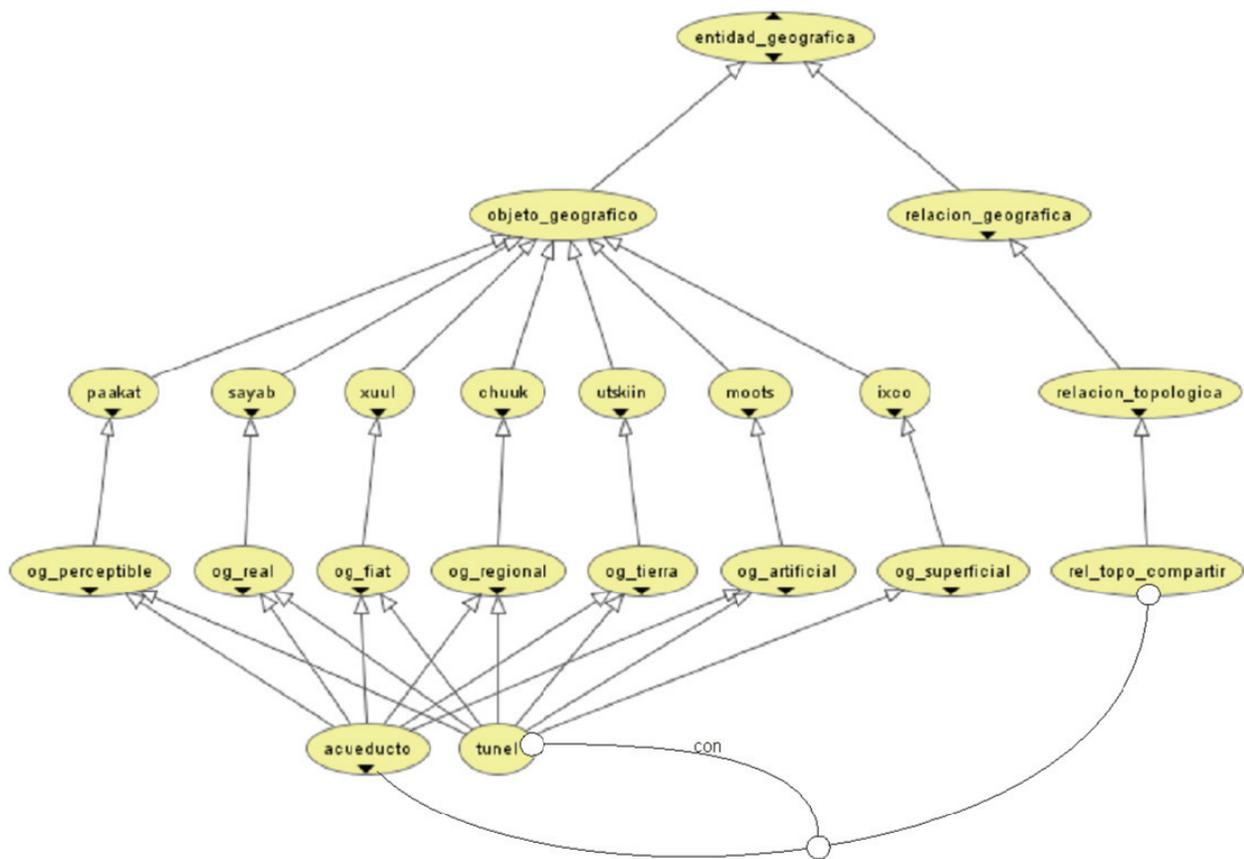


Figura 37. Asignación de la relación geográfica “comparte con túnel” a “acueducto” como una habilidad

De manera similar se conceptualizan todos y cada uno de los conceptos mencionados anteriormente.

Instancias de la ontología del dominio geográfico

Al tratarse de una ontología de dominio general, la ontología Kaab no tiene instancias.

5.1.2. Ontología del dominio de formas del terreno

De manera similar a la definición de la ontología kaab, para definir la ontología de las formas del terreno, debemos establecer los mismos elementos que se presentaron en la metodología, es decir, clases, conceptos, instancias, relaciones, propiedades y habilidades.

Clases de la ontología del dominio de formas del terreno

En la Tabla 12 se presenta la definición de las clases de la ontología del dominio de las formas del terreno. En esta tabla y en las subsecuentes se presenta, si está disponible:

- La definición del concepto dada por el diccionario de la Real Academia de la Lengua (Española, 2007) resaltada en fuente negra.
- En el nombre de la clase o concepto en fuente negra cursiva, el concepto en inglés encontrado en las definiciones de datos ambientales de la ISO (ISO, 2005). Así mismo, la definición en inglés de tal concepto resaltada en fuente cursiva.

Las clases principales de esta ontología son “entidad fisiográfica”, “hunxeet”²⁴, “sitio de interés”, “forma del terreno”, “elevación”, “llanura” y “depresión”; en la Figura 38 se muestra la estructura existencial de estas clases.

Tabla 12. Definición de las clases de la ontología del dominio de las formas del terreno

Clase	Definición
Hunxeet	Esta clase es la más general del dominio de las formas del terreno; es a través de esta clase que iniciamos la conceptualización de dicho dominio. Para el presente trabajo, esta clase es la raíz, descende solo de una clase llamada “cosa”.
Propiedad fisiográfica	Esta clase se utiliza para agrupar los conceptos que más tarde se convertirán en propiedades de las formas del terreno. Ejemplos: altitud, extensión, localización, etc.
Sitio de interés	Esta clase se utiliza para agrupar conceptos que más adelante se emplearán como propiedades de las formas del terreno, pero a diferencia de las “propiedades fisiográficas”, estos conceptos constituyen puntos de singular relevancia en la conceptualización. Ejemplos: cima, cumbre, vaguada, etc.

²⁴ Del maya *hunxéet' lu'um* que significa “paisaje”.

Forma del terreno	Esta clase sirve para agrupar las regiones de superficie que por su forma, constituyen un elemento trascendente en el paisaje; tiene tres subclases: "elevación", "llanura" y "depresión".
Elevación	Es la clase de "forma de terreno" cuya región de superficie se levanta sobre el paisaje, formando un rasgo de una altitud notable. Ejemplo: cerro, colina, monte, etc.
	<p>Región con mayor altitud. Ángulo sobre el plano horizontal.</p> <p>La elevación del terreno en un lugar específico.<i>T</i></p>
Llanura	Es la clase de "forma de terreno" que representa regiones de superficie que no se elevan ni decaen marcadamente en el paisaje. Ejemplos: llano, planicie, meseta, etc.
	<p>Campo o terreno igual y dilatado, sin altos ni bajos.</p> <p>Una gran área geográfica con poca o ninguna variación de altitud, como por ejemplo un desierto o un pantano</p> <p>Superficie plana que tiene una altitud menor que 500 metros sobre el nivel del mar.</p>
Depresión Terrain depression	Es la clase de forma de terreno que representa regiones de superficie de terreno de relativamente baja elevación que están completamente rodeadas por otra región más alta.
	<p>En un terreno u otra superficie, concavidad de alguna extensión.</p> <p>Zona honda del terreno, está más baja que su alrededor.</p> <p>Vaguada, hundimiento del terreno.</p> <p>Una región de superficie de terreno de relativamente baja elevación que esta completamente rodeada por otra región más alta.</p> <p><i>A terrain surface region of relatively low terrain elevation that is completely surrounded by another terrain surface region of higher terrain elevation; a terrain depression.</i></p>

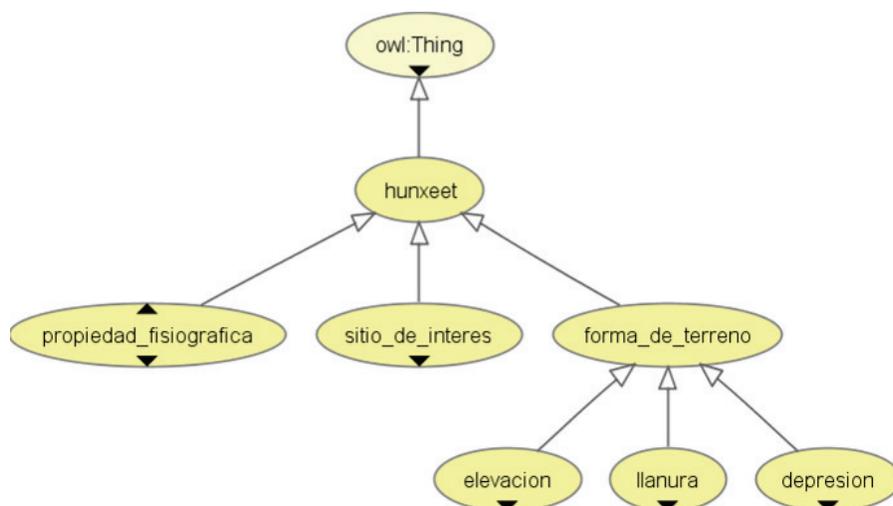


Figura 38. Clases de la ontología del dominio de las formas del terreno

Conceptos de la ontología del dominio de las formas del terreno

En la Tabla 13 se presentan las definiciones de los conceptos de las formas de terreno. Como ya se ha mencionado, esta conceptualización está basada en las definiciones que proporciona el diccionario de la Real Academia de la Lengua. Asimismo, en la Figura 39 se presenta la estructura existencial de tales conceptos.

Tabla 13. Definición de los conceptos de la clase “forma de terreno”

Concepto	Definición
Abismo Abyss	Profundidad grande, imponente y peligrosa, como la de los mares, la de un tajo, la de una sima, etc. Piso del cuerpo de agua con profundidad mayor que 549 metros (aproximadamente 800 pies). <i>A waterbody floor with depth of waterbody floor greater than 549 meters (approximately 1 800 feet); an abyss.</i>
Aluvión Alluvial fan	Dicho de un terreno que queda al descubierto después de las avenidas, o que se forma lentamente por los desvíos o las variaciones en el curso de los ríos. Una configuración del terreno en forma de abanico suavemente inclinada,

que usualmente se encuentra cerca de la terminación baja de un cañon.

A gently sloping, fan-shaped configuration of the terrain, usually found near the lower termination of a canyon; an alluvial fan.

Altiplanicie **Meseta de mucha extensión, situada a gran altitud.**

Región plana localizada sobre un sistema montañoso.

Canal **Llanura larga y estrecha entre dos montañas.**

Terrain channel Una región de superficie alargada, especialmente una que es relativamente estrecha, cuyos lados más largos son regiones de superficie de más alta elevación.

An elongated terrain surface region, especially one that is relatively narrow, whose longer sides are terrain surface regions of higher terrain elevation; a terrain channel.

Cañón **Paso estrecho o garganta profunda entre dos altas montañas, por donde**
Canyon **suelen correr los ríos.**

Una garganta profunda, frecuentemente con un río en el fondo.

A deep gorge, frequently with a river at its bottom; a canyon.

Cerro **Elevación del terreno que se encuentra aislada y tiene una altitud inferior que**
Hill **los montes, sierras y montañas**

Elevación de tierra aislada y de menor altura que el monte o la montaña.

Una porción de terreno pequeña, aislada y elevada que es más pequeña que una montaña.

A small, isolated, elevated portion of the terrain that is smaller than a mountain; a hill.

Colina **Elevación natural de terreno, menor que una montaña.**

Montaña suave.

Collado **Tierra que se levanta como un cerro, menos elevada que el monte.**

Mountain pass Depresión suave por donde se puede pasar fácilmente de un lado a otro de una sierra.

Una ruta de transporte terrestre entre una región montañosa, donde la elevación del terreno es más baja que la de las montañas adyacentes.

A land transportation route through a localized terrain surface region in a range of mountains, where the terrain surface region is lower in terrain elevation than the adjacent mountains; a mountain pass.

Cordillera

Serie de montañas enlazadas entre sí.

Sucesión de montañas enlazadas entre sí.

Una cadena de montañas.

Cráter**Terrain crater**

Depresión topográfica más o menos circular formada por explosión volcánica y por la cual sale humo, ceniza, lava, fango u otras materias, cuando el volcán está en actividad.

Depresión semejante formada por caída de meteoritos en la superficie de la Tierra y de la Luna.

Una depresión del terreno localizada creada por una explosión y/o el impacto de un proyectil grande o de gran velocidad contra el terreno.

A localized terrain depression created by an explosion and/or the impact of a large or high-velocity projectile upon the terrain; a crater.

Cresta**Ridge**

Elevación de tierra estrecha y alargada con caídas a los lados.

Terreno relativamente alto formado como una cresta larga y estrecha que es rodeada por un terreno más bajo.

Relatively high terrain formed as a long and narrow crest that is surrounded by lower terrain; a ridge.

Cuenca**Geographic basin**

Territorio rodeado de alturas. Territorio cuyas aguas afluyen todas a un mismo río, lago o mar.

Depresión en la superficie de la tierra y territorio cuyas aguas confluyen en un río, lago o mar.

Es una depresión en la superficie de la tierra: un valle rodeado de alturas.

La región de superficie de terreno que es drenada por un río y sus tributarios o

que se drena hacia un mar o lago en particular.

The terrain surface region that is drained by a river and its tributaries or that drains into a particular lake or sea; a geographic basin.

Cuenca que se inunda **Una zona que se inunda durante la crecida más grande de una zona de potencial inundación.**

Flood basin *A tract that submerges during the highest known flood in a terrain surface region or has the potential for being flooded; a flood basin.*

Cuesta **Terreno en pendiente.**

Terrain slope La cuesta (asenso) del terreno.

The slope (rise/run) of the terrain.

Duna **Colina de arena movediza que en los desiertos y en las playas forma y empuja el viento.**

Sand dune

Un montón de arena floja formado por el viento.

A mound or ridge of loose sand formed by the wind; a sand dune.

Escarpa **Declive áspero del terreno.**

Escarpment Una cara escarpada, vertical o sobresaliente de la roca y/o el suelo en la tierra.

A steep, vertical, or overhanging face of rock and/or soil in the land; an escarpment.

Falla **Quiebra que los movimientos geológicos han producido en un terreno.**

Geologic fault Una fractura en el terreno con desplazamiento en un lado de la fractura con respecto a la otra.

A fracture in the terrain with displacement on one side of the fracture relative to the other; a geologic fault.

Foso Una larga y estrecha excavación en el terreno.

Trench *A long and narrow excavation in the terrain; a trench or a ditch.*

Llano <i>Level</i>	Terreno igual y extendido, sin altos ni bajos.
Loma Butte	Altura pequeña y prolongada. Una colina o una altiplanicie separada, con la cima plana más pequeña que una “mesa” y que se levanta escarpadamente del terreno llano circundante. <i>A detached, flat-topped hill or tableland that is smaller than a mesa and rises steeply from a surrounding terrain plain; a butte.</i>
Mesa Mesa	Una colina de cima plana o altiplanicie que es más grande que una loma que se eleva escarpadamente de sus alrededores. <i>A flat-topped hill or tableland that is larger than a butte and rises steeply from a surrounding terrain plain; a mesa.</i>
Meseta Plateau	Planicie extensa situada a considerable altura sobre el nivel del mar. Una región de la superficie de terreno elevada que es comparativamente plana en su cima y de considerable extensión. <i>An elevated terrain surface region that is comparatively flat-topped and of considerable extent; a plateau.</i>
Montaña Mountain	Gran elevación natural del terreno. Una elevación natural del terreno de una altitud que es impresionante o notable con relación a las elevaciones adyacentes. <i>A natural elevation of terrain of an altitude that is impressive or notable relative to adjacent elevations; a mountain.</i>
Monte Mount	Gran elevación natural de terreno. Tierra inculta cubierta de árboles, arbustos o matas. Una colina alta.
Peña	Piedra grande sin labrar, según la produce la naturaleza. Monte o cerro peñascoso.
Peñasco	Peña grande y elevada.

Pico	Cúspide aguda de una montaña. Montaña de cumbre puntiaguda.
Peak	Cúspide de una montaña que ha sido esculpida por la erosión hasta formar un punto. <i>Mountaintop that has been sculpted by erosion to form a point.</i>
Planicie	Terreno llano de alguna extensión.
Terrain plain	Una región de superficie de terreno que es comparativamente plana. <i>A terrain surface region that is comparatively flat; a terrain plain.</i>
Sierra	Parte de una cordillera. Cordillera de montes o peñascos cortados.
Sima	Cavidad grande y muy profunda en la tierra.
Valle	Llanura de tierra entre montes o alturas.
Valley region	Depresión de la superficie terrestre, de forma alargada e inclinada hacia un lago, mar o cuenca endorreica, habitualmente ocupada por un río. Generalmente se forma por la erosión fluvial y la meteorización mecánica. Llanura entre montañas o alturas. Cuenca de un río. Zona larga y baja de tierra, suele estar entre lomas o montañas atravesadas por un río. Una región de superficie baja localizada entre Colinas y/o montañas y que muchas veces tiene un río que fluye a través de él. <i>A low-lying terrain surface region located between hills and/or mountains and often having a river flowing through it; a valley region.</i>

En la Figura 39 se presenta la estructura existencial de los conceptos de las formas de terreno.



Figura 39. Conceptos de la clase “forma de terreno”

De manera similar, en la Tabla 14 se presentan las definiciones de los conceptos que pertenecen a la clase "sitio de interés". En la Figura 40 se presenta la estructura de los conceptos que conforman esta clase.

Tabla 14. Conceptos de la clase "sitio de interés"

Concepto	Definición
Acantilado	Escarpa casi vertical en un terreno.
Cliff	Una cara de roca vertical o cerca de serlo en un terreno de considerable altura. <i>A vertical or nearly vertical face of rock in the terrain of considerable height; a cliff.</i>
Cima	Punto más alto de los montes, cerros y collados. La cumbre de una colina o una montaña que acaba en punta. Cresta de una montaña. Parte más alta de los montes, cerros, etcétera.
Cumbre	Cima o parte superior de un monte. Mayor elevación de algo.
Vaguada Valley bottom line	Línea que marca la parte más honda de un valle, y es el camino por donde van las aguas de las corrientes naturales. Una línea que representa la parte más baja de un valle. <i>A line representing of the lowest part of a valley region; a valley bottom line.</i>
Zima	Parte más baja de una depresión.

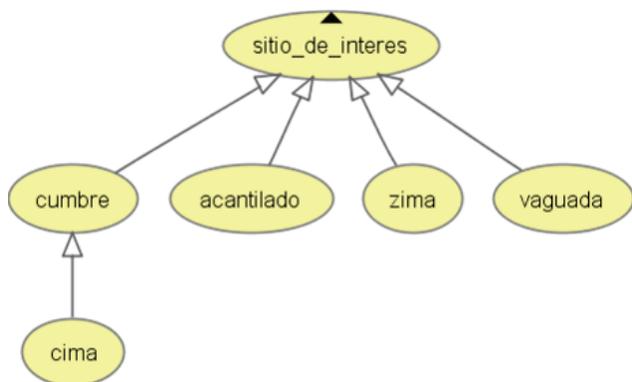


Figura 40. Conceptos de la clase “sitio de interés”

Continuando con la definición de conceptos de la ontología de las formas de terreno, y de manera similar que en las clases anteriores, en la Tabla 15 se presentan las definiciones de los conceptos que pertenecen a la clase “propiedad fisiográfica”. En la Figura 41 se presenta la estructura de los conceptos que conforman esta clase.

Tabla 15. Conceptos de la clase “propiedad fisiográfica”

Concepto	Definición
Cuesta	Terreno en pendiente.
Terrain slope	La cuesta (asenso) del terreno. <i>The slope (rise/run) of the terrain.</i>
Escarpa	Declive áspero del terreno.
Escarpment	Una cara escarpada, vertical o sobresaliente de la roca y/o el suelo en la tierra. <i>A steep, vertical, or overhanging face of rock and/or soil in the land; an escarpment.</i>
Localización	Ubicación de un “objeto geográfico” con respecto al sistema de referencia.
Altitud	Distancia vertical de un punto de la tierra respecto al nivel medio del mar.
Altitud mínima	Distancia vertical más baja de una “forma de terreno” con respecto al nivel medio del mar.
Altitud media	Distancia vertical promedio de una “forma de terreno” con respecto al nivel

	medio del mar.
Altitud máxima	Distancia vertical más alta de una “forma de terreno” con respecto al nivel medio del mar.
Extensión	Medida del espacio ocupado por un cuerpo.

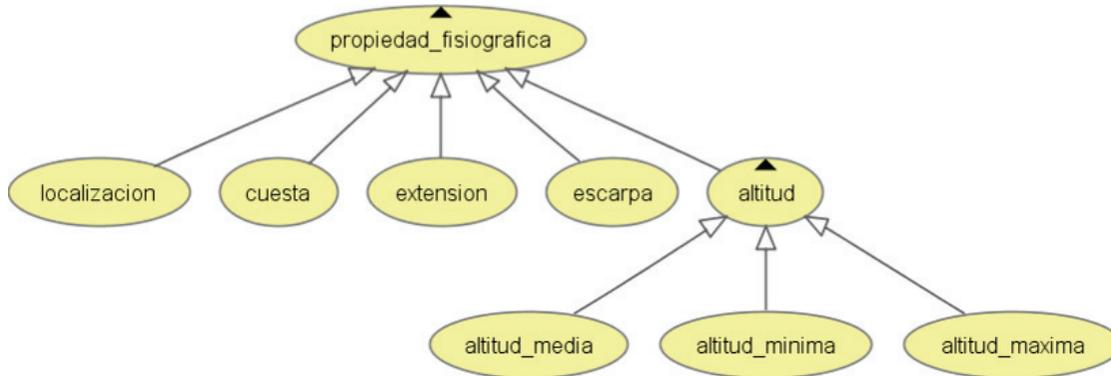


Figura 41. Conceptos de la clase “propiedad fisiográfica”

Finalmente, en la Tabla 16 se introducen algunos conceptos utilizados en las definiciones dadas anteriormente. Tales conceptos no forman parte de la ontología del dominio de las formas de terreno, sino más bien formarían parte de una ontología más general. Para los fines de este trabajo solo presentamos estas definiciones solo como elementos clarificadores del significado de los conceptos que sí pertenecen a la ontología del dominio de las formas del terreno.

Tabla 16. Otros conceptos utilizados en la conceptualización

Concepto	Definición
Aislado	Sola, suelta, individual.
Alargado	Considerablemente más largo que ancho.
Confluir	Dicho de dos o más ríos u otras corrientes de agua cuando se unen.
Cubierto	Tener extendido algo sobre la superficie.
Dilatado	Extenso, vasto, numeroso.
Enlazado	Junto.
Entre	Denota la situación o estado en medio de dos o más cosas.
Erizado	Lleno o rodeado de obstáculos, asperezas, inconvenientes.
Extendido	Esparcido.

Inclinado	Apartada de su posición perpendicular a otra cosa o al horizonte.
Inculto	Que no tiene cultivo.
Prolongado	Más largo que ancho.
Sobre	Encima de.
Región Region	<p>Porción de territorio determinada por caracteres étnicos o circunstancias especiales de clima, producción, topografía, administración, gobierno, etc.</p> <p>Un subconjunto usualmente contiguo del medio ambiente que debe incluir el terreno, cuerpos de agua, la atmósfera y/o el espacio.</p> <p><i>A usually contiguous subset of the environment that may include terrain, a waterbody, the atmosphere, and/or space; a region.</i></p>
Región de superficie de terreno Terrain surface Region	<p>Una región de terreno, especialmente una que se distingue por sus características naturales, su clima, su fauna o flora.</p> <p><i>A region of terrain, especially one distinguished by natural features, climate, fauna or flora; a terrain surface region.</i></p>
Terreno Terrain	<p>Sitio o espacio de tierra.</p> <p>La superficie de la porción sólida de un planeta, que puede o no estar cubierta por agua.</p> <p><i>The surface of the solid portion of a planet, which may or may not be covered by water; terrain or ground.</i></p>
Tierra Land	<p>Parte superficial del planeta Tierra no ocupada por el mar.</p> <p>Terreno que no está permanentemente cubierto por agua.</p> <p><i>Terrain that is not permanently covered by water; land.</i></p>
Suelo Soil	<p>Superficie de la Tierra.</p> <p>Un material compuesto por partículas hechas de fragmentos de roca, de plantas descompuestas, de agua y de aire, que abarca típicamente una capa superior fina de la tierra.</p>

A material composed of fragmented particles of rock, decomposed plants, water, and air that typically comprises a thin top layer of the land; soil.

Región inclinada Una región de superficie donde la inclinación o cuesta del terreno se encuentra dentro de un rango de valores.

Slope region

A terrain surface region where the terrain slope is within a set range of values.

Región montañosa Una región de superficie de terreno cuya fisiografía consiste predominantemente de montañas.

Mountainous

region

A terrain surface region, the physiography of which consists predominantly of mountains; a mountainous region.

Sobresalir Exceder a otras en figura, tamaño, etc.

Relaciones, propiedades y habilidades en la ontología del dominio de las formas del terreno

Las relaciones existenciales en esta ontología ya se han ido presentando conforme se han introducido las clases y los conceptos que la conforman. Hasta este punto falta por definir las relaciones “tiene” y “hace”, propiedades y habilidades, respectivamente.

Las propiedades de los conceptos son los conceptos pertenecientes a la clase “propiedad fisiográfica”, o como en algunos casos, otras “formas del terreno”. Como ya se ha visto, se realiza la agregación de tales propiedades por medio de una relación “tiene”. En la tabla se presentan las propiedades asociadas a los conceptos de la ontología del dominio de las formas del terreno (por ende, las relaciones “tiene” en la misma ontología).

Tabla 17. Propiedades de los conceptos de la ontología del dominio de las formas del terreno

Concepto	Propiedades
Todos los de la clase “forma del terreno”	“localización”, “extensión” y “altitud media”.
Todos los de la clase “depresión”	Los heredados de la clase “forma del terreno” más “zima”, “altitud mínima” y “altitud máxima”.

Todos los de la clase "elevación"	Los heredados de la clase "forma del terreno" más "cima", "altitud mínima" y "altitud máxima".
Todos los de la clase "llanura"	Los heredados de la clase "forma del terreno".
"abismo"	"acantilado"
"cañón"	"acantilado"
"cordillera"	"collado"

La ontología del dominio de las formas del terreno no tiene relaciones "hace", o lo que es lo mismo, los conceptos de la ontología no tienen habilidades²⁵.

Instancias de la ontología del dominio de las formas del terreno

La ontología del dominio de las formas del terreno no tiene instancias.

5.1.3. Ontología de la aplicación

Clases de la ontología de la aplicación

La ontología de la aplicación no tiene clases, pues como mencionamos en la metodología, se trata de una jerarquía de extractos en la cual (al menos en un principio) todos los conceptos de la ontología pueden o podrían tener instancias. Si recordamos, en la metodología se establece que una de las restricciones es que las instancias no pueden ser heredadas de clases; entonces está claro que ninguno de los conceptos de la ontología de la aplicación puede ser una clase.

Conceptos de la ontología de la aplicación

Ahora bien, la ontología de la aplicación es una conceptualización en forma de jerarquía de los extractos que se obtienen con los algoritmos de extracción. Así, está claro entonces que ésta ontología depende de la funcionalidad de los algoritmos seleccionados. Para el caso de estudio de las formas del terreno, utilizaremos un

²⁵ En un trabajo de tesis de maestría derivado del presente, se está trabajando en la conceptualización de las relaciones geográficas existentes entre diferentes formas del terreno. Ver el apartado Trabajos Futuros en el siguiente capítulo.

algoritmo de extracción cuya funcionalidad arroja tres tipos de extracto: depresión, llanura y elevación. En las siguientes secciones se explica más detalladamente dicho algoritmo de extracción. En este punto nos interesa conceptualizar los resultados de tal algoritmo. En la Figura 42 se ilustra la jerarquía que se obtiene al ejecutar una vez el algoritmo en cuestión al DEM que está siendo analizado.

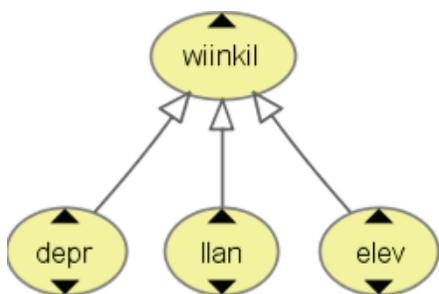


Figura 42. Extractos que genera el algoritmo de extracción

En la misma figura se ilustran cuatro conceptos:

- “wiinkil”²⁶, que es el concepto que representa al CDER que está siendo analizado.
- “llan”, es el concepto que representa a los extractos que han sido identificados como llanuras. La etiqueta “llan” la utilizaremos como firma de tales extractos.
- “elev”, es el concepto que representa a los extractos que han sido identificados como elevaciones.
- “depr”, es el concepto que representa a los extractos que han sido identificados como depresiones.

En este nivel, los extractos solo se han diferenciado como elevaciones, depresiones o llanuras, los que se corresponden con sendos conceptos en la ontología del dominio de las formas del terreno. Un siguiente nivel de ejecución del algoritmo de extracción revelará más detalles sobre los extractos y su comportamiento para poder relacionarse con el concepto de la ontología del dominio de las formas del terreno que mejor los

²⁶ Del maya *wiinkil* que significa forma

describe. En la Figura 43 se presenta una parte de la jerarquía que resulta de la conceptualización del algoritmo de extracción.

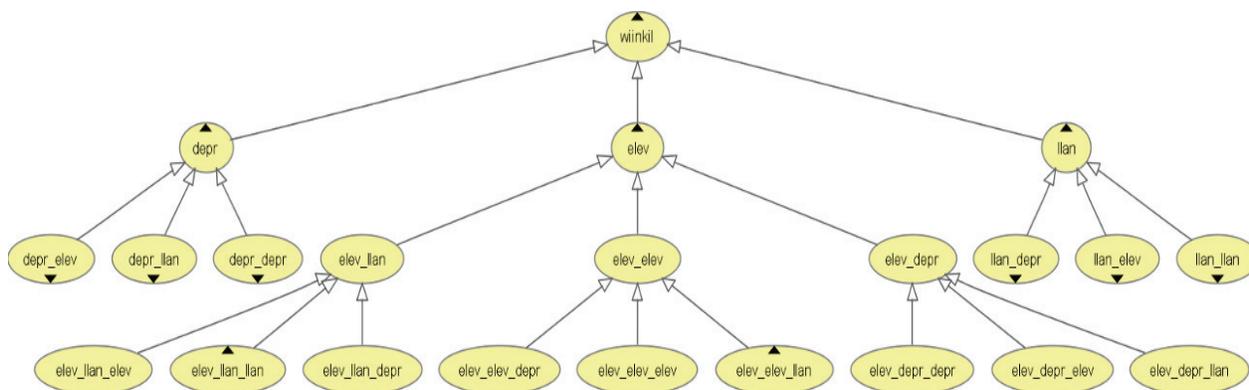


Figura 43. Fragmento de la jerarquía de extractos que describe la ontología de la aplicación

Relaciones, propiedades y habilidades de la ontología de la aplicación

Al igual que en secciones anteriores, las relaciones “es” se ilustran por medio de las diferentes figuras en las que se muestra la estructura existencial de los conceptos. Ahora bien, en el caso de la ontología de la aplicación y como se mencionó en la metodología, las propiedades y habilidades de los conceptos son heredados de los conceptos de la ontología del dominio de las formas del terreno con los cuales se relacionan. Más adelante se definirán estas relaciones, así como las relaciones de la ontología del dominio de las formas del terreno con la ontología del dominio geográfico.

Instancias de la ontología de la aplicación

Las instancias de la ontología de la aplicación son dinámicas, pues se trata de los extractos que se obtienen en el proceso de síntesis mediante el algoritmo de extracción utilizado. Adicionalmente, como se presentó en la metodología de descripción, cada extracto es calificado en cuanto a las propiedades que posee. De esta manera, habrá tantas instancias como extractos, además de las instancias de las propiedades de cada extracto. En la Figura 44 se muestra un ejemplo del uso de instancias en la ontología de la aplicación; en tal figura se muestran la instancia de extracto llamada “ext_1”, así como las instancias de las propiedades asociadas a él

(“exten_1” para su extensión, “loc_1” para su localización y “altm_1” para su altitud media). De acuerdo a la integración de la ontología de la aplicación con la ontología hunxeet, que se presenta más adelante, un extracto cuya firma es “llan_llan_llan” se relaciona con el concepto llano; quien a su vez tiene tres propiedades: “extensión”, “localización” y “altitud media”.

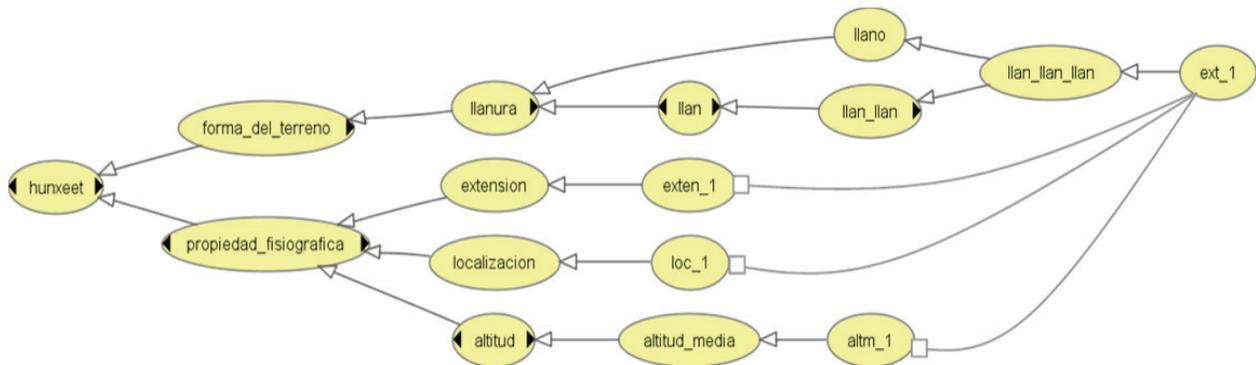


Figura 44. Ejemplos de instancias en la ontología de la aplicación

5.1.4. Integración de las ontologías

Como se ha mencionado previamente, las tres ontologías desarrolladas se deben integrar para utilizar y enriquecer el conocimiento intrínseco de cada una²⁷. En primer lugar, en la Figura 45 se muestra la integración de la ontología kaab con la ontología hunxeet, mediante la clasificación²⁸ de la clase “forma del terreno” como un “objeto geográfico”. Como se puede apreciar en la misma figura, “forma del terreno” se clasifica como “og perceptible”, “og real”, “og tierra”, “og superficial”, “og regional”, “og natural” y “og bona fide”. De esta manera caracterizamos las formas del terreno existencialmente de acuerdo a la ontología del dominio geográfico.

²⁷ Esta integración **no es automática**.

²⁸ Asignar los conceptos a las clases correspondientes en la ontología kaab.

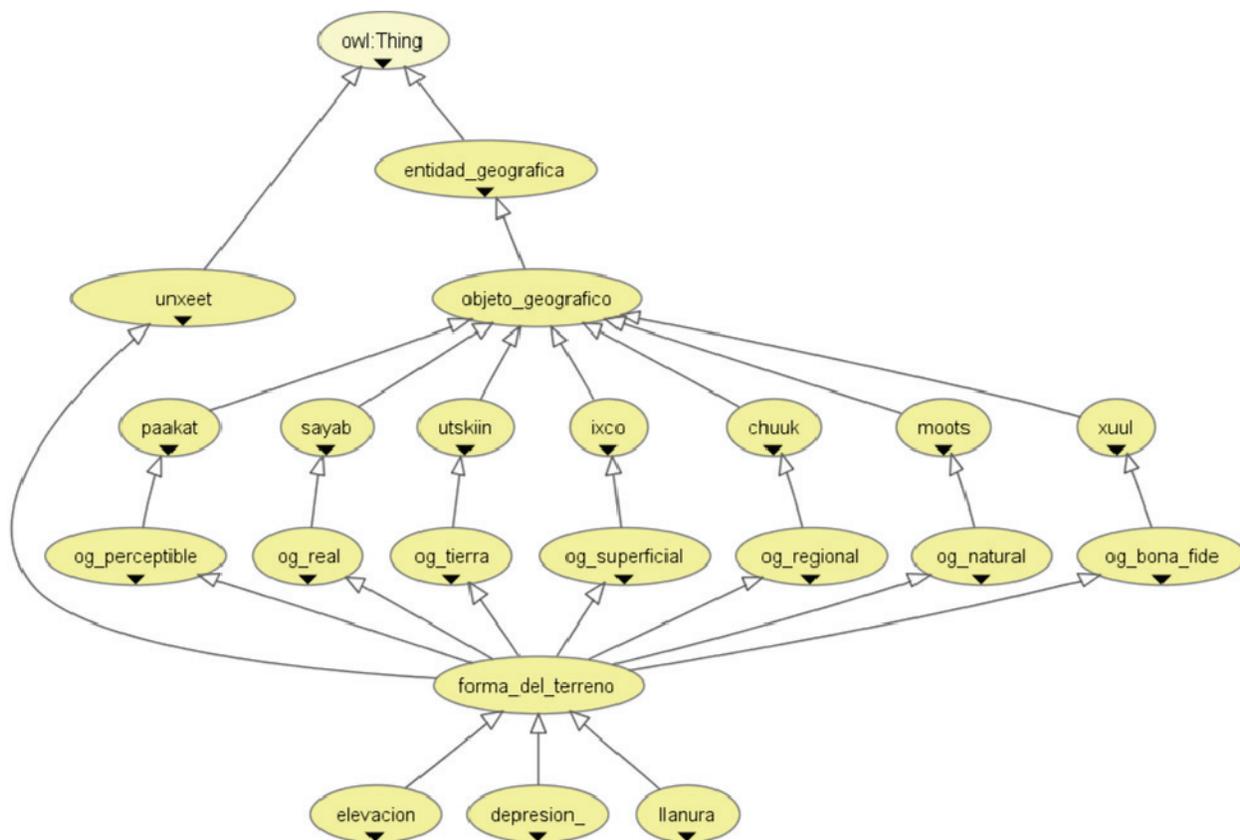


Figura 45. Integración de la ontología hunxeet con la ontología kaab

De manera análoga, debemos integrar la ontología de la aplicación con la ontología de las formas del terreno. Esta es un poco más compleja que en el caso anterior, pues se deben relacionar cada uno de los conceptos en la jerarquía de extractos con conceptos en la ontología hunxeet. En la Figura 46 se ilustran algunos ejemplos de dicha integración. En dicha figura podemos apreciar, por ejemplo, que los extractos cuya firma es “elev” se relacionan con el concepto “elevación”; otro ejemplo, no tan trivial, lo podemos apreciar con los extractos con firma “depr_llan_llan”, los cuales se vinculan con el concepto “cuenca”. De esta manera, se vincula cada concepto en la ontología wiits con algún concepto en la ontología hunxeet.

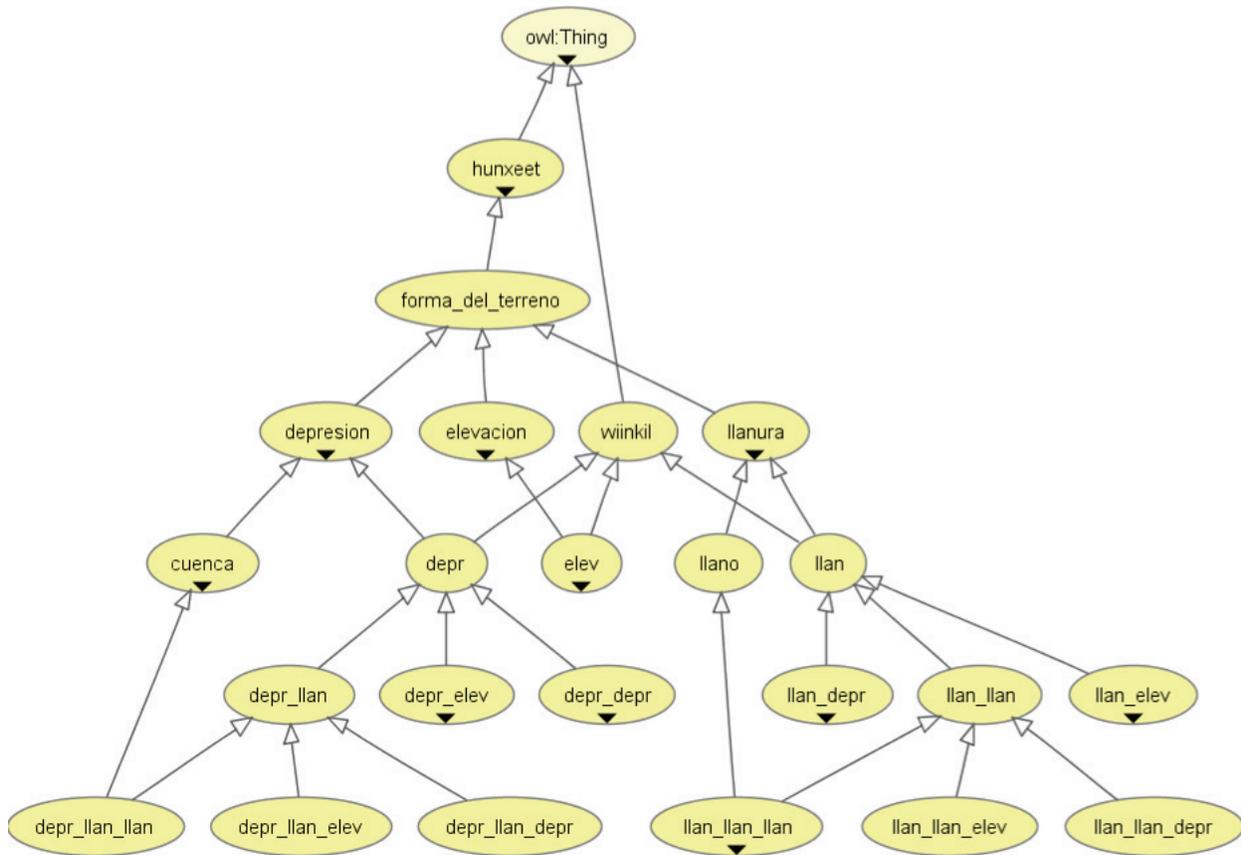


Figura 46. Integración de la ontología wiinkil con la ontología hunxeet

En la Tabla 18 se listan las relaciones entre los conceptos de la ontología de la aplicación y la ontología del dominio de las formas del terreno. Debemos recordar que, de acuerdo a la metodología propuesta, estas relaciones siempre serán del tipo “es” en su forma simple.

Tabla 18. Lista de relaciones entre la ontología wiinkil y la ontología hunxeet

Concepto wiits	Concepto hunxeet	Concepto wiits	Concepto hunxeet
elev	elevación	depr_elev_elev	-
depr	depresión	depr_elev_depr	-
llan	llanura	depr_elev_llan	-
elev_elev	elevación	depr_depr_elev	-
elev_depr	elevación	depr_depr_depr	-

elev_llan	elevación	depr_depr_llan	-
depr_elev	-	depr_llan_elev	-
depr_depr	cuenca	depr_llan_depr	-
depr_llan	valle	depr_llan_llan	cuenca
llan_elev	llano	llan_elev_elev	-
llan_depr	meseta	llan_elev_depr	-
llan_llan	planicie	llan_elev_llan	-
elev_elev_elev	montaña	llan_depr_elev	-
elev_elev_depr	monte	llan_depr_depr	-
elev_elev_llan	cordillera	llan_depr_llan	-
elev_depr_elev	-	llan_llan_elev	-
elev_depr_depr	-	llan_llan_depr	-
elev_depr_llan	-	llan_llan_llan	planicie
elev_llan_elev	-		
elev_llan_depr	-		
elev_llan_llan	-		

5.2. Síntesis

En esta sección se presentan los resultados de la etapa de síntesis. Como hemos mencionado previamente en la metodología, la etapa de síntesis genera las instancias que pueblan a la ontología de la aplicación por medio de la ontología de extractos. Así que, fundamentalmente, en esta sección abordaremos el algoritmo de extracción propuesto para el caso de estudio.

En la sección anterior mencionamos ya que el algoritmo propuesto genera tres tipos de extracto “elev”, “depr” y “llan”; al aplicar sucesivamente el mismo algoritmo a los extractos se generan las combinaciones de estos tipos que a la postre se constituyen

en lo que llamamos la firma del extracto. En la Figura 47 se presenta un ejemplo de descomposición de un CDER en extractos, utilizando el algoritmo de extracción propuesto que arroja tres extractos.

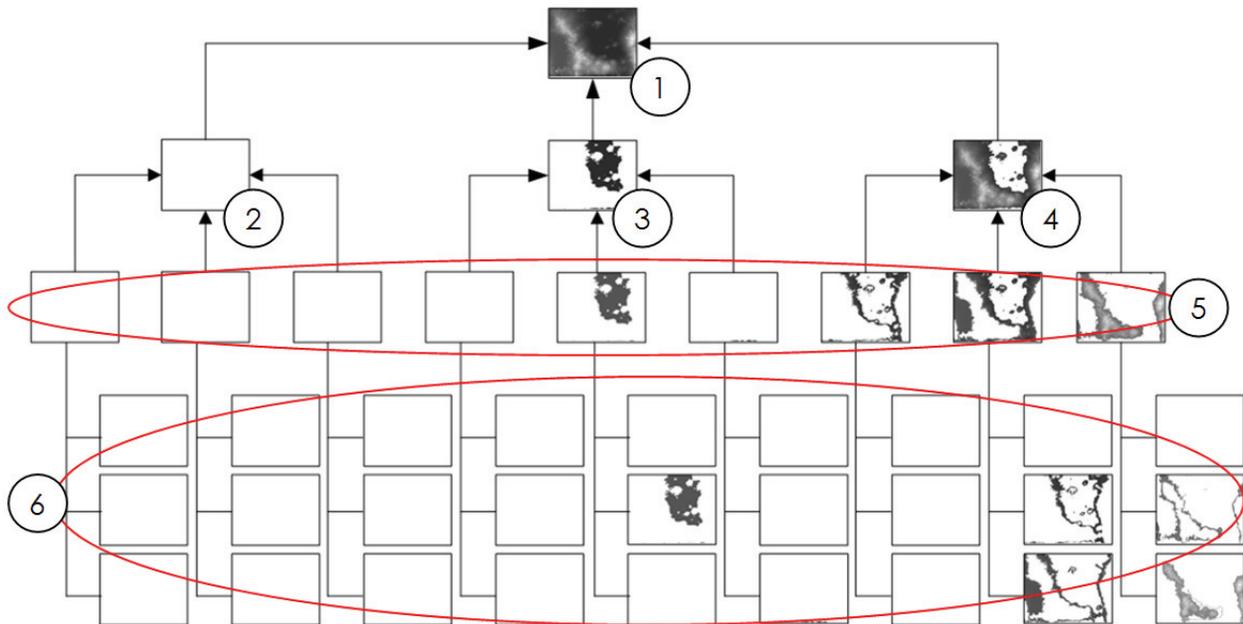


Figura 47. Un ejemplo de jerarquía de extractos

En la parte superior presentamos el DEM original (1), el cual, al pasar a un nivel más abajo, lo encontramos descompuesto en tres modelos de elevación con los segmentos que han sido clasificados como depresiones (2), llanuras (3) y elevaciones (4). Si bajamos otro nivel en la figura vemos la división de cada segmento en depresiones, llanuras y elevaciones (5), y así sucesivamente (6).

El algoritmo de extracción que proponemos comienza con la búsqueda de la Zona Llana más Extensa (ZLE), que se trata de la región o porción del modelo que tiene una altitud más o menos similar y que además es la más grande de todo el modelo. Una vez que se ha encontrado la ZLE, el siguiente paso consiste en etiquetar esa ZLE como "llan"; posteriormente todos los datos del DEM con altitud mayor a la altitud máxima del ZLE se etiquetan con "elev" y los que tienen una altitud menor a la altitud mínima del ZLE se etiquetan como "depr".

Añadiendo una etapa de preprocesamiento, en la cual se realiza un suavizado de los datos, y una de posprocesamiento, en la que se separan los grupos con una misma etiqueta, podemos decir que el algoritmo de extracción consta de cuatro etapas: suavizado (preprocesamiento), cálculo de la ZLE, etiquetado y extracción (posprocesamiento).

5.2.1. Suavizado

Como etapa de preprocesamiento es deseable realizar un suavizado de los datos pues, como se verá más adelante, se utiliza la segunda derivada de los datos para encontrar la ZLE (aplicando un filtro Laplaciano). Al realizar el suavizado de los datos reducimos la sensibilidad de esta derivada a las componentes de alta frecuencia.

Para ejemplificar lo anterior, en la Figura 48 se presenta una captura de pantalla de una aplicación que nos permite analizar el comportamiento de un conjunto de datos y sus primeras dos derivadas ante un suavizado mediante un filtro mediana. En la misma figura se presentan seis gráficas: (1) los datos originales, (2) los datos después del suavizado, (3) la primera derivada, (4) la segunda derivada, (5) el número de cruces por cero de la primera derivada (máximos y mínimos) contra el número de iteración y (6) el número de cruces por cero de la segunda derivada (puntos de inflexión o cambios de concavidad) contra el número de iteración. Además se presenta el número de iteración (paso), el número de cruces por cero de la primera derivada en la iteración actual (cruces dx) y el número de cruces por cero de la segunda derivada en la iteración actual (cruces ddx).

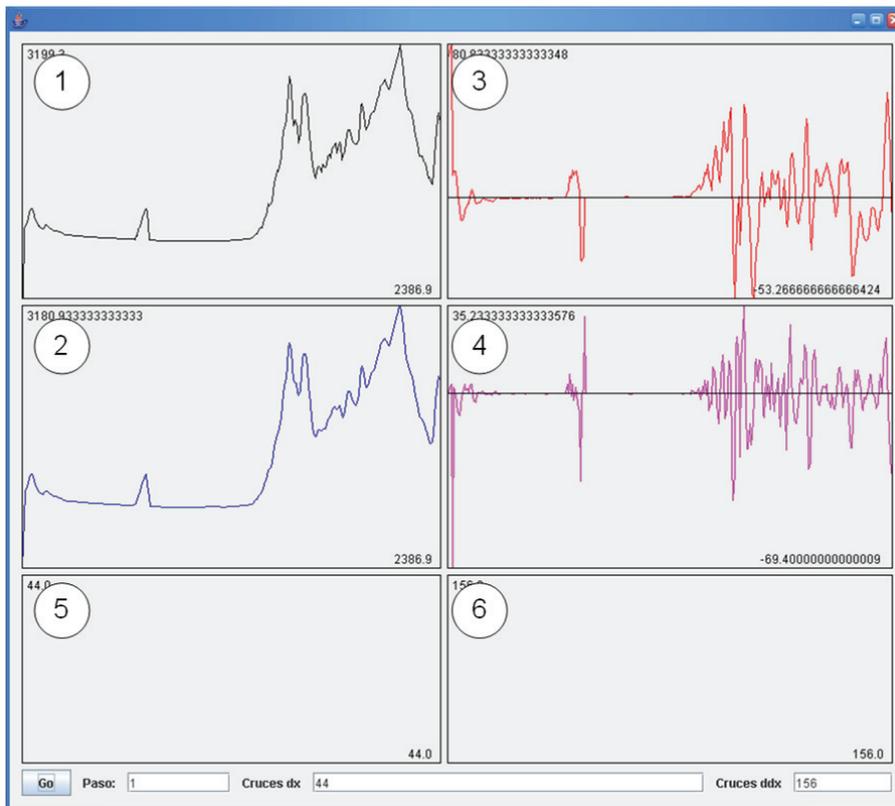


Figura 48. Aplicación de análisis de derivadas, iteración 1

En la Figura 49 se muestra la misma aplicación con el mismo conjunto de datos al haber realizado 30 iteraciones. En el caso de este conjunto de datos, podemos apreciar que el número de cruces por cero de la primera derivada se estabiliza²⁹ hacia un valor de 8, lo que significa que existen ocho características fundamentales en el conjunto de datos. De manera similar, el número de cruces por cero de la segunda derivada converge, en este caso hacia un valor de 23.

²⁹ En un valor que nos permite eliminar los elementos de alta frecuencia (textura) y conservar aquellos que representan las formas del terreno.

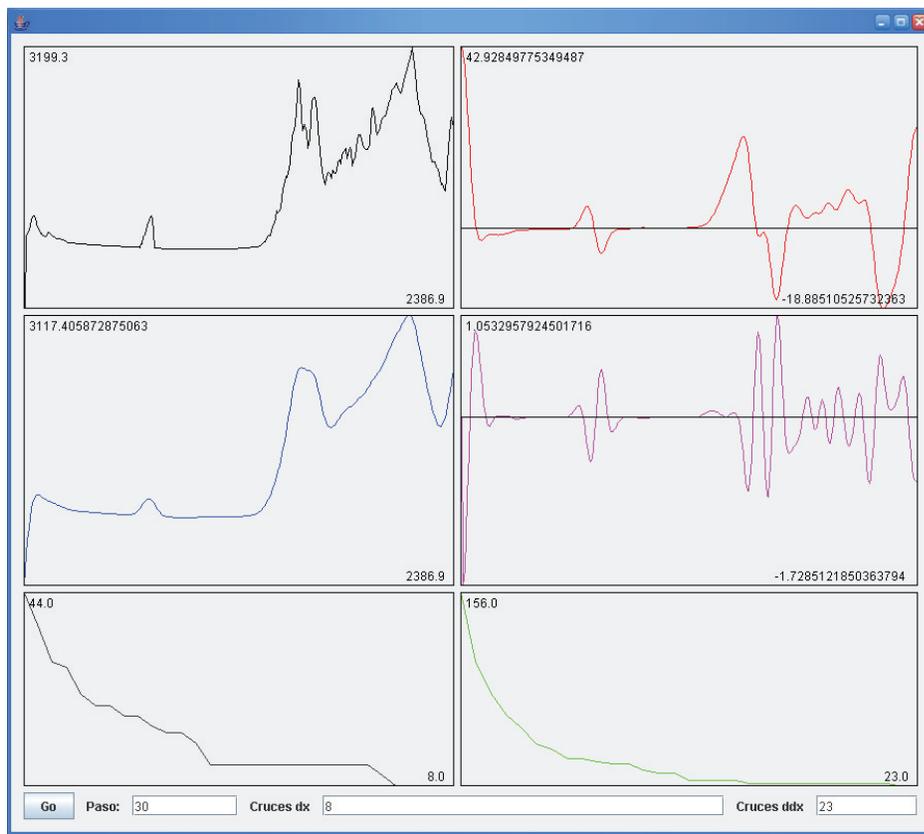


Figura 49. Aplicación de análisis de derivadas, iteración 30

Ahora bien, si siguiéramos iterando encontraríamos que el número de cruces por cero de la primera derivada se mantiene constante durante bastantes iteraciones, lo cual es un indicativo de que el número de veces que se debe suavizar está acotado por este valor de convergencia del número de cruces por cero de la primera derivada, ya que el tamaño del conjunto de datos es finito.

Obviamente, los CDER tienen una dimensión más que los ejemplos que se presentaron arriba, pero la aplicación de análisis de derivadas sirve para ilustrarnos respecto de la importancia del suavizado como etapa de preprocesamiento.

Adelantándonos un poco, en la Figura 50 presentamos los efectos que el suavizado tiene en las etapas posteriores. En tal figura se presenta (1) el DEM original, (2) el DEM suavizado con una máscara de 3x3, (3) el resultado del etiquetado utilizando el DEM original y (4) el resultado del etiquetado utilizando el DEM suavizado. Como se puede

apreciar, la ZLE encontrada es la misma en ambos casos, pero en términos de desempeño es más rápido encontrarla con el uso del DEM suavizado.

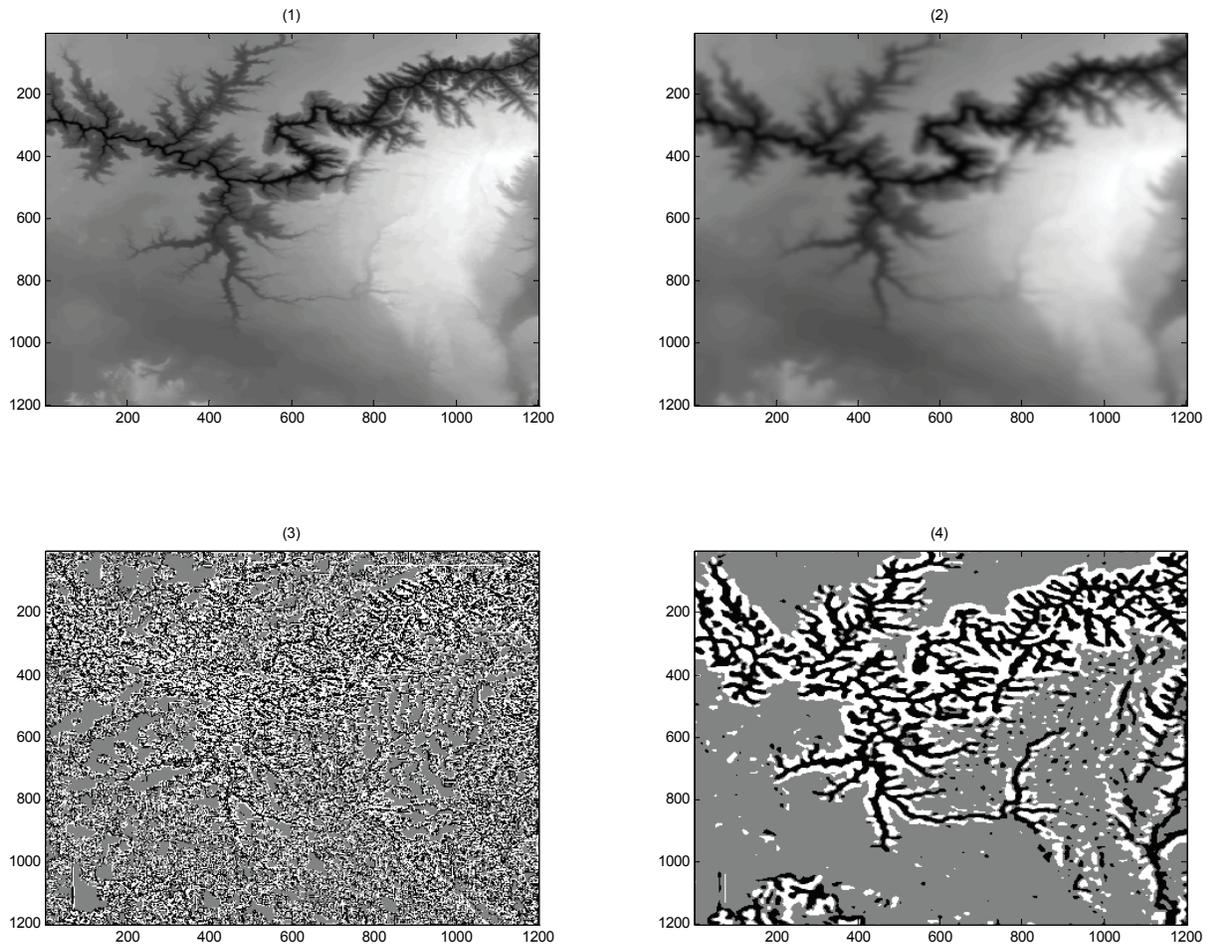


Figura 50. Ejemplo de suavizado y sus efectos en el etiquetado

Aprovechando este ejemplo, caracterizaremos los datos que estamos utilizando para realizar estas pruebas. El ejemplo de la Figura 50 utiliza los datos de elevación correspondientes a la zona del Gran Cañón de Colorado en los Estados Unidos, obtenidos del directorio de modelos digitales de elevación de los Estados Unidos (USGS, 2007b), con una escala de 1:250000. En la Figura 51 se muestra la imagen en escala de grises obtenida a partir del DEM mencionado y en la Figura 52 se muestra el mismo modelo con una vista tridimensional para una mejor apreciación de las formas del terreno. Sus coordenadas extremas en proyección Geográfica son (129600, -406800) y

(133200, -403200) segundos-arco, el número de columnas es de 1201 y el número de filas es de 1201.



Figura 51. Modelo Digital de Elevación de la Zona del Gran Cañón de Colorado, E.U.

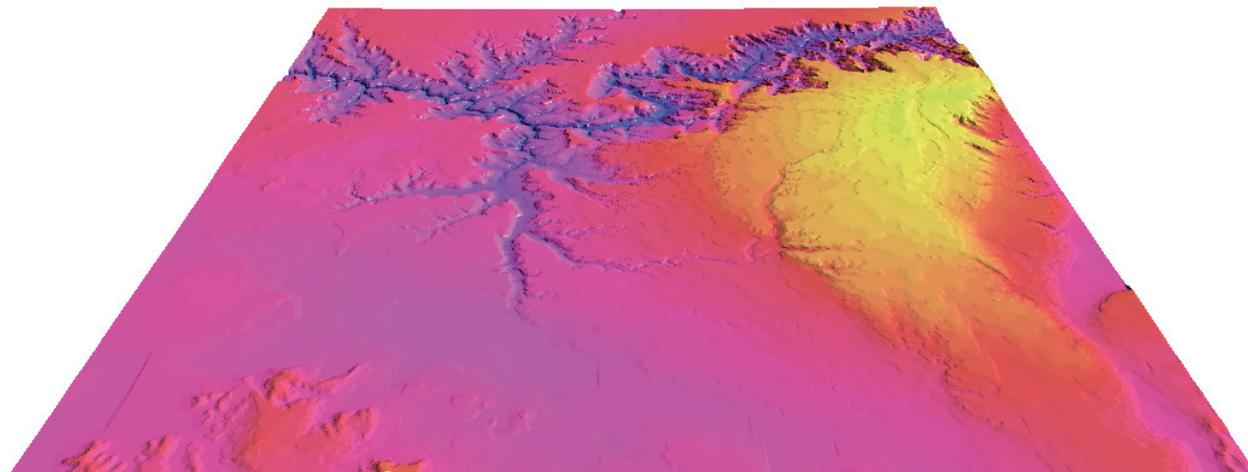


Figura 52. Vista tridimensional del Modelo Digital de Elevación de la Zona del Gran Cañón de Colorado, E.U.

5.2.2. Cálculo de la zona llana más extensa

Para calcular la ZLE utilizaremos, como ya lo habíamos adelantado, la segunda derivada del DEM, pues el valor de ésta nos indica la concavidad o convexidad que tiene el modelo en un punto dado. Si dicho valor es cero, podemos decir que ese sitio es un punto de inflexión, es decir, un punto donde el terreno pasa de ser cóncavo a convexo o viceversa. En otras palabras, un valor de cero en la segunda derivada del DEM nos indica que en esa posición el terreno no se va elevando ni bajando, es decir, se mantiene llano.

Utilizando lo anterior, buscamos la ZLE obteniendo la segunda derivada del DEM. Para ello, le aplicamos a éste un filtro Laplaciano. En la Figura 53 se muestra la imagen de contornos del resultado de aplicar un filtro Laplaciano al DEM de prueba. Esta forma de representar el resultado, sirve también para ilustrar el siguiente paso en esta etapa: encontrar la zona más extensa cuyo valor del Laplaciano sea cero³⁰, para lo cual utilizamos un algoritmo de crecimiento de regiones de 8-conexas.

³⁰ Un valor cercano a cero: en este caso consideramos un valor en el rango de -0.05 a 0.05.

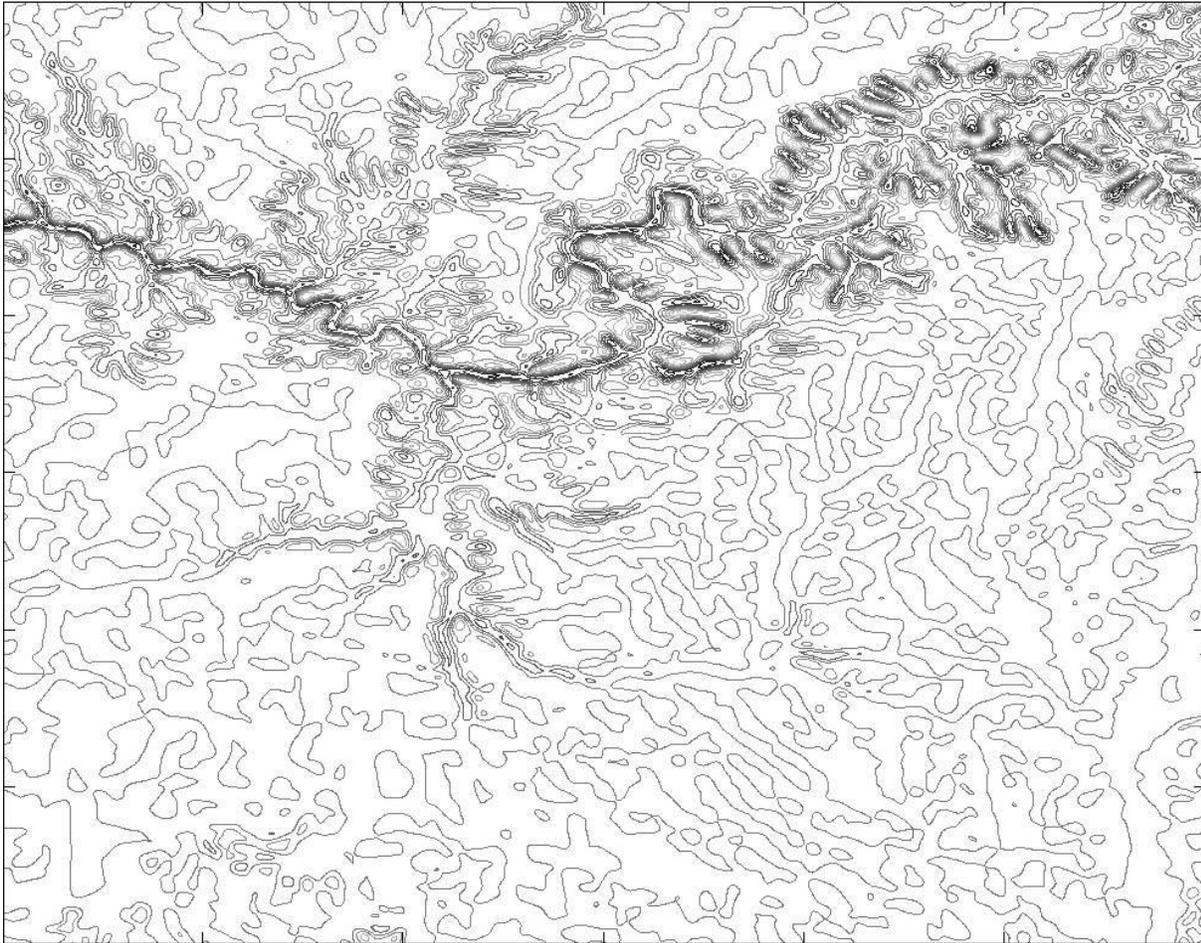


Figura 53. Imagen de los contornos del resultado de aplicar el filtro Laplaciano al DEM

En la Figura 54 se muestra el algoritmo utilizado para buscar la ZLE, en donde podemos ver que, como habíamos mencionado, aplicamos el filtro Laplaciano utilizando (FILTRO-LAPLACIANO) en primer lugar.

BUSCAR-ZLE(*dem*)

1	<i>lap</i> ← FILTRO-LAPLACIANO(<i>dem</i>)
2	<i>zle</i> ←
3	for <i>i</i> =0 to NUM-COLUMNAS(<i>lap</i>)
4	for <i>j</i> =0 to NUM_FILAS(<i>lap</i>)
5	if <i>lap</i> _{<i>ij</i>} = ZERO() then
6	<i>zona</i> ← REG8CONN(<i>lap</i> , <i>i</i> , <i>j</i>)
7	if MAS-GRANDE(<i>zona</i> , <i>zle</i>) then
8	<i>zle</i> ← <i>zona</i>

Figura 54. Algoritmo de búsqueda de la Zona Llana más Extensa

Como consideración inicial, establecemos una zona vacía como la ZLE candidata. Con el resultado del filtro Laplaciano buscamos el primer píxel cuya curvatura podamos considerarla como cero (ZERO), y tomando este píxel como semilla, buscamos la región 8-conexa a la que pertenece dicho píxel utilizando para ello un algoritmo de crecimiento de regiones de conectividad 8 (REG8CONN). Una vez calculada la región, se compara con la ZLE candidata actual, si la nueva región calculada es más grande que ésta, entonces la nueva región calculada pasa a ser la ZLE candidata. En algoritmo continúa hasta haber inspeccionado todos los píxeles. Al finalizar, la ZLE candidata pasa a ser la ZLE del DEM que estamos procesando.

5.2.3. Etiquetado

Una vez que se ha encontrado la ZLE, el siguiente paso consiste en separar los datos de acuerdo a la relación que guardan con ésta. Se obtienen a partir de la ZLE los valores mínimo y máximo de la altitud; a todos los valores del DEM en ese rango se les da la etiqueta de "llan". Los valores que se encuentran por arriba de la altitud máxima de la ZLE se les etiquetan como "elev"; y, finalmente, a los valores por debajo de la altitud mínima de la ZLE se les da la etiqueta de "depr".

En la Figura 55 se muestra el algoritmo de etiquetado utilizado para realizar el proceso que hemos propuesto en el párrafo anterior, en donde *zle* es la zona llana más extensa, *dem* es el modelo digital de elevación que se está procesando y *etq* es el conjunto resultante donde se colocan las etiquetas que se han asignado. Además, NUM-

COLUMNAS y NUM-FILAS son procedimientos que permiten conocer, respectivamente, el número de columnas y de filas que tiene un DEM; ALTITUD-MINIMA y ALTITUD-MAXIMA son procedimientos que permiten conocer, respectivamente, la altitud mínima y máxima de una ZLE.

ETIQUETAR(*zle,dem,etq*)

```

1  for i=0 to NUM-COLUMNAS(dem)
2  for j=0 to NUM_FILAS(dem)
3  if demij< ALTITUD-MINIMA(zle) then
4  etqij←"depr"
5  else if demij> ALTITUD-MAXIMA(zle) then
6  etqij←"elev"
7  else
8  etqij←"llan"

```

Figura 55. Algoritmo de etiquetado

En la Figura 56 se muestra el resultado de este proceso de etiquetado realizado sobre los datos de ejemplo que describimos anteriormente. En ella se muestra el DEM de ejemplo sin modificaciones (1), y el resultado del proceso de etiquetado (2).

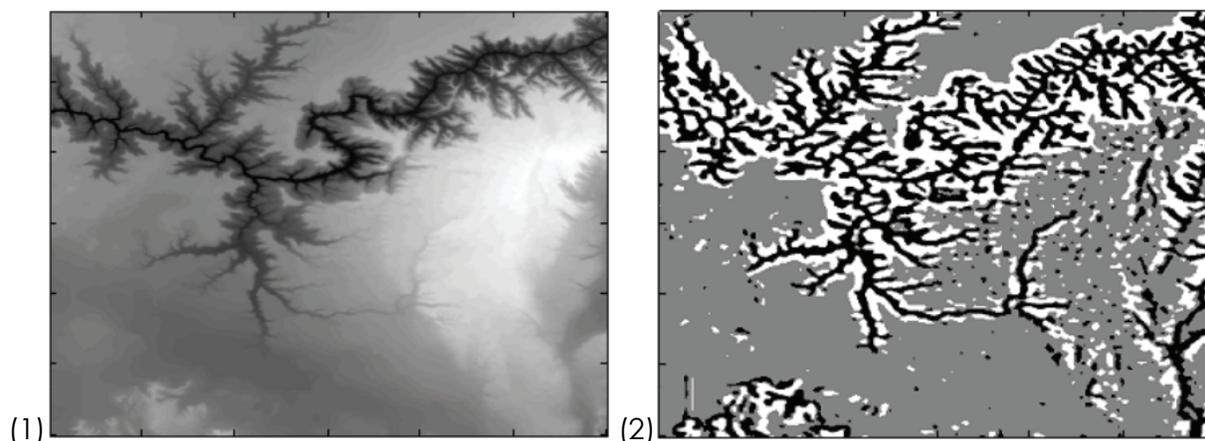


Figura 56. Ejemplo de etiquetado de un DEM

Se muestra el DEM original (1) y el resultado del etiquetado (2) donde solo se aprecian tres valores: negro para "depr", gris para "llan" y blanco para "elev".

5.2.4. Segmentación

El proceso de segmentación consiste en separar del DEM las regiones de acuerdo a la etiqueta que les fue asignada en el proceso anterior, con lo cual obtenemos tres “nuevos” DEM; cada uno con la información correspondiente al tipo de forma relacionada con la etiqueta que se había aplicado. El algoritmo que se utiliza es sencillo, consiste en crear tres DEM con base en el DEM original y en las etiquetas obtenidas en el proceso de etiquetado. En la Figura 57 se muestra el algoritmo para la segmentación, donde, *dem* y *etq* se utilizan en el mismo sentido que ya se ha explicado; *elev*, *llan* y *depr* son los conjuntos de datos de salida. Además, NO-DATA es un procedimiento que inicializa un conjunto de datos con un *token* especial llamado de la misma manera³¹.

SEGMENTAR(*dem,etq,elev,llan,depr*)

```

1  elev←NO-DATA()
2  llan←NO-DATA()
3  depr←NO-DATA()
4  for i=0 to NUM-COLUMNAS(dem)
5  for j=0 to NUM-FILAS(dem)
6  if etqij="elev" then
7    elevij←demij
8  else if etqij="llan" then
9    llanij←demij
10 else if etqij="depr" then
11  deprij←demij

```

Figura 57. Algoritmo para la segmentación

En la Figura 58 y en la Figura 59 se muestra el resultado de la aplicación de este proceso al conjunto de datos presentado en la Figura 51. En la primera se muestran en tonos rojos lo que se ha clasificado como elevaciones, en tonos verdes, lo que se ha

³¹ Es común que en los DEM se utilice este *token* que significa que una celda del conjunto de datos no contiene un valor válido. Esto se debe a que algunas veces, dependiendo de la proyección, la zona y el tamaño del modelo, el conjunto de datos no forma un rectángulo sino un romboide o trapezoide, por lo cual es necesario rellenar los huecos que se generan al modelarlo como una matriz.

clasificado como llanuras y en tonos azules se muestran las zonas clasificadas como depresiones. En la Figura 59 se pueden apreciar la ZLE y los segmentos de cada una de las firmas para un primer nivel de ejecución ("elev", "llan" y "depr"). En la sección de anexos se puede encontrar el programa escrito en *MATLAB*, correspondiente al proceso de extracción.

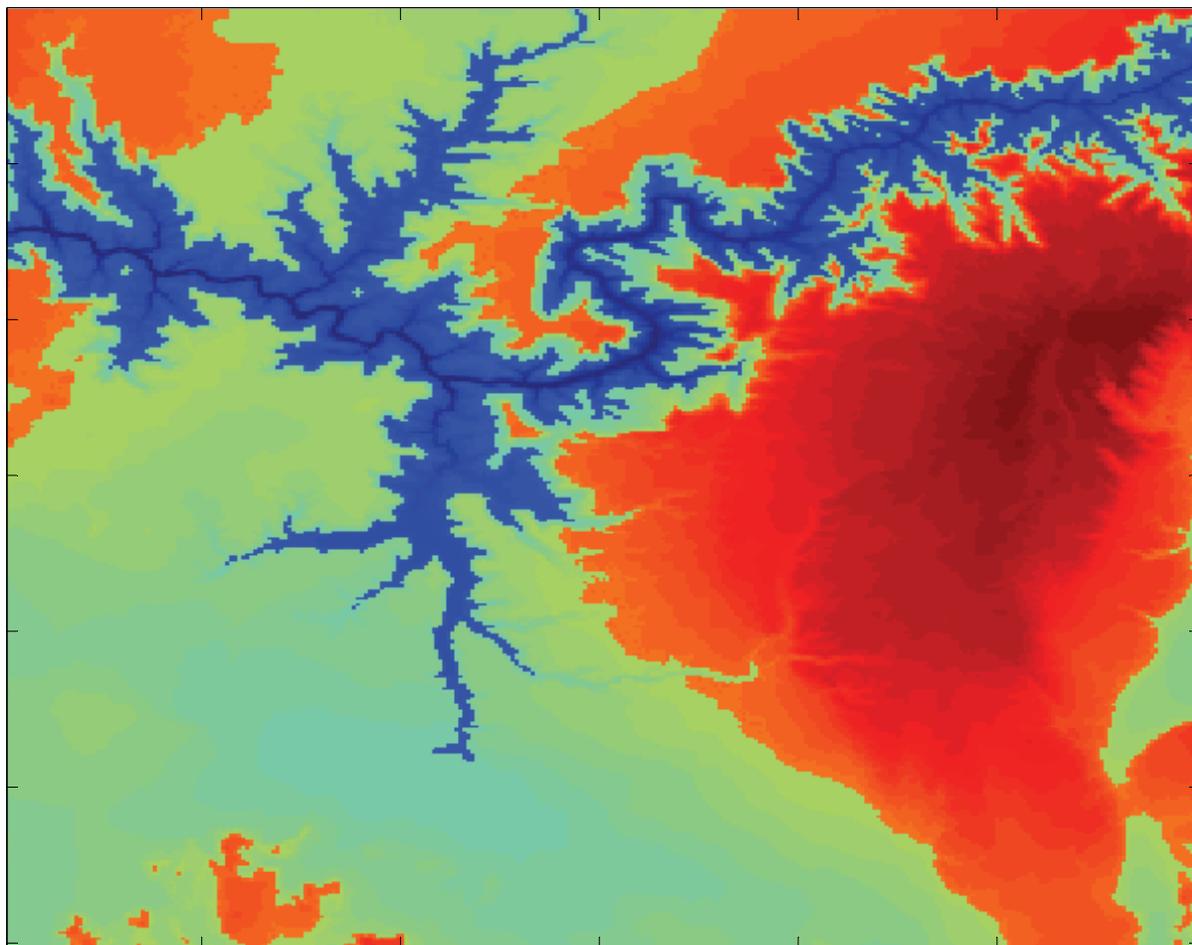


Figura 58. Resultados del proceso de segmentación aplicado al DEM original

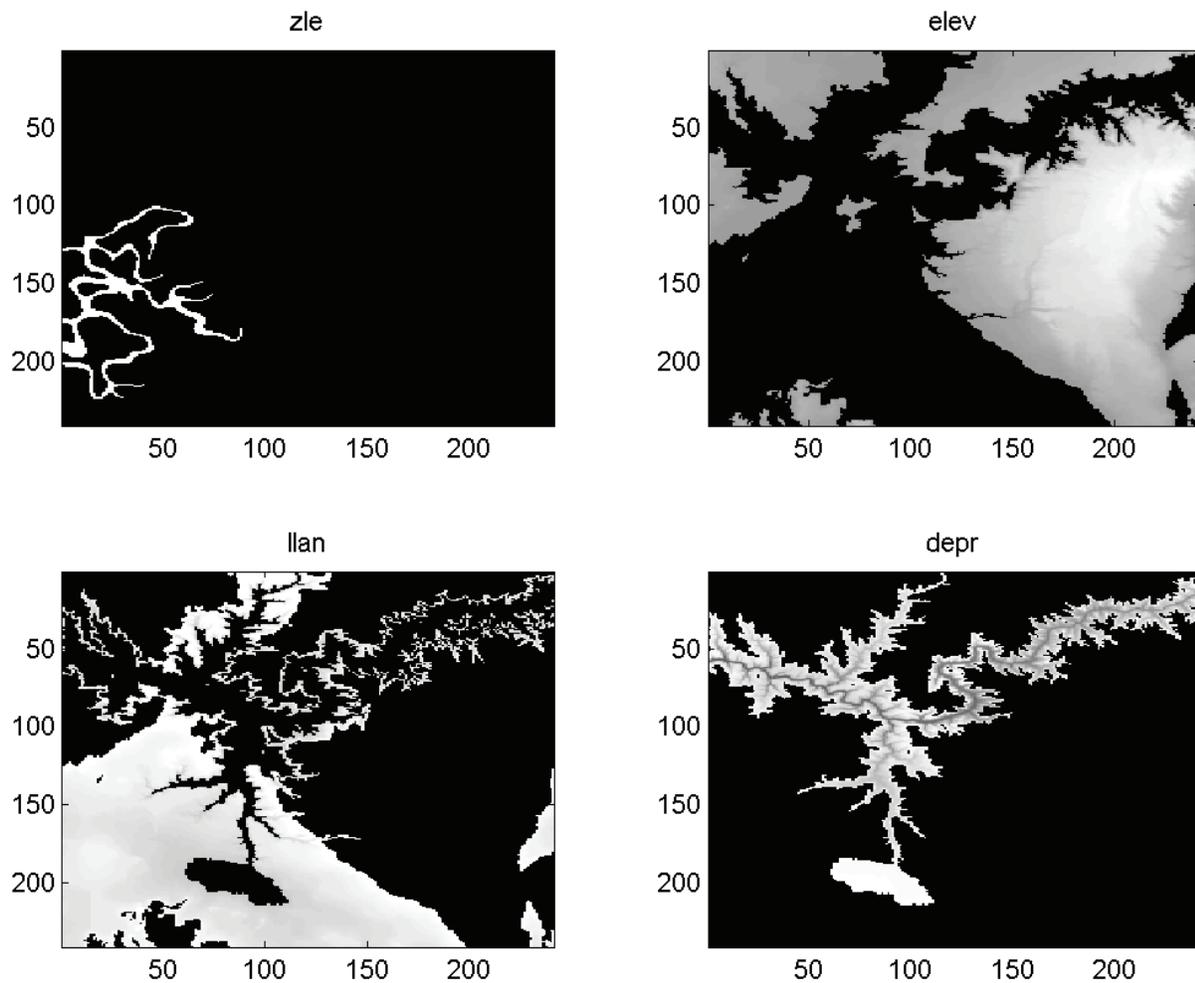


Figura 59. Resultados por separado del proceso de segmentación aplicado al DEM original

Este proceso de segmentación es aplicado nuevamente a cada segmento que se ha obtenido, es decir se aplica a los segmentos "elev", "llan" y "depr" (marcados con el mismo nombre en la Figura 59) para obtener los segmentos "elev-elev", "elev-llan" y "elev-depr" a partir del segmento "elev" (Figura 60); obtenemos los segmentos "llan-elev", "llan-llan" y "llan-depr" al aplicar el proceso al segmento "llan" (Figura 61). Los segmentos "depr-elev", "depr-llan" y "depr-depr" son obtenidos al aplicar el proceso de segmentación al segmento "depr" (Figura 62).

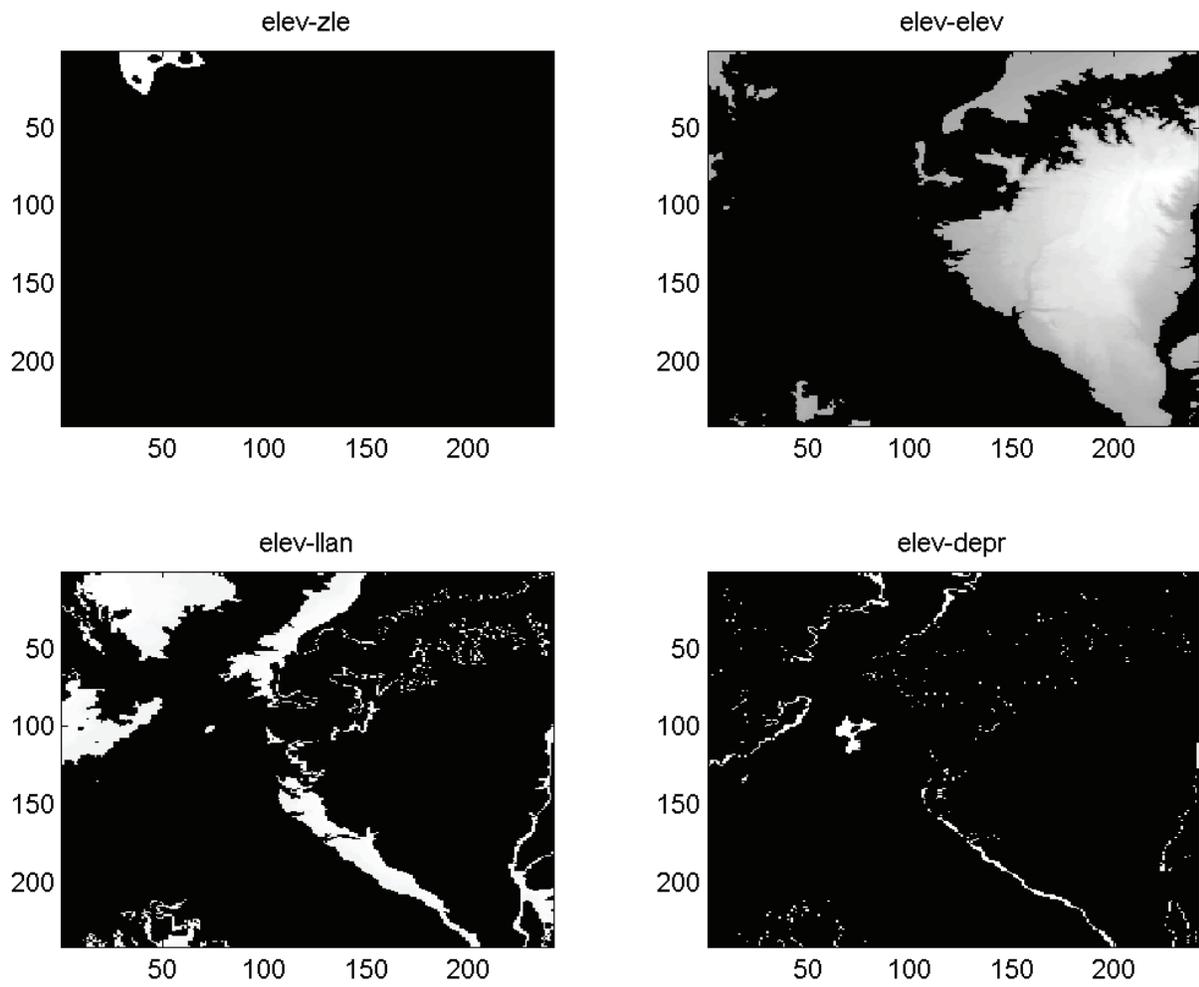


Figura 60. Resultados del proceso de segmentación aplicado al segmento "elev"

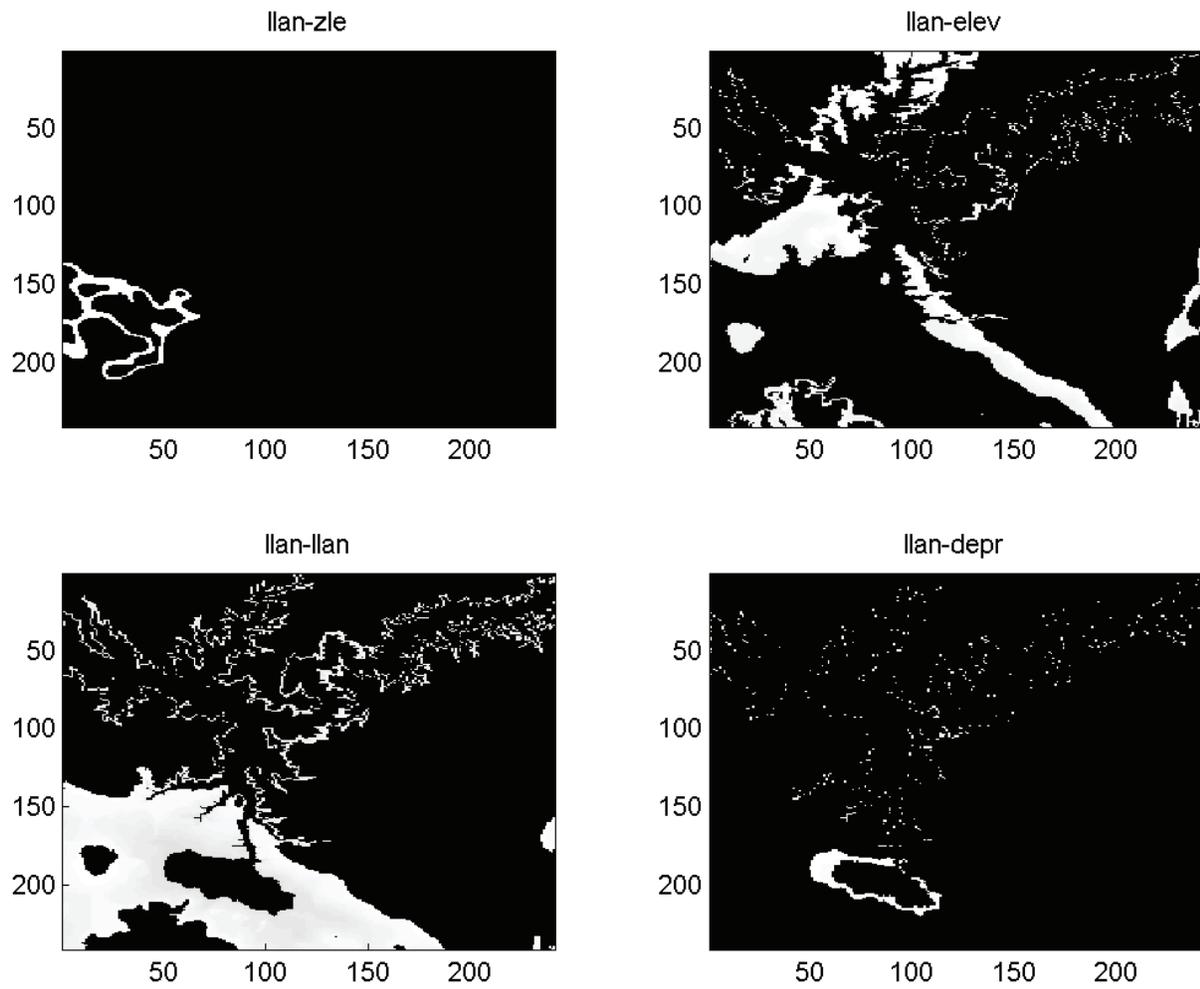


Figura 61. Resultados del proceso de segmentación aplicado al segmento "llan"

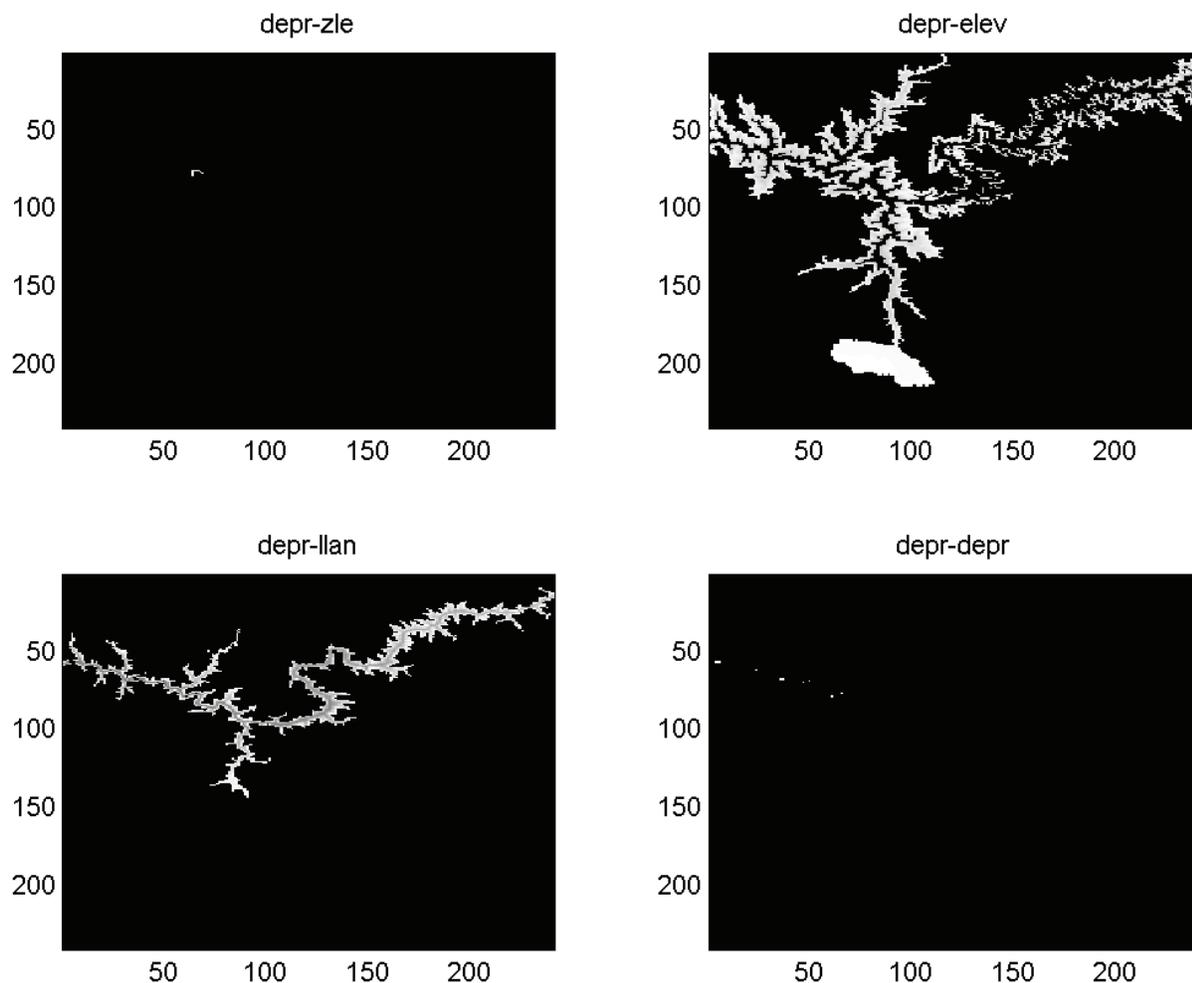


Figura 62. Resultados del proceso de segmentación aplicado al segmento “depr”

Conforme a lo que hemos mencionado en la conceptualización, existen más formas del terreno que podemos considerar como elevaciones, por lo que el proceso de segmentación se aplica una vez más a los segmentos de elevación que hemos obtenido. Así pues, obtendremos los segmentos “elev-elev-elev”, “elev-elev-llan” y “elev-elev-depr” al segmentar “elev-elev” (Figura 63); segmentando “elev-llan” obtenemos los segmentos “elev-llan-elev”, “elev-llan-llan” y “elev-llan-depr” (Figura 64). Finalmente, los segmentos “elev-depr-elev”, “elev-depr-llan” y “elev-depr-depr” se obtienen al aplicar el proceso de segmentación al segmento “elev-depr” (Figura 65).

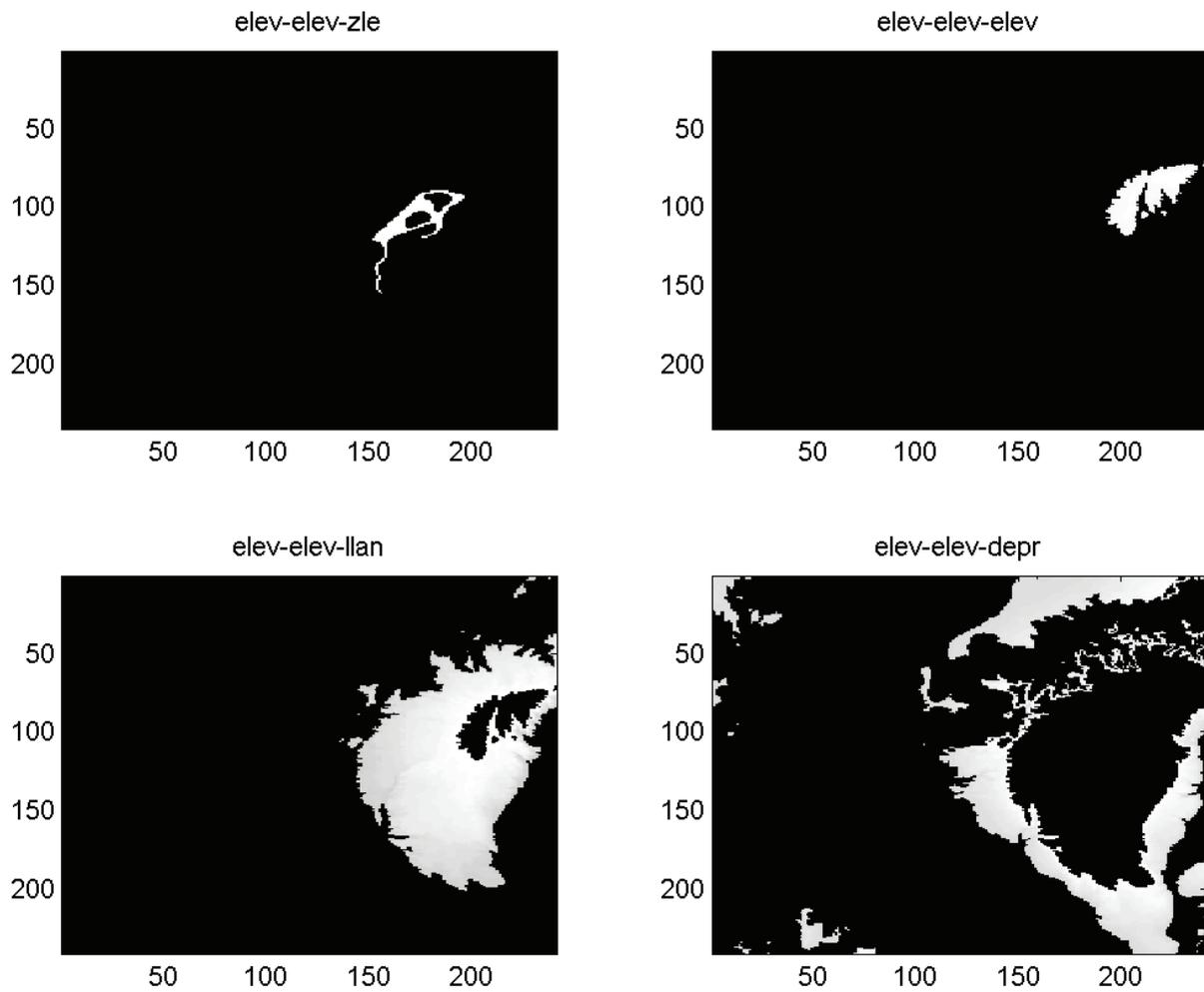


Figura 63. Resultados del proceso de segmentación aplicado al segmento "elev-elev"

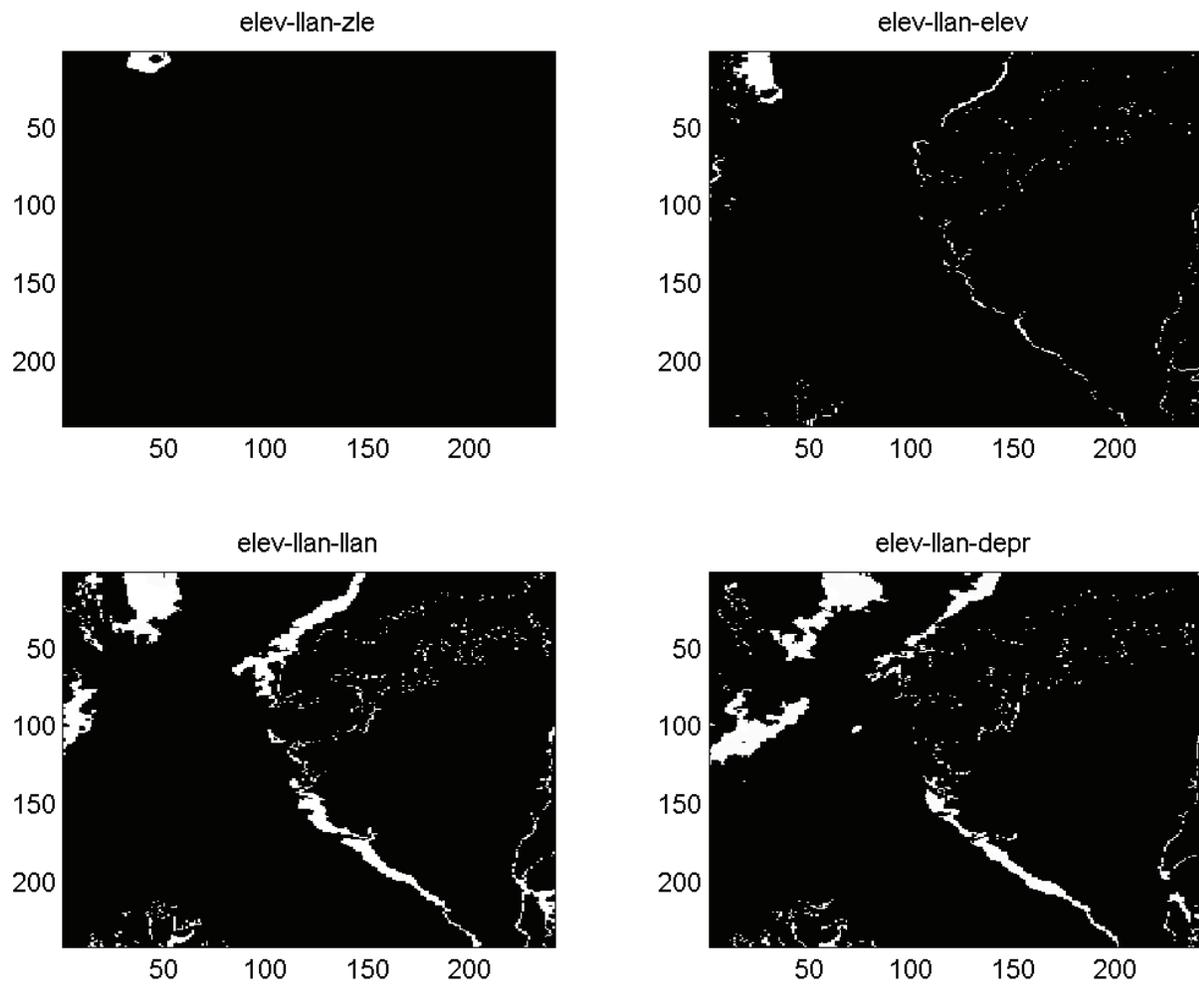


Figura 64. Resultados del proceso de segmentación aplicado al segmento "elev-llan"

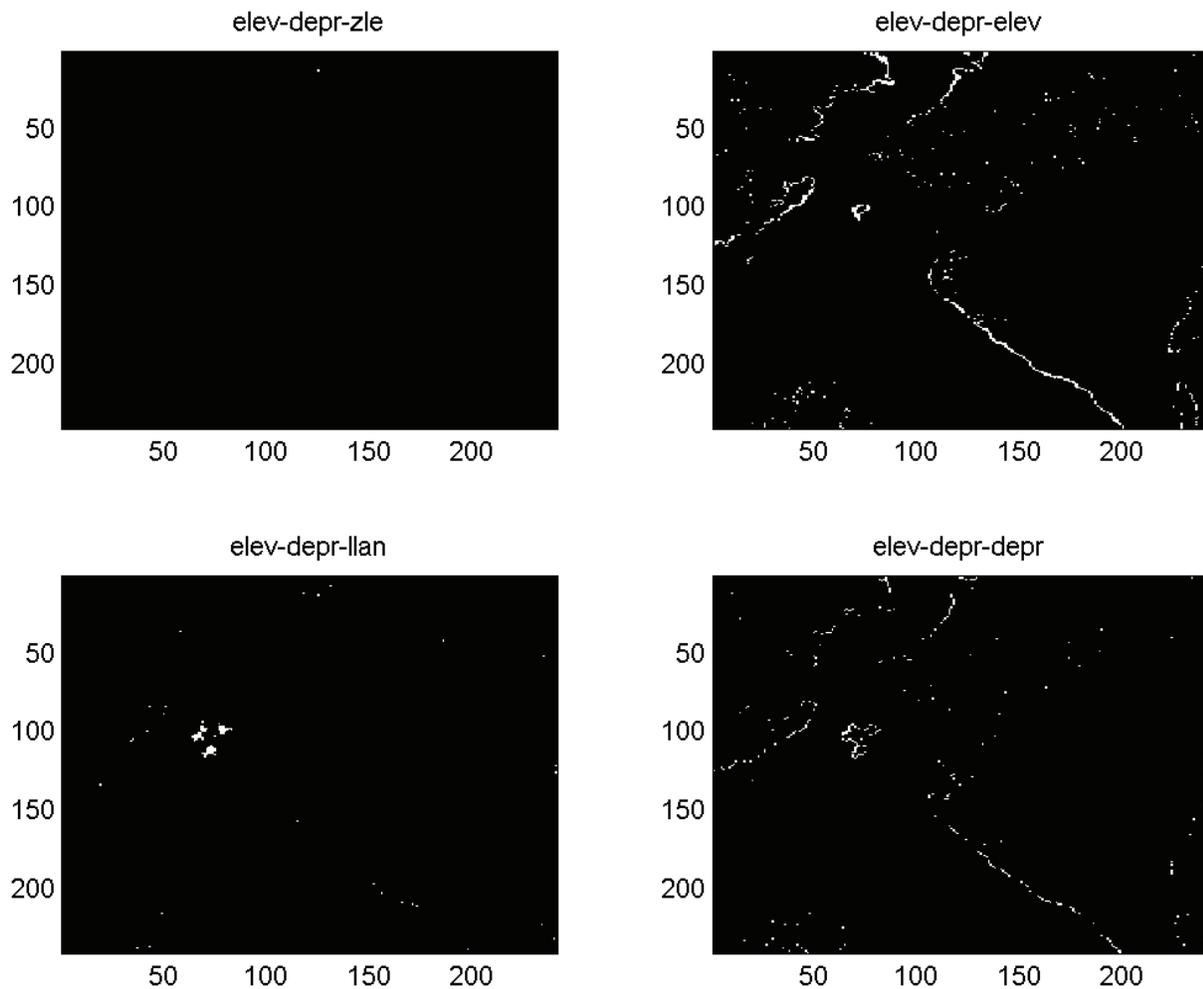


Figura 65. Resultados del proceso de segmentación aplicado al segmento “elev-depr”

5.2.5. Extracción

La parte final del proceso de síntesis consiste en obtener los extractos que resultan. Es decir, obtener de un modelo que ya ha sido etiquetado y segmentado, los diferentes conglomerados que se encuentren bajo la misma firma, es decir que se encuentran en el mismo CDER segmentado, como los que hemos mostrado en la sección anterior. En la Figura 66 se presenta el algoritmo de extracción que proponemos.

EXTRAER(*dem*)

```

1  ext←
2  nExt←0
3  for i=0 to NUM-COLUMNAS(dem)
4  for j=0 to NUM_FILAS(dem)
5  if dem(i,j)>0 then
6  nExt++
7  e←REG8CONN(i,j,nExt)
8  MARCA(e, nExt)
9  ext←ext U {e}

```

Figura 66. Algoritmo para la extracción

Dicho algoritmo consiste en buscar los conglomerados de píxeles que tengan datos válidos (mayor que cero). Existe un conjunto de extractos encontrados *ext* que inicialmente se encuentra vacío (1). La parte iterativa del algoritmo comienza con la búsqueda del primer píxel con un dato válido (3, 4 y 5) y se usa como semilla para un algoritmo de crecimiento de regiones basado en una vecindad de orden 8 llamado REG8CONN (7). La región encontrada es marcada, mediante la función MARCA, como extracto con un identificador único y consecutivo, con lo que se elimina de la búsqueda (8). Después, el extracto se agrega a la colección de extractos encontrados bajo la misma firma (9). El algoritmo continúa buscando hasta que ya no haya píxeles con datos válidos. Al finalizar el algoritmo, contamos con la colección de extractos encontrados bajo la misma firma.

En la Figura 67 se muestra un ejemplo de los extractos que se obtienen utilizando el algoritmo descrito. Por ejemplo, en la firma "elev-elev-elev" (identificado como "ext-eee") se encuentran cuatro extractos, identificados del -1 al -4³². En el caso de la firma "elev-elev-depr" (identificado como "ext-eed") se localizan 50 extractos.

³² Los identificadores se muestran negativos por cuestiones de la implementación.

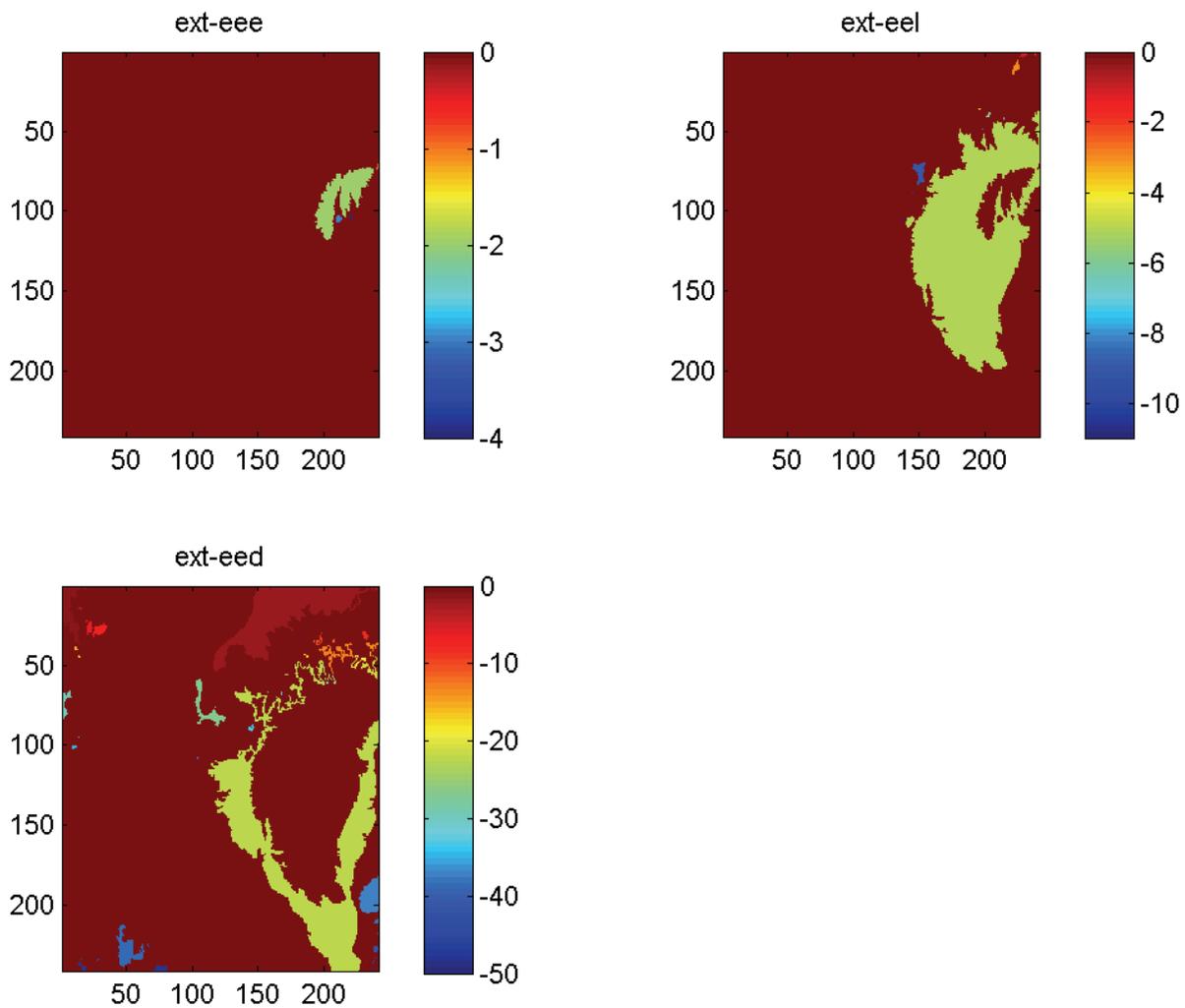


Figura 67. Extractos obtenidos de las firmas “elev-elev-elev”, “elev-elev-llan” y “elev-elev-depr”

Se puede apreciar que algunos extractos pueden llegar a ser muy pequeños, por lo que sería útil contar con algún criterio para eliminar aquellos que por su escaso tamaño puedan considerarse como ruido. Sin embargo, en este punto no es posible discriminarlos de forma alguna sin tomar en cuenta la resolución espacial de los datos, pues en escalas grandes los objetos pueden tener un tamaño pequeño relativo al tamaño del CDER, es decir, en cierta escala una montaña puede ocupar una buena porción de los datos, mientras que en una escala más grande, la misma montaña podría apreciarse como un simple píxel. Lo anterior queda fuera de los alcances del presente trabajo, sin embargo, en el siguiente capítulo se presentan algunas ideas de lo que se puede hacer para lidiar con este y otros problemas relacionados con la

resolución espacial de los datos. Por el momento, solo aclaramos que todos los segmentos que son encontrados por el algoritmo, son considerados en la descripción. En las ilustraciones que van de la Figura 68 a la Figura 71 se muestran los resultados del algoritmo de extracción aplicado a los segmentos obtenidos en etapa de segmentación y que han sido mencionados en párrafos anteriores.

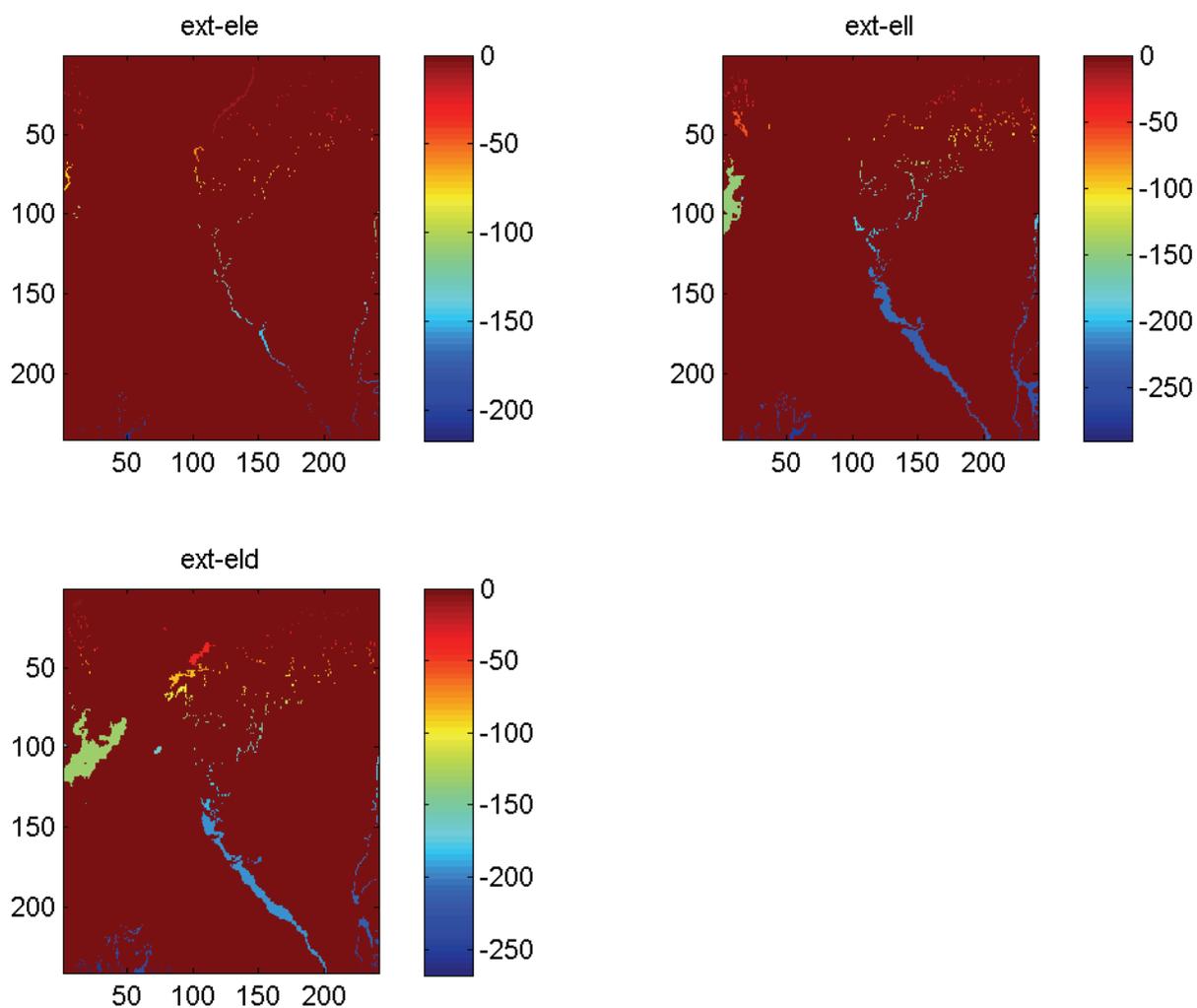


Figura 68. Extractos obtenidos de las firmas “elev-llan-elev”, “elev-llan-llan” y “elev-llan-depr”

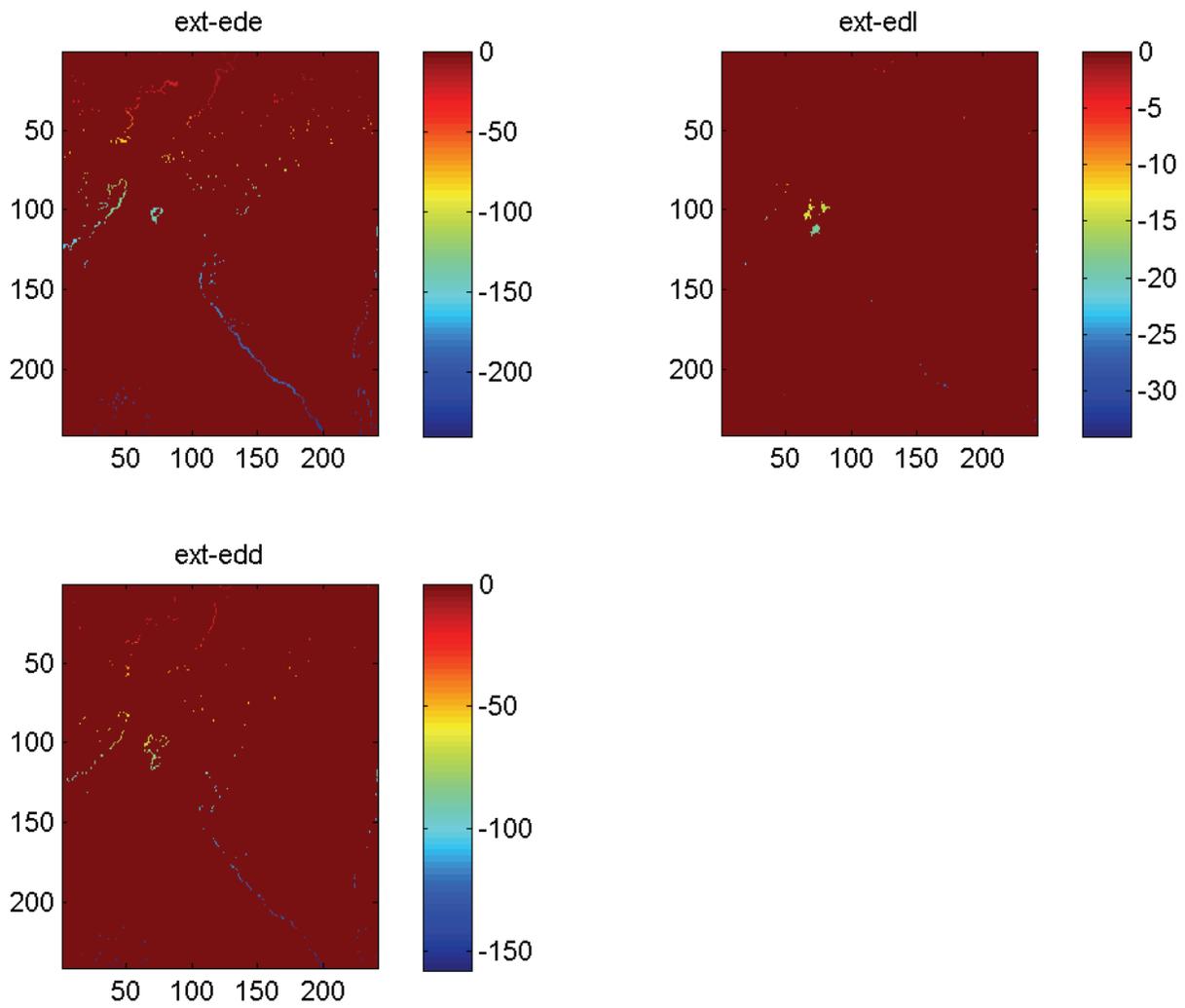


Figura 69. Extractos obtenidos de las firmas “elev-depr-elev”, “elev-depr-llan” y “elev-depr-depr”

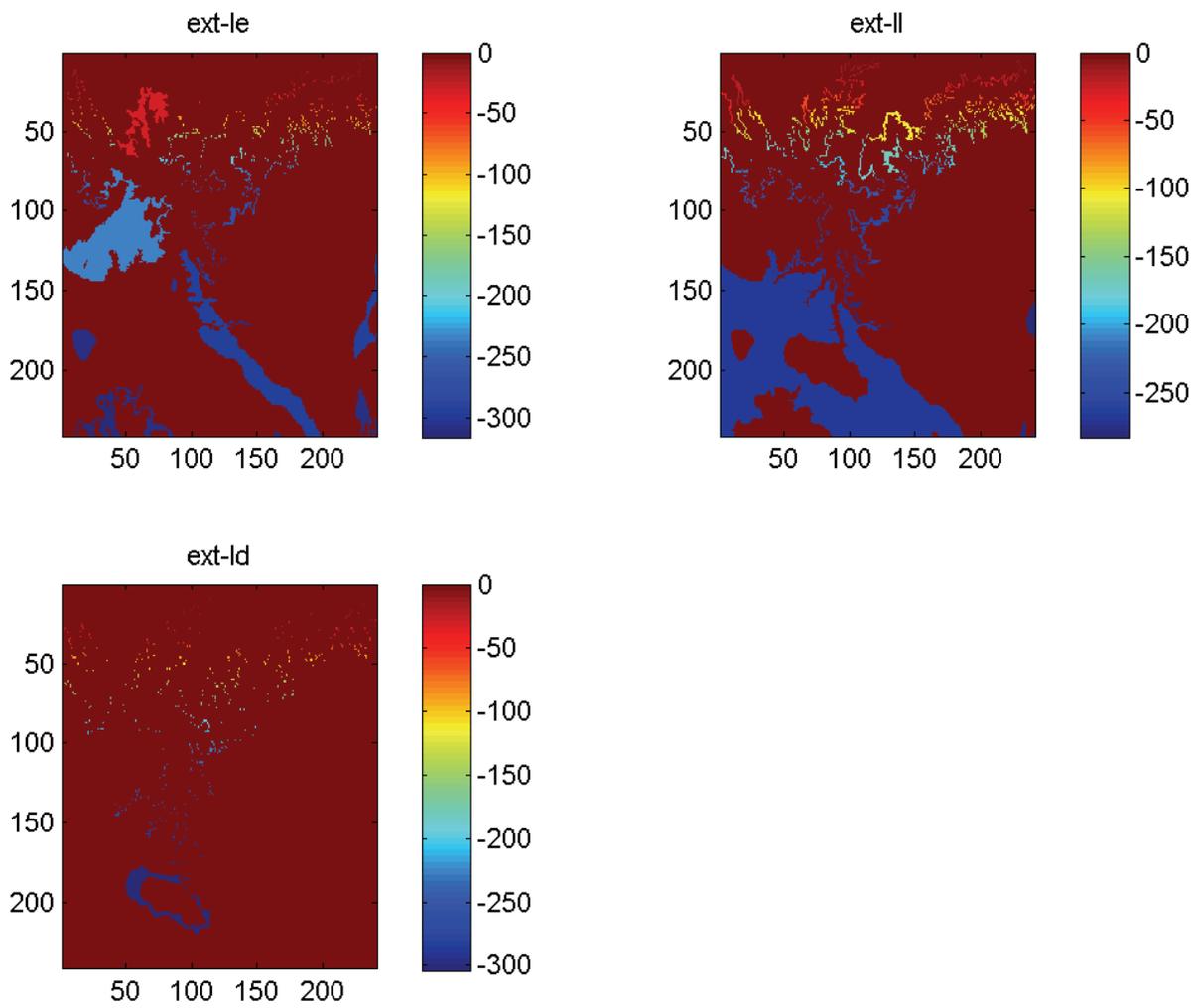


Figura 70. Extractos obtenidos de las firmas "llan-elev", "llan-llan" y "llan-depr"

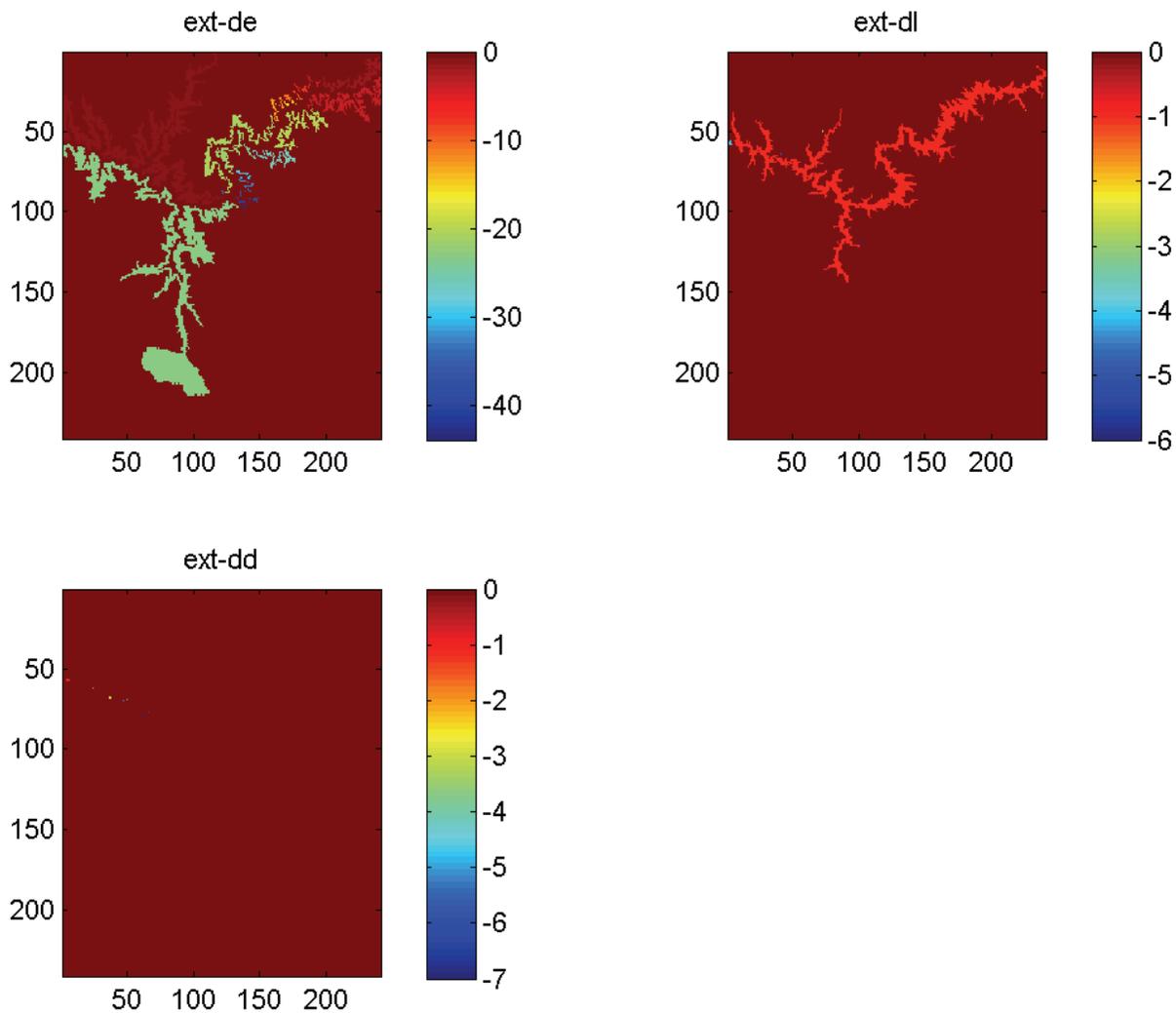


Figura 71. Extractos obtenidos de las firmas “depr-elev”, “depr-llan” y “depr-depr”

En cada una de las figuras anteriores se muestra de un color distinto cada extracto. Esto se hace sólo con fines ilustrativos, pues lo que se obtiene es un CDER que tiene solo un objeto identificado bajo la firma del segmento del cual fue extraído. Como ejemplo, en la Figura 72 se muestra la gráfica tridimensional del extracto más grande de la firma “elev-elev-llan”.

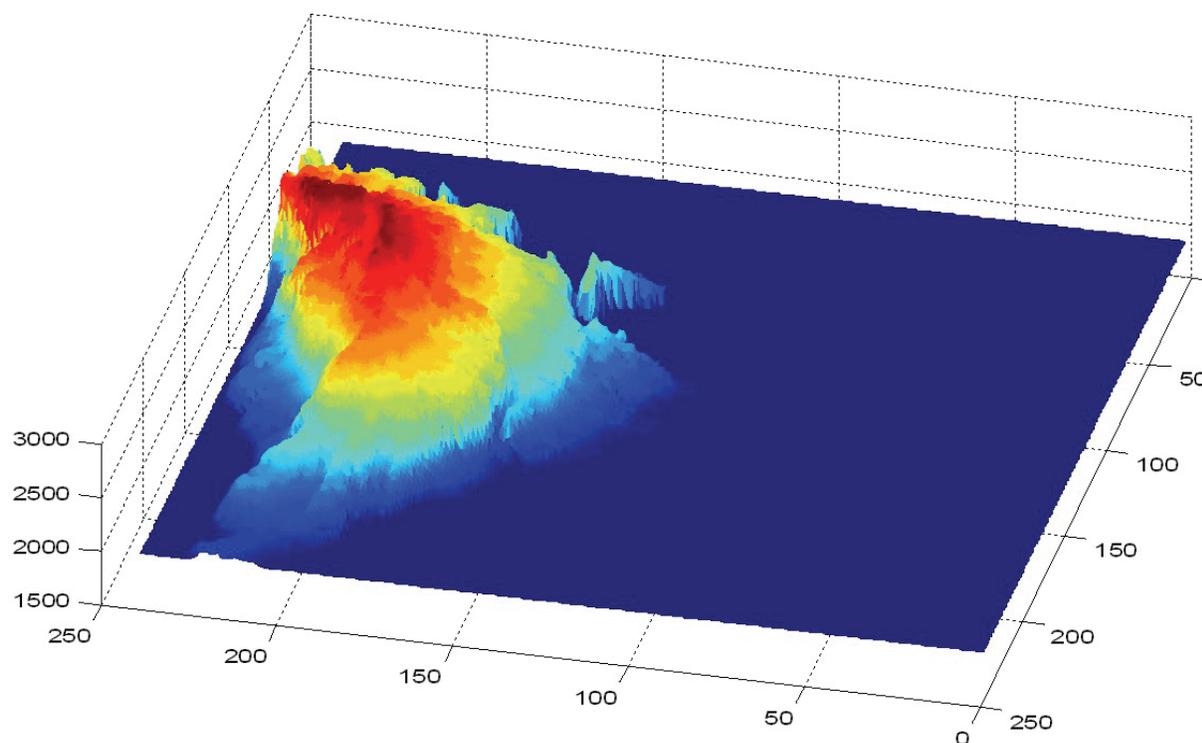


Figura 72. Gráfica tridimensional de un extracto de la firma "elev-elev-llan"

5.3. Descripción

Una vez obtenidos los extractos el siguiente paso consiste en describir cada uno de ellos. El proceso de descripción se encarga de esta tarea, la cual dividimos en dos etapas: la medición y la descripción en sí. En la Figura 73 se presenta el algoritmo para la descripción de un extracto, en donde *ext* y *firma* son el extracto y su firma respectivamente. Además, *OBT-PROPS* es un procedimiento que determina, de acuerdo a la conceptualización y a la firma, las propiedades que corresponden al extracto en cuestión (1). El procedimiento *MEDIR-PROPS* calcula las propiedades directamente de los valores numéricos del extracto (2). *OBT-PLANT* es un procedimiento que arroja la plantilla correspondiente para describir un extracto con una firma dada (3). Finalmente, la plantilla mencionada se llena con los valores que se han obtenido de la medición por medio del procedimiento *LLENAR-PLANT*, el cual arroja la descripción correspondiente al extracto (4).

Un par de aclaraciones caben en este punto: la primera y la más importante es que no se realiza la reasignación descrita en la sección 4.4, primeramente por cuestiones técnicas, ya que se han programado los algoritmos utilizando *MATLAB* y la conceptualización se ha realizado utilizando la herramienta *Protegé*, herramientas que hasta la fecha no es posible hacer interoperar (automáticamente); en segundo lugar porque es necesario (en algunos casos) conocer las relaciones topológicas y/o geométricas que los extractos guardan con relación de los demás.

La segunda aclaración es que, por la cuestión técnica ya mencionada, tanto las propiedades como las plantillas para la descripción son fijas, obtenidas a partir de la conceptualización, pero fijas en la implementación.

DESCRIPCION(*ext, firma*)

```

1  props←OBT-PROPS(firma)
2  vals←MEDIR-PROPS(ext,props)
3  plant←OBT-PLANT(firma)
4  desc←LLENAR-PLANT(plant,vals)
5  return desc
    
```

Figura 73. Algoritmo para realizar la descripción de un extracto

En la Figura 74 se muestra el resultado de la descripción de los extractos obtenidos en las etapas anteriores. En la sección de anexos se puede encontrar el programa escrito en *MATLAB* correspondiente al proceso de descripción.

```

1  CONJUNTO DE DATOS ESPACIALES RASTER DEL MODELO DIGITAL DE ELEVACIÓN DE "GRAND CANYON - E           AZ           " QUE TIENE
2  UNA RESOLUCION ESPACIAL DE 30 SEGUNDOS-ARCO. SU ALTITUD MINIMA ES DE 548.000000 METROS Y SU ALTITUD MAXIMA ES DE 2838.000000
3  METROS. LAS COORDENADAS EXTREMAS SON (129600.000000, -406800.000000) Y (133200.000000, -403200.000000) SEGUNDOS-ARCO EN
4  PROYECCION GEOGRAFICA Y CONTIENE:
5  UNA MONTAÑA CON AREA DE 168300 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 2699.301185 METROS Y MÁXIMA DE 2774.861966 METROS,
6  SUS COORDENADAS EXTREMAS SON (132540.000000, -405660.000000) Y (136680.000000, -401460.000000) SEGUNDOS-ARCO. SU CIMA SE
7  ENCUENTRA EN LA COORDENADA (132930.000000, -405540.000000) A UNA ALTITUD DE 2774.861966 METROS
8  ...
9  UNA CORDILLERA CON AREA DE 9000 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 2690.729503 METROS Y MÁXIMA DE 2699.001900 METROS,
10 SUS COORDENADAS EXTREMAS SON (132720.000000, -405390.000000) Y (136500.000000, -401520.000000) SEGUNDOS-ARCO. SU CIMA SE
11 ENCUENTRA EN LA COORDENADA (132750.000000, -405180.000000) A UNA ALTITUD DE 2699.001900 METROS
12 ...
13 UN MONTE CON AREA DE 96300 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 2656.953259 METROS Y MÁXIMA DE 2690.657480 METROS, SUS
14 COORDENADAS EXTREMAS SON (132480.000000, -405720.000000) Y (136740.000000, -401310.000000) SEGUNDOS-ARCO. SU CIMA SE
15 ENCUENTRA EN LA COORDENADA (132570.000000, -405060.000000) A UNA ALTITUD DE 2690.657480 METROS
16 ...
17 UN CERRO CON AREA DE 7200 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 2640.503674 METROS Y MÁXIMA DE 2650.418398 METROS, SUS
18 COORDENADAS EXTREMAS SON (133050.000000, -405750.000000) Y (136800.000000, -402090.000000) SEGUNDOS-ARCO. SU CIMA SE
19 ENCUENTRA EN LA COORDENADA (133170.000000, -405720.000000) A UNA ALTITUD DE 2650.418398 METROS
20 ...
21 UN ALTIPLANO CON AREA DE 150300 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 2602.933170 METROS Y MÁXIMA DE 2640.289818 METROS,
    
```

22	SUS COORDENADAS EXTREMAS SON (132390.000000, -405810.000000) Y (136830.000000, -401070.000000) SEGUNDOS-ARCO. SU CIMA SE
23	ENCUENTRA EN LA COORDENADA (132480.000000, -404970.000000) A UNA ALTITUD DE 2640.289818 METROS
24	...
25	UNA COLINA CON AREA DE 10800 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 2580.942524 METROS Y MÁXIMA DE 2601.143053 METROS,
26	SUS COORDENADAS EXTREMAS SON (132840.000000, -405840.000000) Y (136710.000000, -402150.000000) SEGUNDOS-ARCO. SU CIMA SE
27	ENCUENTRA EN LA COORDENADA (133050.000000, -405810.000000) A UNA ALTITUD DE 2601.143053 METROS
28	...

Figura 74. Parte de la descripción obtenida a partir del DEM de ejemplo

De esta manera, se realiza la descripción de los extractos contenidos en un DEM, lo cual a su vez constituye la representación semántica del modelo de elevación.

5.4. Pruebas adicionales

A continuación se presentan más pruebas realizadas de los algoritmos implementados. En la sección de anexos se pueden encontrar más imágenes de las pruebas realizadas con otros conjuntos de datos.

5.4.1. Pruebas con el DEM del Valle de México

Datos de elevación correspondientes a la zona del Valle de México y Valle de Toluca generados a partir de los datos vectoriales del INEGI de escala 1:50000, correspondientes a las cartas E14A28, E14A29, E14A38, E14A39, E14A48, E14A49, E14B21, E14B31 y E14B41; con una resolución espacial de 100 metros. En la Figura 75 se muestra la imagen en escala de grises obtenida a partir del DEM mencionado, mientras que en la Figura 76 se presenta la vista tridimensional del mismo conjunto de datos. Sus coordenadas extremas en proyección UTM zona 14 son (424569.35, 2094463.61701), el número de columnas es de 1158 y el número de filas es de 956, abarcando una zona de 115.8 x 95.6 kilómetros.

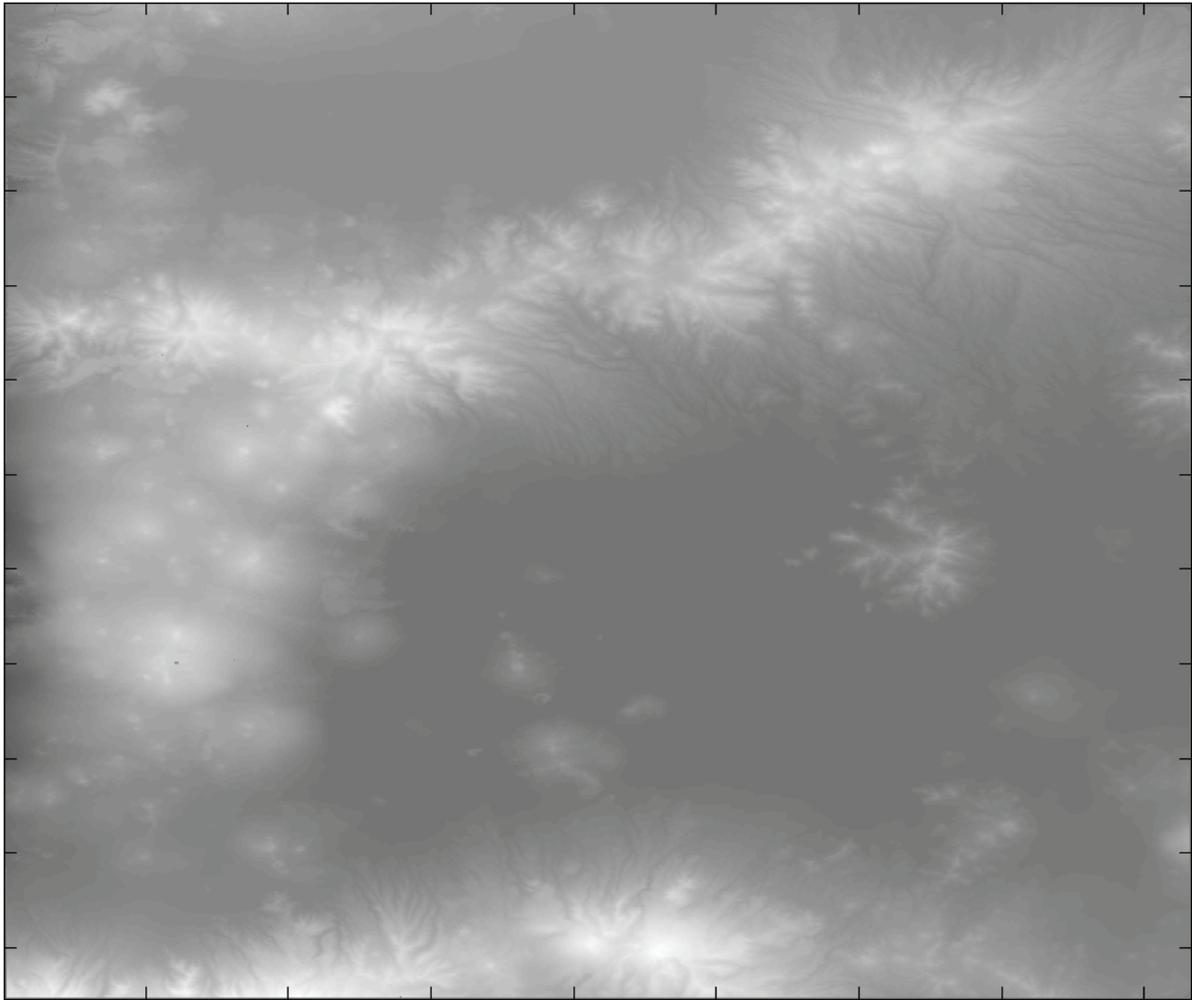


Figura 75. Modelo Digital de Elevación del Valle de México

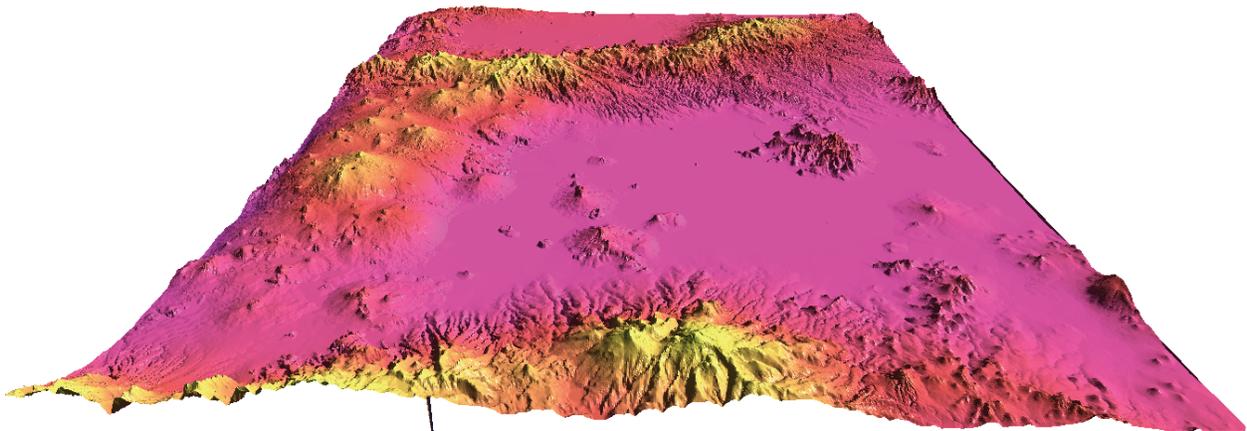


Figura 76. Vista tridimensional del Modelo Digital de Elevación del Valle de México

En las figuras que van de la Figura 77 a la Figura 83 se presentan los resultados de aplicar el proceso de segmentación a los datos de elevación del Valle de México.

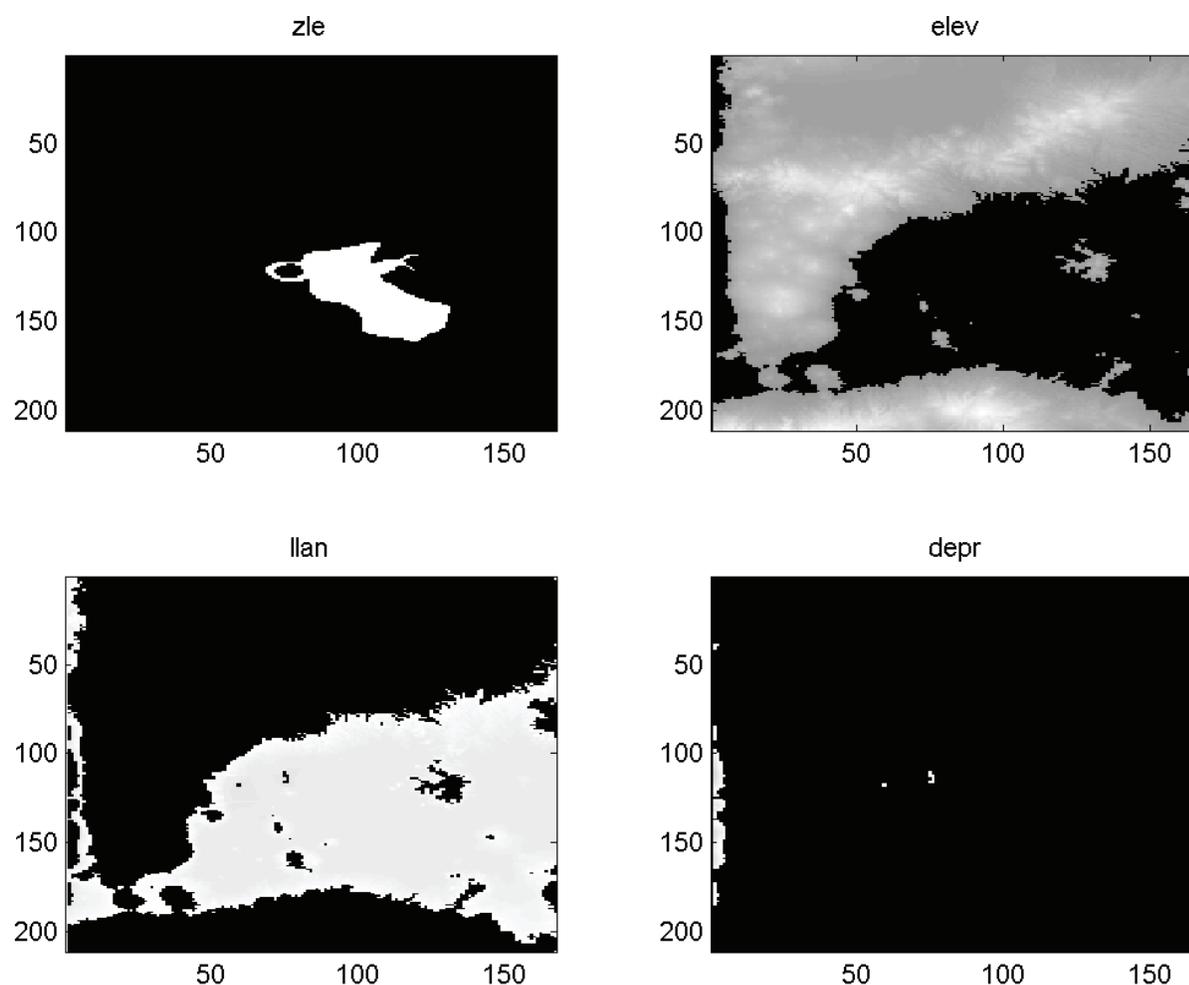


Figura 77. Resultados del proceso de segmentación aplicado al DEM del Valle de México

Se aprecia que en este caso casi no existen zonas que se consideren como depresiones (depr), lo cual se explica pues se trata de un valle (zona llana rodeada por alturas), la zona llana se aprecia en (llan) casi de un color blanco uniforme. La mayor parte de los datos se concentran en las elevaciones alrededor del valle. Cabe mencionar que en estas figuras el norte se encuentra hacia la parte derecha. Así pues, podemos mencionar que la zona de elevaciones que se aprecia en la parte baja de las figuras se trata de las partes bajas de los volcanes Popocatepetl e Iztacciatl y corresponden a las zonas altas de Texcoco y Amecameca.

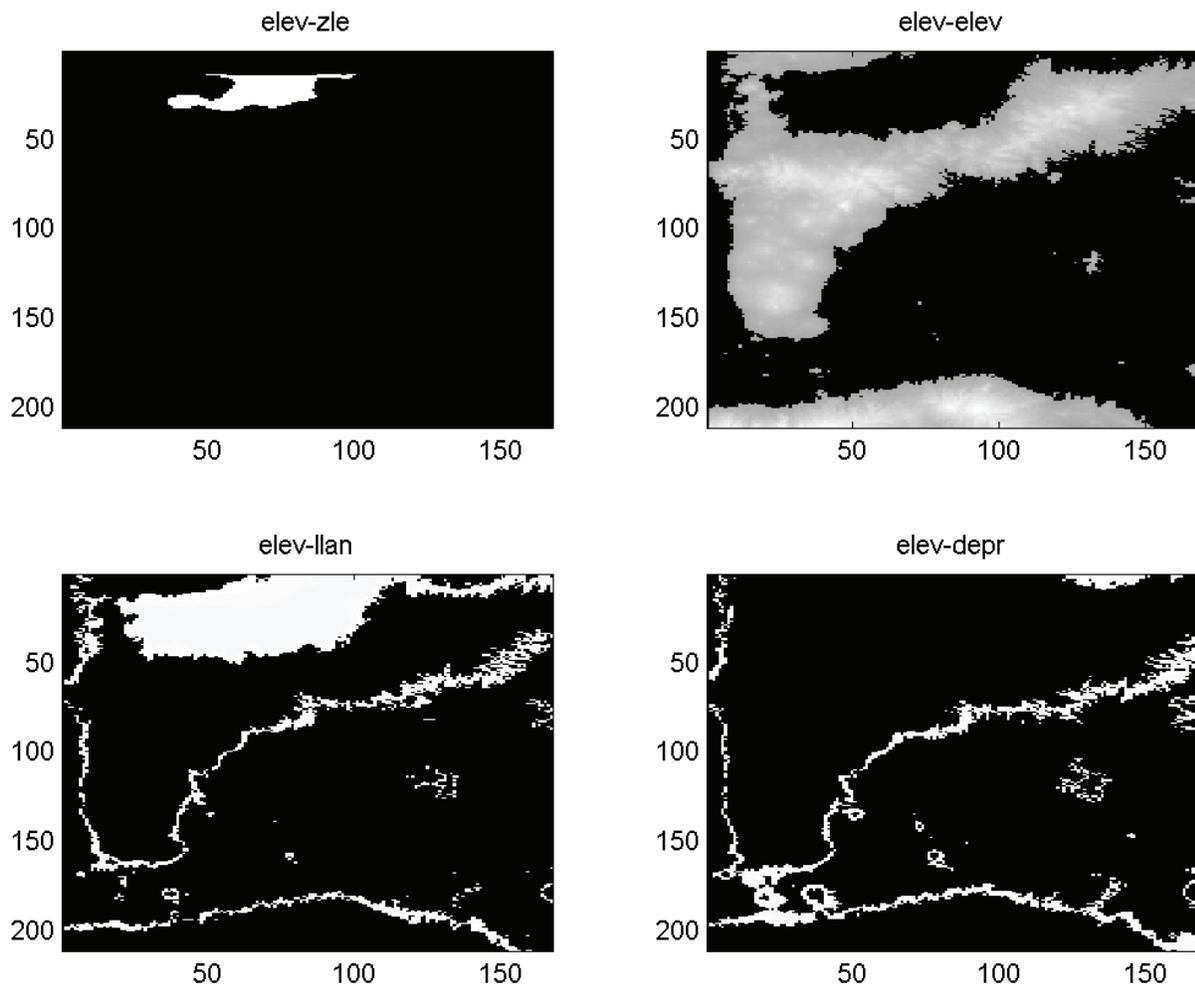


Figura 78. Resultados del proceso de segmentación aplicado al segmento “elev” DEM del Valle de México

Algo interesante a apreciar en esta figura es que aparece en la firma “elev-llan” una zona llana bastante amplia en la parte superior. Esta zona corresponde a parte del llamado Valle de Toluca, que al igual que el Valle de México se trata de una zona llana rodeada de elevaciones, solo que en el caso de Toluca se encuentra a una altura superior.

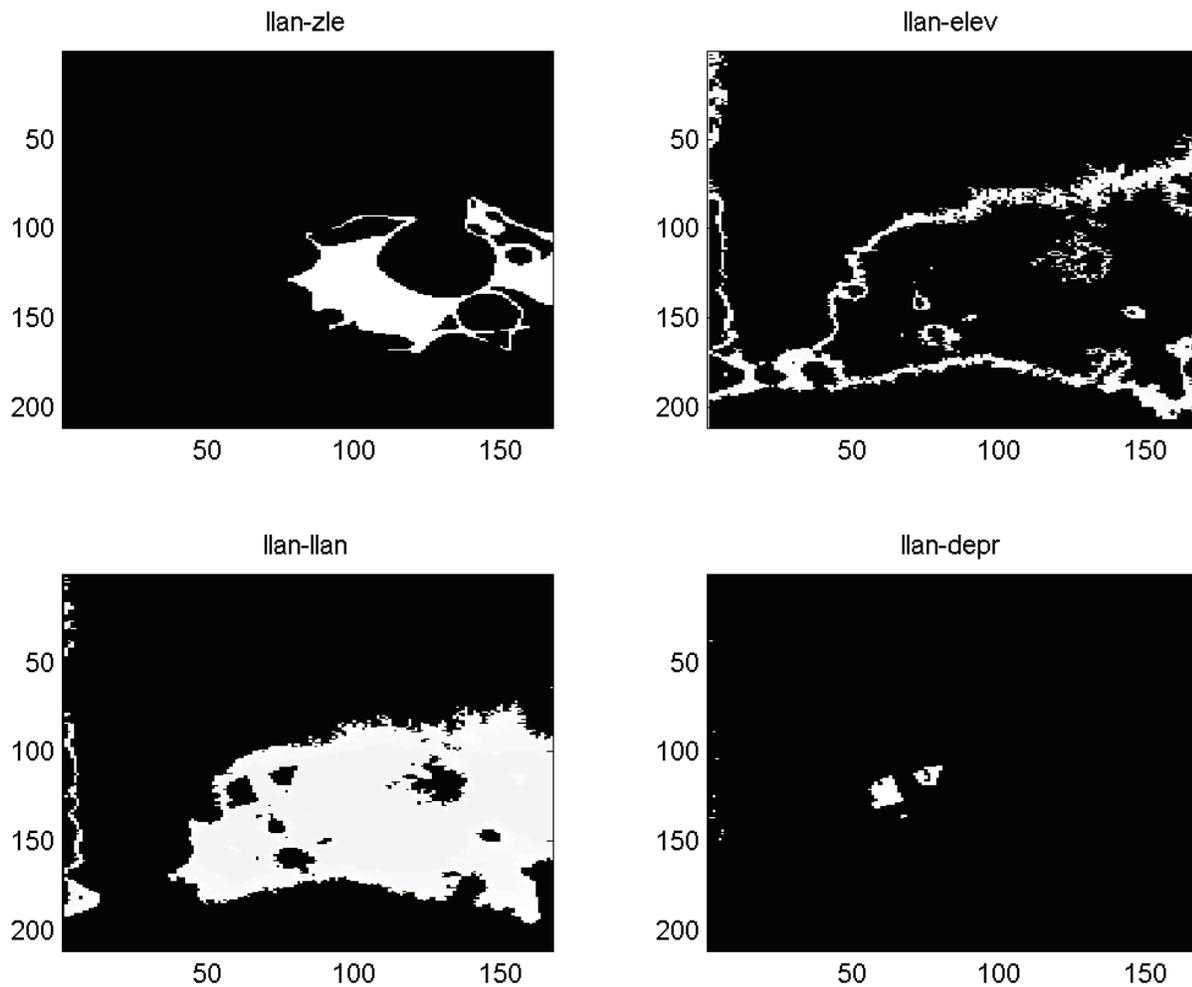


Figura 79. Resultados del proceso de segmentación aplicado al segmento “llan” DEM del Valle de México

En esta figura se aprecia que la mayor parte del valle se comporta como una zona llana (llan-llan), aunque se presentan algunas zonas más bajas que corresponden a bajos en la zona del Pedregal de San Ángel.

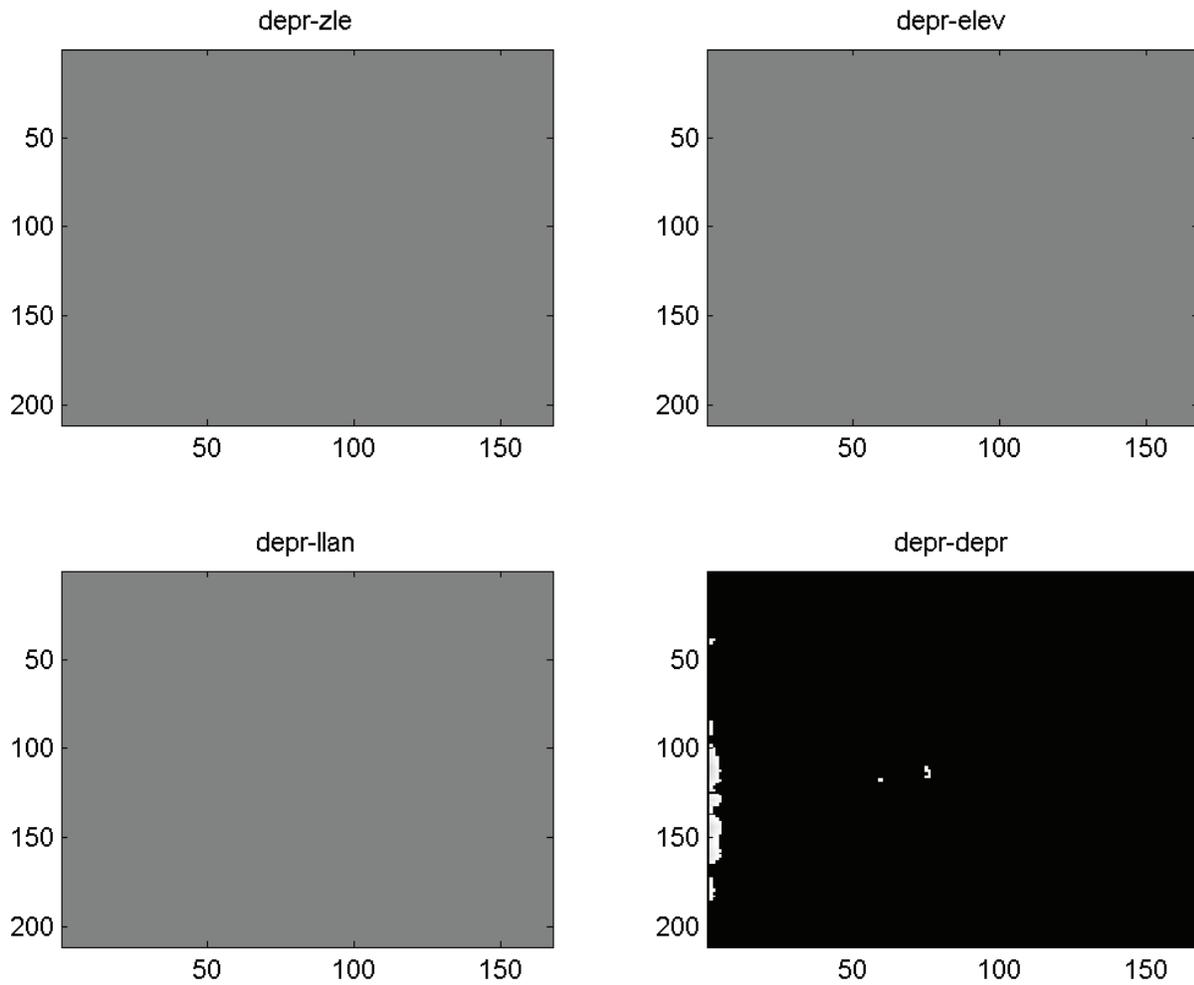


Figura 80. Resultados del proceso de segmentación aplicado al segmento “depr” DEM del Valle de México

Las pocas depresiones que se encontraron desde el inicio se mantienen como tal, lo que nos indica que se trata de zonas por debajo del nivel del valle. Las zonas que aparecen del lado izquierdo (al sur) de la firma “depr-depr” corresponden a las zonas bajas cercanas a Cuernavaca en el estado de Morelos. Podemos ver también, que las otras firmas se encuentran vacías, lo que significa que no tenemos objetos que correspondan a tales firmas.

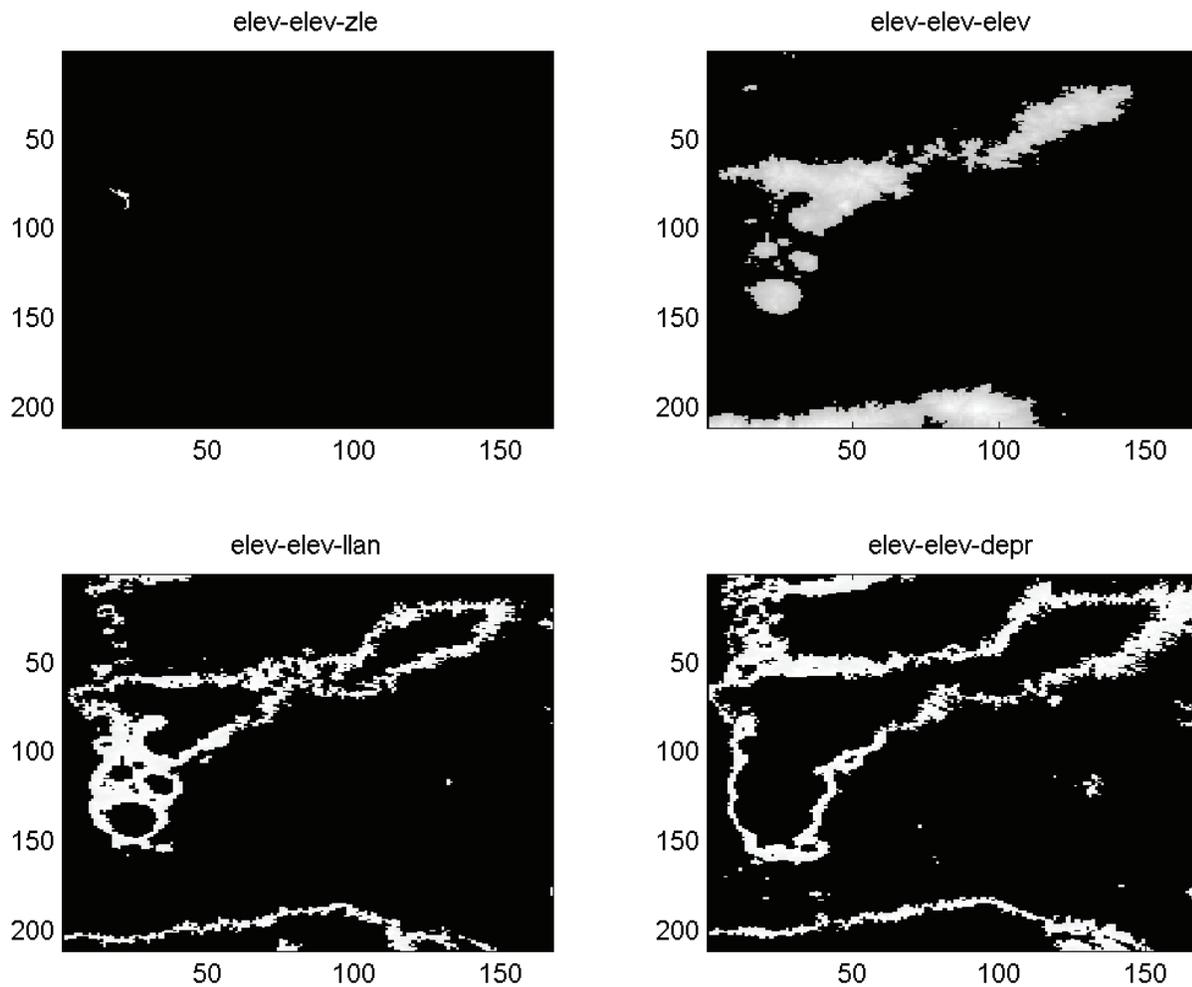


Figura 81. Resultados del proceso de segmentación aplicado al segmento "elev-elev" DEM del Valle de México

En esta figura es posible apreciar claramente la zona del Ajusco (en la parte izquierda de la firma "elev-elev"), así como los pasos que comunican la Ciudad de México con Cuernavaca (Tres Marías) y Toluca (La Marquesa) en las firmas "elev-elev-llan" y "elev-elev-depr".

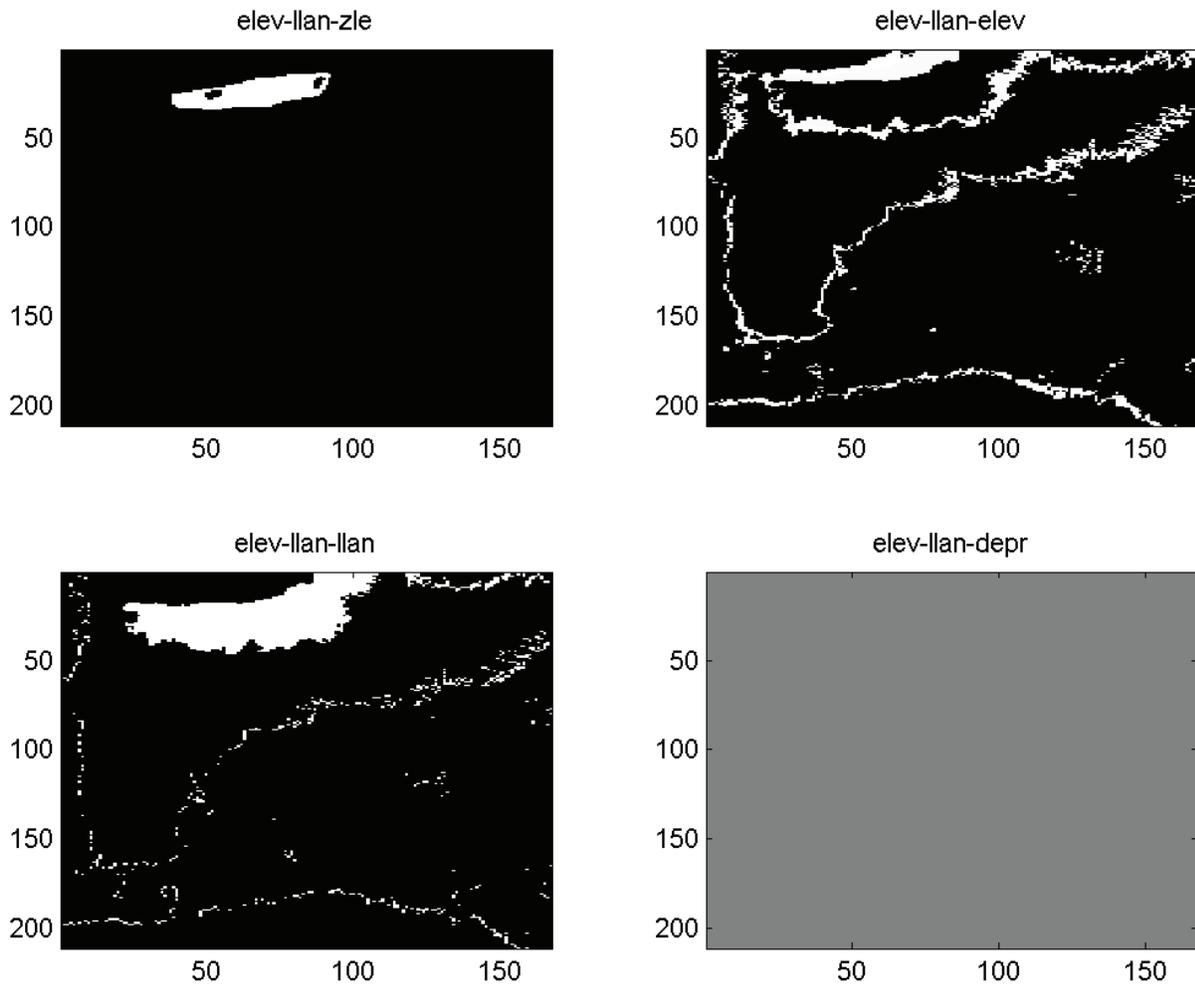


Figura 82. Resultados del proceso de segmentación aplicado al segmento “elev-llan” DEM del Valle de México

En esta figura podemos apreciar nuevamente la presencia de una zona llana bastante amplia en la parte superior de la firma “elev-llan-llan”, que como ya habíamos mencionado, corresponde al Valle de Toluca.

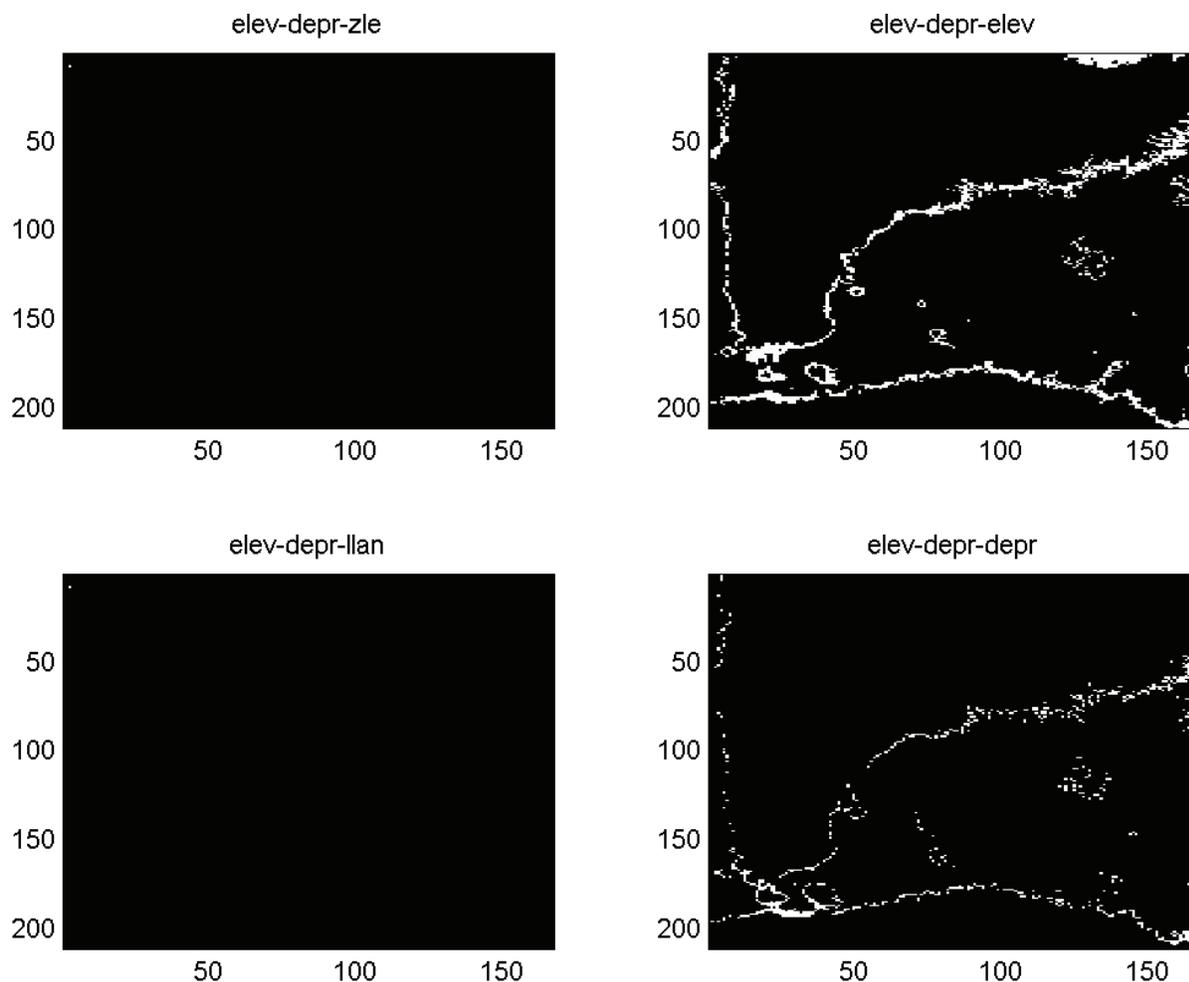


Figura 83. Resultados del proceso de segmentación aplicado al segmento “elev-depr” DEM del Valle de México

En esta figura podemos ver un caso en el que la ZLE corresponde a una zona muy pequeña (pequeño puntito blanco en la esquina superior izquierda de la ZLE) y que además es la única zona llana en este conjunto de datos (esto lo podemos ver en la firma “elev-depr-llan”, pues es esa misma zona la que califica con esta firma).

En las figuras que van de la Figura 84 a la Figura 88 se presentan los resultados del proceso de extracción de los datos que hemos presentado hasta este punto. Finalmente, en la Figura 89 se presenta parte de la descripción obtenida.

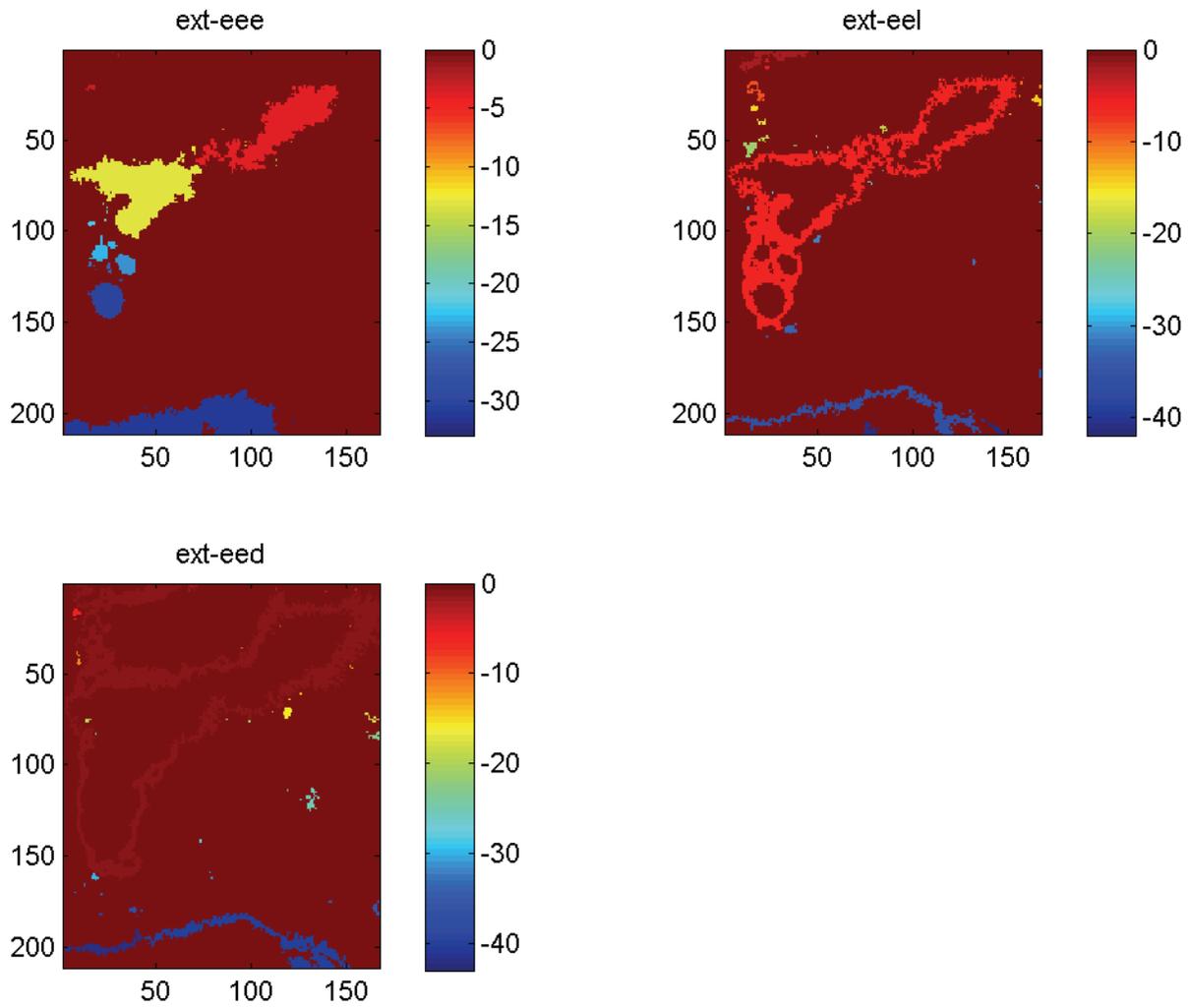


Figura 84. Extractos obtenidos de las firmas “elev-elev-elev”, “elev-elev-llan” y “elev-elev-depr” del DEM del Valle de México

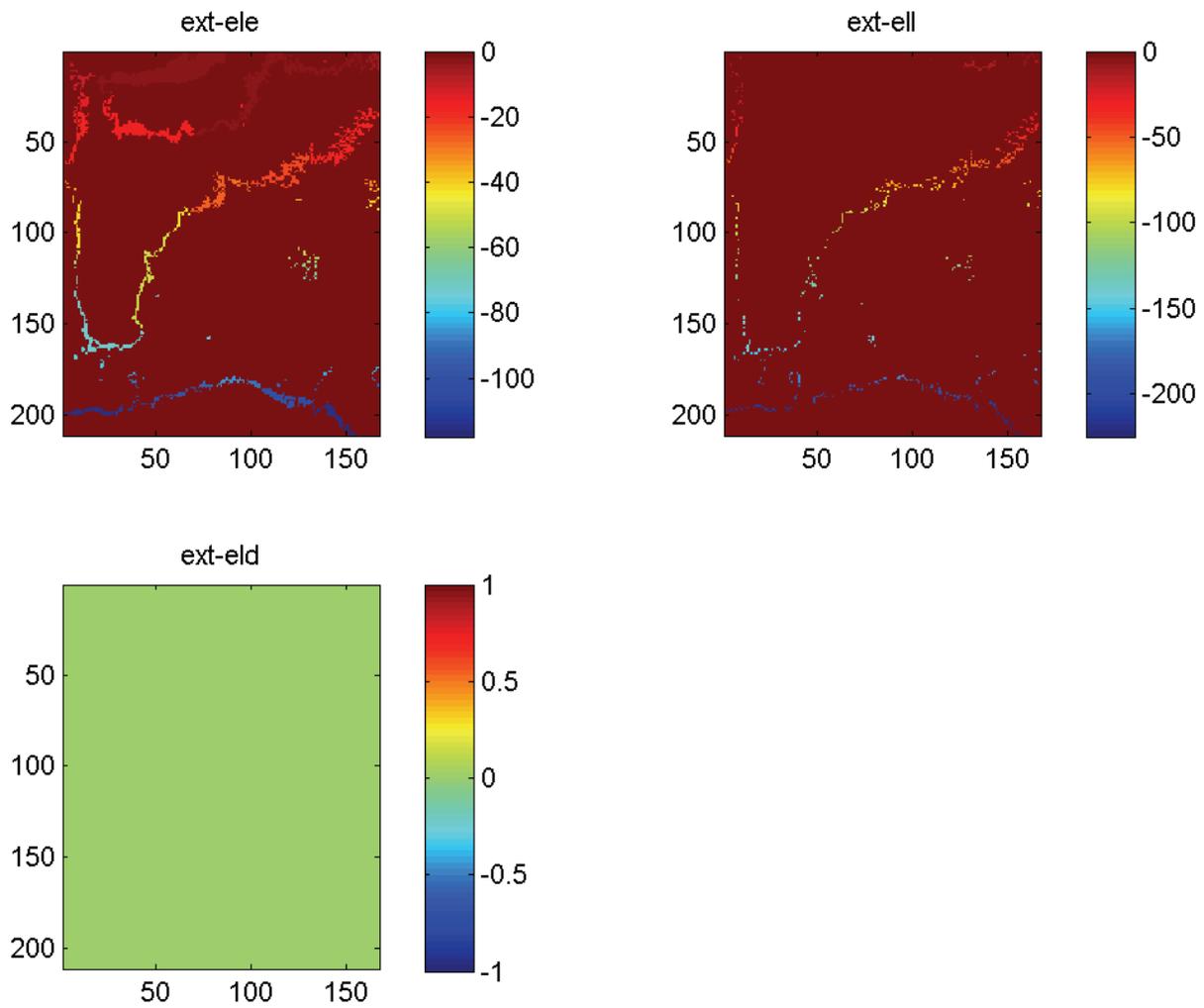


Figura 85. Extractos obtenidos de las firmas “elev-llan-elev”, “elev-llan-llan” y “elev-llan-depr” del DEM del Valle de México

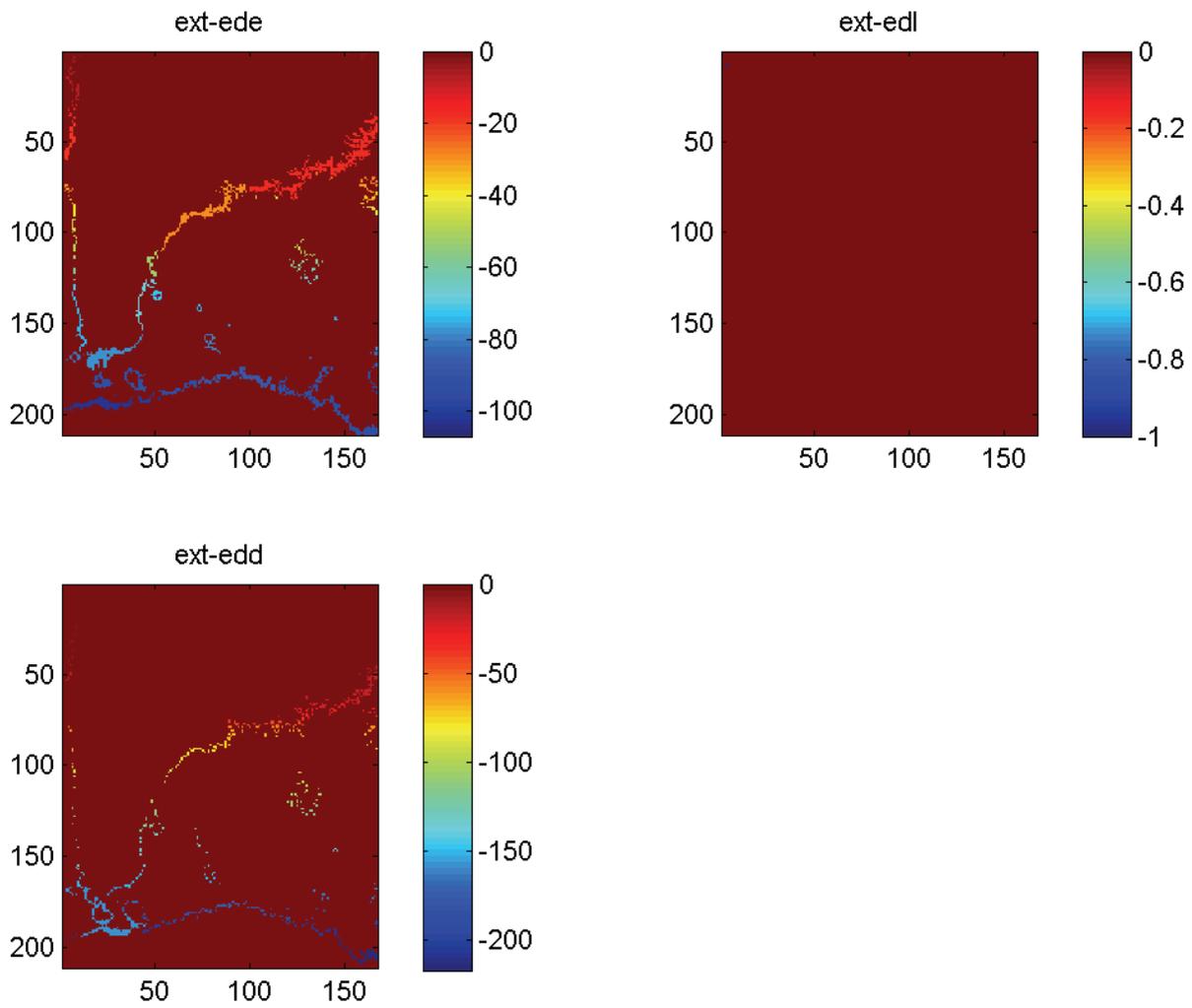


Figura 86. Extractos obtenidos de las firmas “elev-depr-elev”, “elev-depr-llan” y “elev-depr-depr” del DEM del Valle de México

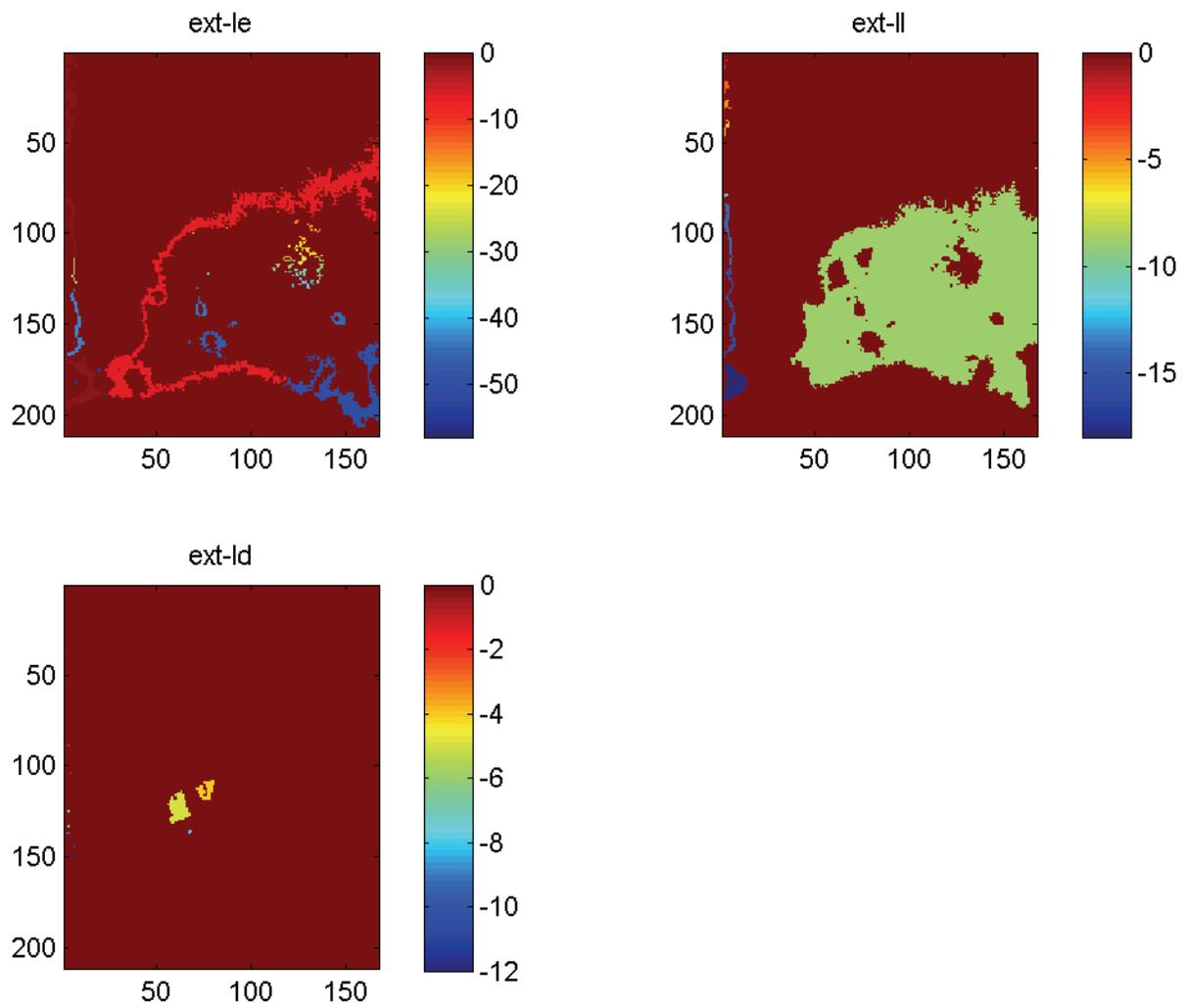


Figura 87. Extractos obtenidos de las firmas "llan-elev", "llan-llan" y "llan-depr" del DEM del Valle de México

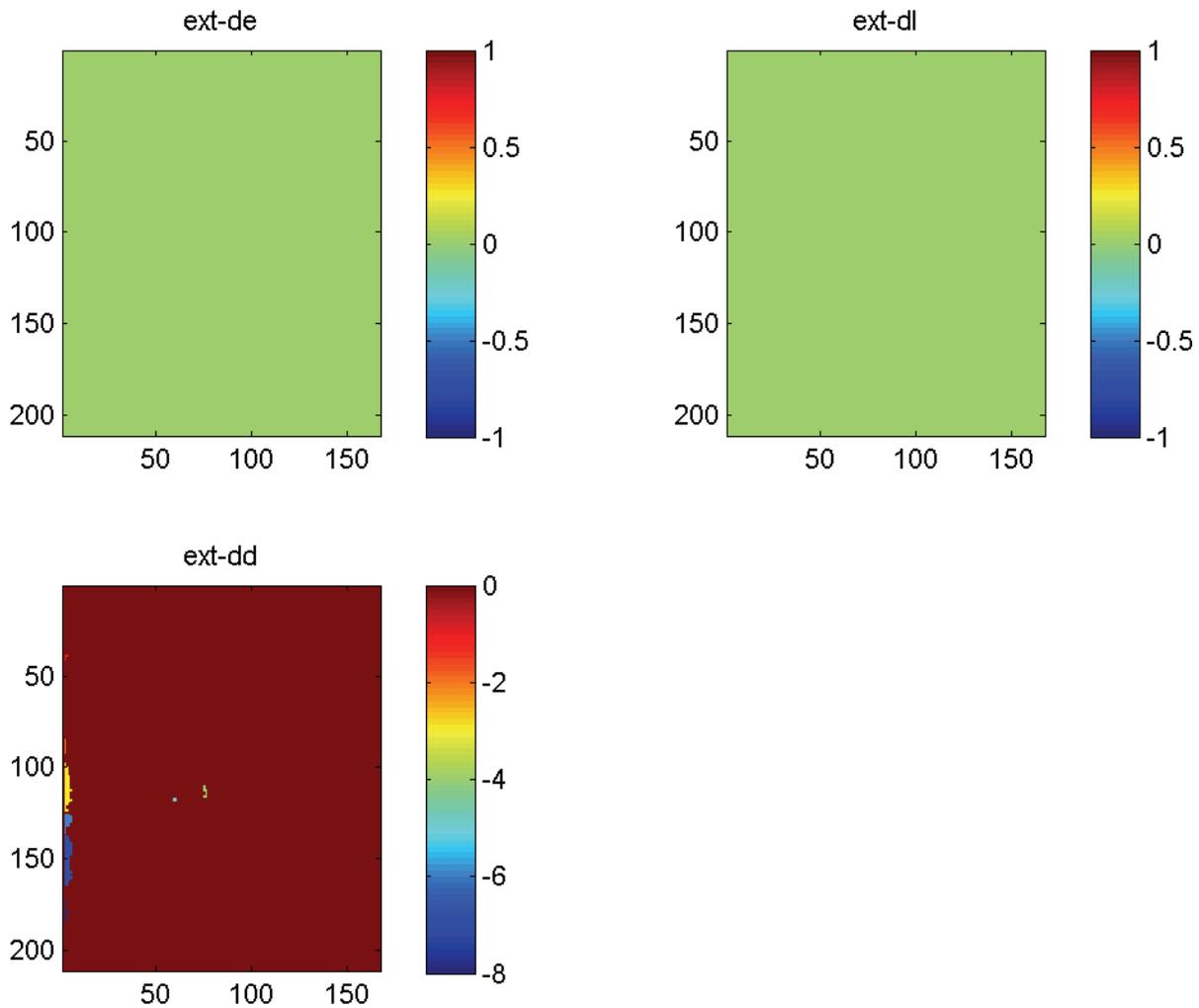


Figura 88. Extractos obtenidos de las firmas “depr-elev”, “depr-llan” y “depr-depr” del DEM del Valle de México

En la Figura 89 se muestra parte de la descripción del DEM del Valle de México utilizando los extractos obtenidos en las etapas anteriores.

```

CONJUNTO DE DATOS ESPACIALES RASTER DEL MODELO DIGITAL DE ELEVACIÓN DE "VALLE DE MEXICO" QUE TIENE UNA RESOLUCION ESPACIAL DE 500 METROS. SU ALTITUD MINIMA
ES DE 613.802000 METROS Y SU ALTITUD MAXIMA ES DE 4160.000000 METROS. LAS COORDENADAS EXTREMAS SON (424569.350000, 2094463.617010) Y (540369.350000,
2190063.617010) METROS EN PROYECCION UTM 14 Y CONTIENE:

UNA MONTAÑA CON AREA DE 3000000 METROS CUADRADOS, ALTITUD MÍNIMA DE 3130.000000 METROS Y MÁXIMA DE 3283.229000 METROS, SUS COORDENADAS EXTREMAS SON
(431069.350000, 2104463.617010) Y (548869.350000, 2201563.617010) METROS. SU CIMA SE ENCUENTRA EN LA COORDENADA (432569.350000, 2104963.617010) A UNA
ALTITUD DE 3283.229000 METROS

UNA MONTAÑA CON AREA DE 276750000 METROS CUADRADOS, ALTITUD MÍNIMA DE 3114.913000 METROS Y MÁXIMA DE 3743.360000 METROS, SUS COORDENADAS EXTREMAS SON
(460069.350000, 2104463.617010) Y (612869.350000, 2224063.617010) METROS. SU CIMA SE ENCUENTRA EN LA COORDENADA (486069.350000, 2110463.617010) A UNA
ALTITUD DE 3743.360000 METROS

...

UNA CORDILLERA CON AREA DE 31000000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2877.426000 METROS Y MÁXIMA DE 3099.840000 METROS, SUS COORDENADAS EXTREMAS SON
(429069.350000, 2095463.617010) Y (562369.350000, 2196563.617010) METROS. SU CIMA SE ENCUENTRA EN LA COORDENADA (439069.350000, 2095463.617010) A UNA
ALTITUD DE 3099.840000 METROS
    
```

UNA CORDILLERA CON AREA DE 724750000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2877.419000 METROS Y MÁXIMA DE 3114.086000 METROS, SUS COORDENADAS EXTREMAS SON (426069.350000, 2101963.617010) Y (617369.350000, 2267063.617010) METROS. SU CIMA SE ENCUENTRA EN LA COORDENADA (434569.350000, 2133963.617010) A UNA ALTITUD DE 3114.086000 METROS

...

UN MONTE CON AREA DE 2500000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2690.092000 METROS Y MÁXIMA DE 2797.486000 METROS, SUS COORDENADAS EXTREMAS SON (458069.350000, 2095463.617010) Y (575869.350000, 2192563.617010) METROS. SU CIMA SE ENCUENTRA EN LA COORDENADA (459569.350000, 2095463.617010) A UNA ALTITUD DE 2797.486000 METROS

UN MONTE CON AREA DE 3250000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2678.419000 METROS Y MÁXIMA DE 2825.393000 METROS, SUS COORDENADAS EXTREMAS SON (427569.350000, 2101963.617010) Y (545369.350000, 2199063.617010) METROS. SU CIMA SE ENCUENTRA EN LA COORDENADA (428069.350000, 2102963.617010) A UNA ALTITUD DE 2825.393000 METROS

...

UN CERRO CON AREA DE 145750000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2596.979000 METROS Y MÁXIMA DE 2675.452000 METROS, SUS COORDENADAS EXTREMAS SON (434069.350000, 2095463.617010) Y (583369.350000, 2202063.617010) METROS. SU CIMA SE ENCUENTRA EN LA COORDENADA (452569.350000, 2101463.617010) A UNA ALTITUD DE 2675.452000 METROS

UN CERRO CON AREA DE 115000000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2596.879000 METROS Y MÁXIMA DE 2675.347000 METROS, SUS COORDENADAS EXTREMAS SON (460069.350000, 2095463.617010) Y (620369.350000, 2213063.617010) METROS. SU CIMA SE ENCUENTRA EN LA COORDENADA (487569.350000, 2099963.617010) A UNA ALTITUD DE 2675.347000 METROS

...

UN ALTIPLANO CON AREA DE 7750000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2570.000000 METROS Y MÁXIMA DE 2596.478000 METROS, SUS COORDENADAS EXTREMAS SON (483569.350000, 2095463.617010) Y (603369.350000, 2194063.617010) METROS. SU CIMA SE ENCUENTRA EN LA COORDENADA (485569.350000, 2095463.617010) A UNA ALTITUD DE 2596.478000 METROS

UN ALTIPLANO CON AREA DE 6750000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2570.000000 METROS Y MÁXIMA DE 2596.568000 METROS, SUS COORDENADAS EXTREMAS SON (497069.350000, 2095463.617010) Y (617369.350000, 2195063.617010) METROS. SU CIMA SE ENCUENTRA EN LA COORDENADA (498569.350000, 2097963.617010) A UNA ALTITUD DE 2596.568000 METROS

...

UN ELEV-DEPR-ELEV CON AREA DE 2000000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2550.000000 METROS Y MÁXIMA DE 2569.593000 METROS, SUS COORDENADAS EXTREMAS SON (504069.350000, 2095463.617010) Y (621369.350000, 2192563.617010) METROS. SU CIMA SE ENCUENTRA EN LA COORDENADA (504069.350000, 2096963.617010) A UNA ALTITUD DE 2569.593000 METROS

UN ELEV-DEPR-ELEV CON AREA DE 2000000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2502.649000 METROS Y MÁXIMA DE 2561.326000 METROS, SUS COORDENADAS EXTREMAS SON (426569.350000, 2098963.617010) Y (543869.350000, 2196563.617010) METROS. SU CIMA SE ENCUENTRA EN LA COORDENADA (427569.350000, 2100463.617010) A UNA ALTITUD DE 2561.326000 METROS

...

UN COLLADO CON AREA DE 9750000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2471.109000 METROS Y MÁXIMA DE 2496.271000 METROS, SUS COORDENADAS EXTREMAS SON (502069.350000, 2119963.617010) Y (623869.350000, 2221063.617010) METROS. SU CIMA SE ENCUENTRA EN LA COORDENADA (506069.350000, 2120963.617010) A UNA ALTITUD DE 2496.271000 METROS

UN COLLADO CON AREA DE 4000000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2474.296000 METROS Y MÁXIMA DE 2495.320000 METROS, SUS COORDENADAS EXTREMAS SON (499569.350000, 2125963.617010) Y (619369.350000, 2223063.617010) METROS. SU CIMA SE ENCUENTRA EN LA COORDENADA (503069.350000, 2126463.617010) A UNA ALTITUD DE 2495.320000 METROS

...

UN LLANO CON AREA DE 442750000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2364.748000 METROS Y MÁXIMA DE 2470.902000 METROS, SUS COORDENADAS EXTREMAS SON (436069.350000, 2120463.617010) Y (623869.350000, 2285563.617010) METROS.

UN LLANO CON AREA DE 6000000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2369.628000 METROS Y MÁXIMA DE 2459.766000 METROS, SUS COORDENADAS EXTREMAS SON (487569.350000, 2145463.617010) Y (607369.350000, 2246063.617010) METROS.

...

UNA PLANICIE CON AREA DE 3250000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2241.869000 METROS Y MÁXIMA DE 2347.249000 METROS, SUS COORDENADAS EXTREMAS SON (425569.350000, 2102963.617010) Y (542369.350000, 2201563.617010) METROS.

UNA PLANICIE CON AREA DE 2750000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2280.940000 METROS Y MÁXIMA DE 2362.898000 METROS, SUS COORDENADAS EXTREMAS SON (425569.350000, 2107963.617010) Y (542869.350000, 2205563.617010) METROS.

...

UNA MESETA CON AREA DE 15250000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2231.864000 METROS Y MÁXIMA DE 2239.860000 METROS, SUS COORDENADAS EXTREMAS SON (460069.350000, 2148463.617010) Y (580369.350000, 2249063.617010) METROS.

UNA MESETA CON AREA DE 32000000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2231.962000 METROS Y MÁXIMA DE 2239.973000 METROS, SUS COORDENADAS EXTREMAS SON (452569.350000, 2151463.617010) Y (573869.350000, 2255563.617010) METROS.

...

```
UNA CUENCA CON AREA DE 2000000 METROS CUADRADOS, ALTITUD MÍNIMA DE 2168.397000 METROS Y MÁXIMA DE 2230.000000 METROS, SUS COORDENADAS EXTREMAS SON (425569.350000, 2136963.617010) Y (541369.350000, 2236063.617010) METROS. SU ZIMA SE ENCUENTRA EN LA COORDENADA (425569.350000, 2138963.617010) A UNA ALTITUD DE 2168.397000 METROS

UNA CUENCA CON AREA DE 17000000 METROS CUADRADOS, ALTITUD MÍNIMA DE 1907.440000 METROS Y MÁXIMA DE 2224.643000 METROS, SUS COORDENADAS EXTREMAS SON (425569.350000, 2143463.617010) Y (542869.350000, 2252063.617010) METROS. SU ZIMA SE ENCUENTRA EN LA COORDENADA (425569.350000, 2150463.617010) A UNA ALTITUD DE 1907.440000 METROS

...
```

Figura 89. Parte de la descripción obtenida a partir del DEM del Valle de México

5.4.2. Pruebas con el DEM de Chicago

Datos de elevación correspondientes a la zona este de Chicago en el estado de Illinois, Estados Unidos, obtenidos del directorio de modelos digitales de elevación de los Estados Unidos (USGS, 2007b), con una escala de 1:250000. En la Figura 90 se muestra la imagen en escala de grises obtenida a partir del DEM mencionado, en la Figura 91 se muestra el mismo modelo con un sombreado para una mejor apreciación de las formas del terreno. Sus coordenadas extremas en proyección Geográfica son $-309600,1476$ y $-313200,1512$ segundos-arco, el número de columnas es de 1201 y el número de filas es de 1201.

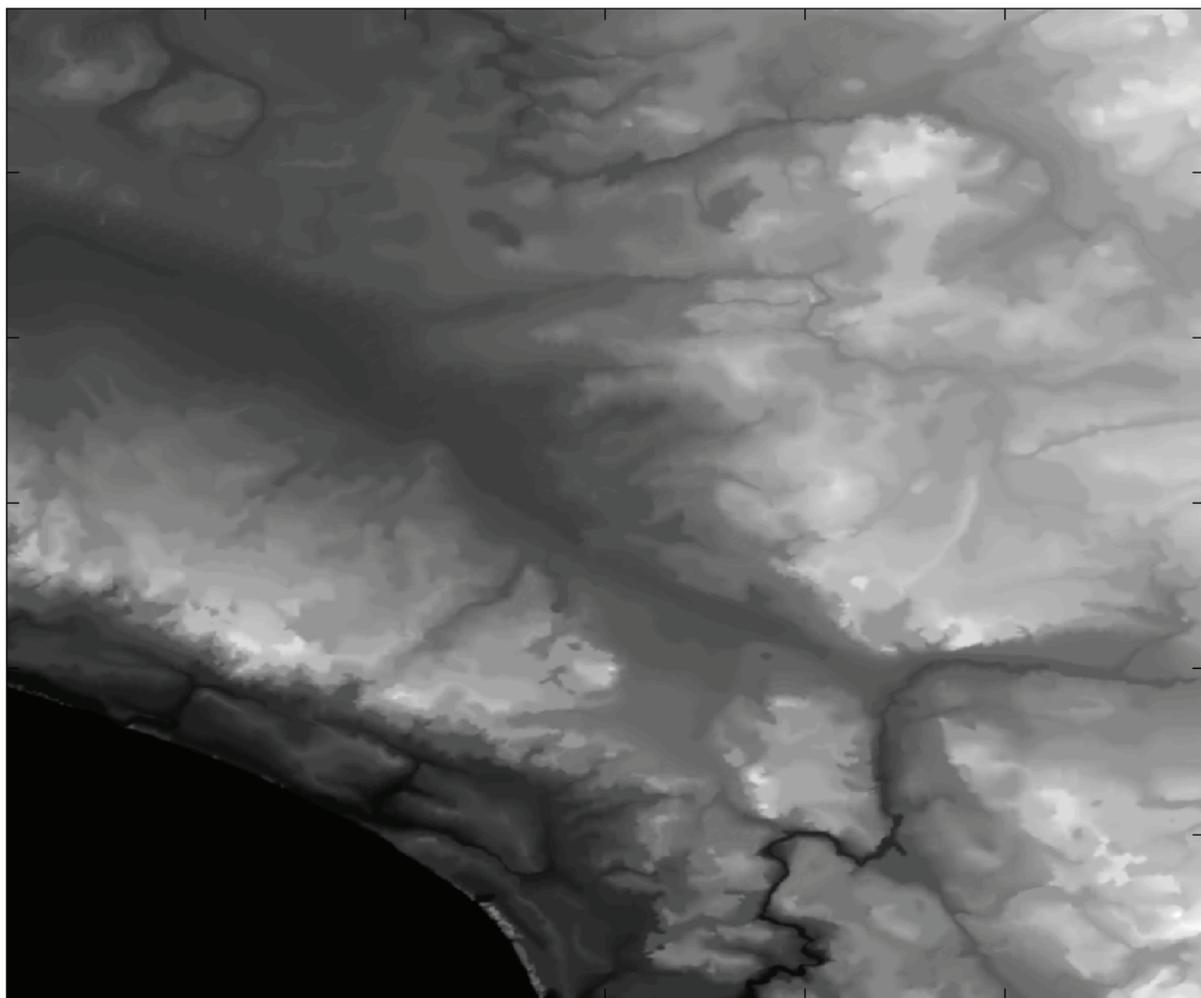


Figura 90. Modelo Digital de Elevación de Chicago

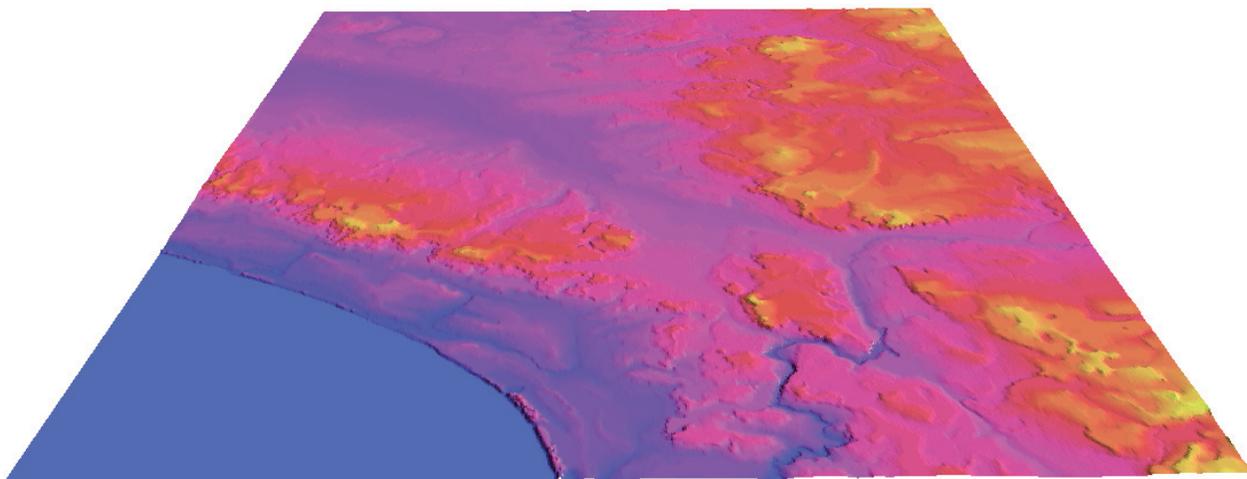


Figura 91. Vista tridimensional del Modelo Digital de Elevación de Chicago

En las figuras que van de la Figura 92 a la Figura 98 se presentan los resultados de aplicar el proceso de segmentación a los datos de elevación de Chicago. En este ejemplo veremos que hay una clasificación errónea de los datos, debido a la presencia de un cuerpo de agua en el modelo que se aprecia en la esquina inferior izquierda de la Figura 91, en color azul.

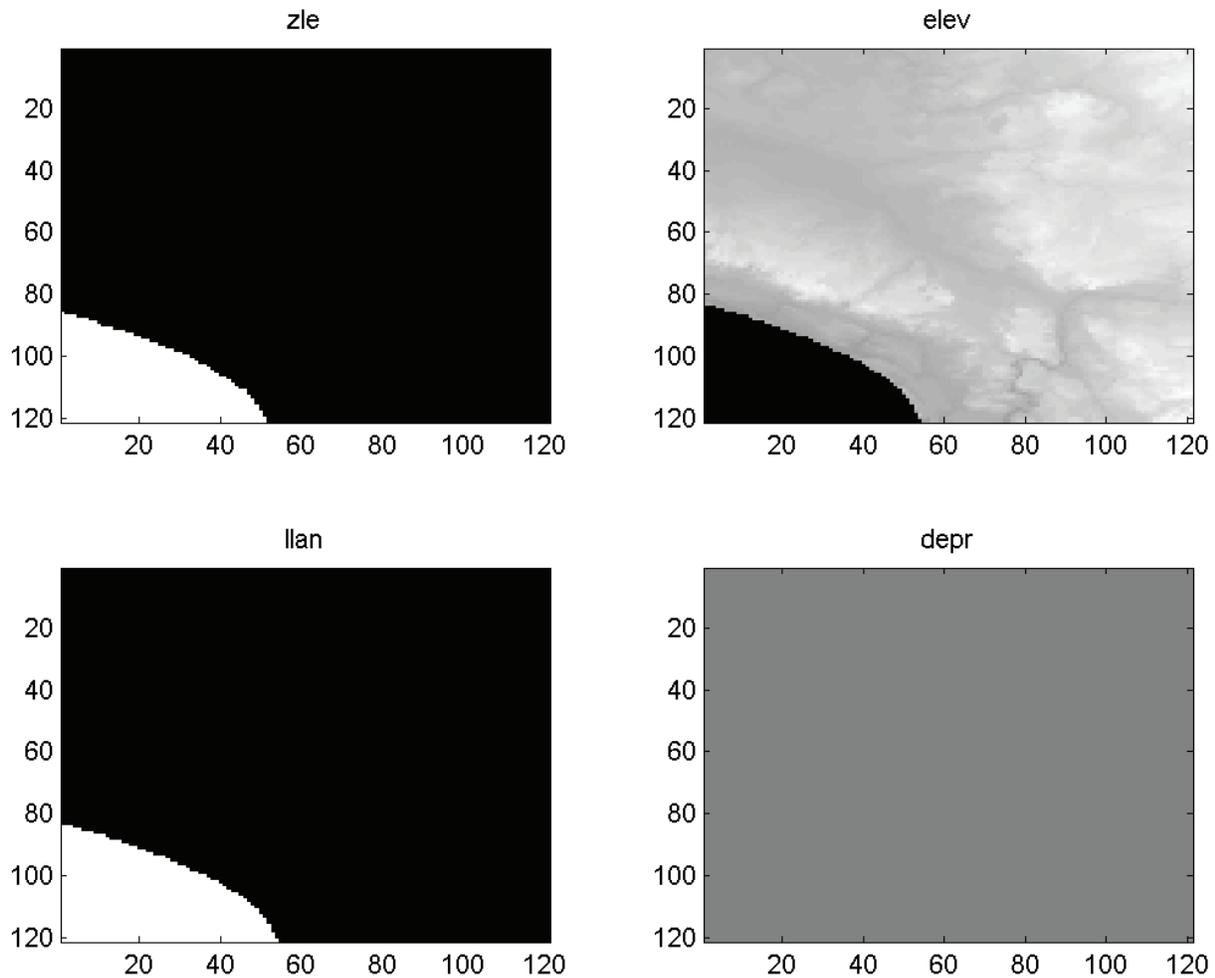


Figura 92. Resultados del proceso de segmentación aplicado al DEM de Chicago

En esta figura podemos apreciar que se detecta como la ZLE una zona llana que es bastante pareja; con tan solo la información de elevación el algoritmo se comporta correctamente. Sin embargo, con información adicional nos damos cuenta que esa ZLE es errónea, ya que esa zona es más bien parte de un lago, del Lago Michigan. Como consecuencia, podemos apreciar que no existen zonas que se clasifiquen como depresiones (depr).

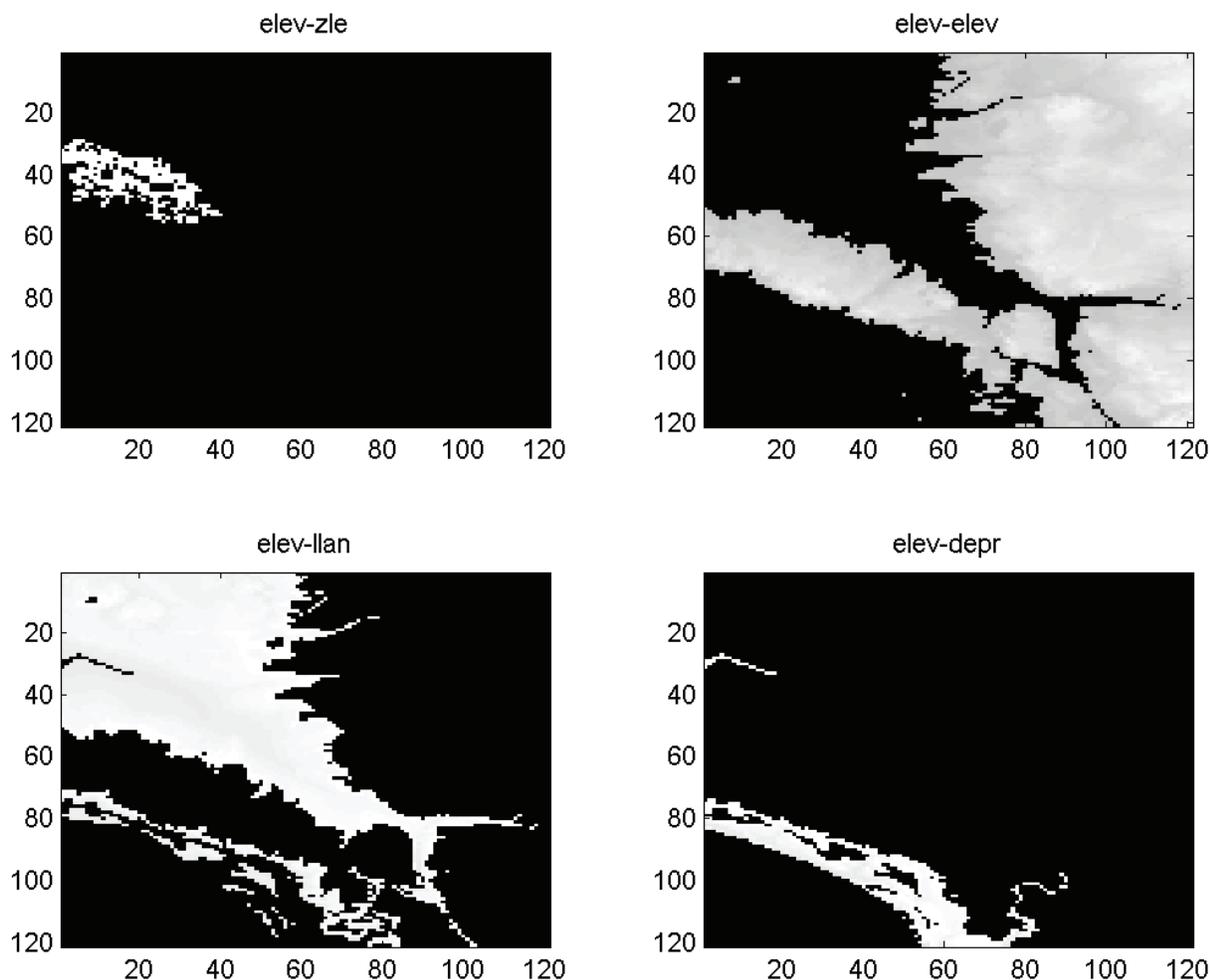


Figura 93. Resultados del proceso de segmentación aplicado al segmento “elev” del DEM de Chicago

Si tuviésemos forma de discriminar los datos de elevación del lago, esta figura debería ser la correspondiente a la primera segmentación. Sin embargo, esto no es posible puesto que solo tenemos información de elevación.

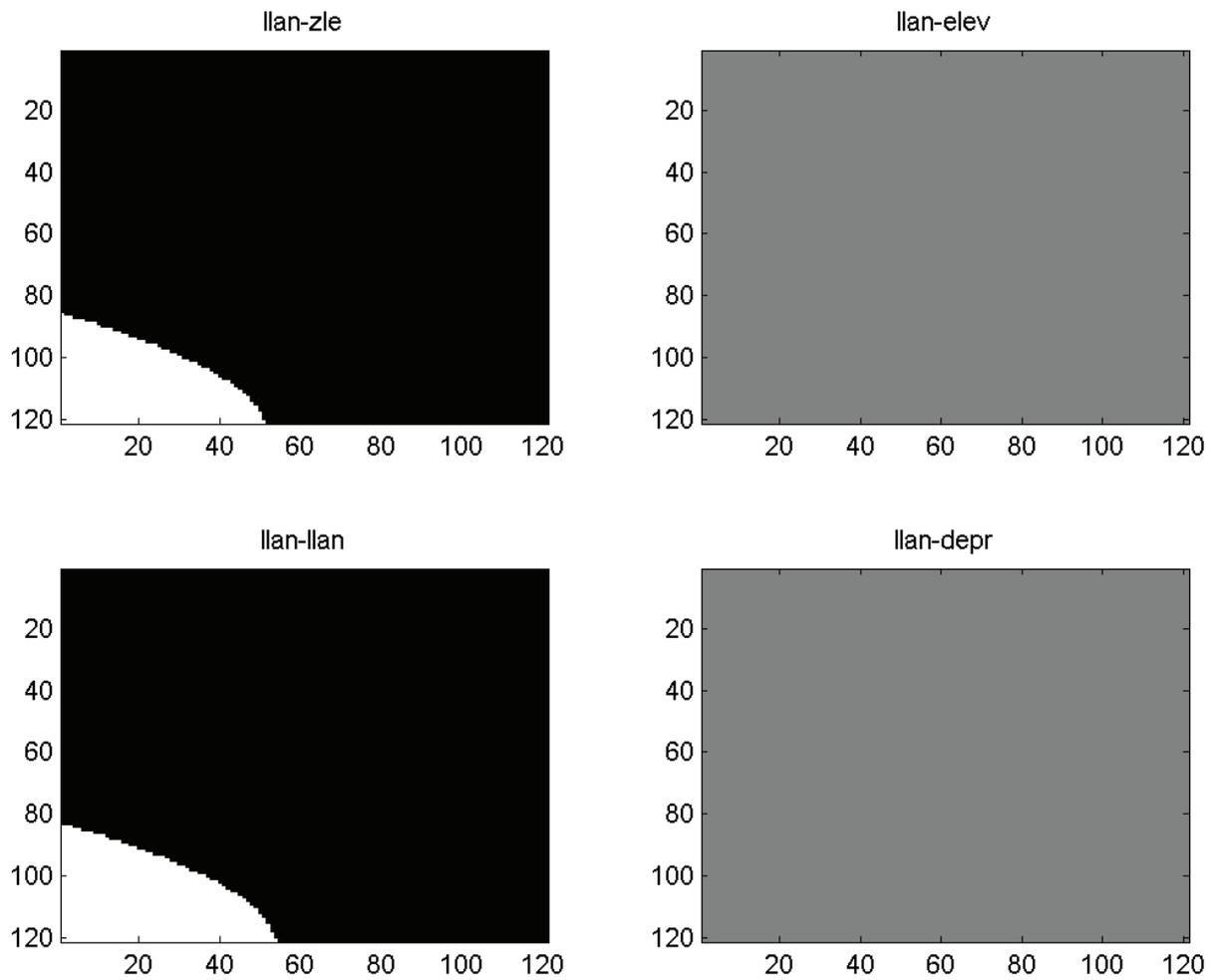


Figura 94. Resultados del proceso de segmentación aplicado al segmento “llan” del DEM de Chicago

En esta figura podemos ver que el lago es clasificado siempre como llanura, dejando vacías las firmas “llan-elev” y “llan-depr”.

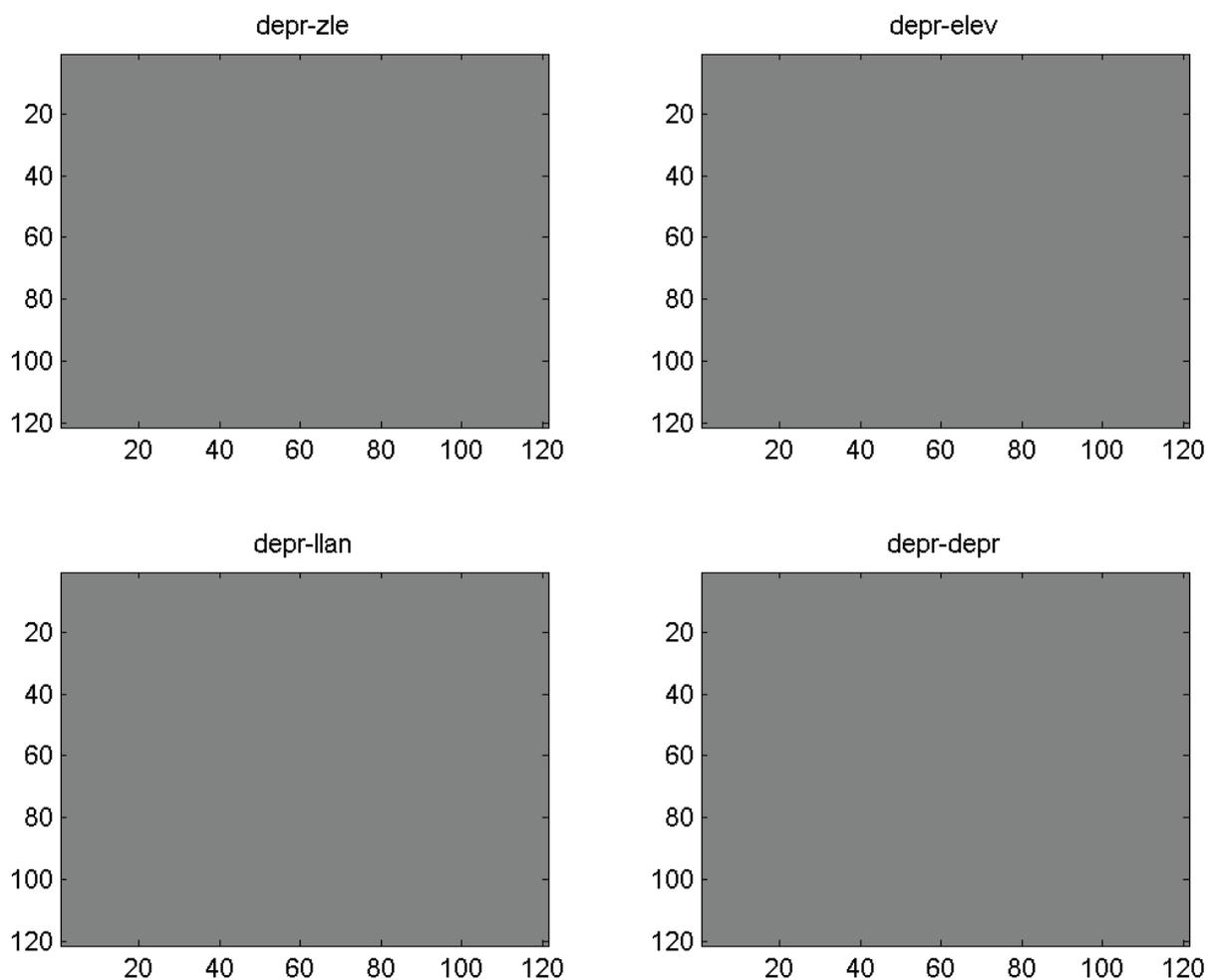


Figura 95. Resultados del proceso de segmentación aplicado al segmento "depr" del DEM de Chicago

Debido a la clasificación falsa del lago como una llanura y a que normalmente los cuerpos de agua son las zonas más bajas de una región, la metodología no detecta ninguna depresión en el conjunto de datos.

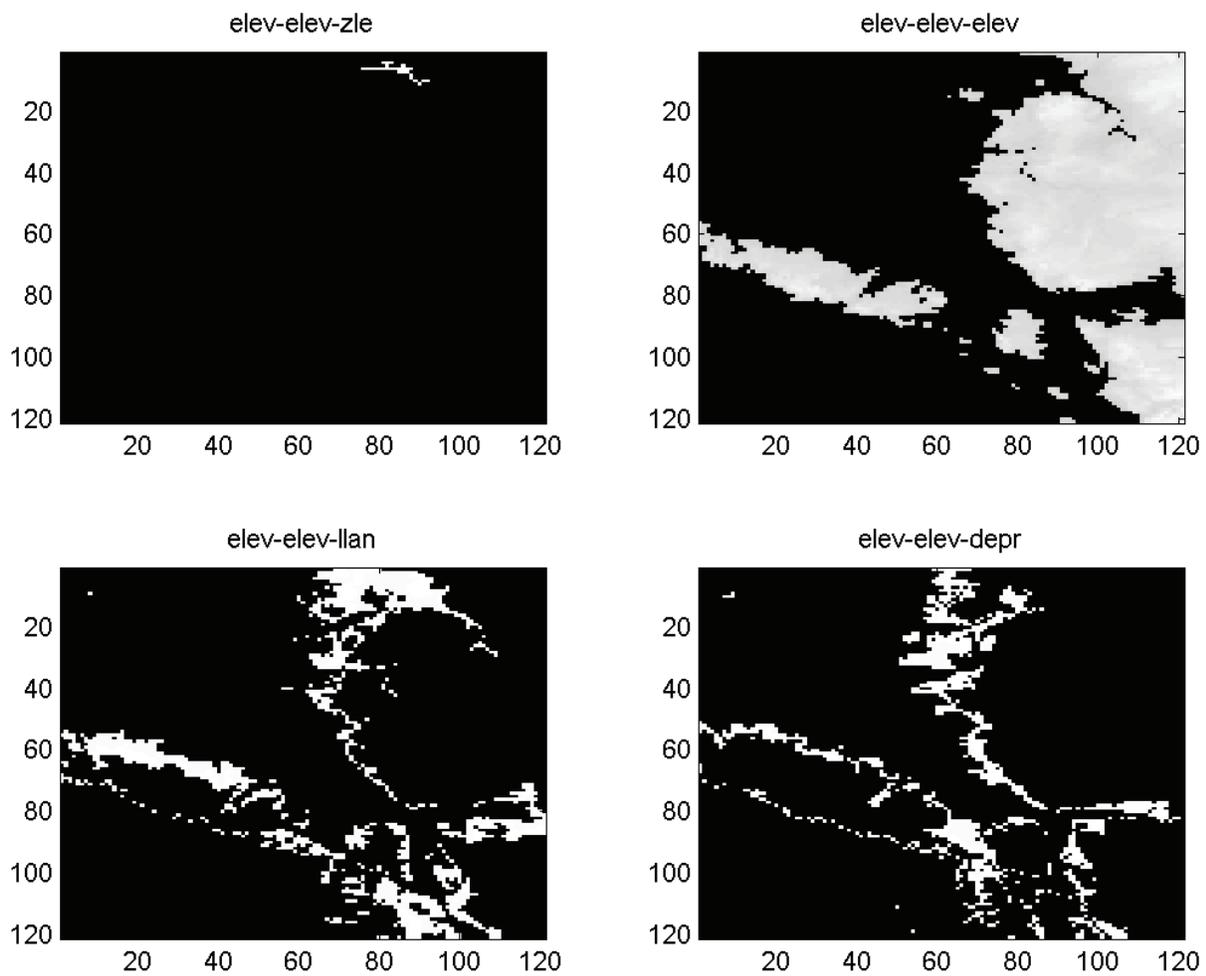


Figura 96. Resultados del proceso de segmentación aplicado al segmento “elev-elev” del DEM de Chicago

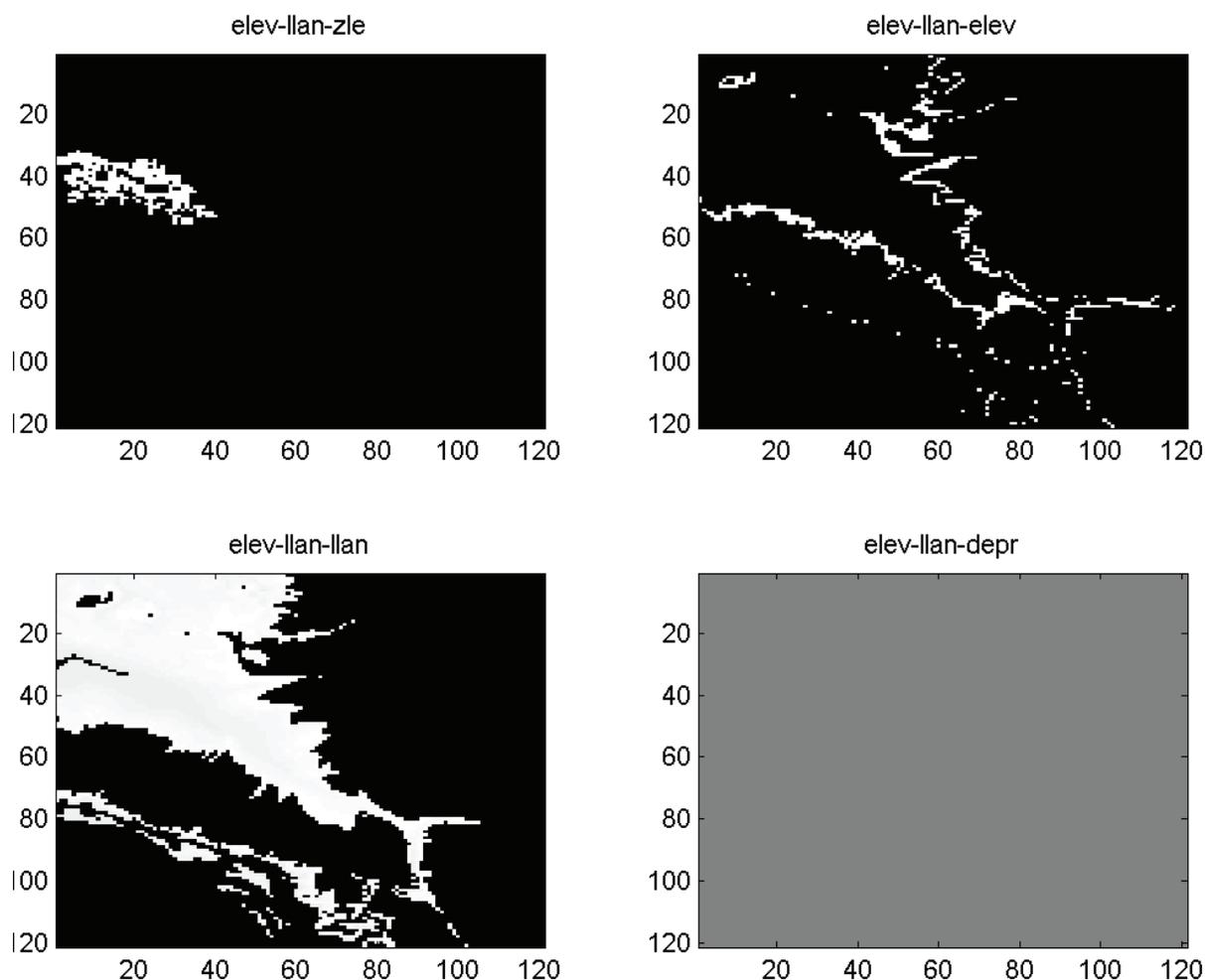


Figura 97. Resultados del proceso de segmentación aplicado al segmento “elev-llan” del DEM de Chicago

En esta figura podemos ver que no hay datos en la firma “elev-llan-depr”, pero esto no se debe a la clasificación incorrecta del lago, sino a que en realidad no existen depresiones por debajo de las zonas llanas en esta región.

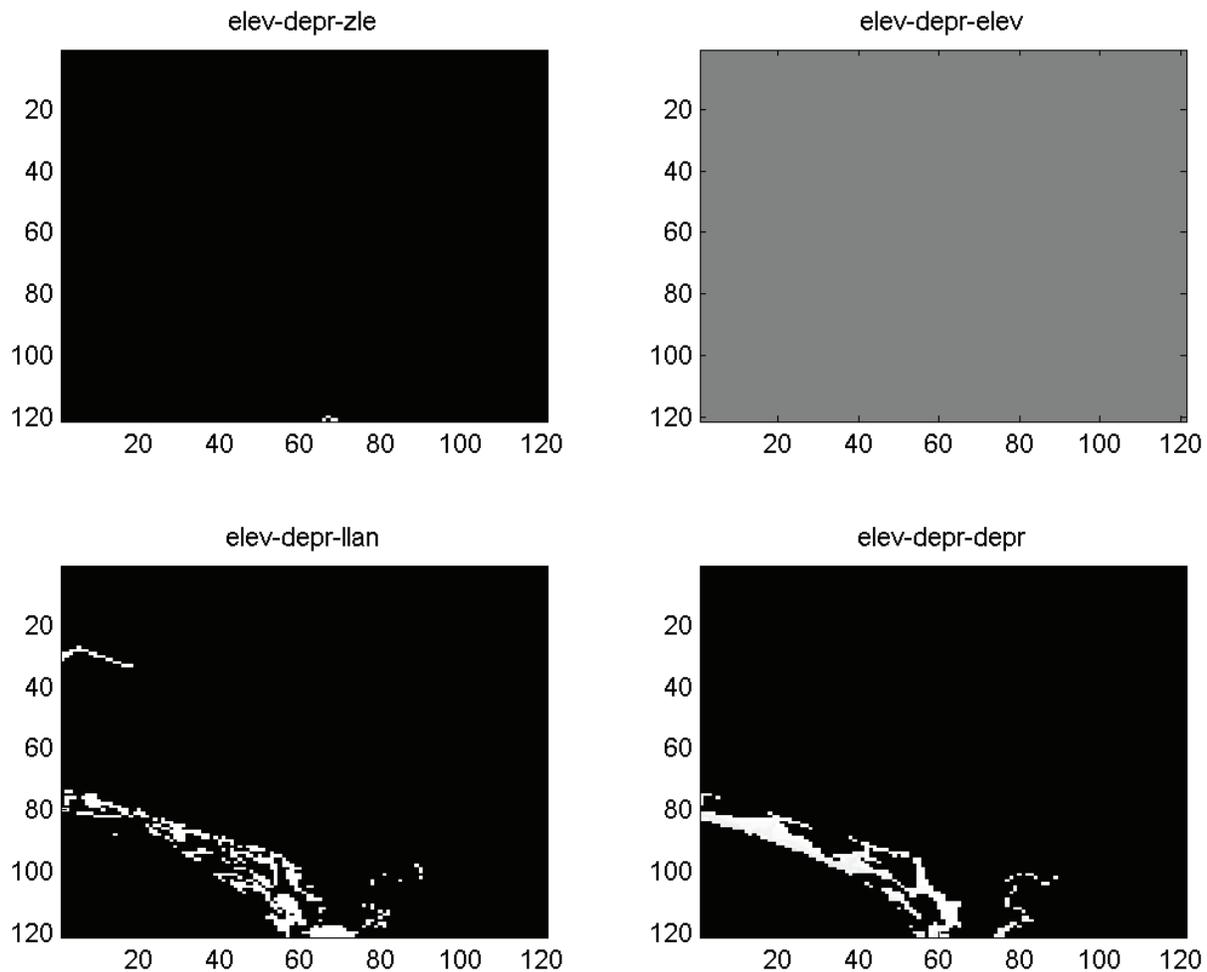


Figura 98. Resultados del proceso de segmentación aplicado al segmento "elev-depr" del DEM de Chicago

Debido a que la clasificación errónea del lago como ZLE ocurre desde las primeras etapas del proceso, los resultados subsecuentes no son confiables hasta cierto punto. Esta falla en la clasificación ocurrirá siempre que exista en la región un cuerpo de agua bastante grande en comparación con las zonas llanas. En el caso de las zonas costeras esta falla no ocurre si el cuerpo de agua es el mar, pues toda la zona que ocupe el mar en el conjunto de datos tendrá un valor de cero, valor que es ignorado por los programas realizados. En resumen, esta falla se presentará siempre que existan cuerpos de agua en el conjunto de datos y no hay forma de discriminarlos con solo los datos de elevación.

En las figuras que van de la Figura 99 a la Figura 103 se presentan los resultados del proceso de extracción de los datos que hemos presentado hasta este punto. Finalmente, en la Figura 104 se presenta parte de la descripción obtenida; al igual que en el ejemplo anterior es una descripción compacta, pero no la pondremos completa por cuestión de espacio.

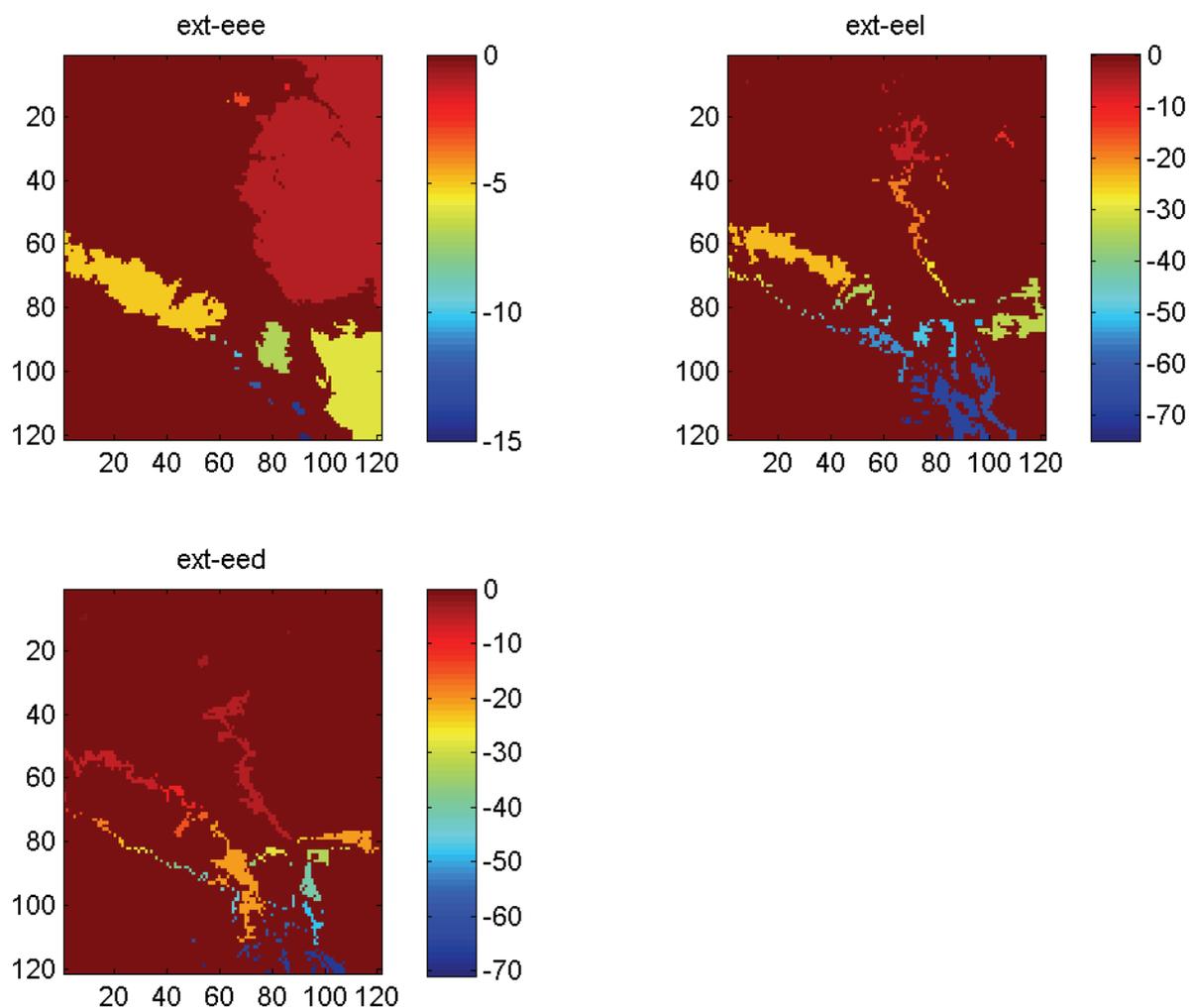


Figura 99. Extractos obtenidos de las firmas “elev-elev-elev”, “elev-elev-llan” y “elev-elev-depr” del DEM de Chicago

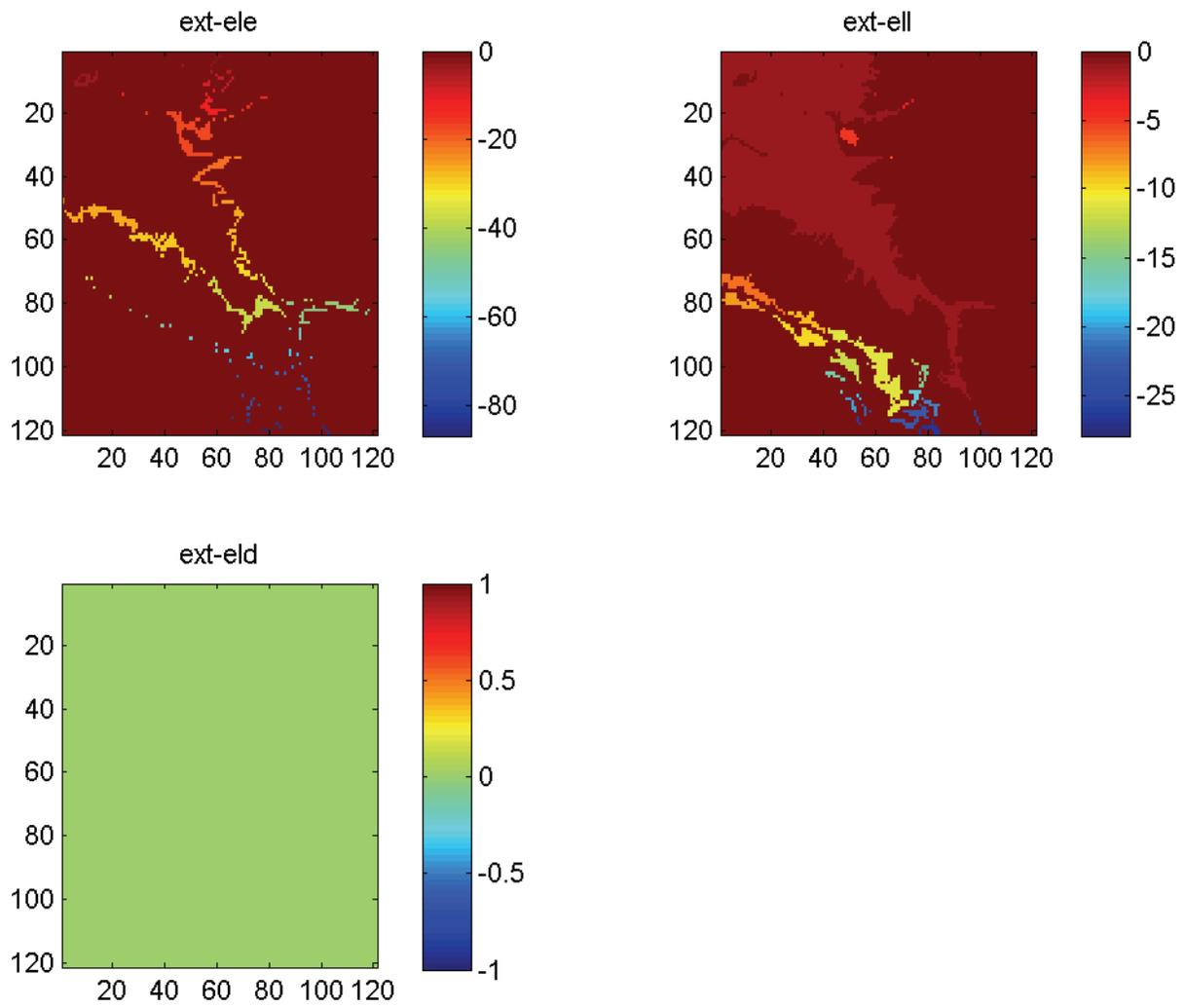


Figura 100. Extractos obtenidos de las firmas “elev-llan-elev”, “elev-llan-llan” y “elev-llan-depr” del DEM de Chicago

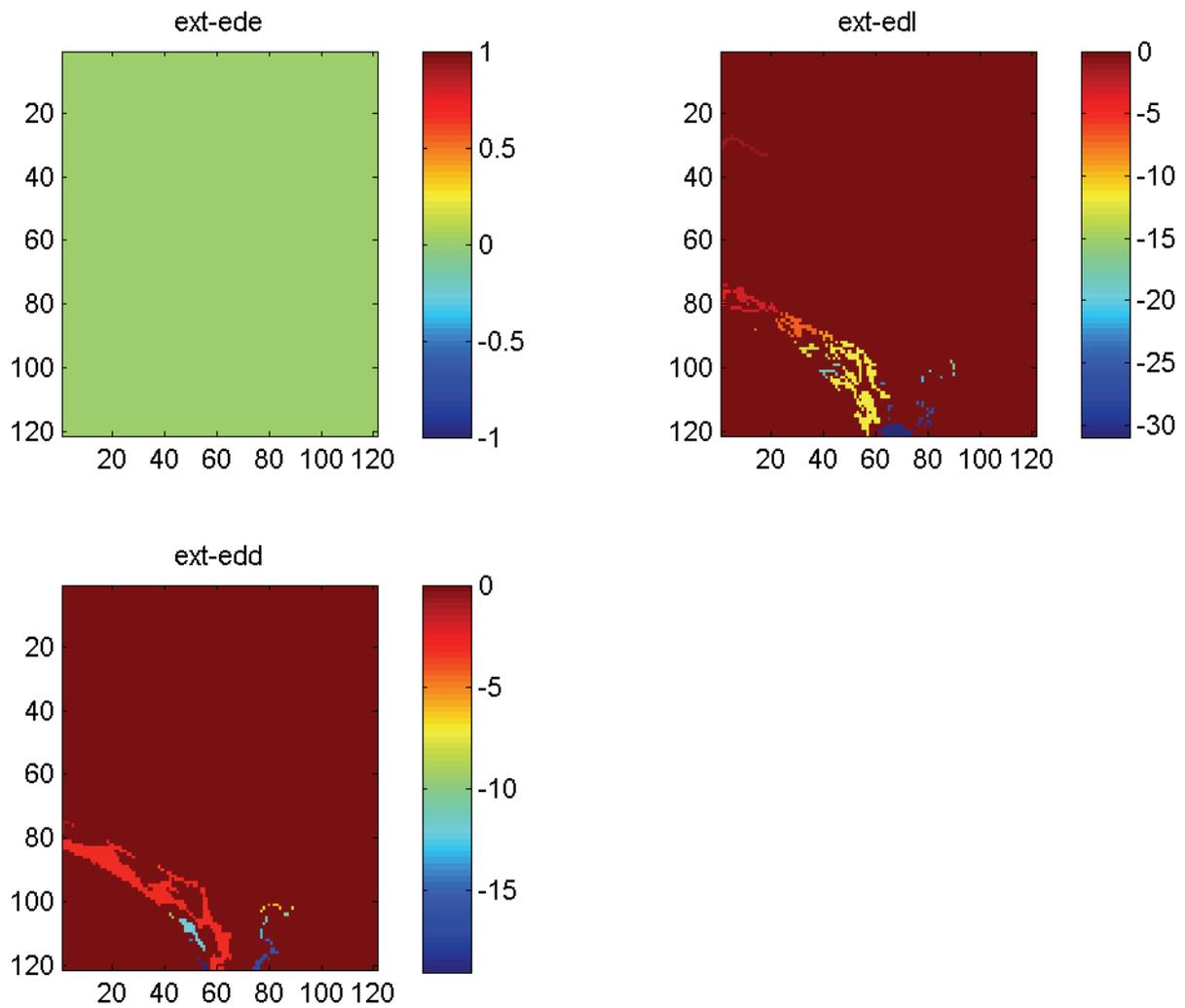


Figura 101. Extractos obtenidos de las firmas “elev-depr-elev”, “elev-depr-llan” y “elev-depr-depr” del DEM de Chicago

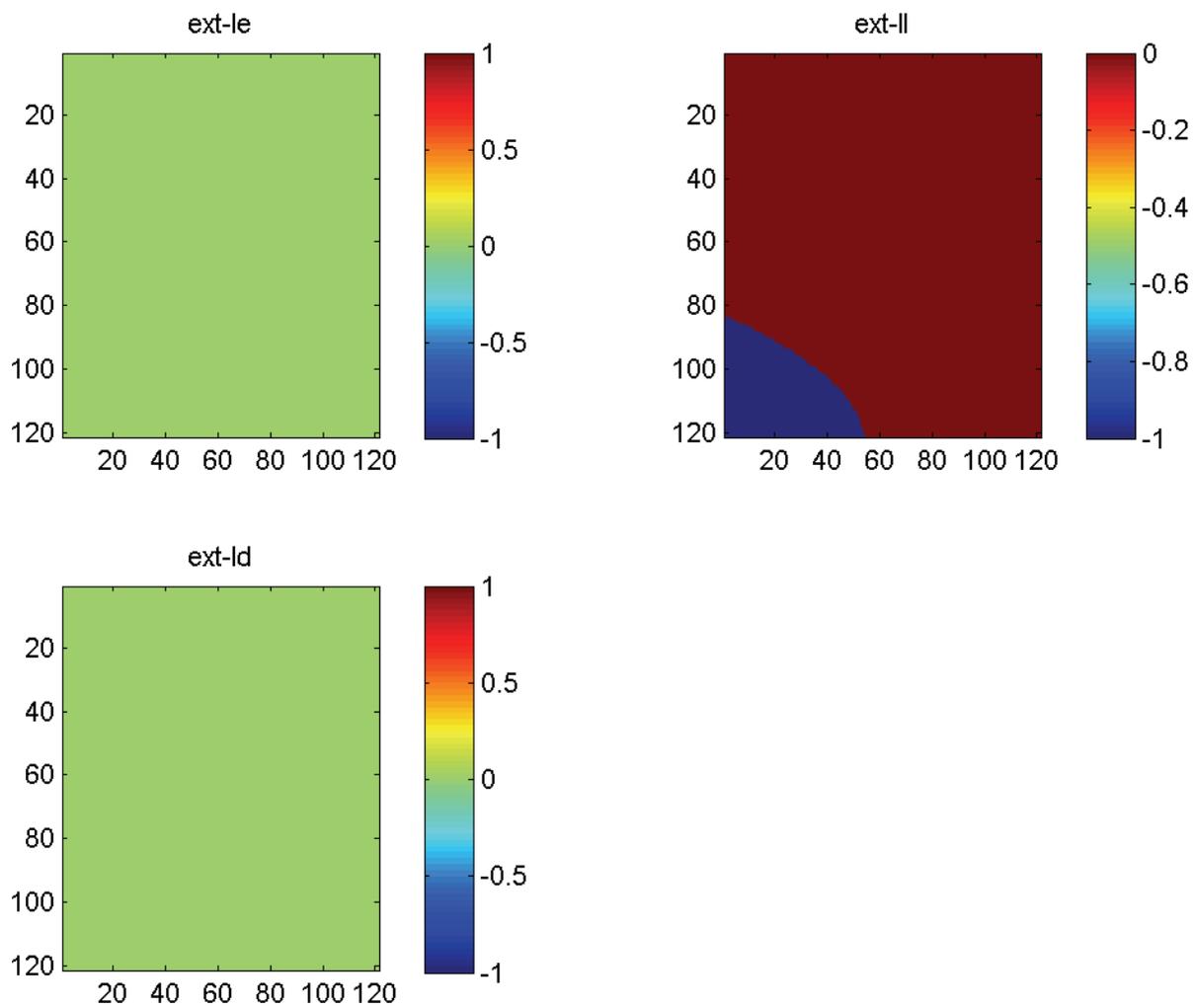


Figura 102. Extractos obtenidos de las firmas “llan-elev”, “llan-llan” y “llan-depr” del DEM de Chicago

Podemos ver que el único extracto que se obtiene en las firmas de llanura es el erróneamente clasificado Lago Michigan (ext-ll).

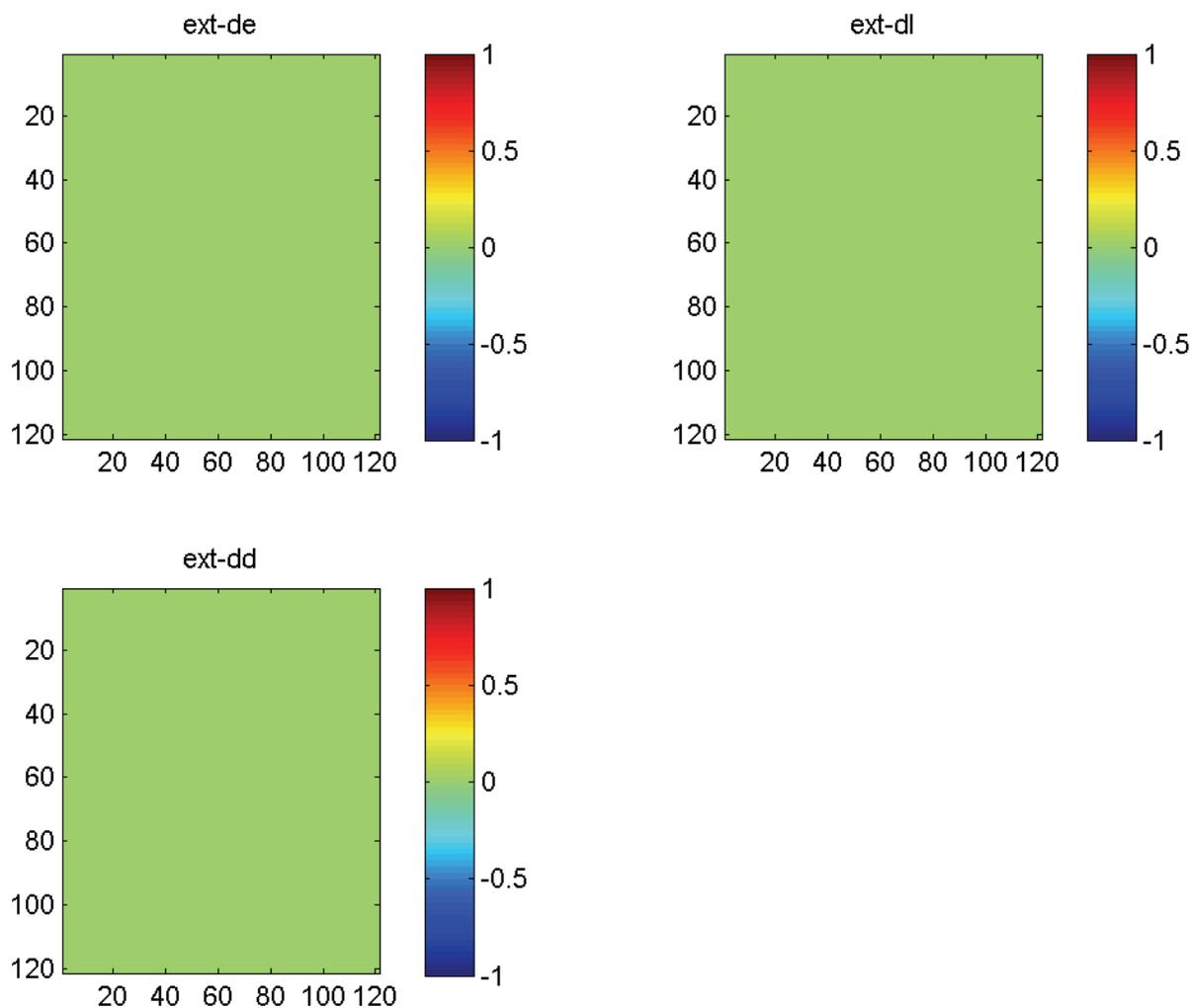


Figura 103. Extractos obtenidos de las firmas “depr-elev”, “depr-llan” y “depr-depr” del DEM de Chicago

En la Figura 104 se muestra parte de la descripción del DEM de Chicago utilizando los extractos obtenidos en las etapas anteriores.

```

CONJUNTO DE DATOS ESPACIALES RASTER DEL MODELO DIGITAL DE ELEVACIÓN DE "CHICAGO - E
IN,MI " QUE TIENE UNA RESOLUCION ESPACIAL DE 30
SEGUNDOS-ARCO. SU ALTITUD MINIMA ES DE 177 METROS Y SU ALTITUD MAXIMA ES DE 290 METROS. LAS COORDENADAS EXTREMAS SON (-309600,147600) Y (-313200,151200)
SEGUNDOS-ARCO EN PROYECCION GEOGRAFICA Y CONTIENE:

UNA MONTAÑA CON AREA DE 15300 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 237.000000 METROS Y MÁXIMA DE 246.000000 METROS, SUS COORDENADAS EXTREMAS SON
(NaN, NaN) Y (NaN, NaN) SEGUNDOS-ARCO. SU CIMA SE ENCUENTRA EN LA COORDENADA (NaN, NaN) A UNA ALTITUD DE 246.000000 METROS

UNA MONTAÑA CON AREA DE 617400 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 237.000000 METROS Y MÁXIMA DE 274.000000 METROS, SUS COORDENADAS EXTREMAS SON
(NaN, NaN) Y (NaN, NaN) SEGUNDOS-ARCO. SU CIMA SE ENCUENTRA EN LA COORDENADA (NaN, NaN) A UNA ALTITUD DE 274.000000 METROS

...

UNA CORDILLERA CON AREA DE 77400 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 229.000000 METROS Y MÁXIMA DE 236.000000 METROS, SUS COORDENADAS EXTREMAS SON
(NaN, NaN) Y (NaN, NaN) SEGUNDOS-ARCO. SU CIMA SE ENCUENTRA EN LA COORDENADA (NaN, NaN) A UNA ALTITUD DE 236.000000 METROS

UNA CORDILLERA CON AREA DE 8100 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 235.000000 METROS Y MÁXIMA DE 236.000000 METROS, SUS COORDENADAS EXTREMAS SON
(NaN, NaN) Y (NaN, NaN) SEGUNDOS-ARCO. SU CIMA SE ENCUENTRA EN LA COORDENADA (NaN, NaN) A UNA ALTITUD DE 236.000000 METROS
    
```

```
...  
UN MONTE CON AREA DE 9900 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 222.000000 METROS Y MÁXIMA DE 226.000000 METROS, SUS COORDENADAS EXTREMAS SON (NaN,  
NaN) Y (NaN, NaN) SEGUNDOS-ARCO. SU CIMA SE ENCUENTRA EN LA COORDENADA (NaN, NaN) A UNA ALTITUD DE 226.000000 METROS  
  
UN MONTE CON AREA DE 182700 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 222.000000 METROS Y MÁXIMA DE 228.000000 METROS, SUS COORDENADAS EXTREMAS SON (NaN,  
NaN) Y (NaN, NaN) SEGUNDOS-ARCO. SU CIMA SE ENCUENTRA EN LA COORDENADA (NaN, NaN) A UNA ALTITUD DE 228.000000 METROS  
  
...  
  
UN CERRO CON AREA DE 18900 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 219.000000 METROS Y MÁXIMA DE 221.000000 METROS, SUS COORDENADAS EXTREMAS SON (NaN,  
NaN) Y (NaN, NaN) SEGUNDOS-ARCO. SU CIMA SE ENCUENTRA EN LA COORDENADA (NaN, NaN) A UNA ALTITUD DE 221.000000 METROS  
  
UN CERRO CON AREA DE 20700 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 219.000000 METROS Y MÁXIMA DE 221.000000 METROS, SUS COORDENADAS EXTREMAS SON (NaN,  
NaN) Y (NaN, NaN) SEGUNDOS-ARCO. SU CIMA SE ENCUENTRA EN LA COORDENADA (NaN, NaN) A UNA ALTITUD DE 221.000000 METROS  
  
...  
  
UN ALTIPLANO CON AREA DE 19800 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 212.000000 METROS Y MÁXIMA DE 218.000000 METROS, SUS COORDENADAS EXTREMAS SON  
(NaN, NaN) Y (NaN, NaN) SEGUNDOS-ARCO. SU CIMA SE ENCUENTRA EN LA COORDENADA (NaN, NaN) A UNA ALTITUD DE 218.000000 METROS  
  
UN ALTIPLANO CON AREA DE 65700 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 202.000000 METROS Y MÁXIMA DE 218.000000 METROS, SUS COORDENADAS EXTREMAS SON  
(NaN, NaN) Y (NaN, NaN) SEGUNDOS-ARCO. SU CIMA SE ENCUENTRA EN LA COORDENADA (NaN, NaN) A UNA ALTITUD DE 218.000000 METROS  
  
...  
  
UNA LOMA CON AREA DE 43200 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 198.000000 METROS Y MÁXIMA DE 201.000000 METROS, SUS COORDENADAS EXTREMAS SON (NaN,  
NaN) Y (NaN, NaN) SEGUNDOS-ARCO. SU CIMA SE ENCUENTRA EN LA COORDENADA (NaN, NaN) A UNA ALTITUD DE 201.000000 METROS  
  
UNA LOMA CON AREA DE 39600 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 198.000000 METROS Y MÁXIMA DE 201.000000 METROS, SUS COORDENADAS EXTREMAS SON (NaN,  
NaN) Y (NaN, NaN) SEGUNDOS-ARCO. SU CIMA SE ENCUENTRA EN LA COORDENADA (NaN, NaN) A UNA ALTITUD DE 201.000000 METROS  
  
...  
  
UN COLLADO CON AREA DE 326700 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 178.000000 METROS Y MÁXIMA DE 197.000000 METROS, SUS COORDENADAS EXTREMAS SON  
(NaN, NaN) Y (NaN, NaN) SEGUNDOS-ARCO. SU CIMA SE ENCUENTRA EN LA COORDENADA (NaN, NaN) A UNA ALTITUD DE 197.000000 METROS  
  
UN COLLADO CON AREA DE 22500 SEGUNDOS-ARCO CUADRADOS, ALTITUD MÍNIMA DE 183.000000 METROS Y MÁXIMA DE 197.000000 METROS, SUS COORDENADAS EXTREMAS SON (NaN,  
NaN) Y (NaN, NaN) SEGUNDOS-ARCO. SU CIMA SE ENCUENTRA EN LA COORDENADA (NaN, NaN) A UNA ALTITUD DE 197.000000 METROS  
  
...
```

Figura 104. Parte de la descripción obtenida a partir del DEM de Chicago

6.

Conclusiones y trabajo futuro

En el presente trabajo de tesis hemos presentado una metodología para realizar una representación semántica de datos espaciales *raster*. Dicha metodología consta de tres etapas: Conceptualización, Síntesis y Descripción.

Como parte del caso de estudio se desarrolló una ontología del dominio geográfico (utilizando los diccionarios de datos del INEGI), en la cual se discernen algunas clases importantes para el dominio; el resultado es una ontología a la que le llamamos ontología *Kaab*. De manera similar, utilizamos el diccionario de la Real Academia de la Lengua para obtener las definiciones de los conceptos del dominio de las formas del terreno; también utilizamos como fuente de información la Especificación de Códigos para Datos Ambientales dada por la Organización Internacional de Estándares (ISO – por sus siglas en inglés) (ISO, 2005), el resultado de esto es la ontología que llamamos *Hunxeet*. Para ello proponemos una metodología de conceptualización con un mínimo de relaciones axiomáticas, lo que permite trasladar las relaciones directamente a la conceptualización con la ventaja de una mayor semántica en la definición de tales relaciones. La reducción de relaciones axiomáticas se logra al utilizar solamente un par³³ de conjuntos pequeños de relaciones axiomáticas, con las cuales es posible definir el resto de las posibles relaciones presentes en el dominio geográfico.

³³ El primer conjunto consta de las relaciones "es", "tiene" y "hace". El segundo conjunto es el conjunto de las preposiciones del español. Ambos conjuntos, utilizados en la forma simple o compleja, permiten realizar la conceptualización del dominio geográfico.

Para la conceptualización de la aplicación, así como para el proceso de síntesis, propusimos una metodología basada en la diferenciación de criterios de agrupamiento. Esto permite obtener extractos de información, que cumplen con los criterios establecidos hasta cierto nivel de abstracción. El resultado de esta conceptualización de la aplicación es una ontología que llamamos *Wiinkil*.

La etapa de síntesis es llevada a cabo en la forma en que comúnmente se procesan las imágenes, con fases de pre-procesamiento, procesamiento y post-procesamiento. Como resultado de esta etapa encontramos segmentos del CDER a los que llamamos extractos, los cuales son interpretados como una instancia de algún concepto descrito en la conceptualización.

La siguiente etapa (la de descripción), se lleva a cabo utilizando la conceptualización, que indica las propiedades que se deben medir de cada concepto. Para ello establecemos algunas plantillas que se llenan de acuerdo a las mediciones que se realizan. En realidad, la descripción puede ser hecha hacia cualquier tipo de representación que se desee³⁴. En el trabajo propusimos, aunque no desarrollamos, la utilización de formas de referencia, a las que denominamos bloques constructivos, las cuales al ser combinadas pueden formar un modelo sintético de los datos *raster* que se analizan.

Se están desarrollando un par de tesis de maestría enfocadas a explotar los conceptos descritos en este trabajo. La primera de ellas desarrolla un sistema para realizar la descripción semántica de modelos digitales de elevación. El segundo trabajo de tesis tiene que ver con algunos elementos que quedan pendientes en la presente tesis; se trata de la conceptualización de las relaciones topológicas entre los conceptos existentes en el dominio de las formas del terreno. Es, desde cierto punto de vista, una extensión de la primera tesis, al incluir en la conceptualización las relaciones topológicas. Esto permitirá realizar descripciones más ricas y útiles de los modelos de elevación.

³⁴ Obviamente, siempre y cuando cumpla con la conceptualización.

Otro asunto que queda pendiente en este trabajo, es el análisis semántico sobre la descripción obtenida. Esto es, una vez que contamos con la descripción semántica de CDER, debemos responder a preguntas sobre ellos, lo que es lo mismo, cómo podemos hacer consultas sobre las descripciones.

Otro aspecto importante es la validación que se pueda hacer de los resultados: con base en la conceptualización pudiéramos conocer de antemano que un tipo de segmento no puede ser vecino (estar dentro de, o rodeado de) de otro tipo distinto. Con esta validación semántica se podrían encontrar errores en los algoritmos de segmentación, aunque para ello es necesario contar con una buena conceptualización del dominio al que se han de mapear los segmentos.

6.1. Logros

Haciendo un resumen de los logros que se han obtenido con el desarrollo de este trabajo, podemos puntualizar lo siguiente:

- Una metodología para la representación semántica de datos espaciales *raster*, la cual consta de tres etapas bien identificadas: conceptualización, síntesis y descripción.
- Una metodología para la conceptualización de los resultados de los algoritmos de segmentación, mediante la cual se puede conocer de antemano la interpretación de los segmentos de datos generados; con la posible ventaja de poder realizar validaciones semánticas de los resultados.
- Una ontología del dominio geográfico basada en la metodología de conceptualización propuesta y en los diccionarios de datos provistos por el INEGI.
- Una ontología del dominio de las formas del terreno basada en las definiciones dadas por el Diccionario de la Lengua, así como de la Especificación de Códigos para Datos Ambientales dada por ISO.
- Un conjunto de programas que mejoran una implementación previa de la metodología propuesta para el caso de estudio de los DEM (Villegas, 2007), que permiten realizar una descripción semántica de un modelo de este tipo.

6.2. Alcances y limitaciones

Como es de esperarse, la metodología propuesta tiene sus limitaciones. Por ello a continuación presentamos puntualmente los alcances de las diferentes partes de la metodología:

- La metodología para la representación semántica de DER que ha sido propuesta está pensada para funcionar sobre los DER en general, pero fue probada exclusivamente en DEM, por lo que es posible que algunas adecuaciones sean necesarias para aplicarla a otros tipos de DER.
- Aunque la metodología para la conceptualización del dominio geográfico ha sido creada con una visión general de la representación del conocimiento, en este trabajo la presentamos para cubrir las necesidades específicas de dicho dominio; aunque pudiera ser aplicada para otros, su funcionalidad fuera del entorno geográfico no ha sido probada.
- La conceptualización de los resultados de los algoritmos está supeditada a que se debe saber de antemano el número de segmentos posibles que ha de generar cada algoritmo; a que se cuenta con una conceptualización del dominio sobre el cual se puedan realizar las descripciones de los segmentos; a que se debe saber de antemano, por medio de un experto, el significado de los distintos resultados posibles.
- Las ontologías creadas reflejan el conocimiento contenido exclusivamente en las fuentes que ya se han mencionado; a excepción de la de aplicación, la información contenida en las ontologías ha sido tomada tal y como se encontró. Otras posibles fuentes de información pudieran ser utilizadas para enriquecerlas. Sin embargo, nuevas fuentes de conocimiento requieren ser integradas, lo cual puede provocar cambios significativos en las ontologías que hemos creado.
- Los programas mencionados anteriormente han sido creados exclusivamente para demostrar la viabilidad de la metodología propuesta, con las limitantes que ya han sido expresadas anteriormente. Además, tales códigos no han sido optimizados ni en tiempo, ni en memoria; por lo que al utilizarlos con grandes

volúmenes de datos pudiera ocurrir una larga demora o incluso algún fallo en el sistema.

6.3. Trabajo futuro

Como consecuencia de los alcances y limitaciones que hemos presentado, es posible puntualizar algunos aspectos que en un futuro deben ser desarrollados:

- Como se ha mencionado previamente, una vez que se han identificado los conceptos (y sus respectivas instancias) presentes en un CDER, es necesario analizar y conceptualizar las relaciones (por ejemplo topológicas y geométricas) existentes entre los conceptos identificados y descritos en este trabajo.
- El proceso de síntesis permite obtener una representación del CDER, la cual tiene cierta calidad. Para determinar dicha calidad es necesario definir en qué consiste y cómo es posible medirla.
 - Para ello podríamos contar con métricas (y algoritmos para determinarlas); obviamente estas métricas dependerán del contexto (caso de estudio). Se propone que el uso de los bloques constructivos o formas de referencia para comparar la fidelidad de la representación con estas formas de referencia y así obtener una medida de la calidad. La idea de este enfoque consiste en que una vez que se ha construido el modelo sintético, con base en las formas de referencia, obtendremos un CDER que podemos comparar numéricamente con el CDER original, y así obtener una medida del error contenido en el modelo sintético.
 - Otra idea en el sentido de medir la calidad de la representación obtenida, es realizar la representación del CDER a través de una lógica de primer orden. Así, al realizar la representación lógica del CDER original y del sintético es posible, mediante la comparación de las lógicas resultantes, medir la fidelidad de la representación sintética comparativamente con la de los datos originales.

- En la parte de pruebas y resultados, resulta claro que se debe hacer algo para poder interoperar las ontologías implementadas en *Protegé* (que generan código OWL) con los algoritmos implantados en *MATLAB*. Vislumbramos claramente al menos dos opciones:
 - La primera opción es reescribir los algoritmos en un lenguaje, por ejemplo *Java*, capaz de inter-operar con la conceptualización generada en *Protegé*.
 - Una segunda opción consiste en crear algún programa *bridge* capaz de interactuar con ambas herramientas, dejando a *MATLAB* el procesamiento de los datos y a *Protegé* la explotación de las ontologías.
- Como ya se ha expuesto, queda como tarea pendiente la reasignación de conceptos a los extractos con base en la conceptualización y en las propiedades intrínsecas de los extractos.

Bibliografía

- Able-Software.** (2000). "R2V." en <http://www.ablesw.com/r2v/>.
- Ackermann, F.** (1993). Automatic generation of digital elevation models. OEEPE Commision B, DTM Accuracy Meeting. Southampton.
- Bezdek, J. C.** (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. New York, Plenum Press.
- Bodansky, E.** (2002). System Approach to *raster-to-vector* conversion: From research to commercial system. International Workshop on Semantic Processing of Spatial Data Mexico City, National Politechnical Institute.
- Brunsdén, D.** (1996). "Geomorphological events and landform change." Zeitschrift für Geomorphologie **40**: 273-288.
- Burrough, P. A.** (1986). Principles of Geographical Information Systems for Land Resources Assessment OUP. Methods of interpolation, Oxford.
- Buttenfield, B. P. y McMaster, R. B.** (1991). Map Generalization: Making Rules for Knowledge Representation, Harlow: Longman, Scientific and Technical.
- Castleman, K. R.** (1996). Digital Image Processing. NJ, USA, Prentice Hall.
- Chase, C. G.** (1992). "Fluvial landsculpting and the fractal dimension of topography." Geomorphology **5**: 39-57.
- Chevriaux, Y., Saux, E. y Claramunt, C.** (2005). A Landform-based Approach for the Representation of Terrain Silhouettes. 13th annual ACM international workshop on Geographic information systems Bremen, Germany, ACM Press.
- Chorowicz, J., Parrot, J. y Taud, H.** (1995). "Automated pattern-recognition of geomorphic features from DEMs and satellite images." Zeitschrift Für Geomorphologie N.F.(101): 69-84.
- De Boer, D. H.** (1992). "Hierarchies and spatial scale in process geomorphology: a review." Geomorphology **4**: 303-318.
- Dietler, G. y Zhang, Y.** (1992). "Fractal aspects of the Swiss landscape." Physica A(191): 213-219.

- Dikau, R.** (1989). The application of a digital relief model to landform analysis in geomorphology. Three dimensional applications in GIS, London, Taylor & Francis.
- Dikau, R.** (1990). "Derivates from detailed geoscientific maps using computer methods." *Zeitschrift fur Geomorphologie* **80**: 45-55.
- Dikau, R.** (1994). "Computergestützte Geomorphographie und ihre Anwendung in der Regionalisierung des Reliefs." *Petermanns Geographische Mitteilungen* **138**: 99-114.
- Doermann, D.** (1998). "An introduction to vectorization & segmentation." *Lecture Notes in Computer Science* **1389**: 1-8.
- Dunn, J. C.** (1973). "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters." *Journal of Cybernetics* **3**: 32-57.
- e-Cognition, D. I. G.** (2003). "Object oriented image analysis." en <http://www.definiens-imaging.com/ecognition/>.
- Española, R. A.** (2007). *Diccionario de la Lengua Española*.
- Etzel Müller, B. y Sulebak, J. R.** (2000). "Developments in the use of digital elevation models in periglacial geomorphology and glaciology." *Physische Geographie* **41**: 35-58.
- Evans, I. S.** (1972). *General geomorphometry, derivatives of altitude, and descriptive statistics. Spatial Analysis in Geomorphology*, Methuen, London.
- Evans, I. S.** (1979). *An integrated system of terrain analysis and slope mapping*, University of Durham, England.
- Evans, I. S.** (1980). *An integrated system of terrain analysis and slope mapping. Zeitschrift fur Geomorphologie, Suppl-Bd.*
- Evans, I. S.** (1984). *Correlation structures and factor analysis in the investigation of data dimensionality: statistical properties of the Wessex land surface. International Symposium on Spatial Data Handling, Zurich.*
- Evans, I. S. y McClean, C. J.** (1995). "The land surface is not unifractal: variograms, cirque scale and allometry." *Zeitschrift fur Geomorphologie* **101**: 127-147.
- Feder, J.** (1988). *Fractals*, Plenum Press.
- Felicísimo, A. M.** (2004). *Modelos Digitales del Terreno, Introducción y aplicaciones en las ciencias ambientales*. Oviedo, Universidad de Oviedo.
- Fernández-López, M., Gómez-Pérez, A. y Jurista, N.** (1997). *METHONTOLOGY: From ontological art towards ontological engineering*, Stanford.
- Frank, A. U.** (1996). *The prevalence of objects with sharp boundaries in GIS. Geographic Objects with Indeterminate Boundaries*, Taylor & Francis.
- Friedrich, K.** (1996). *Digital relief arrangement procedures for the derivative of relevant unit areas boundaries*. Frankfurt, University of Frankfurt **Ph.D.**

- Gallant, J. C. y Hutchinson, M. F.** (1996). Towards an understanding of landscape scale and structure. Third International Conference on Integrating GIS and Environmental Modelling. Santa Barbara: National Center for Geographic Information and Analysis.
- Gangemi, A., Pisanelli, D. y Steve, G.** (1998). Ontology Alignment: Experiences With Medical Terminologies. Formal Ontology in Information Systems. N. Guarino.
- Garbrecht, J. y Martz, L.** (1993) "Grid size dependency of parameters from digital elevation models." 12 DOI:
- Gómez-Pérez, A., Fernández-López, M. y De-Vicente, M.** (1996). Towards a method to conceptualize domain ontologies. Working notes of the workshop on Ontological Engineering, ECAI'96, ECCAI.
- González-Gómez, E. y Levachkine, S.** (2004). "Color Cartographic Pattern Recognition using Croase to Fine Scale Method." Lecture Notes in Computer Science **3287**: 533-540.
- Gonzalez, R. C. y Woods, R. E.** (1992). Digital Image Processing, Prentice Hall.
- Goodchild, M. F.** (1997). "Scale in a Digital World." Geographical & Environmental Modelling **1**: 5-23.
- Graff, L. H. y Usery, E. L.** (1993). "Automated classification of generic terrain features in digital elevation models." Photogrammetric Engineering and Remote Sensing **59**: 1409-1417.
- Graps, A.** (1995). "An introduction to wavelets." IEEE Computational Science and Engineering **2**(2): 18.
- Gruber, T. R.** (1993). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. Formal Ontology in Conceptual Analysis and Knowledge Representation, Padova, Italy, Kluwer Academic Publishers.
- Guarino, N.** (1998). Formal Ontology and Information Systems. Formal Ontology in Information Systems, Trento, Italy, IOS Press.
- Guzmán, G.** (2007). Un algoritmo general para la descomposición semántica de geo-imágenes. Laboratorio de Procesamiento Inteligente de Información Geoespacial. Ciudad de México, Instituto Politécnico Nacional. **Doctorado en Ciencias de la Computación**: 150.
- Hammond, E. H.** (1964). "Analysis of properties in land form geography: An application to broad-scale land form mapping." Annals of the Association of American Geographers **54**: 11-19.
- Haralick, R. M.** (1983). "Ridge and valley on digital images." Computer Vision, Graphics and Image Processing **22**(1): 28-38.
- Heerdegen, R. G. y Beran, M. A.** (1982). "Quantifying source areas through land surface curvature and slope." Journal of Hydrology **57**(3): 359-373.

- Hodgson, M. E.** (1995). "What cell size does the computed slope / aspect angle represent?" *Photogrammetric Engineering and Remote Sensing* **61**(5): 513-517.
- Horton, R.** (1982). *Tradition and Modernity Revisited. Rationality and Relativism*, Oxford, Blackwell.
- Huang, J. y Trucotte, D. L.** (1990). "Reply to comment by J. A. Goff on "Fractal mapping of digitized images"." *Journal of Geophysical Research* **95**(B): 5161.
- INEGI** (1996). *Diccionario de datos topográficos 1:50 000 (Vectorial)*. Aguascalientes, Instituto Nacional de Estadística Geografía e Informática: 115.
- INEGI** (2000a). *Normas Técnicas para la Elaboración de Ortofotos Digitales*, Instituto Nacional de Estadística, Geografía e Informática.
- INEGI** (2000b). *Modelos Digitales de Elevación, Generalidades y Especificaciones*, Instituto Nacional de Estadística, Geografía e Informática.
- ISO** (2005). *Information technology — Environmental Data Coding Specification (EDCS). 18025:2005(E)*. International.
- Jain, A. K.** (1984). *Fundamentals of digital image* NJ, USA, Prentice Hall.
- Kidd, A.** (1987). *Knowledge Acquisition for Expert Systems, A Practical Handbook*, Plenum Press.
- Klinkenberg, B. y Goodchild, M. F.** (1992). "The fractal properties of topography: a comparison of methods." *Earth Surface Processes and Landforms* **17**: 217-234.
- Kuhn, W.** (2003). *Implementing Semantic Reference Systems*. 6th Agile conference on Geographic Information Science, Lyon, France, AGILE.
- Kumler, M. P.** (1994). "An intensive comparison of triangulated irregular networks (TINs) and Digital Elevation Models (DEMs)." *Cartographica* **31**(2).
- Levachkine, S. y Polchkov, E.** (2000) "Integrated Technique for Automated Digitalization of Raster Maps." *Revista Digital Universitaria* **1**, DOI:
- Levachkine, S.** (2003). "System Approach to R2V Conversion for Analytical GIS." *Research on Computing Science(Semantic Processing of Spatial Data)*: 22-33.
- MacQueen, J. B.** (1967). *Some Methods for classification and Analysis of Multivariate Observations*. 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, University of California Press.
- Mahesh, K.** (1996). *Ontology Development for Machine Translation: Ideology and Methodology*. New Mexico, New Mexico State University, Computing Research Laboratory.
- Mandelbrot, B. B.** (1967). "How long is the coastline of Britain? Statistical self-similarity and fractional dimension." *Science* **156**: 636-638.
- Mandelbrot, B. B.** (1977). *Fractals: Form, Chance and Dimension*. San Fransisco, WH Freeman.
- Mandelbrot, B. B.** (1982). *The Fractal Geometry of Nature*. New York, WH Freeman.

- Mandelbrot, B. B.** (1986). Self-affine fractal sets. *Fractals in Physics*, Holland.
- Mark, D., Smith, B. y Tversky, B.** (1999). "Ontology and Geographic Objects: An Empirical Study of Cognitive Categorization." *Lecture Notes in Computer Science 1661*(Spatial Information Theory: A Theoretical Basis for GIS): 283-298.
- Mark, D. M.** (1975). "Geomorphometric parameters: a review and classification." *Geografiska Annaler* **57**(A): 165-177.
- Mark, D. M. y Aronson, P. B.** (1984). "Scale dependent fractal dimensions of topographic surfaces an empirical investigation, with applications in geomorphology and computer mapping." *Mathematical Geology* **16**: 671-682.
- Mark, D. M. y Turk, A. G.** (2003). "Landscape Categories in Yindjibarndi: Ontology, Environment, and Language.
- McClellan, C. J.** (1991). The scale-free and scale-bound properties of land surfaces: fractal analysis and specific geomorphometry from digital terrain models. Durham, University of Durham.
- McCray, A. T.** (1993). Representing biomedical knowledge in the UMLS semantic network. High performance medical libraries: Advances in information management for the virtual era: 45-55.
- Mizen, H., Dolbear, C. y Hart, G.** (2005). "Ontology Ontogeny: Understanding how an Ontology is created and developed." *Lecture Notes in Computer Science 3799*(GeoSpatial Semantics): 15-29.
- Montello, D. R., Goodchild, M. F., Gottesegen, J. y Fohl., P.** (2003). "Where's downtown? Behavioral methods for determining referents of vague spatial queries." *Spatial Cognition and Computation* **3**(2&3): 185-204.
- Moreno, M.** (2007). Similitud Semántica entre Sistemas de Objetos Geográficos Aplicada a la Generalización de Datos Geo-espaciales. Laboratorio de Procesamiento Inteligente de Información Geoespacial. Ciudad de México, Instituto Politécnico Nacional. **Doctorado en Ciencias de la Computación**: 150.
- Openshaw, S.** (1983). The modifiable area unit problem, Norwich.
- Outcalt, S. I. y Melton, M. A.** (1992). "Geomorphic application of the Hausdorff-Besicovich dimension." *Earth Surface Processes and Landforms* **17**: 775-787.
- OwlSeek.** (2007). "Ontologies As A Mechanism For Specifications In Computer Science." en <http://owlseek.com/>.
- Parrot, J. y Taud, H.** (1992). "Detection and classification of circular structures on SPOT images." *IEEE Trans. Geoscience and Remote Sensing* **30**(5): 996-1005.
- Pentland, A.** (1984). "Fractal-based description of natural scenes." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 661-674.
- Peucker, T. K.** (1978). Data structures for digital terrain models - discussion and comparison. *Harvard Papers on Geographic Information Systems*.

- Pike, R. J. y Rozema, W. J.** (1975). "Spectral analysis of landforms." *Annals of the Association of American Geographers*(65): 499-516.
- Pike, R. J.** (1988). "The geometric signature: Quantifying landslide terrain types from digital elevation models." *Mathematical Geology* **20**(5): 491-511.
- Pike, R. J., Acevedo, W. y Card, D. H.** (1989). Topographic grain automated from digital elevation models. 1989 ASPRS/ACSM Auto-Carto9 Annual Convention, Baltimore.
- Polidori, L., Chorowicz, J. y Guillande, R.** (1991). "Description of terrain as a fractal surface and application to digital elevation model quality assessment." *Photogrammetric Engineering and Remote Sensing* **57**: 1329-1332.
- Quintero, R., Guzmán, G., Torres, M. y Moreno, M.** (2002). Real-Time Elevation Map Viewer for 3D Spatial Analysis. International Workshop on Semantic Processing of Spatial Data (GEOPRO 2002), Mexico City, National Polytechnical Institute.
- Quintero, R., Guzmán, G., Torres, M. y Moreno, M.** (2003a). Spatial Analysis based on 3D Rendering Techniques. 6th AGILE Conference on Geographic Information Science, Lyon, France.
- Quintero, R., Moreno, M., Levachkine, S. y Torres, M.** (2003b). Scale Changing in Digital Elevation Models using Bicubic Parametric Patches for Real-time Simulations. 8th Iberoamerican Congress on Pattern Recognition, Havana, Cuba.
- Rees, W. G.** (1995). Characterisation and imaging of fractal topography. *Fractals in Geosciences and Remote Sensing*. G. Wilkinson, I. Kanellopoulos and J. Mégier, Joint Research Centre: 298-324.
- Romstad, B.** (2001). Improving Relief Classification with Contextual Merging. *Proceedings. ScanGIS 2001*, Ås, Norway.
- Rosen, K. H.** (1995). *Discrete Mathematics and its Applications*. New York, McGraw-Hill.
- Salge, F.** (1995). *Semantic accuracy. Elements of Spatial Data Quality*, Elsevier.
- Smith, B.** (1995a). "The Structures of the Commonsense World." *Acta Philosophica Fennica* **58**(290-317): 290.
- Smith, B.** (1995b). "Formal Ontology, Common Sense, and Cognitive Science." *International Journal of Human-Computer Studies* **43**: 641-667.
- Smith, B. y Mark, D.** (2001). "Geographic categories: An ontological investigation." *International Journal of Geographic Information Science* **15**(7): 591-612.
- Smith, B. y Mark, D. M.** (2003). "Do Mountains Exist? Towards an Ontology of Landforms." *Environment & Planning B (Planning and Design)* **30**(3): 411-427.
- Smith, B.** (2004). Beyond Concepts: Ontology as Reality Representation. International Conference on Formal Ontology and Information Systems, Turin, Italy.
- Sonka, M., Hlavac, V. y Boyle, R.** (1999). *Image Processing, Analysis and Machine Vision*, Thomson-Engineering.

- Sulebak, J. R., Eitzelmüller, B. y Sollid, J. L.** (1997). "Landscape regionalisation by automatic classification of terrain elements." *Norsk geografisk Tidsskrift* **51**: 35-46.
- Sulebak, J. R.** (1999). "Fractal analysis of surface topography." *Norsk geografisk Tidsskrift* **53**: 213-225.
- Sulebak, J. R., Tallaksen, L. M. y Erichsen, B.** (2000). "Estimation of areal soil moisture by use of terrain data." *Geografiska Annaler* **82**(A): 89-105.
- Tomita, F., Shirai, Y. y Tsuji, S.** (1982). "Description of texture by structural analysis." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **4**(2): 183-191.
- Torres, M., Levachkine, S., Moreno, M., Quintero, R. y Guzmán, G.** (2004). TOGWA: Web-Mapping Application to Retrieve Spatial Data by means of Spatial Ontologies. Third International Conference on Geographic Information Science, University of Maryland, Adelphi, USA.
- Torres, M., Quintero, R., Moreno, M. y Guzmán, G.** (2005). "Spatial Semantic Definition to Generate Semantic Description for Spatial Data." *Research on Computing Science* **14**(Advances in: Artificial Intelligence and Computing Science): 297-310.
- Torres, M.** (2007). Representación ontológica basada en descriptores semánticos aplicada a objetos geográficos. Laboratorio de Procesamiento Inteligente de Información Geoespacial. Ciudad de México, Instituto Politécnico Nacional. **Doctorado en Ciencias de la Computación**: 150.
- Tribe, A.** (1990). Towards the automated recognition of landforms (valley heads) from digital elevation models. 4th International Symposium on Spatial Data Handling, Zurich, Switzerland.
- Uschold, M. y King, M.** (1995). A Methodology for Building Ontologies. IJCA195 Workshop on Basic Ontological Issues in Knowledge Sharing. Montreal, Canada.
- USGS.** (2007a). "Digital Elevation Models." en www.usgs.gov.
- USGS.** (2007b). "Directory of 1:250,000-scale Digital Elevation Models." en <http://edcftp.cr.usgs.gov/pub/data/DEM/250/>.
- Villegas, V.** (2007). Identificación de Conceptos Geomorfológicos en Modelos Digitales de Elevación. Centro de Investigación en Computación. México, Instituto Politécnico Nacional. **MSc**: 120.
- Voss, R. F.** (1985). Random fractal forgeries. *Fundamental Algorithms in Computer Graphics*, New York, Springer-Verlag.
- Weibel, R. y DeLotto, J. L.** (1988). Automated terrain classification for GIS modelling. *Proceedings. GIS/LIS 88*, Virginia.
- Weibel, R. y Heller, M.** (1991). *Digital Terrain Modelling. Geographical Information Systems: Principles and Applications*, Longman, London.
- Wood, J.** (1997). The Geomorphological Characterisation of Digital Elevation Models. Department of Geography. Leicester, UK, University of Leicester. **Ph. D.**

Wood, W. F. y Snell, J. B. (1957). The dispersion of geomorphic data around measures of central tendency and its application, US Army Quartermaster Research and Development Center.

A nexos

A.1. Código fuente en MATLAB

El código que implementa la metodología, escrito en MATLAB se presenta en la Figura 105.

```
% Función : tesis
% Desc : dado un conjunto de datos de elevación, genera una
% descripción semántica del mismo
% Parámetros:
% dem = conjunto de datos de elevación, una matriz
% demParam = parámetros del modelo de elevación
% red = factor de reducción
function [ ] = tesis( dem, demParam, red, dir, tipo )
    [m n] = size(dem);

    for i=1:m
        for j=1:n
            if( dem(i,j)>(demParam.no_data-0.5) && dem(i,j)<(demParam.no_data+0.5) )
                dem(i,j) = 0;
            end
        end
    end

    demParam.res = demParam.res * red;
    workDem = dem(1:red:m, 1:red:n);

    disp('-----');
    disp('Segmentacion ...');
    disp('-----');
    disp('Primer nivel ...');
    [elev, llan, depr, zle] = segmentacion(workDem);

    disp('Segundo nivel ...');
    [elev_elev, elev_llan, elev_depr, elev_zle] = segmentacion(elev);
    [llan_elev, llan_llan, llan_depr, llan_zle] = segmentacion(llan);
    [depr_elev, depr_llan, depr_depr, depr_zle] = segmentacion(depr);

    disp('Tercer nivel ...');
    [elev_elev_elev, elev_elev_llan, elev_elev_depr, elev_elev_zle] = segmentacion(elev_elev);
    [elev_llan_elev, elev_llan_llan, elev_llan_depr, elev_llan_zle] = segmentacion(elev_llan);
    [elev_depr_elev, elev_depr_llan, elev_depr_depr, elev_depr_zle] = segmentacion(elev_depr);

    disp('-----');
    disp('Extracción ...');
```

```

disp('-----');
total = 15;
[ext_eee, eee_plot] = extraccion(elev_elev_elev); disp(sprintf('Terminado 1 de %d', total));
[ext_eel, eel_plot] = extraccion(elev_elev_llan); disp(sprintf('Terminado 2 de %d', total));
[ext_eed, eed_plot] = extraccion(elev_elev_depr); disp(sprintf('Terminado 3 de %d', total));
[ext_ele, ele_plot] = extraccion(elev_llan_elev); disp(sprintf('Terminado 4 de %d', total));
[ext_ell, ell_plot] = extraccion(elev_llan_llan); disp(sprintf('Terminado 5 de %d', total));
[ext_eld, eld_plot] = extraccion(elev_llan_depr); disp(sprintf('Terminado 6 de %d', total));
[ext_ede, ede_plot] = extraccion(elev_depr_elev); disp(sprintf('Terminado 7 de %d', total));
[ext_edl, edl_plot] = extraccion(elev_depr_llan); disp(sprintf('Terminado 8 de %d', total));
[ext_edd, edd_plot] = extraccion(elev_depr_depr); disp(sprintf('Terminado 9 de %d', total));
[ext_le, le_plot] = extraccion(llan_elev); disp(sprintf('Terminado 10 de %d', total));
[ext_ll, ll_plot] = extraccion(llan_llan); disp(sprintf('Terminado 11 de %d', total));
[ext_ld, ld_plot] = extraccion(llan_depr); disp(sprintf('Terminado 12 de %d', total));
[ext_de, de_plot] = extraccion(depr_elev); disp(sprintf('Terminado 13 de %d', total));
[ext_dl, dl_plot] = extraccion(depr_llan); disp(sprintf('Terminado 14 de %d', total));
[ext_dd, dd_plot] = extraccion(depr_depr); disp(sprintf('Terminado 15 de %d', total));

disp('-----');
disp('Descripción ...');
disp('-----');

[ tr guardar ] = size(dir);
if guardar
    disp(['CREANDO DIR: ', dir, '... ']);
    mkdir(dir);
    disp(sprintf('GUARDANDO DESCRIPCION EN %s\\desc.txt ...', dir));
    fid = fopen(sprintf('%s\\desc.txt', dir), 'w');
    fprintf(fid, descripcionDEM(demParam));
    fprintf(fid, descripcionMultiple(ext_eee, 'eee', demParam));
    fprintf(fid, descripcionMultiple(ext_eel, 'eel', demParam));
    fprintf(fid, descripcionMultiple(ext_eed, 'eed', demParam));
    fprintf(fid, descripcionMultiple(ext_ele, 'ele', demParam));
    fprintf(fid, descripcionMultiple(ext_ell, 'ell', demParam));
    fprintf(fid, descripcionMultiple(ext_eld, 'eld', demParam));
    fprintf(fid, descripcionMultiple(ext_ede, 'ede', demParam));
    fprintf(fid, descripcionMultiple(ext_edl, 'edl', demParam));
    fprintf(fid, descripcionMultiple(ext_edd, 'edd', demParam));
    fprintf(fid, descripcionMultiple(ext_le, 'le', demParam));
    fprintf(fid, descripcionMultiple(ext_ll, 'll', demParam));
    fprintf(fid, descripcionMultiple(ext_ld, 'ld', demParam));
    fprintf(fid, descripcionMultiple(ext_de, 'de', demParam));
    fprintf(fid, descripcionMultiple(ext_dl, 'dl', demParam));
    fprintf(fid, descripcionMultiple(ext_dd, 'dd', demParam));
    fclose(fid);

    disp('GENERANDO FIGURAS... ');
    figure();
    imagesc(dem), colorbar, colormap gray;
    figure();
    subplot(2,2,1); imagesc(zle), title('zle');
    subplot(2,2,2); imagesc(elev), title('elev');
    subplot(2,2,3); imagesc(llan), title('llan');
    subplot(2,2,4); imagesc(depr), title('depr');
    colormap gray;
    figure();
    subplot(2,2,1); imagesc(elev_zle), title('elev-zle');
    subplot(2,2,2); imagesc(elev_elev), title('elev-elev');
    subplot(2,2,3); imagesc(elev_llan), title('elev-llan');
    subplot(2,2,4); imagesc(elev_depr), title('elev-depr');
    colormap gray;
    figure();
    subplot(2,2,1); imagesc(llan_zle), title('llan-zle');
    subplot(2,2,2); imagesc(llan_elev), title('llan-elev');
    subplot(2,2,3); imagesc(llan_llan), title('llan-llan');
    subplot(2,2,4); imagesc(llan_depr), title('llan-depr');
    colormap gray;
    figure();

```

```

subplot(2,2,1); imagesc(depr_zle), title('depr-zle');
subplot(2,2,2); imagesc(depr_elev), title('depr-elev');
subplot(2,2,3); imagesc(depr_llan), title('depr-llan');
subplot(2,2,4); imagesc(depr_depr), title('depr-depr');
colormap gray;
figure();
subplot(2,2,1); imagesc(elev_elev_zle), title('elev-elev-zle');
subplot(2,2,2); imagesc(elev_elev_elev), title('elev-elev-elev');
subplot(2,2,3); imagesc(elev_elev_llan), title('elev-elev-llan');
subplot(2,2,4); imagesc(elev_elev_depr), title('elev-elev-depr');
colormap gray;
figure();
subplot(2,2,1); imagesc(elev_llan_zle), title('elev-llan-zle');
subplot(2,2,2); imagesc(elev_llan_elev), title('elev-llan-elev');
subplot(2,2,3); imagesc(elev_llan_llan), title('elev-llan-llan');
subplot(2,2,4); imagesc(elev_llan_depr), title('elev-llan-depr');
colormap gray;
figure();
subplot(2,2,1); imagesc(elev_depr_zle), title('elev-depr-zle');
subplot(2,2,2); imagesc(elev_depr_elev), title('elev-depr-elev');
subplot(2,2,3); imagesc(elev_depr_llan), title('elev-depr-llan');
subplot(2,2,4); imagesc(elev_depr_depr), title('elev-depr-depr');
colormap gray;
figure();
subplot(2,2,1); imagesc(eee_plot), colorbar, title('ext-eee');
subplot(2,2,2); imagesc(eel_plot), colorbar, title('ext-eel');
subplot(2,2,3); imagesc(eed_plot), colorbar, title('ext-eed');
figure();
subplot(2,2,1); imagesc(ele_plot), colorbar, title('ext-ele');
subplot(2,2,2); imagesc(ell_plot), colorbar, title('ext-ell');
subplot(2,2,3); imagesc(eld_plot), colorbar, title('ext-eld');
figure();
subplot(2,2,1); imagesc(ede_plot), colorbar, title('ext-ed');
subplot(2,2,2); imagesc(edl_plot), colorbar, title('ext-edl');
subplot(2,2,3); imagesc(edd_plot), colorbar, title('ext-edd');
figure();
subplot(2,2,1); imagesc(le_plot), colorbar, title('ext-le');
subplot(2,2,2); imagesc(ll_plot), colorbar, title('ext-ll');
subplot(2,2,3); imagesc(ld_plot), colorbar, title('ext-ld');
figure();
subplot(2,2,1); imagesc(de_plot), colorbar, title('ext-de');
subplot(2,2,2); imagesc(dl_plot), colorbar, title('ext-dl');
subplot(2,2,3); imagesc(dd_plot), colorbar, title('ext-dd');

disp('GUARDANDO FIGURAS... ');
for i=1:13
    saveas( i, sprintf('%s\\fig%02d.%s', dir, i, tipo) );
end
disp('OK');
else
disp('GUARDAR ESTÁ DESACTIVADO');
disp(descripcionDEM(demParam));
disp(descripcionMultiple(ext_eee, 'eee', demParam));
disp(descripcionMultiple(ext_eel, 'eel', demParam));
disp(descripcionMultiple(ext_eed, 'eed', demParam));
disp(descripcionMultiple(ext_ele, 'ele', demParam));
disp(descripcionMultiple(ext_ell, 'ell', demParam));
disp(descripcionMultiple(ext_eld, 'eld', demParam));
disp(descripcionMultiple(ext_ed, 'ede', demParam));
disp(descripcionMultiple(ext_edl, 'edl', demParam));
disp(descripcionMultiple(ext_edd, 'edd', demParam));
disp(descripcionMultiple(ext_le, 'le', demParam));
disp(descripcionMultiple(ext_ll, 'll', demParam));
disp(descripcionMultiple(ext_ld, 'ld', demParam));
disp(descripcionMultiple(ext_de, 'de', demParam));
disp(descripcionMultiple(ext_dl, 'dl', demParam));
disp(descripcionMultiple(ext_dd, 'dd', demParam));
end

```

```

close all;
end

%-----
% SEGMENTACION
%-----
% Función      : segmentacion
% Desc        : dado un conjunto de datos de elevación lo separa en tres
%              segmentos: elevaciones, llanuras y depresiones
% Parámetros:
%             dem      = conjunto de datos de elevación, una matriz
%             min y max = valor mínimo y máximo de curvatura para
%                       considerarla como "llano"
%             elev     = segmento de elevaciones, una matriz
%             llan     = segmento de llanuras, una matriz
%             depr     = segmento de depresiones, una matriz
%             zle      = zona llana mas extensa, una matriz
function [ elev, llan, depr, zle ] = segmentacion( dem )
    avgFilter = fspecial('average', 3);
    lapFilter = [-1, -1, -1; -1, 8, -1; -1, -1, -1]; % Mascara Laplaciano
    [m, n] = size(dem);

    sua = max( m, n ) / (100);
    outDem = dem;
    for i=1:sua
        outDem = imfilter( outDem, avgFilter, 'replicate');
    end
    outLap = imfilter( outDem, lapFilter, 'replicate');
    outClas = dividir( outLap);

    % eliminando regiones sin datos, es decir se marcan como una cuarta
    % clase los pixeles marcados como cero o como no_data

    for i=1:m
        for j=1:n
            if (dem(i,j) <= 0 )
                outClas(i,j) = 4;
            end
        end
    end

    % buscar la zle
    zle = findZLE(outClas);

    % calcular la altitud mínima y máxima de la zle
    mini = 10000;
    maxi = -10000;
    for i=1:m
        for j=1:n
            if zle(i,j) == 1
                if dem(i,j) > maxi
                    maxi = dem(i,j);
                end
                if dem(i,j) < mini
                    mini = dem(i,j);
                end
            end
        end
    end

    % iniciar las matrices con los segmentos
    elev = zeros(size(dem));
    llan = zeros(size(dem));
    depr = zeros(size(dem));

    % repartir los pixeles entre los segmentos, de acuerdo a su altitud
    for i=1:m
        for j=1:n

```

```

        if (dem(i,j) > 0 )           % es un dato válido ?
            if dem(i,j) < mini       % esta por debajo de la alt min de la zle ?
                depr(i,j) = dem(i,j); % mandarlo al segmento de depresiones
            elseif dem(i,j) > maxi    % esta por arriba de la alt max de la zle ?
                elev(i,j) = dem(i,j); % mandarlo al segmento de elevaciones
            else
                llan(i,j) = dem(i,j); % de lo contrario, mandarlo al segmento de llanuras
            end
        end
    end
end

% Función : dividir
% Desc : etiqueta los valores de curvatura de acuerdo a unos rangos
% : datos por un vector
% Parámetros:
% inData = conjunto de datos de entrada, una matriz
% range = rangos para el etiquetado, un vector
% outData = datos etiquetados, una matriz
function [ outData ] = dividir(inData)
    [m, n] = size(inData); % tamaño de la matriz
    outData = zeros(m, n); % crear datos de salida
    for i=1:m % para cada elemento de la matriz
        for j=1:n
            if inData(i, j) < -1
                outData(i, j) = 1;
            elseif inData(i, j) > 1
                outData(i, j) = 3;
            else
                outData(i, j) = 2;
            end
        end
    end
end

% Función : findZLE
% Desc : Calcula la Zona Llana mas Extensa
% Parámetros:
% clas = conjunto de datos de entrada, una matriz
% zle = la Zona Llana mas Extensa, una matriz
function [ zle ] = findZLE( clas )
    zle = zeros(size(clas)); % iniciar la zle con ceros
    [m n] = size(clas); % tamaño de la matriz
    curLabel = 10; % etiqueta para marcar las regiones
    curSize = 0; % tamaño de la zle, en píxeles
    theLabel = -1; % etiqueta de la zle

    % buscar regiones y marcarlas, al final la región de mayor tamaño será
    % la zle
    for i=1:m % para cada píxel en clas
        for j=1:n
            if( clas(i,j) == 2) % el valor de la curvatura del píxel es "llan" (2)
                [clas, tam] = regGrow8(clas, i, j, 2, curLabel); % crecer esa región llana
                if( curSize < tam ) % es la mas grande que se ha encontrado ?
                    curSize = tam; % guardar ese tamaño
                    theLabel = curLabel;% guardar la etiqueta de la región mas grande
                end
                curLabel = curLabel + 1;
            end
        end
    end

    % marcar la zle
    for i=1:m
        for j=1:n
            if( clas(i,j) == theLabel )
                zle(i,j) = 1;
            end
        end
    end
end

```

```

        end
    end
end

% Función : regGrow8
% Desc : Calcula la región 8 conexas dada una semilla
% Parámetros:
% clas = conjunto de datos de entrada, una matriz
% row y col = posición relativa a clas de la semilla
% sourceLabel = valor de la semilla mediante el cual se distingue
% la región.
% targetLabel = valor con el que se etiqueta la región.
% outClas = datos con la región etiquetada
% tam = tamaño de la región (en número de píxeles)
function [ outClas, tam ] = regGrow8( clas, row, col, sourceLabel, targetLabel )
    sp = 1; % stack pointer
    stack(1).x = row; % iniciar la pila con la semilla
    stack(1).y = col;
    [nRow nCol] = size(clas); % tamaño de la matriz clas
    tam = 0; % iniciar el tamaño de la región a cero

    while sp > 0 % mientras haya datos en la pila
        i = stack(sp).x; % sacar el dato en el tope de la pila
        j = stack(sp).y;
        sp = sp - 1;
        if( clas(i, j) == sourceLabel ) % tiene el valor que estamos buscando?
            clas(i, j) = targetLabel; % marcarlo
            tam = tam + 1; % incrementar el tamaño de la region en uno

            % Verificar los 8 vecinos, añadiendolos a la lista
            if( i>1 ) % hay vecino a la izquierda?
                sp = sp + 1; % meterlo a la pila
                stack(sp).x = i-1;
                stack(sp).y = j;
            end
            if( i<nRow ) % hay vecino a la derecha?
                sp = sp + 1; % meterlo a la pila
                stack(sp).x = i+1;
                stack(sp).y = j;
            end
            if( j>1 ) % hay vecino arriba?
                sp = sp + 1; % meterlo a la pila
                stack(sp).x = i;
                stack(sp).y = j-1;
            end
            if( j<nCol ) % hay vecino abajo?
                sp = sp + 1; % meterlo a la pila
                stack(sp).x = i;
                stack(sp).y = j+1;
            end
            if( i>1 && j>1 ) % hay vecino arriba a la izquierda ?
                sp = sp + 1; % meterlo a la pila
                stack(sp).x = i-1;
                stack(sp).y = j-1;
            end
            if( i>1 && j<nCol ) % hay vecino abajo a la izquierda ?
                sp = sp + 1; % meterlo a la pila
                stack(sp).x = i-1;
                stack(sp).y = j+1;
            end
            if( i<nRow && j>1 ) % hay vecino arriba a la derecha?
                sp = sp + 1; % meterlo a la pila
                stack(sp).x = i+1;
                stack(sp).y = j-1;
            end
            if( i<nRow && j<nCol ) % hay vecino abajo a la derecha ?
                sp = sp + 1; % meterlo a la pila

```

```

        stack(sp).x = i+1;
        stack(sp).y = j+1;
    end
end

end
outClas = clas; % copiar el resultado a la variable de salida
end

%-----
% EXTRACCION Y MEDICION
%-----
% Función : extraccion
% Desc : obtiene los extractos de un conjunto de datos segmentados
% Parámetros:
% dem = conjunto de datos de elevacion, una matriz
% extractos = conjunto de datos con extractos etiquetados, una
% matriz
% ext = lista de valores de los extractos (coordenadas extremas,
% tamaño, altitudes minima y maxima)
function [ extractos, ext ] = extraccion( dem )
    extractos = {}; % Lista de objetos que se extraen
    numExtractos = 0; % Número de objetos extraidos
    seg = dem; % Copia de trabajo de los datos de entrada
    [m n] = size(seg); % Tamaño del conjunto de datos

    for i=1:m % Búsqueda de los extractos:
        for j=1:n
            if seg(i,j) > 0 % dato que no sea vacio?
                numExtractos = numExtractos + 1; % Incrementamos la cuenta de objetos
                [seg, tam, extreme] = regGrowBlind8( seg, i, j, -numExtractos);
                % Buscamos todo el extracto etiquetandolo con un num. negativo
                extractos(numExtractos).tam = tam; % Asignamos el tamaño (num. de píxeles)
                del extracto
                extractos(numExtractos).ext = extreme; % Asignamos las coordenadas extremas del
                extracto
            end
        end
    end

    ext = seg; % Copiamos el resultado de las etiquetas

    for i=1:numExtractos % Inicialización de los extractos:
        extractos(i).data = zeros(size(seg)); % creamos matrices con cero para cada
        extracto
        extractos(i).altMin = 10000;
        extractos(i).altMax = -10000;
        extractos(i).altMed = 0;
    end

    for i=1:m % Copiar los valores de elevación:
        for j=1:n
            if seg(i,j) < 0 % Está etiquetado el píxel?
                elevación en el extracto correspondiente
                extractos(-seg(i,j)).data(i,j) = dem(i,j); % Si, copiamos el dato de
                extractos(-seg(i,j)).altMed = ... % Para calcular la altitud media.
                extractos(-seg(i,j)).altMed + dem(i,j);
                if( extractos(-seg(i,j)).altMin > dem(i,j) ) % Para calcular la altitud mínima.
                    extractos(-seg(i,j)).altMin = dem(i,j);
                end
                if( extractos(-seg(i,j)).altMax < dem(i,j) ) % Para calcular la altitud máxima.
                    extractos(-seg(i,j)).altMax = dem(i,j);
                end
            end
        end
    end
end
end
end

```

```

for i=1:numExtractos                                % Calculo final de
las altitudes medias.
    extractos(i).altMed = extractos(i).altMed / tam;
end
end

% Función    : regGrowBlind8
% Desc      : Calcula la región 8 conexas dada una semilla
% Parámetros:
%   clas     = conjunto de datos de entrada, una matriz
%   row y col = posición relativa a clas de la semilla
%   targetLabel = valor con el que se etiqueta la región
%   outClas   = datos con la región etiquetada
%   tam       = tamaño de la región (en número de píxeles)
%   ext       = coordenadas extremas de la región
function [ outClas, tam , ext] = regGrowBlind8( clas, row, col, targetLabel )
    sp = 1;                                         % stack pointer
    stack(1).x = row;                               % iniciar la pila con la semilla
    stack(1).y = col;
    [nRow nCol] = size(clas);                       % tamaño de la matriz clas
    tam = 0;                                        % iniciar el tamaño de la región a cero
    ext.xmin = col;                                 % Coordenadas extremas de la región
    ext.xmax = col;
    ext.ymin = row;
    ext.ymax = row;

    while sp > 0                                    % mientras haya datos en la pila
        i = stack(sp).x;                            % sacar el dato en el tope de la pila
        j = stack(sp).y;
        sp = sp - 1;
        if( clas(i, j) > 0 )                        % tiene el valor que estamos buscando?
            clas(i, j) = targetLabel;              % marcarlo
            tam = tam + 1;                          % incrementar el tamaño de la region en uno
            if( ext.xmax < j )                      % Y calculamos las coord extremas
                ext.xmax = j;
            elseif( ext.xmin > j )
                ext.xmin = j;
            end
            if( ext.ymax < i )
                ext.ymax = i;
            elseif( ext.ymin > i )
                ext.ymin = i;
            end
        end

        % Verificar los 8 vecinos, añadiendolos a la lista
        if( i>1 )                                   % hay vecino a la izquierda?
            sp = sp + 1;                             % meterlo a la pila
            stack(sp).x = i-1;
            stack(sp).y = j;
        end
        if( i<nRow )                                % hay vecino a la derecha?
            sp = sp + 1;                             % meterlo a la pila
            stack(sp).x = i+1;
            stack(sp).y = j;
        end
        if( j>1 )                                   % hay vecino arriba?
            sp = sp + 1;                             % meterlo a la pila
            stack(sp).x = i;
            stack(sp).y = j-1;
        end
        if( j<nCol )                                % hay vecino abajo?
            sp = sp + 1;                             % meterlo a la pila
            stack(sp).x = i;
            stack(sp).y = j+1;
        end
        if( i>1 && j>1 )                            % hay vecino arriba a la izquierda ?
            sp = sp + 1;                             % meterlo a la pila
            stack(sp).x = i-1;

```

```

        stack(sp).y = j-1;
    end
    if( i>1 && j<nCol)                % hay vecino abajo a la izquierda ?
        sp = sp + 1;                  % meterlo a la pila
        stack(sp).x = i-1;
        stack(sp).y = j+1;
    end
    if( i<nRow && j>1)                % hay vecino arriba a la derecha?
        sp = sp + 1;                  % meterlo a la pila
        stack(sp).x = i+1;
        stack(sp).y = j-1;
    end
    if( i<nRow && j<nCol)              % hay vecino abajo a la derecha ?
        sp = sp + 1;                  % meterlo a la pila
        stack(sp).x = i+1;
        stack(sp).y = j+1;
    end
end
end

end
outClas = clas;                       % copiar el resultado a la variable de salida
end

%-----
% DESCRIPCION
%-----
% Función      : descripcion
% Desc        : genera una descripción de un extracto
% Parámetros:
%   extracto   = extracto a describir
%   firma      = firma del extracto
%   demParam   = parámetros del modelo de elevación
%   desc       = descripción textual
function [desc] = descripcion( extracto, firma, demParam )
    desc = 'UN OBJETO GEOGRÁFICO';

    if( extracto.tam > 5 )
        if( strcmp(firma, 'eee') )
            desc = 'UNA MONTAÑA';
        elseif( strcmp(firma, 'eel') )
            desc = 'UNA CORDILLERA';
        elseif( strcmp(firma, 'eed') )
            desc = 'UN MONTE';
        elseif( strcmp(firma, 'ele') )
            desc = 'UN CERRO';
        elseif( strcmp(firma, 'ell') )
            desc = 'UN ALTIPLANO';
        elseif( strcmp(firma, 'eld') )
            desc = 'UNA COLINA';
        elseif( strcmp(firma, 'ede') )
            desc = 'UN ELEV-DEPR-ELEV';
        elseif( strcmp(firma, 'edl') )
            desc = 'UNA LOMA';
        elseif( strcmp(firma, 'edd') )
            desc = 'UN COLLADO';
        elseif( strcmp(firma, 'le') )
            desc = 'UN LLANO';
        elseif( strcmp(firma, 'll') )
            desc = 'UNA PLANICIE';
        elseif( strcmp(firma, 'ld') )
            desc = 'UNA MESETA';
        elseif( strcmp(firma, 'de') )
            desc = 'UN DEPR-ELEV';
        elseif( strcmp(firma, 'dl') )
            desc = 'UNA VALLE';
        elseif( strcmp(firma, 'dd') )
            desc = 'UNA CUENCA';
        end
    end
end

```

```

desc = sprintf('%s CON AREA DE %d %s,', desc, ...
  extracto.tam*demParam.res*demParam.res, ...
  demParam.unidadArea );
desc = sprintf('%s ALTITUD MÍNIMA DE %f %s Y MÁXIMA DE %f %s,', desc, ...
  extracto.altMin, demParam.unidadAltitud, ...
  extracto.altMax, demParam.unidadAltitud );
desc = sprintf('%s SUS COORDENADAS EXTREMAS SON (%f, %f) Y (%f, %f) %s.', desc, ...
  extracto.ext.xmin*demParam.res+demParam.xmin,...
  extracto.ext.ymin*demParam.res+demParam.ymin,...
  extracto.ext.xmax*demParam.res+demParam.xmax,...
  extracto.ext.ymax*demParam.res+demParam.ymax,...
  demParam.unidadProy);

if( firma(1) == 'e' )
  [x, y, a] = buscarCima(extracto, demParam);
  desc = sprintf('%s SU CIMA SE ENCUENTRA EN LA COORDENADA (%f, %f) A UNA ALTITUD DE %f
%s', ...
    desc, x, y, a, demParam.unidadAltitud);
elseif( firma(1)=='d')
  [x, y, a] = buscarZima(extracto, demParam);
  desc = sprintf('%s SU ZIMA SE ENCUENTRA EN LA COORDENADA (%f, %f) A UNA ALTITUD DE %f
%s', ...
    desc, x, y, a, demParam.unidadAltitud);
end
else
  desc = '';
end
end

% Función : descripcionMultiple
% Desc : genera una descripción de un conjunto de extractos
% Parámetros:
% ext = conjunto de extractos
% firma = firma del conjunto de extractos
% demParam = parámetros del modelo de elevación
% desc = descripción textual
function [ desc ] = descripcionMultiple( ext, firma, demParam )
  [m n] = size(ext);
  desc = '';
  for i=1:n
    descTemp = descripcion(ext(1,i), firma, demParam);
    if( strcmp(descTemp, '') == 0 )
      if(i==1)
        desc = sprintf('%s', descTemp);
      else
        desc = sprintf('%s\n%s', desc, descTemp);
      end
    end
  end
end

% Función : descripcionDEM
% Desc : genera una descripción de un DEM con base en los parámetros
% que se han leído del archivo que los contiene
% Parámetros:
% demParam = parámetros del modelo de elevación
% desc = descripción textual
function [desc ] = descripcionDEM( demParam )
  desc = sprintf('CONJUNTO DE DATOS ESPACIALES RASTER DEL MODELO DIGITAL DE ELEVACIÓN DE "%s"',
demParam.nombre);
  desc = sprintf('%s QUE TIENE UNA RESOLUCION ESPACIAL DE %d %s.', desc, demParam.res,
demParam.unidadProy );
  desc = sprintf('%s SU ALTITUD MINIMA ES DE %f %s', desc, demParam.altMin,
demParam.unidadAltitud );
  desc = sprintf('%s Y SU ALTITUD MAXIMA ES DE %f %s.', desc, demParam.altMax,
demParam.unidadAltitud );
  desc = sprintf('%s LAS COORDENADAS EXTREMAS SON (%f, %f) Y (%f, %f) %s', desc,...

```

```

        demParam.xmin, demParam.ymin, demParam.xmax, demParam.ymax, demParam.unidadProy );
    desc = sprintf('%s EN PROYECCION %s', desc, demParam.proy);
    desc = sprintf('%s Y CONTIENE:', desc);
end

% Función      : buscarCima
% Desc        : calcula la cima o punto más alto de una elevación, si dos o
%              más puntos comparten la misma altitud, esta función regresa
%              el primero que se encuentre
% Parámetros:
%      ext      = extracto de elevación
%      demParam = parámetros del modelo de elevación
%      x, y     = posición de la cima
%      alt      = altitud de la cima
function [ x, y, alt ] = buscarCima( ext, demParam)
    [m n] = size(ext.data);           % tamaño del extracto
    x = -1;                           % coordenadas iniciales de la cima
    y = -1;
    alt = 0;                           % altitud inicial de la cima
    for i=1 : m
        for j=1 : n
            if( ext.data(i, j)>0 && alt < ext.data(i, j) ) % la altitud es la mayor ?
                alt = ext.data(i, j); % guardar como posible cima
                x = j;
                y = i;
            end
        end
    end
end

% cambiar a coordenadas mundiales
x = demParam.xmin + x*demParam.res;
y = demParam.ymin + y*demParam.res;
end

% Función      : buscarZima
% Desc        : calcula la zima o punto más bajo de una depresión, si dos o
%              más puntos comparten la misma altitud, esta función regresa
%              el primero que se encuentre
% Parámetros:
%      ext      = extracto de depresión
%      demParam = parámetros del modelo de elevación
%      x, y     = posición de la zima
%      alt      = altitud de la zima
function [ x, y, alt ] = buscarZima( ext, demParam)
    [m n] = size(ext.data);           % tamaño del extracto
    x = -1;                           % coordenadas iniciales de la zima
    y = -1;
    alt = 10000;                       % altitud inicial de la zima
    for i=1 : m
        for j=1 : n
            if( ext.data(i, j)>0 && alt > ext.data(i, j) ) % la altitud es la menor ?
                alt = ext.data(i, j); % guardar como posible zima
                x = j;
                y = i;
            end
        end
    end
end

% cambiar a coordenadas mundiales
x = demParam.xmin + x*demParam.res;
y = demParam.ymin + y*demParam.res;
end

```

Figura 105. Código fuente en MATLAB

A.2. Pruebas con el DEM de Las Vegas

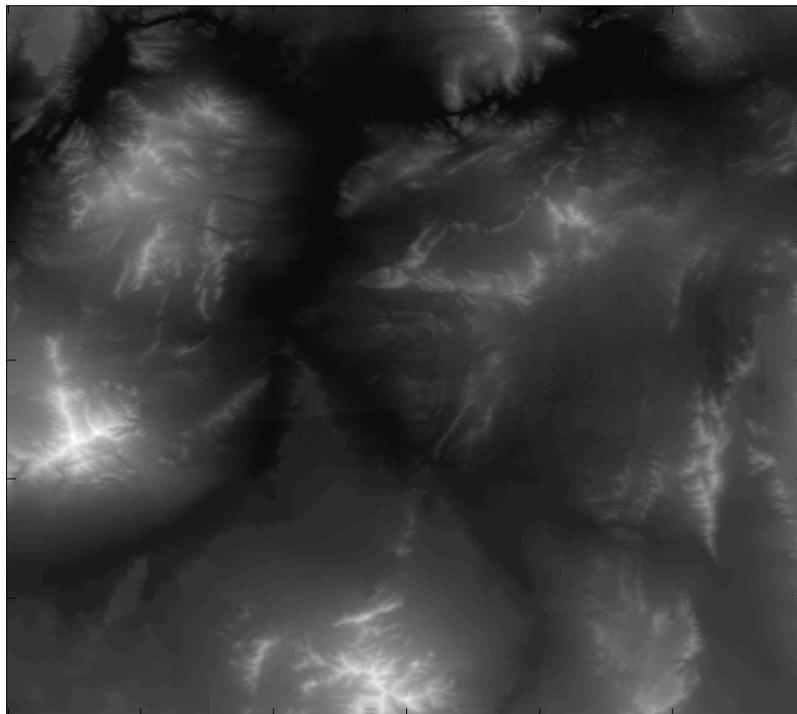


Figura 106. Modelo Digital de Elevación de Las Vegas

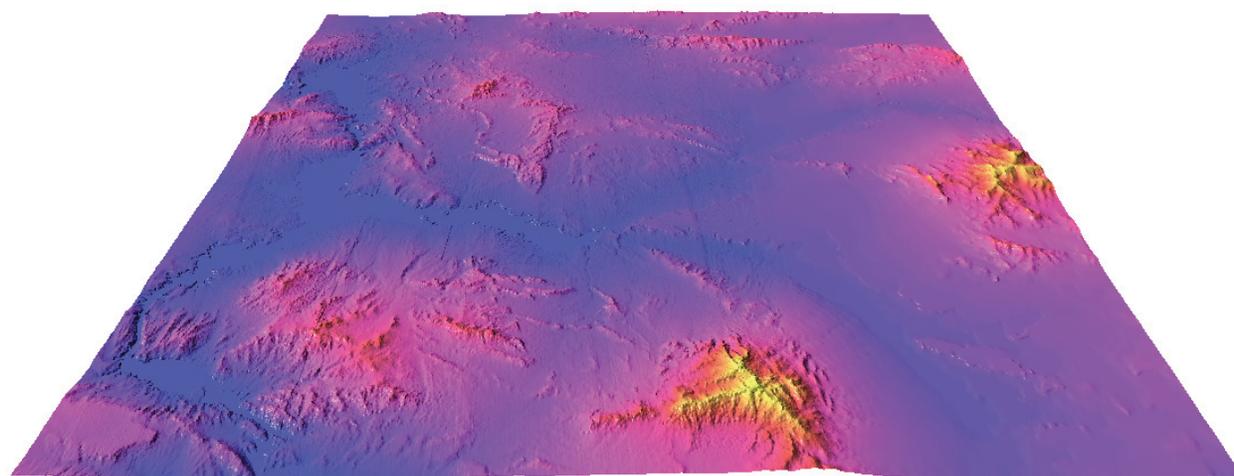


Figura 107. Vista tridimensional del Modelo Digital de Elevación de Las Vegas

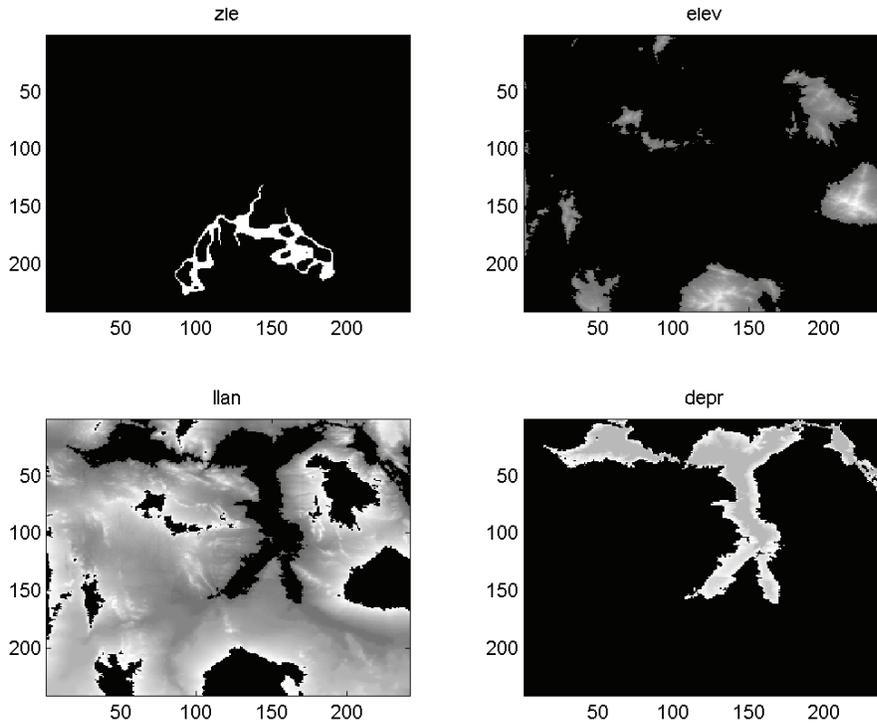


Figura 108. Resultados del proceso de segmentación aplicado al DEM de Las Vegas

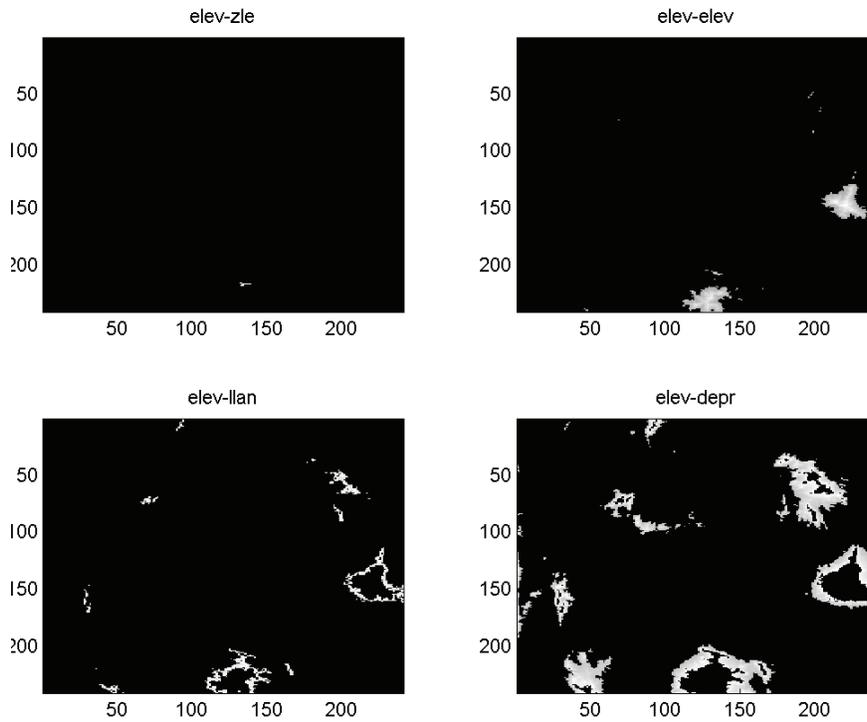


Figura 109. Resultados de la segmentación aplicada al segmento "elev" del DEM de Las Vegas

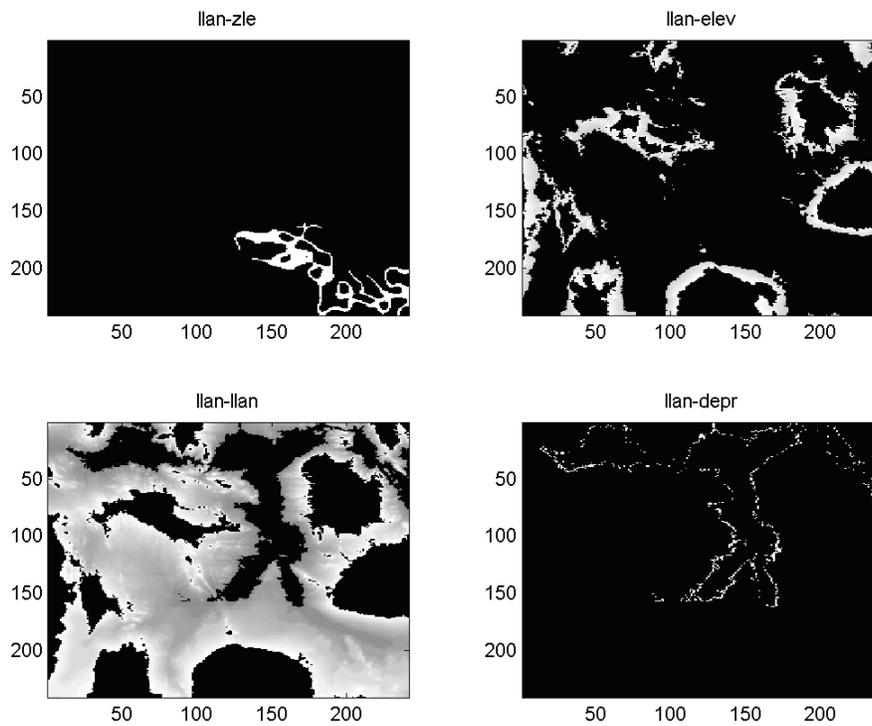


Figura 110. Resultados de la segmentación aplicada al segmento “llan” del DEM de Las Vegas

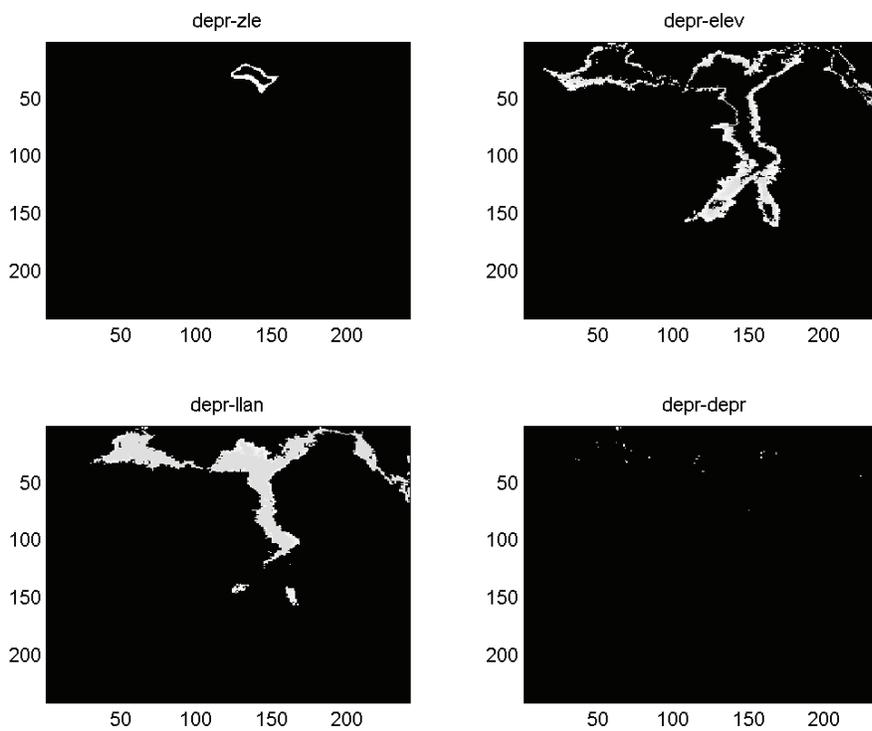


Figura 111. Resultados de la segmentación aplicada al segmento “depr” del DEM de Las Vegas

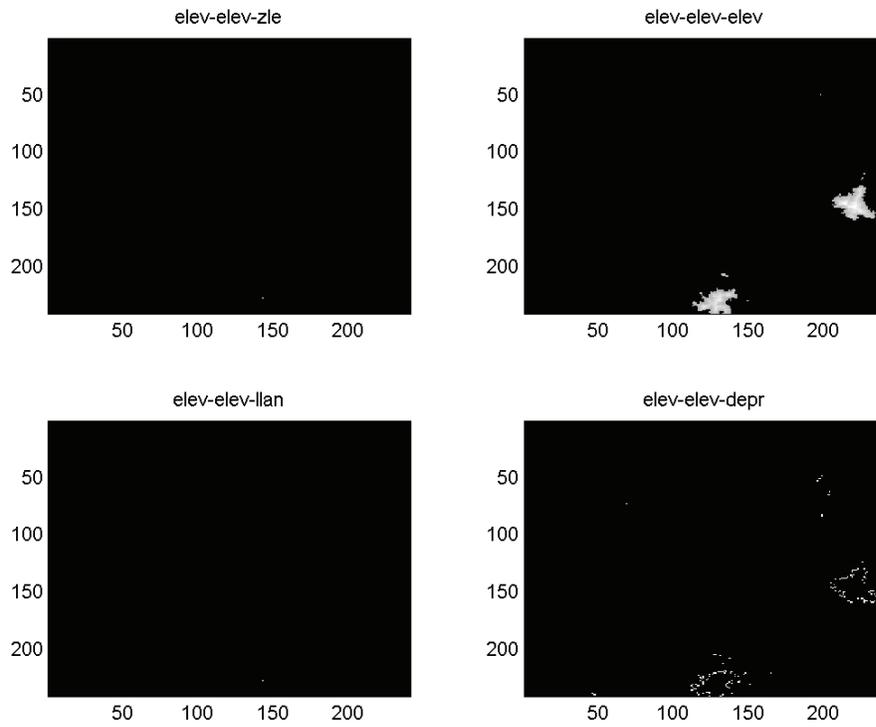


Figura 112. Resultados de la segmentación aplicada al segmento “elev-elev” de Las Vegas

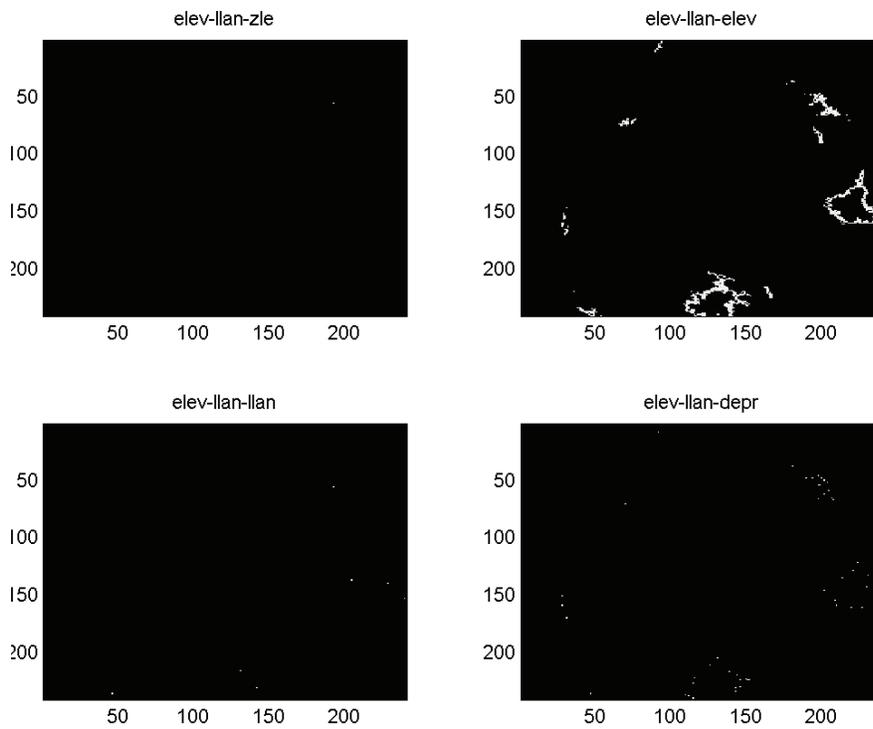


Figura 113. Resultados de la segmentación aplicada al segmento “elev-llan” de Las Vegas

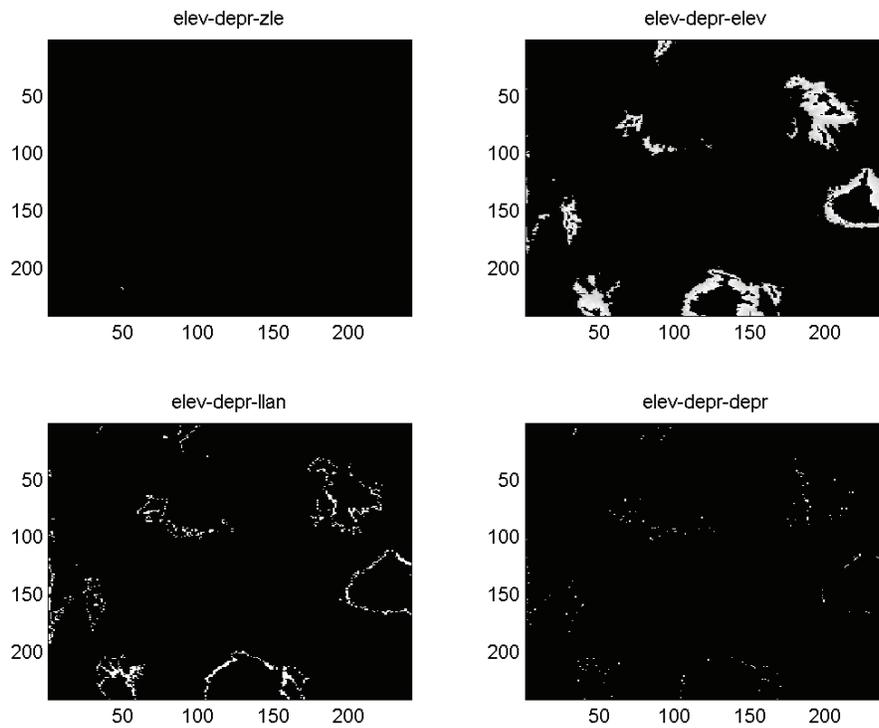


Figura 114. Resultados de la segmentación aplicada al segmento “elev-depr” de Las Vegas

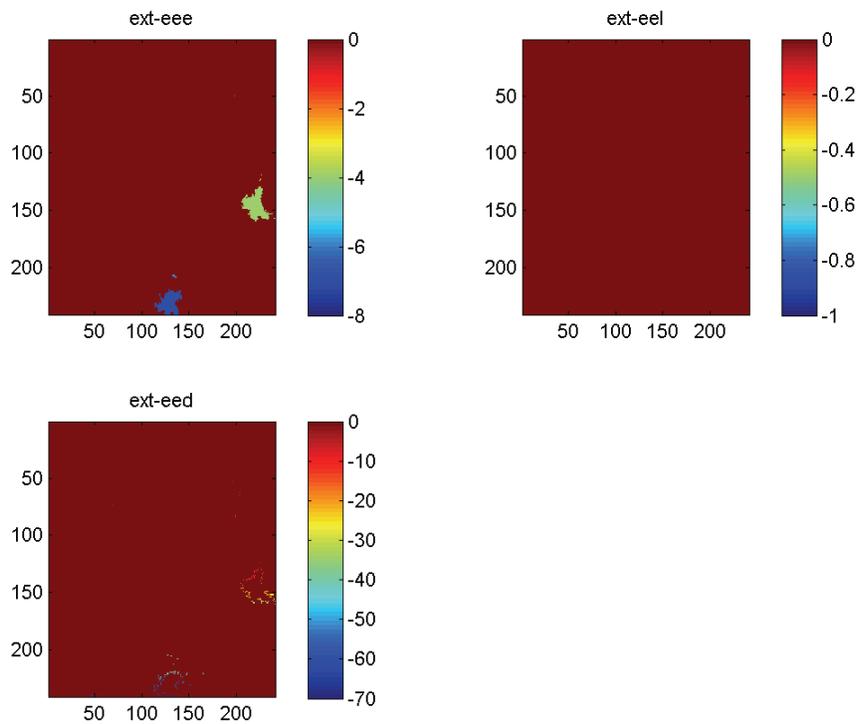


Figura 115. Extractos obtenidos de las firmas “elev-elev” del DEM de Las Vegas

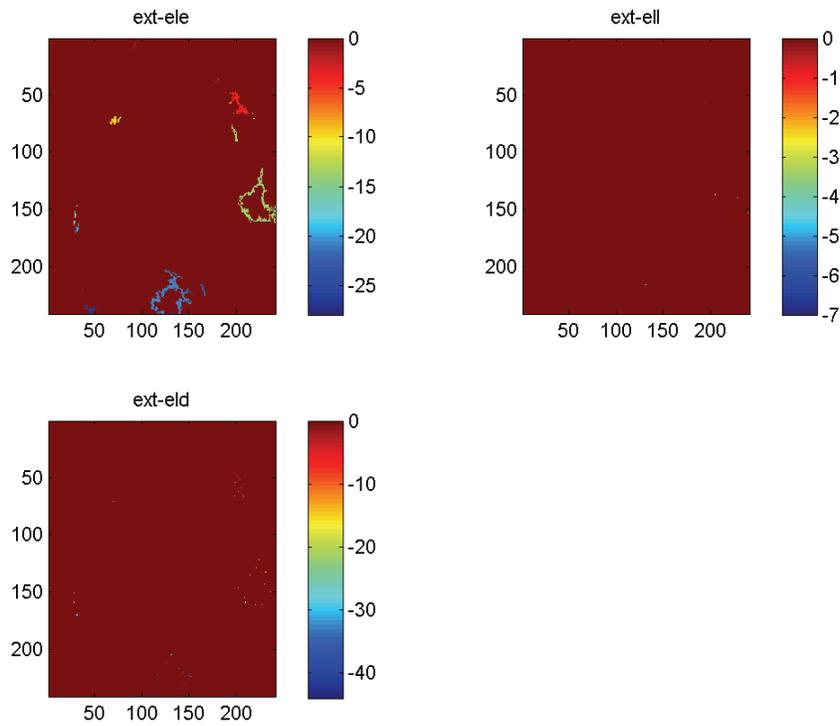


Figura 116. Extractos obtenidos de las firmas “elev-llan” del DEM de Las Vegas

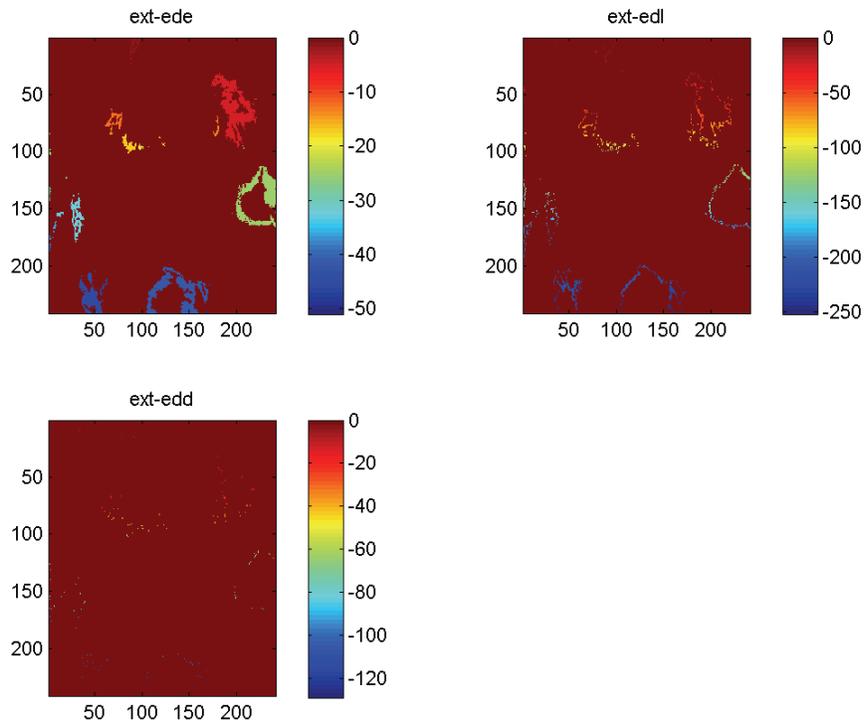


Figura 117. Extractos obtenidos de las firmas “elev-depr” del DEM de Las Vegas

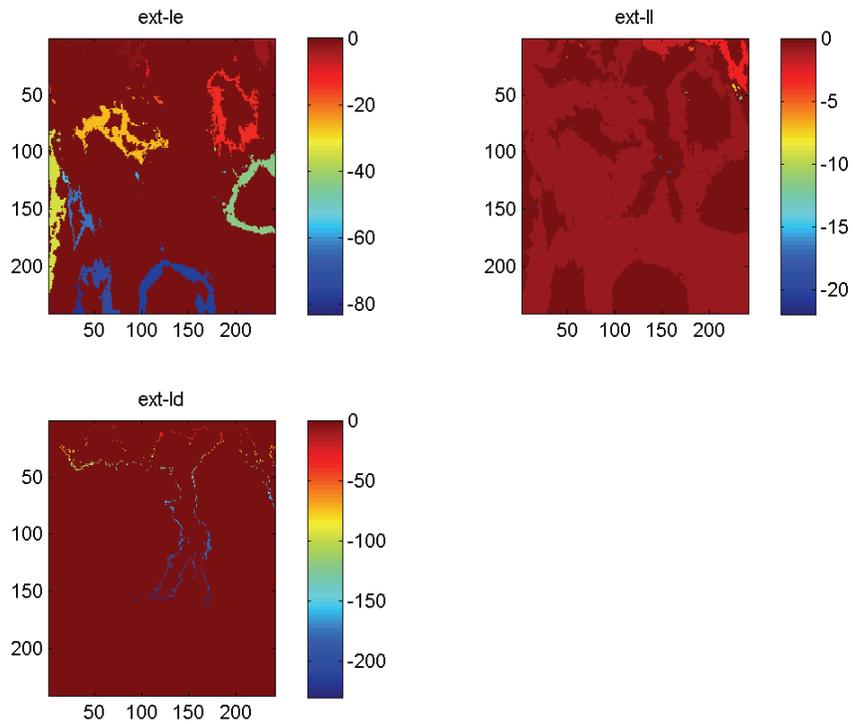


Figura 118. Extractos obtenidos de las firmas “llan” del DEM de Las Vegas

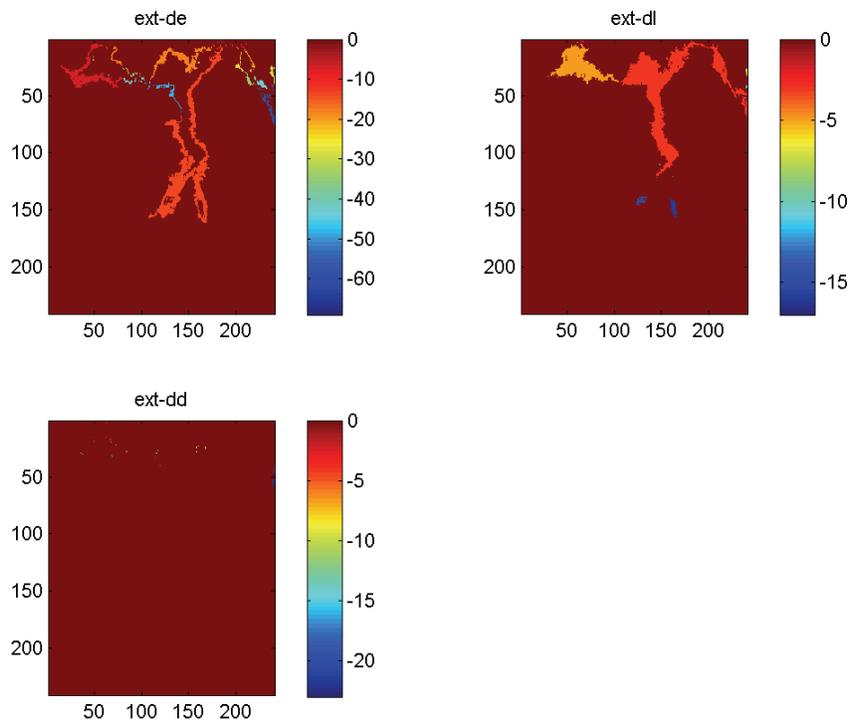


Figura 119. Extractos obtenidos de las firmas “depr” del DEM de Las Vegas

A.3. Pruebas con el DEM de Salt Lake City

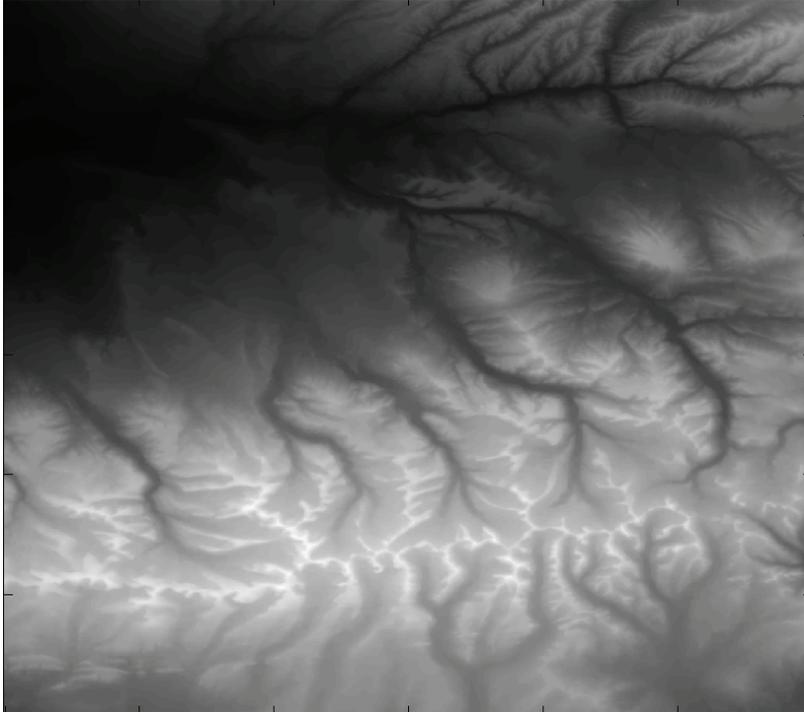


Figura 120. Modelo Digital de Elevación de Salt Lake City

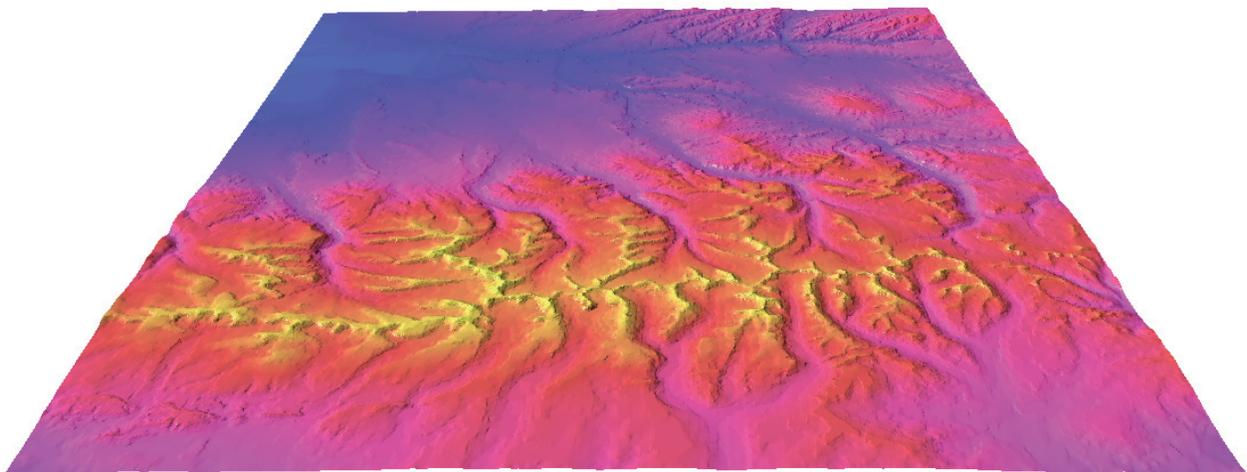


Figura 121. Vista tridimensional del Modelo Digital de Elevación de Salt Lake City

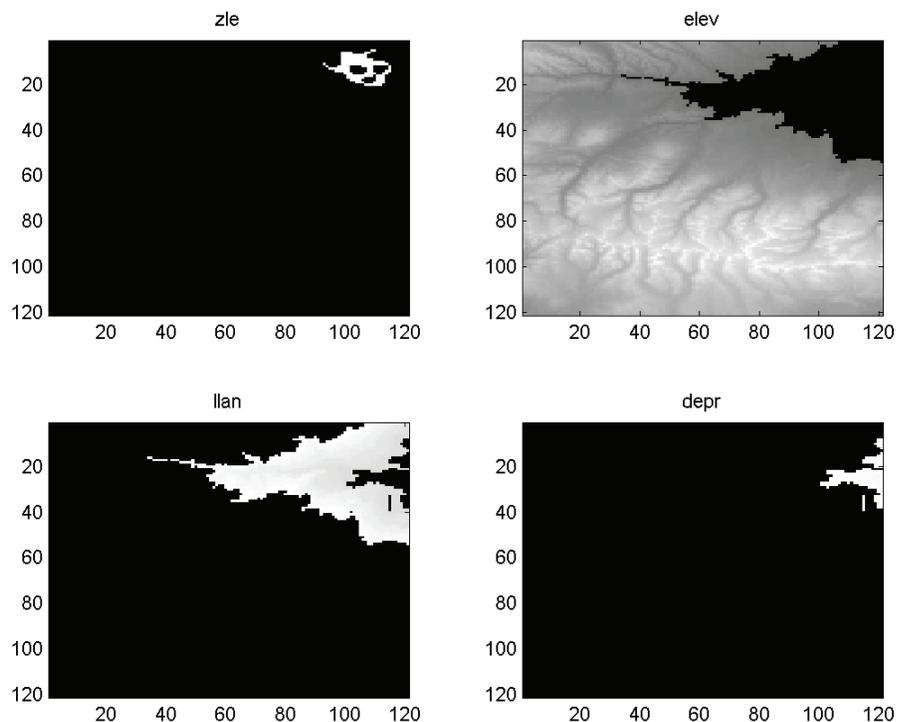


Figura 122. Resultados del proceso de segmentación aplicado al DEM de Salt Lake City

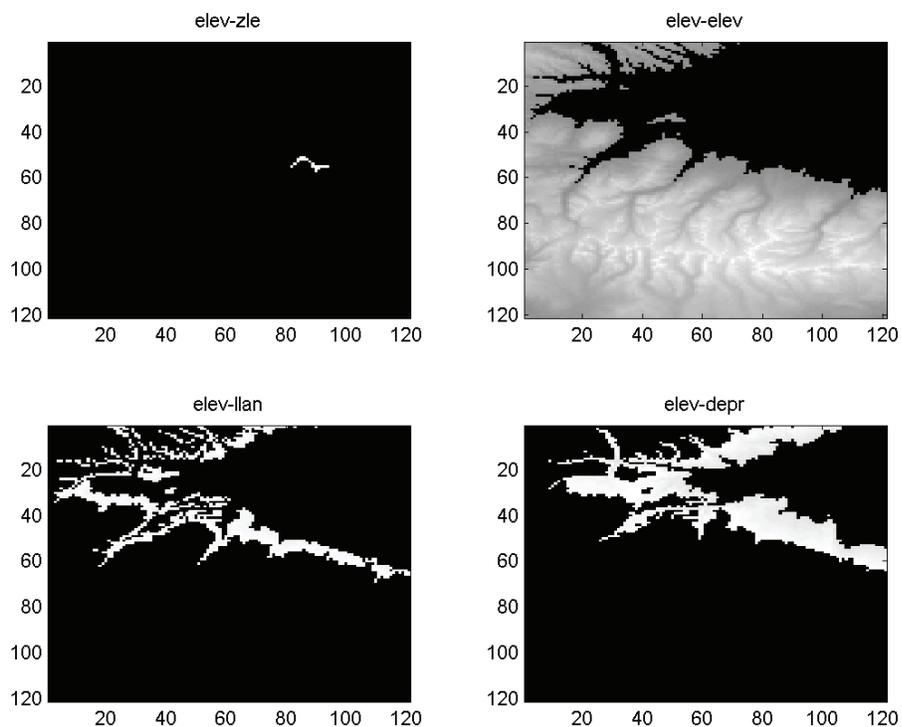


Figura 123. Resultados de la segmentación aplicada al segmento "elev" de Salt Lake City

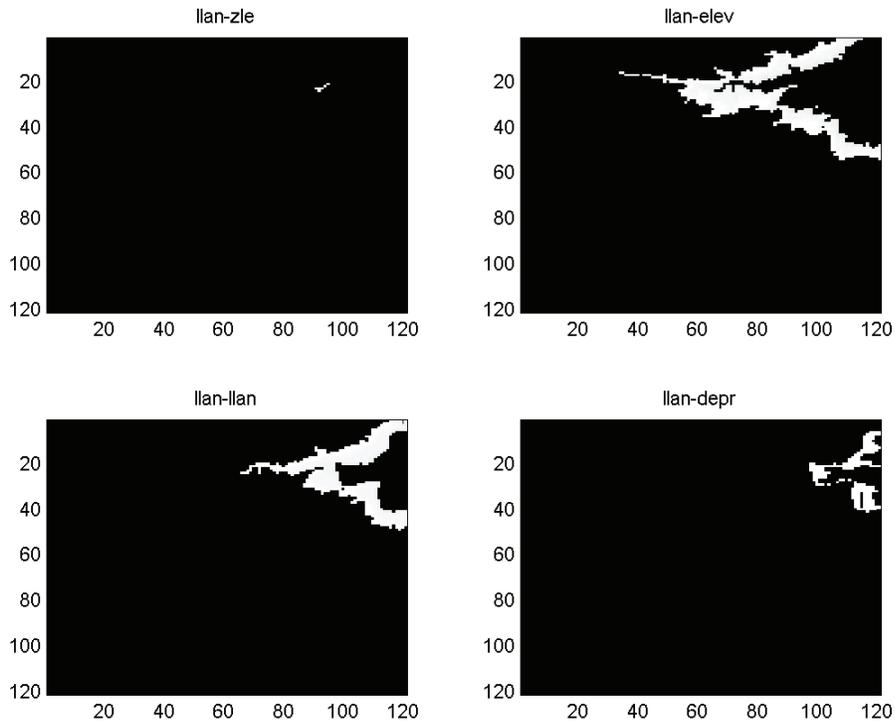


Figura 124. Resultados de la segmentación aplicada al segmento “llan” de Salt Lake City

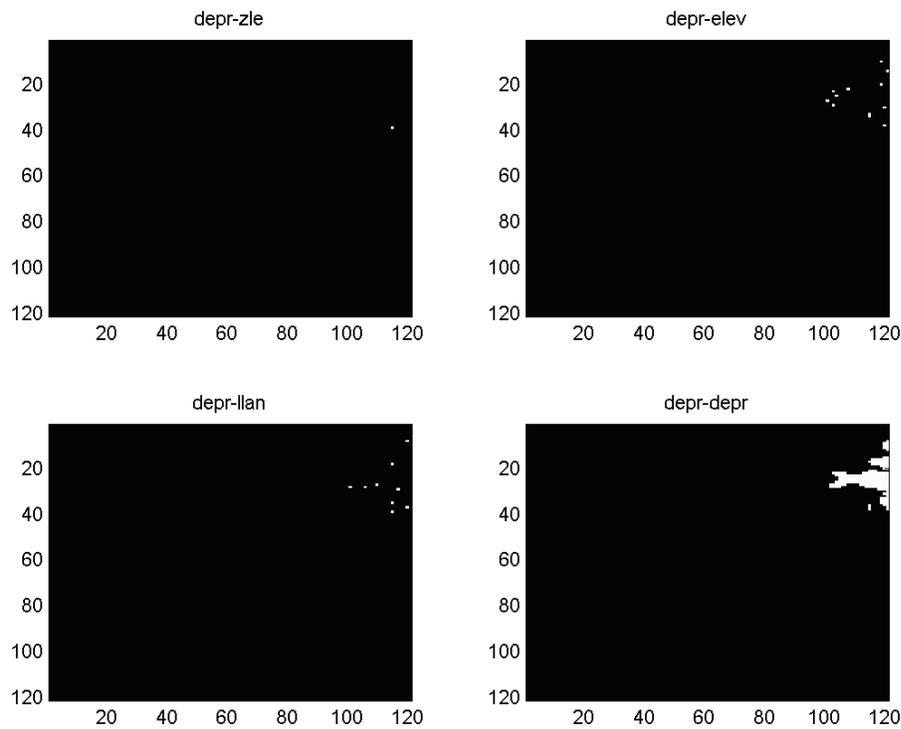


Figura 125. Resultados de la segmentación aplicada al segmento “depr” de Salt Lake City

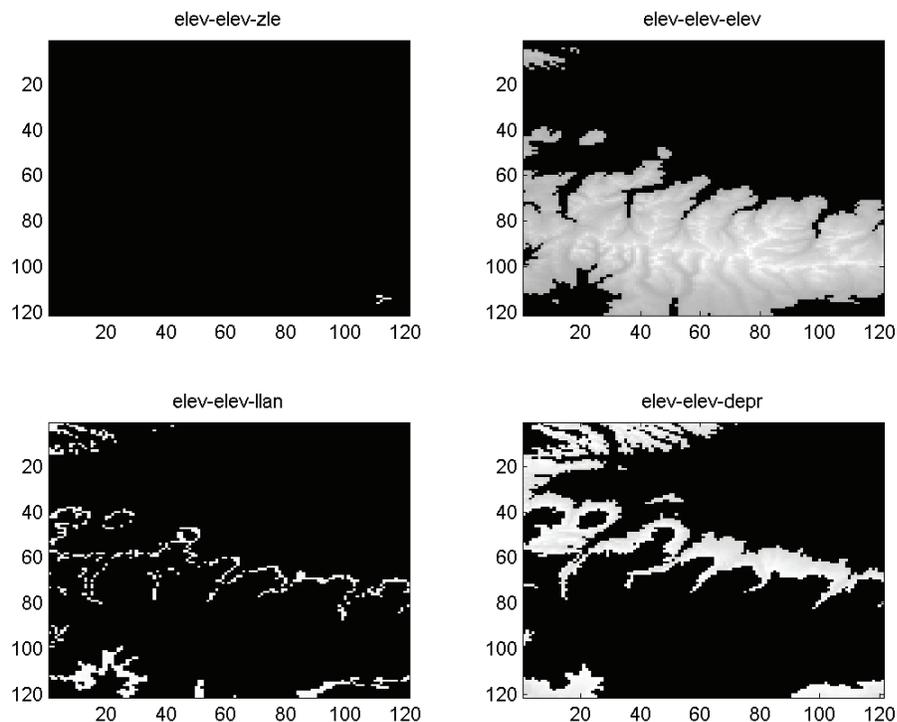


Figura 126. Resultados de la segmentación aplicada al segmento “elev-elev” de Salt Lake City

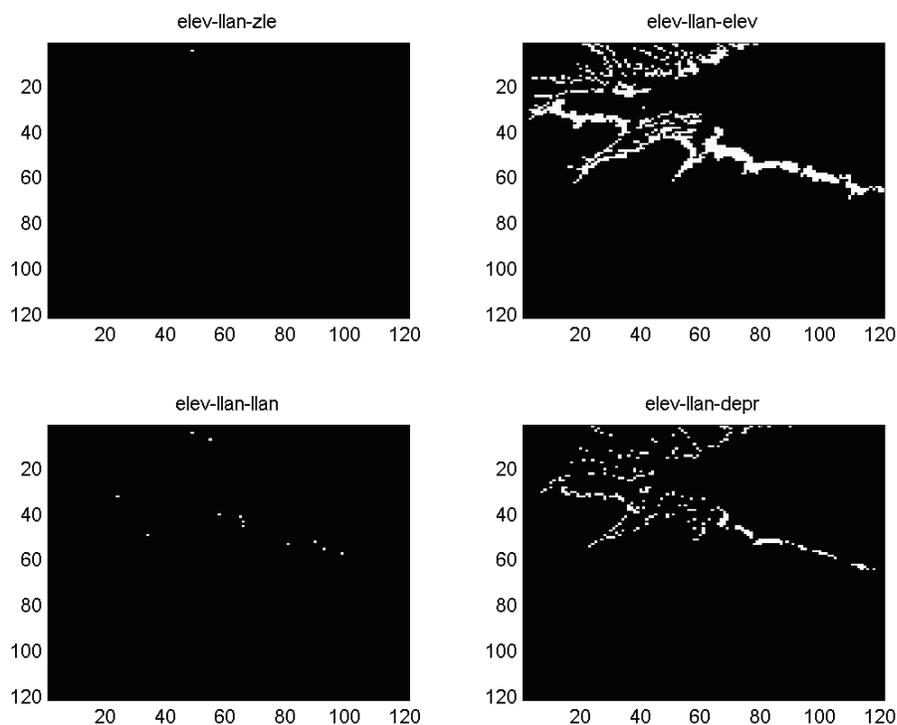


Figura 127. Resultados de la segmentación aplicada al segmento “elev-llan” de Salt Lake City

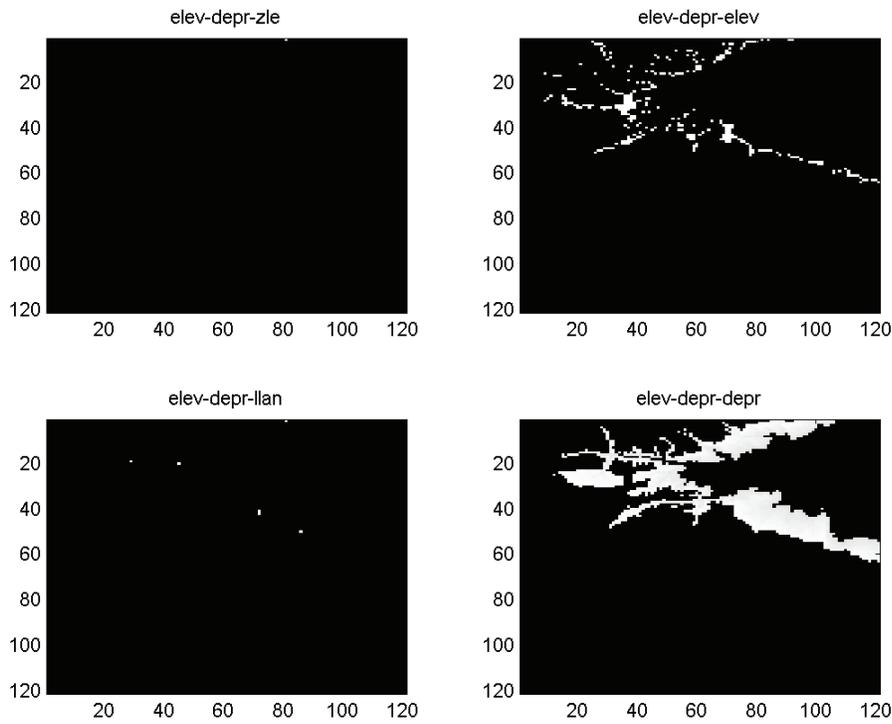


Figura 128. . Resultados de la segmentación aplicada al segmento “elev-depr” de Salt Lake City

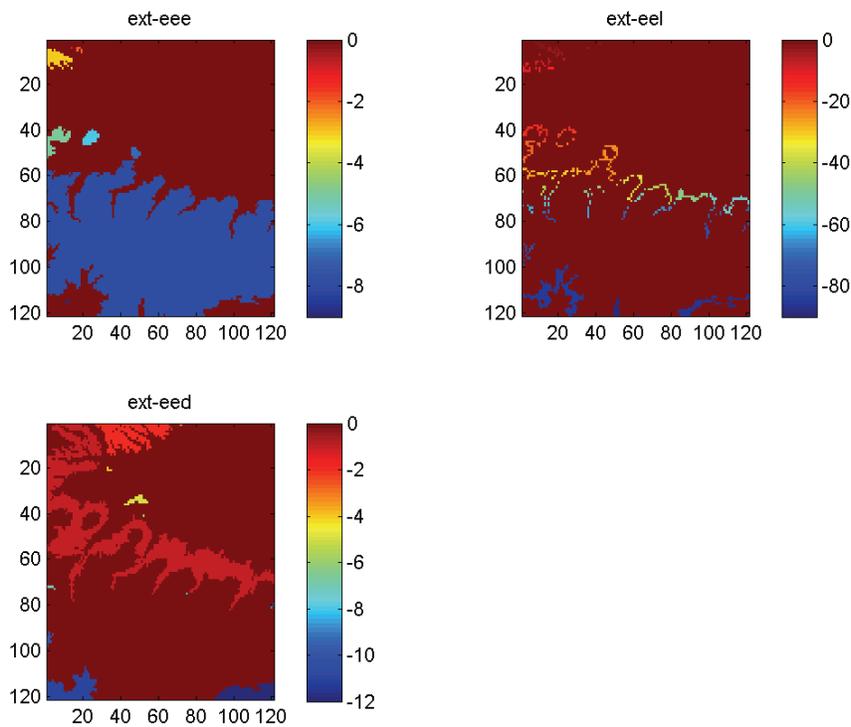


Figura 129. Extractos obtenidos de las firmas “elev-elev” del DEM de Salt Lake City

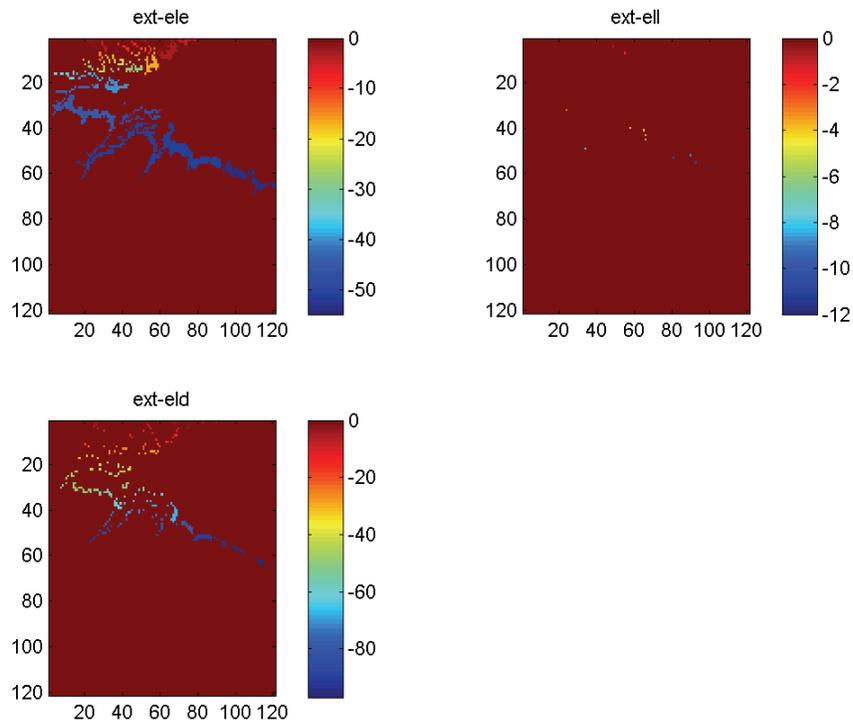


Figura 130. Extractos obtenidos de las firmas “elev-llan” del DEM de Salt Lake City

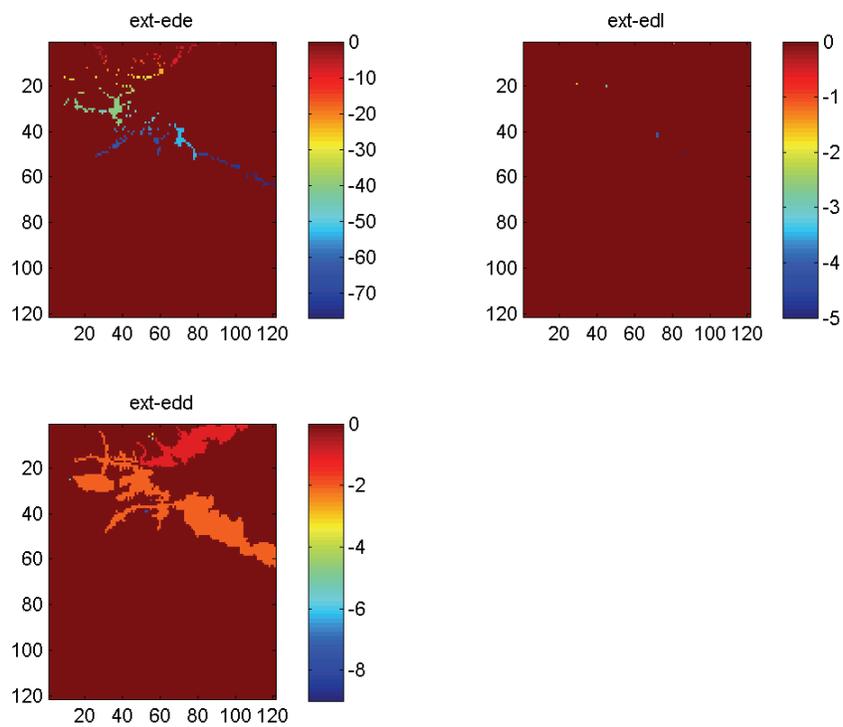


Figura 131. Extractos obtenidos de las firmas “elev-depr” del DEM de Salt Lake City

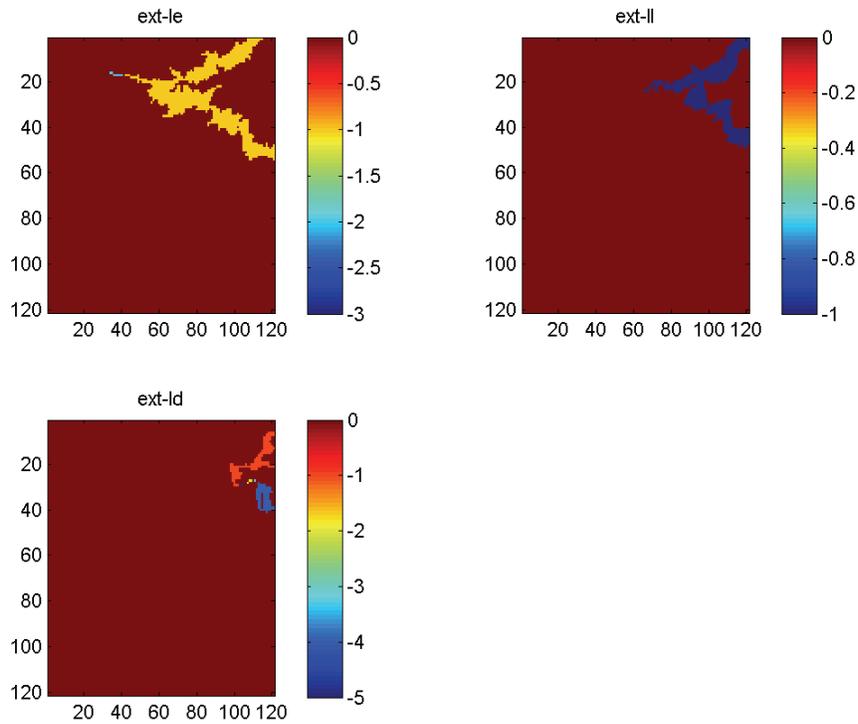


Figura 132. Extractos obtenidos de las firmas "llan" del DEM de Salt Lake City

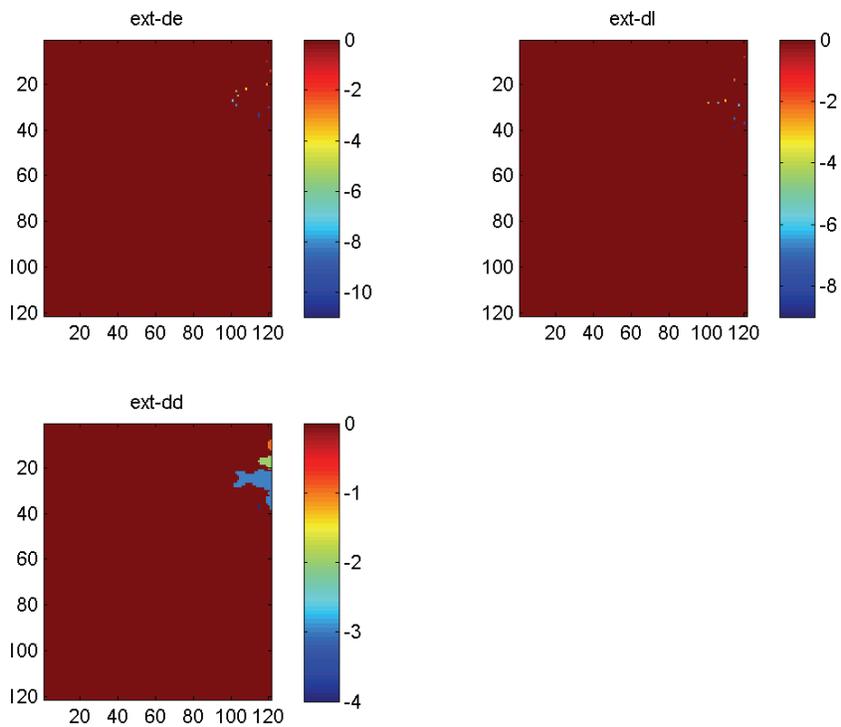


Figura 133. Extractos obtenidos de las firmas "depr" del DEM de Salt Lake City